

Sergey Subbotin
Editor

Computer Modeling and Intelligent Systems

Proceedings of The Third International Workshop
CMIS-2020



Zaporizhzhia, Ukraine
April 27- May 1, 2020

Subbotin S., (Ed.): The Third International Workshop on Computer Modeling and Intelligent Systems (CMIS-2020). Zaporizhzhia, Ukraine, April 27 - May 1, 2020, CEUR-WS.org, online

This volume represents the proceedings of the Third International Workshop on Computer Modeling and Intelligent Systems (CMIS-2020), held in Zaporizhzhia, Ukraine, in April-May, 2020. It comprises 83 contributed papers that were carefully peer-reviewed and selected from 177 submissions.

Copyright © 2020 for the individual papers by the papers' authors.
Copying permitted only for private and academic purposes.
This volume is published and copyrighted by its editors.

Preface

It is my pleasure to present you the proceedings of The Third International Workshop on Computer Modeling and Intelligent Systems (CMIS-2020), held in Zaporizhzhia, Ukraine on 27 of April - 1 of May 2020.

The CMIS workshop was held in 2020 for the third time at the National University "Zaporizhzhia Polytechnic" (<http://www.zp.edu.ua>), which is the largest and oldest regional center for science and higher education in the Zaporizhzhia region of south-eastern Ukraine.

More than ten thousand students of Bachelor, Master, and PhD programs study also as more than seven hundred research and teaching staff work at the university. In the last decade, the role of IT specialties has significantly increased at the university. The leading place in the training of specialists in the fields of Software Engineering and Computer Science is occupied by the Department of Software Tools (<http://pz.zntu.net>), which is the largest IT department in the region and the largest university's department in terms of the number of students. The Department of Software Tools of the National University "Zaporizhzhia Polytechnic" was the organizer and base for all CMIS workshops, including the CMIS-2020 workshop.

The National University "Zaporizhzhia Polytechnic" celebrates its 120th anniversary in 2020. The Third International Workshop on Computer Modeling and Intelligent Systems (CMIS-2020) is timed to coincide with this significant event.

The main purpose of the CMIS-2020 workshop is a discussion of the recent research results in all areas of Computer Sciences and Information Technologies.

The workshop is soliciting literature review, survey and research papers including, whilst not limited to, the following areas of interest:

- Information Technologies for Decision Making, Optimization and Control;
- Information Systems and Data Processing;
- Advanced Info-Communicational Technologies;
- Information Technologies for Computer Modeling;
- Computing in Applications;
- Intelligent Information Technologies;
- Information Technologies for Data Mining, Signal and Image Processing.

The workshop joined scientists from Belarus, Belgium, China, Czech Republic, Germany, Indonesia, Poland, Russia, Slovak Republic, Spain, Ukraine, and United States of America.

The language of CMIS-2020 Workshop is English. The workshop took the form of oral presentations of peer-reviewed regular papers. The video of presentations is available at <https://www.facebook.com/groups/cmisis.workshop/>.

The CMIS-2020 Organizing Committee have received 177 paper submissions (instead of 130 submissions in 2019), out of which 83 were accepted for presentation as a regular papers (instead of 81 in 2019) in the result of reviewing process, where the editor, program committee members, external and peer reviewers sought to ensure a high scientific level, as well as a high-quality presentation of the selected papers.

The paper acceptance ratio in 2020 was 47% (instead of 62% in 2019). These papers were published in this volume of CMIS-2020 proceedings.

The workshop would not have been possible without the support of many people. I would like to thank the Workshop Program Committee Members, organization staff, and external reviewers for their excellent and tireless work. I sincerely wish that all attendees benefited scientifically from the workshop and wish them every success in their research. It is the humble wish of the workshop organizers that the professional dialogue among the researchers, scientists, and engineers continues beyond the event and that the friendships and collaborations forged will linger and prosper for many years to come.

May, 2020

Sergey Subbotin, Dr. Sc., Prof.,
CMIS-2020 PC Chair and Proceedings Editor
Head of the Department of Software Tools,
National University "Zaporizhzhia Polytechnic"

Programme Committee

Chair

Prof. **Sergey Subbotin**, National University "Zaporizhzhia Polytechnic", Ukraine

Members

Prof. **Yevgeniy Bodyanskiy**, Kharkiv National University of Radio Electronics, Ukraine
Prof. **Martin Drahansky**, Brno University of Technology, Czech Republic
Prof. **Eugene Fedorov**, Cherkasy State Technological University, Ukraine
Prof. **Iurii Krak**, Taras Shevchenko National University of Kyiv, Ukraine
Prof. **Vitaly Levashenko**, University of Zilina, Slovak Republic
Prof. **David Luengo**, Universidad Politecnica de Madrid, Spain
Prof. **Volodymyr Lytvynenko**, Kherson National Technical University, Ukraine
Prof. **Nataliya Pankratova**, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine
Prof. **Anatoliy Sachenko**, Ternopil National Economic University, Ukraine
Prof. **Natalya Shakhovska**, Lviv Polytechnic National University, Ukraine
Prof. **Serhiy Shtovba**, Vinnytsia National Technical University, Ukraine
Prof. **Vitaliy Snytyuk**, Taras Shevchenko National University of Kyiv, Ukraine
Prof. **Carsten Wolff**, Dortmund University of Applied Sciences and Arts, Germany
Prof. **Karsten Henke**, Ilmenau University of Technology, Germany
Prof. **Elena Zaitseva**, University of Zilina, Slovak Republic
Dr. **Peter Arras**, Katholieke Universiteit Leuven, Belgium
Dr. **Ivan Izonin**, Lviv Polytechnic National University, Ukraine
Dr. **Igor Korobiichuk**, Warsaw University of Technology, Poland
Dr. **Anzhelika Parkhomenko**, National University "Zaporizhzhia Polytechnic", Ukraine
Dr. **Andrii Oliynyk**, National University "Zaporizhzhia Polytechnic", Ukraine
Dr. **Galyna Tabunshchyk**, National University "Zaporizhzhia Polytechnic", Ukraine
Dr. **Heinz-Dietrich Wuttke**, Ilmenau University of Technology, Germany

Local Organizational Committee

Prof. **Sergey Subbotin**, National University "Zaporizhzhia Polytechnic", Ukraine
MSc **Serhii Leoschenko**, National University "Zaporizhzhia Polytechnic", Ukraine
Maxim Andreev, National University "Zaporizhzhia Polytechnic", Ukraine

Reviewers

Sergey Ablameyko
Antonina Andrieieva
Volodymyr Artemchuk
Artem Artyukhov
Nadiia Artyukhova
Vitalii Babak
Sergey Babchuk
Vira Babenko
Vitalina Babenko
Elena Balalayeva
Oleksandr Bauzha
Maksym Bekuzarov
Olexander Belej
Oleksandr Bezsonov
Peter Bidyuk
Oleg Bisikalo
Vsevolod Bohaienko
Rykhard Bohush
Natalia Brynza
Ivan Burlachenko
Yevhen Burov
Taras Chaikivskyi
Olga Cherednichenko
Igor Cherevko
Varvara Chernenko
Elena Chernetsova
Andriy Dashkevich
Oleg Dekusha
Maksym Dosta
Oleksandr Drozd
Andriy Dudnik
Aleksandr Dykha
Emil Faure
Vladimir Gerasimov
Oleksandr Gerasin
Olga Gladkova
Dmitry Glukhov
Igor Golinko
Vyacheslav Gorev
Volodymyr Gorokhovatskyi
Alexander Gurko

Kateryna Hazdiuk
Svitlana Hrushko
Taras Humeniuk
Kateryna Isirova
Nataliia Istomina
Tetiana Kapliienko
Kostiantyn Karavaiev
Nadiia Karpenko
Dmytro Kasatkin
Dmytro Kavrin
Galina Kirichek
Lyudmyla Kirichenko
Valeriy Kolesnikov
Tetiana Kolpakova
Oksana Konchakovska
Yuriy Kondratenko
Andrii Kopp
Sergey Koshman
Anatolii Kosolapov
Konstantin Kostenko
Alexander Kostikov
Kateryna Kostyk
Gennadiy Kostyuk
Oleksiy Koyfman
Igor Kozin
Andrii Kozyriev
Victor Krasnobayev
Petro Kravets
Yuri Kubrak
Vladislav Kuchanskyi
Nataliia Kulykovska
Alexander Kutsenko
Lyudmyla Kuzmych
Alexander Kuznetsov
Kateryna Kuznetsova
Maryna Kuzniatsova
Valeriy Lakhno
Elizaveta Lazareva
Serhii Leoshchenko
Alexander Litvinov
Valerii Lovkin

Vasyl Lytvyn
Nataliia Maksyshko
Heorhii Marhiiev
Gennadii Martynenko
Iryna Mashkina
Oksana Mazurova
Sergii Minukhin
Victoria Miroshnichenko
Andrii Morozov
Iryna Myronets
Olena Nazarova
Oleg Nikonov
Pavlo Nosov
Roman Odarchenko
Dmytro Orlovskiy
Artem Ostapenko
Sergey Ostapov
Anatolii Pashko
Andriy Pertsov
Mykhailo Poliakov
Nikolay Poluyanenko
Igor Povkhan
Olha Pronina
Maksym Proskurin
Nataliya Pryadko
Oleg Pursky
Lyudmyla Rozova
Kostiantyn Rudakov
Oleg Rudenko
Nickolay Rudnichenko
Kseniia Serdiuk
Anastasiia Serhiienko
Alina Shafronenko
Mariya Shapovalova
Anton Sharypanov
Volodymyr Sherstjuk
Anatoly Shevchenko
Igor Shevchenko
Alexandr Shimanovsky
Mariya Shirokopetleva
Tatiana Shmeleva
Ihor Shostko
Ievgen Sidenko

Maxim Sidorov
Andrii Siriak
Stepan Skrupsky
Alexey Smirnov
Oleksii Smirnov
Natalya Sokolova
Alexander Sosnitsky
Hanna Stadnikova
Alexander Stepanenko
Grygorii Steshenko
Iryna Strubyt'ska
Bohdan Sus
Anton Sysoienko
Dmytro Sytnikov
Kateryna Ternova
Nataliia Titova
Kostiantyn Tkachenko
Oleksandr Tkachenko
Alexander Tonkoshkur
Andriy Topalov
Iryna Tvoroshenko
Vladimir V. Barannik
Tetiana Vakaliuk
Fastovec Valentyna
Morozov Viktor
Vitaliy Vishnevskiy
Oleksii Vodka
Mariia Voronenko
Vladimir Vychuzhanin
Victoria Vysotska
Alina Yanko
Georgiy Yaskov
Hanna Yegoshyna
Oleksandra Yeremenko
Dmitry Zaitsev
Ievgen Zaitsev
Artur Zaporozhets
Irina Zeleneva
Maryna Zharikova
Vladimir Zhikharevich
Inna Zhulkovska
Oleg Zhulkovskiy
Iryna Zhuravska

Organizers



National University "Zaporizhzhia Polytechnic",
Zaporizhzhia, Ukraine

<http://www.zp.edu.ua>



Department of Software Tools
of the National University "Zaporizhzhia Polytechnic",
Zaporizhzhia, Ukraine

<http://pz.zntu.net>

Partners



International project "Innovative Multidisciplinary Curriculum in Artificial Implants for Bio-Engineering BSc/MSc Degrees" (BIOART) of Erasmus+ Programme co-funded by the European Union (Ref. No. 586114-EPP-1-2017-1-ES-EPPKA2-CBHE-JP).



International project " "Internet of Things: Emerging Curriculum for Industry and Human Applications" (ALIoT) of Erasmus+ Programme co-funded by the European Union (Ref. No. 573818-EPP-1-2016-1-UK-EPPKA2-CBHE-JP).



International project «Virtual Master Cooperation Data Science (ViMaCs)» of DAAD.

Algorithmic Support for the Detection Characteristics Improving of the Monitoring Object

Elena Chernetsova¹[0000-0001-5805-3111] and Anatoly Shishkin²[0000-0003-1992-5663]

¹ Russian State Hydrometeorological University, Metallistov Av.,3, St-Peterburg, 195027,
Russia

chernetsova@list.ru

² Russian State Hydrometeorological University, Metallistov Av.,3, St-Peterburg, 195027,
Russia

an.dm.shishkin@mail.ru

Abstract. The paper presents: an object detection algorithm for coherent reception of signals coming from a monitoring network consisting of several sensors; an algorithm for detecting an extended object by analog signals of sensors of a monitoring network. These algorithms use statistics that take into account the most stable features of the distribution of the source data. They can be implemented in an automated decision support system. At the same time, decisions on the detection of a monitoring object made by an automated system will be more reliable

Keywords: Monitoring, Object, Sensor, Signal, Algorithm, Detection.

1 Introduction

To carry out environmental monitoring, it is necessary to conduct continuous observations over time, based on a well-thought-out distribution of measuring instruments in space, for which it is necessary to use a stationary distributed multi-sensor remote monitoring system [1]. It should work efficiently, preferably at a real time scale. Efficiency also means reducing the time frame for deciding on the classification of the observed object. Therefore, it is necessary to automate not only the data collection process, but also the classification algorithms of the monitoring object in order to attract the attention of the human operator only to objects that actually threaten the ecological state of the observed area and even at the stage of automated data processing to weed out objects that do not threaten the ecological state of the zone of responsibility. A stationary network of stations included in the monitoring system requires the availability of communication channels with a Monitoring Control Point (MCP) [2]. Laying a cable communication network is often unprofitable. Therefore, for communication purposes it is necessary to use a radio channel or satellite communication [3]. Since the sensors of the monitoring network receive energy from the batteries, in order to save energy in the monitoring network, it is often justified not to pre-process the signal on the sensor, but to send analog signals to the MCP, which is

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

charged with processing the sensor signals and detecting the monitoring object [4]. Information exchange over the radio channel raises the problem of detecting an analog signal with an unknown law of fluctuations against the background of noise with an unknown distribution [5]. To solve this problem, in this paper, it is proposed to develop the following algorithms:

- an algorithm for detecting a monitoring object during coherent reception of signals coming from a monitoring network consisting of several sensors;
- the sample size for detecting the object of the analog signals of the sensors of the monitoring network.

2 Theoretical Analysis

2.1 The most powerful accordingly to the signal-to-noise ratio criterion algorithm for processing spatially distributed data from a monitoring network consisting of several sensors

Let us consider the problem of coherent detection of a signal from an object distributed in N resolution elements, which are sensors of a monitoring network. It was shown in [6] that the detector calculates the likelihood ratio:

$$l(X) = \frac{1}{\binom{N}{k}} \left(\frac{\gamma_0}{\gamma_1} \right)^k \sum_{n=1}^N \exp \left[\left(\frac{1}{2\gamma_0} - \frac{1}{2\gamma_1} \right) \sum_{n=1}^k x_n^2 \right]^2 \quad (1)$$

where x_n - detector output envelope samples, $n = 1, 2, \dots, N$

γ_0 and γ_1 - signal variances received from $(N - k)$ sensors, that did not fix the object and k sensors, fixed object accordingly.

From equation (1) we can see, that detector which is most powerful accordingly to the signal-to-noise ratio criterion can be implemented by a rather complex circuit, and, in addition, for its implementation a priori information is required about the parameters of signal (γ_1) and noise (γ_0), which, as a rule, in real monitoring conditions are unknown. Therefore the rule (1) characterizes the potential for detecting an object and cannot be realized in many practical cases.

It is necessary to develop an most powerful by signal / noise criterion algorithm for coherent detection of a signal from a monitoring object received from $(N - k)$ sensors on the background of noise interference provided that the signal and noise parameters, as well as the position of the fixed object k sensors among N sensors of the monitoring network are a priori unknown. Detection is formulated as the statistical task of testing general linear hypotheses [7-10] and the rule is found in the class of so-called invariant rules [11].

We use the following premises:

1. There are statistically independent radio pulses sent by $N(N \gg 1)$ sensors. In the absence of the object of observation, these pulses have the same average power. The law of the distribution of the noise background is considered normal.
2. In the presence of an object of observation, the resulting fluctuation in resolution is the additive sum of the signal with unknown amplitude $\xi_m (m = 1, 2, \dots, k)$ and Gaussian noise with unknown variance σ^2 . Coherent processing is assumed. Independent voltage samples are taken at the output of the linear path of the MCP receiver at time instants following the resolution interval. $x_n (n = 1, 2, \dots, N)$.
3. Processing is carried out during the p periods of the signal, so that each reference element n will correspond to a sample vector $(x_{n1}, \dots, x_{np})$ with multidimensional normal probability density

$$g(X_n) = \frac{\sqrt{|A|}}{(2\pi)^{p/2}} \exp \left[-\frac{1}{2} \sum_{i=1}^p \sum_{j=1}^p a_{ij} (x_{ni} - \xi_{ni})(x_{nj} - \xi_{nj}) \right]. \quad (2)$$

The mean values and the covariance matrix of the vector are determined from the expressions $E(x_{ni}) = \xi_{ni}$; $E(x_{ni} - \xi_{ni})(x_{nj} - \xi_{nj}) = \sigma_{ij}$; $\sigma_{ij} = A^{-1}$, where E - is the sign of mathematical averaging, and $\xi_n = 0$, if $n \in (N-k)$, and $\xi_n > 0$ at $n \in k$. It is also believed that the matrix $A = (a_{ij})$ - is common to all vectors N , having dimension p , but unknown.

The challenge is that by sample

$$X = \begin{pmatrix} x_{11} & . & . & . & x_{1p} \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ x_{N1} & . & . & . & x_{Np} \end{pmatrix}$$

determine the presence or absence of a signal about the existence of a monitoring object. Matrix X consists of p column vectors $(x_{1i}, \dots, x_{Ni})$, and each such vector has its own mean value vector $\bar{\xi}_i = (\xi_{1i}, \dots, \xi_{Ni})$.

Given the accepted assumptions, the task of detection is to test complex hypotheses H_0 and H_1 regarding parameters $\bar{\xi}_i$ and A .

$$H_0 : \bar{\xi}_i = 0 \quad H_1 : \bar{\xi}_i > 0 \quad i = 1, 2, \dots, p \quad A \text{ is unknown} \quad (3)$$

Hypothesis testing (3) fits into the scheme of testing multidimensional linear hypotheses. As follows from the general theory [12], principles of invariance and suffi-

ciency allows you to reduce the sample X when testing hypotheses (3) to maximally invariant statistics of the form

$$T = \frac{\sum_{i=1}^p \sum_{j=1}^p \overline{N x_i x_j}}{\sum_{n=1}^N (x_{ni} - \bar{x}_i)(x_{nj} - \bar{x}_j)} \quad (4)$$

and the set of parameters $\bar{\xi}_i$ and (a_{ij}) - to maximal invariant

$$\delta^2 = N \sum_{i=1}^p \sum_{j=1}^p a_{ij} \eta_i \eta_j \quad (5)$$

In expressions (4), (5) $\bar{x}_{i(j)} = N^{-1} \sum_{n=1}^N x_{ni(j)}$; $\eta_{i(j)} = E(\bar{x}_{i(j)})$. Numerator of the formula (4) has off center χ_p^2 - off-center distribution with the noncentrality parameter δ^2 and p degrees of freedom, and the denominator has central distribution $\chi^2_{(N-p)}$, so the statistics $(N-p)T/p$ has off-central F distribution with p and $(N-p)$ degrees of freedom and with the noncentrality parameter δ^2 .

Regarding the parameter δ^2 of F - distribution initial hypotheses (3) can now be formulated as follows :

$$H_0 : \delta = 0; \quad H_1 : \delta > 0 \quad (6)$$

Using the method of constructing optimal rules [13], it can be shown that the most powerful invariant criterion for testing hypotheses (6) has a critical region of the form

$$T > C. \quad (7)$$

Threshold level C determined by the given probability of false alarm α from the condition

$$\int_C^\infty F_{p,(N-p)}(y) dy = \alpha \quad (8)$$

where $F_{p,(N-p)}$ -is central F distribution with p and $(N-p)$ degrees of freedom.

The expressions (4), (7) determine the functional scheme of the detector with completely unknown correlation properties of vectors $(x_{n1}, \dots, x_{np})$. For practical implementation, expressions (7), (8) can be concretized, for example, in the case of the absence of inter-period correlation. In this case, the discovery rule and parameter δ^2 take the form

$$\sum_{i=1}^p \frac{N \bar{x}_i^2}{\sum (x_{ni} - \bar{x}_i)^2} > C, \quad (9)$$

$$\delta^2 = \sum_{i=1}^p \bar{q}_i, \quad (10)$$

where $\bar{q}_i = (\sum_{n=1}^N \xi_n)^2 / N\sigma^2$ - is the average for all N signal-to-noise ratio for one observation period. The detector efficiency is determined by the power function of rule (7), (8), which shows the dependence of the probability of correct detection on the parameter δ^2 . It can be calculated directly from off-center tables of F - distribution [14].

2.2 A most powerful according to the signal-to-noise ratio criterion algorithm for detecting an extended object by analog signals of monitoring network sensors

To develop an algorithm for classifying an extended object (for example, classifying the observed water surface as clean or polluted by oil emissions) using a distributed multisensor geographic information system, suppose:

- The central post decides to detect / not detect an object (contamination) based on signals received from N sensors under the same observation conditions ;
- The resulting radio signal of each sensor is the additive sum of the non-fluctuating signal of unknown amplitude $\xi_i^{(j)} \geq 0$ ($j=1,2$ - is numbers of object - e.g. clean water surface and dirty water surface, $i \in N$) and Gaussian noise with unknown dispersion σ_i^2 . At the output of the receiver's linear path, the amplitude samples $x_i^{(j)}$ are taken for the signal of each sensor.
- Observation of objects is carried out for some time T , during which readings for the signal of each sensor n are taken. Thus, for each object, the sample space is represented as n sample vectors $x_\alpha^{(1)} = (x_{\alpha 1}^{(1)}, \dots, x_{\alpha n}^{(1)})$; $x_\alpha^{(2)} = (x_{\alpha 1}^{(2)}, \dots, x_{\alpha n}^{(2)})$; $\alpha = \overline{1, n}$.

Vectors $x_\alpha^{(1)}$ and $x_\alpha^{(2)}$ have a normal probability density

$$g(x_\alpha^{(j)}) = \frac{\sqrt{A}}{(2\pi)^{N/2}} \exp\left[-\frac{1}{2} \sum_{i=1}^N \sum_{k=1}^N \alpha_{ik} (x_{\alpha i}^{(j)} - \xi_{\alpha k}^{(j)})\right] \quad (11)$$

The mean values and elements of the covariance matrix are determined from the expressions

$E(x_{\alpha k}^{(j)}) = \xi_{\alpha k}^{(j)}; E(x_{\alpha i}^{(j)} - \xi_{\alpha i}^{(j)})(x_{\alpha k}^{(j)} - \xi_{\alpha k}^{(j)}) = \sigma_{ik}; (\sigma_{ik}) = A^{-1}$, where E – is the sign of mathematical averaging. We consider that the matrix $A = (\alpha_{ik})$ is common to vectors $x_{\alpha}^{(1)}$ and $x_{\alpha}^{(2)}$, but its elements are unknown.

The classification task is by the sample $x_{\alpha}^{(1)}, x_{\alpha}^{(2)}, \alpha = \overline{1, n}$ determine whether objects belong to the same class or belong to different classes.

Based on the assumptions made, this problem can be formulated as two hypotheses - 1. objects are of the same type; 2. objects are of the different type:

$$H_0: \bar{\xi}_i^{(1)} = \bar{\xi}_i^{(2)}; H_1: \bar{\xi}_i^{(1)} \neq \bar{\xi}_i^{(2)} \text{ for all } i = \overline{1, N} \quad (12)$$

In expression (12) the parameter $\bar{\xi}_i^{(j)}$ – is matrix column having dimension $(n \times 1)$ with elements $(\xi_{1i}^{(j)}, \dots, \xi_{ni}^{(j)})$.

As follows from the general theory [15], the principles of invariance and sufficiency allow us to reduce the sample space $x_{\alpha}^{(1)}, x_{\alpha}^{(2)}, \alpha = \overline{1, n}$, when testing hypotheses (12) for maximally invariant statistics (MI) of the form

$$T = \sum_{i=1}^N \sum_{k=1}^N \frac{n(\bar{x}_i^{(1)} - \bar{x}_i^{(2)})(\bar{x}_k^{(1)} - \bar{x}_k^{(2)})}{\sum_{\alpha=1}^n (x_{\alpha 1}^{(1)} - \bar{x}_1^{(1)})(x_{\alpha k}^{(1)} - \bar{x}_k^{(1)}) + \sum_{\alpha=1}^n (x_{\alpha i}^{(2)} - \bar{x}_i^{(2)})(x_{\alpha k}^{(2)} - \bar{x}_k^{(2)})} \quad (13)$$

and the parameter space is to MI

$$\psi^2 = \sum \sum \frac{n}{2} \alpha_{ik} (\eta_i^{(1)} - \eta_i^{(2)})(\eta_k^{(1)} - \eta_k^{(2)}) \quad (14)$$

In expressions (13) and (14) is marked: $\bar{x}_k^{(j)} = n^{-1} \sum_{\alpha=1}^n x_{\alpha k}^{(j)}; \eta_k^{(j)} = E(\bar{x}_k^{(j)}); k = \overline{1, N}$.

It can be shown that there is uniformly the most powerful (UMP) criterion for testing hypotheses (12), (12), which rejects the hypothesis H_0 in case if

$$T > C, \quad (15)$$

where C – is the threshold constant.

The constant C should be determined from the condition that under the hypothesis H_0 ($\psi^2 = 0$) the probability of the fulfillment of condition (15) was no more than a certain predetermined significance level β . Whereas statistics $\nu_1 / \nu_2 T$

under the hypothesis H_0 has central F distribution with $\nu_1 = N$ and $\nu_2 = (2n - N - 1)$ degrees of freedom [16], the constant C can be found from the expression

$$\int_c^\infty F_{\nu_1, \nu_2}(\omega) d\omega = \beta \quad (16)$$

The rule (15) can be specified for the case when the matrix A is diagonal. In this case, it has the form

$$\sum_{i=1}^N \frac{n(\bar{x}_i^{(1)} - \bar{x}_i^{(2)})^2}{\sum_{\alpha=1}^n (x_{\alpha 1}^{(1)} - \bar{x}_i^{(1)})^2 + \sum_{\alpha=1}^n (x_{\alpha i}^{(2)} - \bar{x}_i^{(2)})^2} > C_1 \quad (17)$$

where $C_1 = CN / (2n - N - 1)$.

From the expressions (13) and (17) it can be seen that $\bar{x}_i^{(j)}$ ($j=1,2; i=\overline{1,N}$) - are maximum likelihood estimates for parameters $\xi_i^{(j)}$, calculated for the sensor N signals for the first and second objects, and the value in the denominator is the sum of the parameter σ_i^2 estimates calculated for the signal of the first and second object of i -th sensor. Thus, to distinguish between objects, it is necessary to estimate the amplitudes of the N sensor signals, calculate the square of the distance between the parameters of the signals of the classified objects by the sensors of the same name, and sum them with weights inversely proportional to the noise variance. The amount received is compared with a threshold, in case of exceeding which a decision is made on whether the objects belong to different classes.

Expression (15) can also be used to detect a distributed object, if we put $x_\alpha^{(2)} = 0; \alpha = \overline{1,n}$. Formula (17) in this case takes the form [17]:

$$\sum_{i=1}^N \frac{n\bar{x}_i^2}{\sum_{\alpha=1}^n (x_{\alpha i} - \bar{x}_i)^2} > C_2, \text{ где } C_2 = CN / (n - N - 1). \quad (18)$$

Considering that under the hypothesis H_1 statistics T has off-central F distribution with off-center parameter ψ^2 and ν_1, ν_2 degrees of freedom, the probability of correctly distinguishing between objects is determined by the expression [18]:

$$P(T > C) = \int_0^\infty F_{\nu_1, \nu_2}(\omega, \psi^2) d\omega \quad (19)$$

and can be calculated according to the tables of off-central F distribution [19].

3 Results

Figure 1 shows the curves characterizing the effectiveness of the detector of oil pollution of the water surface depending on the resolving power of the network of contact sensors constructed in accordance with expressions (4), (7). Characteristics calculated for false alarm probability value $\alpha = 10^{-2}$ and the number of received signal periods $p = 2$ provided that the value of the signal-to-noise ratio averaged over all N sensors for one observation period $\bar{q}_i$ is independent of resolution (uniform distribution of translational buoys (contact sensors) along the length of contamination). For comparison, the same figure shows the power function of the potential most powerful rule (MP) of coherent detection of a known signal [20] in the presence of only one sensor ($N = 1$).

It can be seen from the figure 1 that ignorance of the noise and signal levels in the decision elements leads to losses in the signal-to-noise ratio. However, with increasing resolution, the detector's efficiency increases. This is due to the fact that the increase allows a more accurate assessment of noise and signal levels. So, when $N = 8$ the loss in the signal-to-noise ratio is ~ 4 dB, and when $N = 22$ - less than 1 dB.

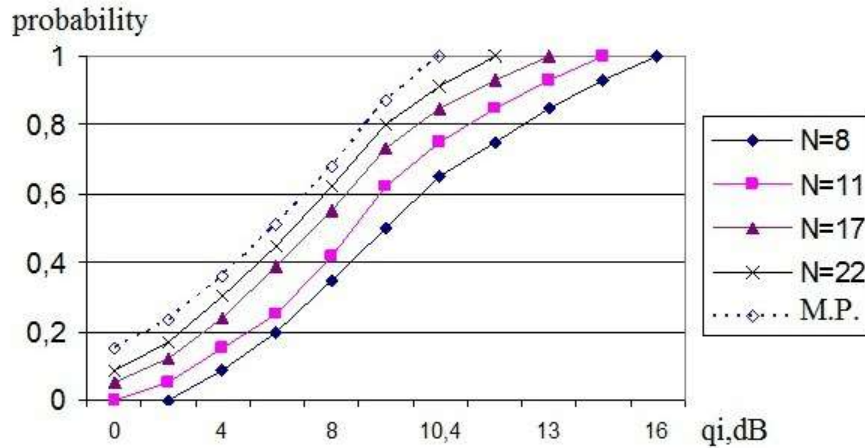


Fig. 1. The probability of detecting oil pollution of the water surface by signals received from a network of contact sensors

Figure 2 shows the dependences of the probability $P(N)$ of correct distinguishing between two objects, calculated by the formula (19), for different values of the signal sample size n .

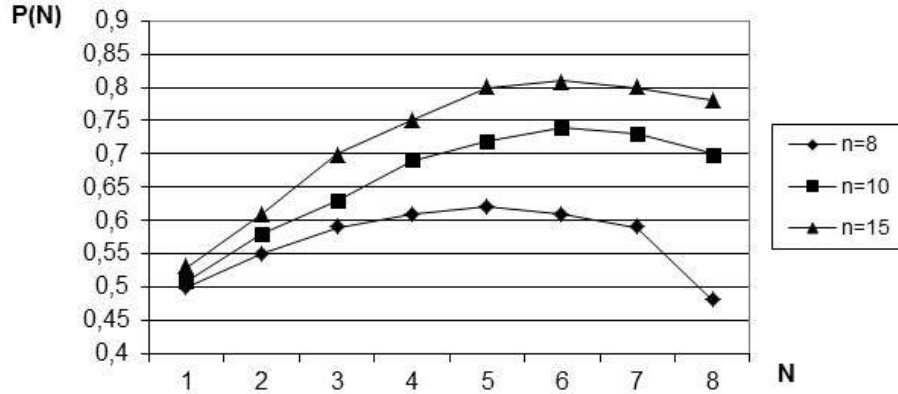


Fig. 2. The dependence of the probability of distinguishing objects $P(N)$ for a different number of sensors in the monitoring system (N) for several values of the signal sample size (n)

In this case, the noncentrality parameter ψ^2 of F distribution was assumed constant, independent of the number N of sensors in the monitoring system. As can be seen from Figure 2, the dependences have an optimum in the probability of distinguishing between objects, and its position depends on the size of the sample n . The presence of an optimum and its position are apparently due to the following reasons. On the one hand, with an increase in the number of sensors in the monitoring system, the difference in signals increases, that is, the “distance” between objects in the parameter space increases. Let us explain what was said by the following example. Let the objects have the same area, but a different distribution of them among the sensors. The value of the parameters of the amplitudes of the signals from the first and second objects $\xi_i^{(1)}$ and $\xi_i^{(2)}$ for $N=3$ is presented in table form 1.

At $N=1$, the distance in the parameter space between objects A and B is $(\sum_{i=1}^3 \xi_i^A - \sum_{i=1}^3 \xi_i^B)^2 = 9 - 9 = 0$ and it's not possible to distinguish between them. At the same time,

for $N=3$ we get $\sum_{i=1}^3 (\xi_i^A - \xi_i^B)^2 = (3-1)^2 + (4-5)^2 + (2-3)^2 = 6$, i.e. the difference in parameters is significant. On the other hand, with a decrease in the number of sensors in the monitoring network, the correlation between the signals of objects of various classes increases. Moreover, the accuracy of parameter estimates can be improved by increasing the accumulation time, i.e., increasing the size of the sample n .

Table 1. The value of the parameters of the amplitudes of the signals from the first and second objects

Object	i	1	2	3	
A	ξ_i^A	3	4	2	$\sum_{i=1}^3 \xi_i^A = 9$
B	ξ_i^B	1	5	3	$\sum_{i=1}^3 \xi_i^B = 9$

4 Conclusion

The proposed algorithm in the sense of signal-to-noise ratio for processing spatially distributed data coming from a monitoring network consisting of several sensors with the following practically important properties: a) does not depend on a priori unknown parameters σ^2 and ξ_n ($n = 1, 2, \dots, N$) and provides a constant probability of false alarm at any noise level; b) is invariant to the location of k sensors that recorded the object and $(N-k)$ sensors that have not fixed the object, among N sensors of the monitoring network; c) has the highest probability of correct detection, depending on the average signal-to-noise ratio and for large $N > p$ close to potential.

The proposed algorithm for detecting an extended object by the analog signal of sensors of the monitoring network can be used to identify objects if, for example, as $x_i^{(2)}, i = \overline{1, N}$, a priori estimates of the parameters of the recognized object are used.

The practical significance of the results lies in the development of analog signal detection algorithms that are resistant to changes in the signal-to-noise ratio in the communication channels of the sensors of the monitoring network with a monitoring and control post. Algorithms can be implemented programmatically using various programming languages and used to automate the process of classifying monitoring objects at a monitoring and control point.

References

1. Krapivin, Vladimir & Shutko, Anatolij.. Information Technologies for Remote Monitoring of the Environment. 1st edn ,Springer, Berlin Heidelberg (2012).
2. R.R.Brooks, P.Ramanathan, A. Sayeed Distributed target tracking and classification in sensor networks IEEE Signal Processing Magazine, vol.19. no.2, pp.17-29 (2002)
3. Wu and J. Hu: Design and Implementation of Production Environment Monitoring System Based on GPRS-Internet. In 4th International Conference on Genetic and Evolutionary Computing, pp. 818-821. Shenzhen (2010).
4. Gabriel Nallathambi, Jose C. Principe Theory and Algorithms for Pulse Signal Processing . CoRR abs/1901.01140 (2019)

5. Dan L., Wong K., Hu Y., Sayeed A. Detection, classification and tracking of targets in distributed sensor networks *IEEE Signal Processing Magazine*, vol.19, no.2, pp.17-29 (2002).
6. P. Venkatasubramanian, S. Adireddy, L.Tong Sensor networks with mobile access: Optimal random access and coding, *IEEE J.Sel.Areas Commun. (Special Issue on Sensor Networks)*, vol.22, pp.1058-1068 (2004).
7. Keith A. McNeil, Isadore Newman, Francis J. Kelly Testing Research Hypotheses with the General Linear Model, 1st edn, SIU Press, Illinois (1996).
8. Lei Song, Hongchang Hu and Xiaosheng Cheng Hypothesis Testing in Generalized Linear Models with Functional Coefficient Autoregressive Processes, *Hindawi Publishing Corporation Mathematical Problems in Engineering*, pp.2-18 (2012).
9. R. Azrak and G. M  lard, "Asymptotic properties of quasi-maximum likelihood estimators for ARMA models with time-dependent coefficients," *Statistical Inference for Stochastic Processes*, vol. 9, no. 3, pp. 279–330 (2006).
10. R. A. Maller, "Asymptotics of regressions with stationary and nonstationary residuals," *Stochastic Processes and Their Applications*, vol. 105, no. 1, pp. 33–67 (2003).
11. Y. Bai, W. K. Fung, and Z. Zhu, "Weighted empirical likelihood for generalized linear models with longitudinal data," *Journal of Statistical Planning and Inference*, vol. 140, no. 11, pp. 3446–3456 (2010).
12. L. Fahrmeir and H. Kaufmann, "Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models," *The Annals of Statistics*, vol. 13, no. 1, pp. 342–368 (1985).
13. F. Carsoule and P. H. Franses, "A note on monitoring time-varying parameters in an autoregression," *International Journal for Theoretical and Applied Statistics*, vol. 57, no. 1, pp. 51–62 (2003).
14. A. Logothetis, A. Isaksson On sensor scheduling via information theoretic criteria, in *Proc. Amer. Control Conf.*, San Diego, CA, pp.2402-2406 (1999).
15. W. A. Fuller, *Introduction to Statistical Time Series*, John Wiley & Sons, New York, NY, USA, 2nd edition (1996).
16. J. D. Hamilton, *Time Series Analysis*, 1st edn , Princeton University Press, Princeton, NJ, USA (1994).
17. Sergio Albeverio, Raphael Hoegh-Krohn, Jens Erik Fenstad, and Tom Lindstrom *Non-standard methods in stochastic analysis and mathematical physics*, 1st edn , Academic Press, Orlando (1990)
18. F. Zhao, J. Liu, J. Liu, L. Guibas, and J. Reich, "Collaborative signal and information processing: An information directed approach," *Proceedings of the IEEE*, vol. 91, no. 8, pp. 199–1209 (2003).
19. The F Distribution and the F-Ratio, <https://openstax.org/books/introductory-statistics/pages/13-2-the-f-distribution-and-the-f-ratio>, last accessed 2020/01/21
20. Data communication systems and their performance :proceedings of the IFIP TC6 Fourth International Conference on Data Communication Systems and Their Performance, Barcelona, Spain (1990).

Decision Support During the Vessel Control at the Time of Negative Manifestation of Human Factor

Pavlo Nosov¹[0000-0002-5067-9766], Serhii Zinchenko¹[0000-0001-5012-5029], Ihor Popovych²[0000-0002-1663-111X], Michail Safonov³[0000-0001-9742-8270], Ihor Palamarchuk¹[0000-0001-9161-1592], Valeriia Blakh²[0000-0001-6349-1183]

¹ Kherson State Maritime Academy, Ushakova avenue 20, Kherson, 73000, Ukraine
pason@ukr.net, srz56@ukr.net, ihor3107@meta.ua

² Kherson State University, Universytetska st., 27, Kherson, 73003, Ukraine
ihorpopovych999@gmail.com, blah@ksu.ks.ua

³ Odessa National Polytechnic University, Nebesnoy sotni st., 23, Kherson, 73013, Ukraine
nemko85@gmail.com

Abstract. The simulation and program implementation approaches are presented in the tasks of determining the periods of loss of control due to the fault of the human factor in the operation of marine transport while carrying the navigation watch. Experiments have been carried out confirming the problem of the negative influence of the human factor on the example of navigational tasks in the Bosphorus and Hong Kong Straits. Automated tools have been developed to identify hazardous areas for navigation on location mapping by analyzing the accident geolocation in the Hong Kong Strait and the Bosphorus, which is a decision support system for emergency situations. A software module has been developed that makes it possible to identify the time periods for the manifestation of the human factor of the navigator by analyzing the ECDIS database in real time. Mathematical models of the triggering of the vessel automated course alteration system (VACA) and the actions of the navigator when controlling/navigating the vessel in difficult maneuvering zones are proposed.

Keywords. human factor, decision support, automated divergence, model of behavior, emergency situations

1 Introduction

During the passage planning, preliminary and executive plotting is performed in ECDIS, but the navigator cannot be fully confident in its effectiveness [1-2]. A number of factors associated with oncoming traffic and random obstacles, weather conditions and the composition of the watch bring in adjustments which leading to emergency situations and catastrophic consequences [3-4]. A lot of research is aimed at analyzing risks due to the human factor at the time of maneuvers and divergences, at obtaining a priori probability of a catastrophe occurring depending on the situation [5-6]. In the most difficult locations, such as Hong Kong, Singapore, the Bosphorus and Dardanelles straits require the help of experienced pilots, but even their preparation

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

also includes the full range of possible situations leading to disasters [7-9]. Despite a lot of research in such area, this study is relevant because the human factor is still the most significant cause of collisions and accidents in maritime transport [10-13].

Analysis of the literature showed that the main problem is the lack of models for identifying the manifestation of the human factor in the early stages. The complexity of constructing such models is that each navigator responds to a particular situation differently, as a result of which it is difficult to predict at what point a “human error” occur [14]. Generalized statistics do not allow for the effective monitoring of ship management/handling processes with a high degree of reliability.

The purpose of the article is to develop formal models and software for determining periods of loss of control due to the human factor of the navigator and methods for switching to automated vessel control in emergency situations.

To achieve the goal of the article it is necessary:

1. To develop means of automated identification and display on the cartographic panel of the most dangerous navigation zones within the framework of the expert decision support system during emergency situations.
2. Determine periods of loss of ship control dependent on individual stable reactions of navigators, affecting control parameters and dependent on location and complexity by automated analysis of ECDIS databases.
3. Carry out a mathematical simulation of the human factor manifestation in the time of vessel management/handling when the vessel automated course alteration (VACA) is triggered in areas of high attention.
4. Develop an algorithm and software for emergency transition to the automated control mode of the vessel in order to automatically diverge with target vessels in critical situations.

The solving of this task will allow at a qualitatively new level to approach the solution of the problem of intellectual data mining in the management/handling of the vessel and prevent catastrophic situations in maritime transport.

2 Materials and method

To build adequate mathematical models, behavioral responses of navigators in the Hong Kong Strait were analyzed to identify signs of occurrence of abnormal and catastrophic situations due to the human factor of navigators.

In order to obtain reliable results for the experiments, the “TRANSAS NAVIGATIONAL SIMULATOR NTPRO 5000” navigation simulator was used to enable utilize information navigation panels of Navi-Sailor 4000 ECDIS Multifunction Display (ECDIS, Radar, Conning) at Kherson State Maritime Academy, Ukraine.

A preliminary analysis of the location showed that the entire passage of Hong Kong Strait is difficult due to several points [15]. The main point is a lot of cross over ways by different type's of fishing boats, special service fleet (tugboats, floating cranes and etc.), coastal fleet and ferries. Additional difficulties which occur during passage are weather condition [16-18].

When analyzing accidents in the Hong Kong Strait, areas near the narrowness of Ma-Wan Island were clearly distinguished (Fig. 2). At Fig. 2(a) stated main traffic route to the East from Ma Wan island which common use by all kind of vessels and applicable for all large merchant vessel of different types. At Fig. 2(b) mentioned traffic route to the West from Ma Wan Island restricted by shallows, bridge high and in use by small and coastal vessels only.

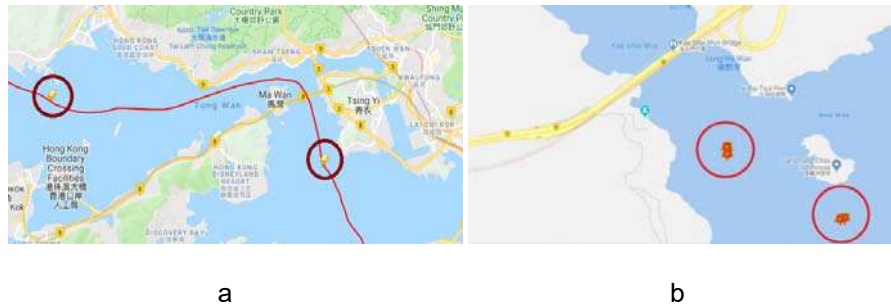


Fig. 2. a Identification trajectories of routes, **b** Determination of collision sites

Also analyzed the routes of the vessel in the Bosphorus Strait for collisions (Fig. 3 a). Bosphorus Strait was divided into zones according to the levels of difficulty of passage. Those zones where up to 2 collisions were recorded within a radius of 0.5 mm were considered simple (green); from 3 to 5 - medium difficulty (yellow); more than 5 - complex (red).

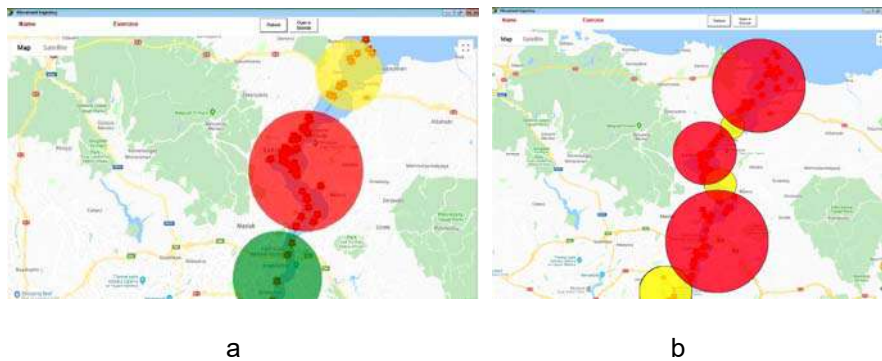


Fig. 3. a In the heavy traffic, **b** In the reduced traffic intensity

Then the process of passing the location was changed – weather conditions deteriorated, which resulted in restricted visibility, while the traffic intensity was reduced by 32% (Fig. 3 a).

As a result of the experiment, an increase in the area of the complex zone was recorded from 37% to 83%. This suggests that for navigators the conditions of restricted visibility are more dangerous than the intensity of the traffic.

There is a need for timely identification of the moment of loss of control over the vessel. To do this, must clearly identify and classify the factors that most often lead to navigation errors.

Among the numerous mistakes made by navigators, there is a clear tendency to their repeatability, which makes it possible to classify the main types of errors by category: the multifactorial situation, false identification, stress, habits, fatigue.

Taking into account this classification, it becomes possible to create a computerized system for monitoring the behavior of the navigator, which will help to identify in time the loss of control over the vessel by analyzing the database of control parameters (Fig. 4).

To identify the moment of loss of control, a software / hardware module was developed that allows tracking control parameters such as the speed and rotation of the steering wheel. Distinctive features are the frequent change of positions, such as rudder shifts for a short time. This fact testifies to the impossibility of taking a firm decision and a strategy for maneuvering the diverge with obstacles. Figure 6 shows a graph indicating a clear loss of control over the situation by the navigator in 69 seconds. The system consists of several modules, each of which will be responsible for controlling certain factors. Information from all modules will be accumulated in a single database, and then processed by the central system, which is responsible for making certain decisions aimed at assisting to the navigator.

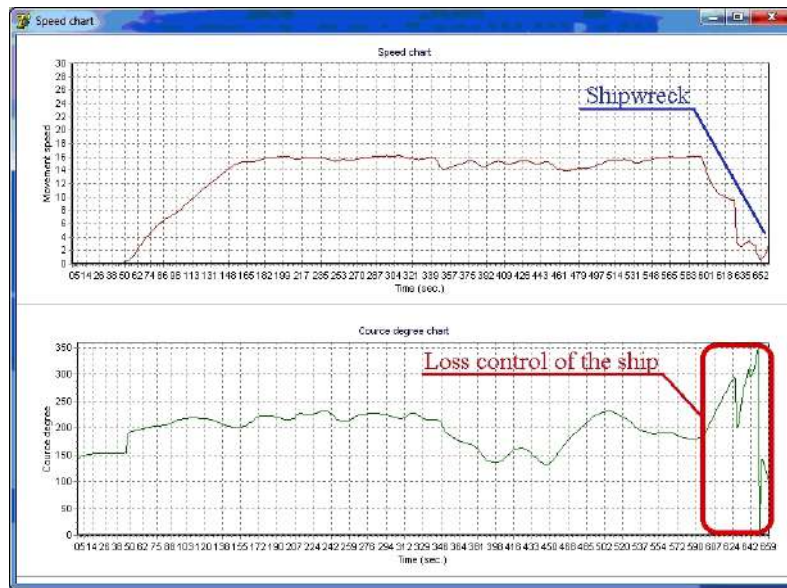


Fig. 4. ECDIS database analysis and management graph analysis

Analysis of the vessel's trajectory through the Bosphorus Strait from the Black Sea shows that incorrect decision making during the divergence maneuver led to vessel crossing the dangerous isobath near the coastline of Rumelihisar, Turkey (Fig. 5 a, b).



Fig. 5. a Satellite view, **b** Display of trajectory on the map

The formalization of control tasks at the time of loss of control over the vessel in the framework of mathematical models can lead to the formulation of the next class of problems of managing complex coordination processes in systems with latency responses of the navigator [20,21].

The process of loss of control over a ship at the moment of an oversupply of information factors may well be described by a system of differential equations [22] with delays containing VACA as a function of time:

$$\frac{ds}{dt} = f(s(t), s(t - \tau_1), \dots, s(t - \tau_s), d(t)) \quad (1)$$

Where condition $s(t) = (s_1(t), \dots, s_n(t))$, and $d(t) = (d_1(t), \dots, d_m(t))$, $\tau_i > 0, i = \overline{1, z}$ is the navigator response to an emergency situation z , $d = d(t)$ - VACA, which takes into account the influence of considered corrective actions process. Corrective actions imply a transition to VACA or the impact of a decision support system at the time of exacerbation of the situation while controlling/handling the vessel [23].

Then system (1) can be considered in conjunction with the following initial conditions:

$$s(t_0) = s_0, s(t) = \varphi(t), t \in \left[t_0 - \max_{1 \leq i \leq z} \tau_i, t_0 \right] \quad (2)$$

Where t_0 - the moment of the beginning of VACA.

The class of functions [24] is considered as admissible ranges of VACA, on the range of values D , on which additional constraints related to the specifics of the task can be imposed. These restrictions may take into account the location mapping, the maneuverability of the vessel and its parameters, weather conditions, ice cover, environmental conditions, etc. [25]. Then in the general case the condition is considered:

$$d(t) \in D \subset Y^m \quad (3)$$

Where D is a compact set in Y^m . It should be noted that the VACA process can be broken down into stages due to the difficulty of getting out of this situation.

Given that the main goal of VACA is formalized in the framework of (1), then a number of conditions must be fulfilled:

$$\eta_{q-i+1}(d) \geq \eta, i = \overline{1, y}, y \leq q \quad (4)$$

Here $\eta_{q-i+1} = \eta_{q-i+1}(d), i = \overline{1, y}$ - the length of the time intervals relative to the stages of VACA $[t_0, T]$, determined from the conditions:

$$c(s(t)) \leq 0, t \in [t_1^i - \eta_i, t_1^i], i = \overline{1, q} \quad (5)$$

Where $t_0^i = t_1^i - \eta_i, t_1^i$ - the boundary points of successive stages on the segment VACA, in which the inequality holds $c(s(t)) \leq 0$. The magnitude $\eta > 0$ and number of stages $y \geq 1$ for which conditions (4) must be satisfied are specified.

The interval at which inequality (5) is fulfilled is interpreted in emergency situations models as a stage of a substantial reduction in the risk of a catastrophe between dangerous sections of the route relative to the location [26,27].

A specific condition may be, for example, a condition $m(t) \leq \overline{m}, t \in [t_1^i - \eta_i, t_1^i], i = \overline{1, q}$ where $m(t)$ is an indicator of the complexity of the situation. Conditions (4) are interpreted as the goal of increasing, to a predetermined value, the intervals of corrective action during a cyclically exacerbating situation, recurring factors leading to loss of control over the vessel.

The specified number of intervals q, y are selected from the analysis of the vessel's passage plan relative to the location and independent of q, y . The selection of q, y values also depends on the capabilities of the VACA, the maneuverability of the vessel, etc., and determined by the restrictions on the VACA.

The task allows the constraints on VACA not only in the form of condition (3), but also in the form of a set of constraints of the type of inequalities imposed on the final state of the system at the time t_1^q , which determines the end moment of the VACA process:

$$J_i(d) = q_i(s(t_1^q)) \leq 0, i = \overline{y+1, y+\eta} \quad (6)$$

As functionals VACA can be considered the time $t_1(d) = t_1^q - t_0$ to achieve a safe state for the transition to manual control. Thus, a task that can be presented in a single form $J_j(d) \leq 0, j = \overline{1, y+\eta}$ is considered.

3 Results

Consider the use of VACA in constrained conditions. Let the sea electronic navigation chart presented on a plane be an orthogonal grid. Then the process of moving the vessel in difficult constrained conditions and narrows can be represented as a grid trajectory.

Let us compare to the waypoints the position of the vessel, which coincides with the set of states depending on the state of the neighboring grid frames.

These waypoints or grid nodes will be represented as a field superimposed on the chart. The neighbors of the frame in which the vessel is currently located are frames that are in contact with the frame of the vessel. A set of target vessels located in proximity and relative to the vessel frame forming the interaction field considered in this model. In the case of the standard cartographic Mercator projection [28], we will call the four frames as the transition framework, which have a common side with the vessel frame. At each time point, the state of the vessel frame varies depending on the state of the interaction field.

In order to apply the developed model in practice, the frames are transformed according to the principles of vessel movement in space. For each specific case, the transition frame will have different outlines. For example, if the speed of the vessel increases, the frame will be extended.

Suppose that each vessel seeks to move in a certain (one of the four) direction. If it is impossible to move in the preferred direction, the presence of insurmountable cartographic obstacles on the way, constraining by the draft vessel or a significant amount of marine traffic, the navigator tries to change the direction of vessel course, choosing one at which the obstacles are minimal.

Accepting the fact that navigators of maritime transport can view the situation through radar and NIS (Nautical Information Systems) [29] in a location at a distance r and choose the direction in which they observe the least amount of marine traffic and the absence of cartographic obstacles (Fig. 6).

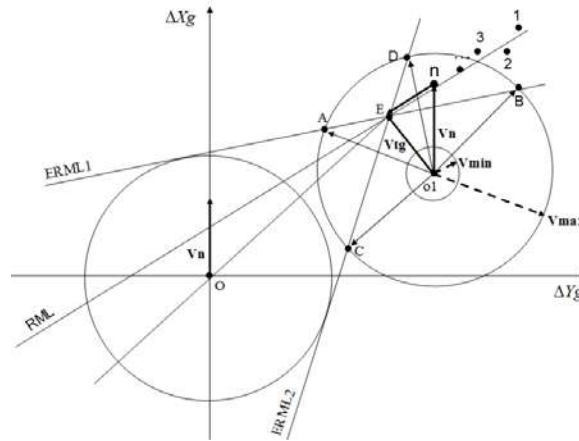


Fig. 6. Definition of productive discrepancy vectors

VACA performs the task of defining the frames on the map and forms the trajectory of the ship for safe divergence $\Omega = \Omega_1 \cap \Omega_2 \cap \dots \cap \Omega_{N_{itg}}$. The approach defines two stages: the use of the on-board controller of the determining constraint (fixed and dynamic); a module that defines a discrete change in time constraints [30].

This approach allows you to track dynamic spatial changes of restrictions on the map by analyzing displacement vectors. The approach creates a trajectory of safe movement of your own ship through conditionally defined grid nodes and test vectors are used: $\bar{V}_T = (V_T \cdot \cos K_T, V_T \cdot \sin K_T)$. The most effective divergence vectors determine the strategy for the automated divergence of the vessel during dynamically changing constraints. The use of VACA allows the identification of the fact of loss of control of the vessel's maneuverability due to the human factor to bring the vessel to the safest cell (area) at location (Fig. 7).

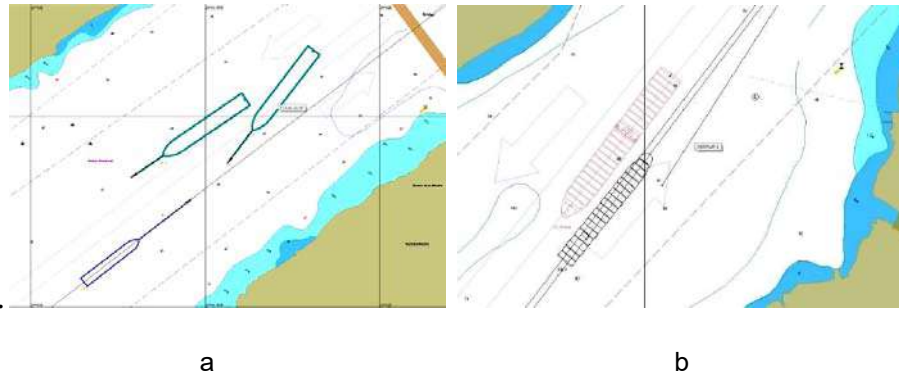


Fig. 7. **a** Before VACA launch, **b** After starting the system

In the Figure 8, the number “1” denotes the areas of permissible discrepancy parameters for many targets, including maneuvering ones, as well as obstacles in the coordinates speed - discrepancy course at different points in the time.

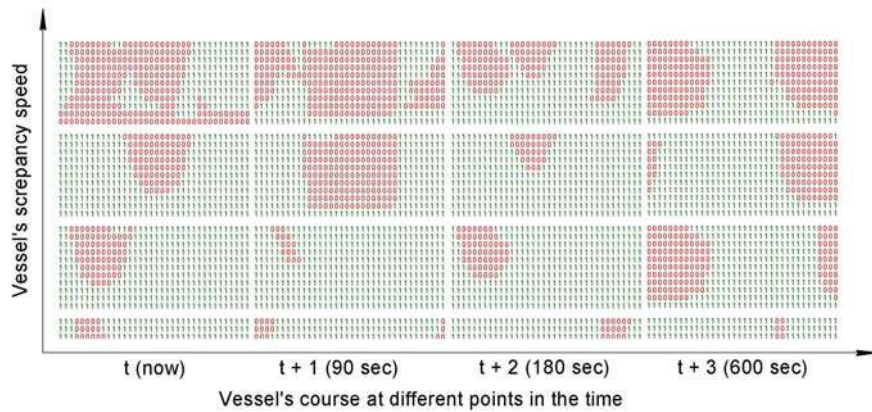


Fig. 8. VACA transition matrix

As can be seen from the presented matrices, fragments with the value “0” prevail after a certain period of time, which indicates the expansion of non safe maneuvering zones. In situations where the matrix elements with a value of "0".

An analysis of the database for the feasibility of using VACA is shown on the graph in the form of the number of accidents along the Bosphorus Strait, 15.66 Nm (Fig. 9). The presented dependence indicates that the accident rate drops by 44% when using VACA.

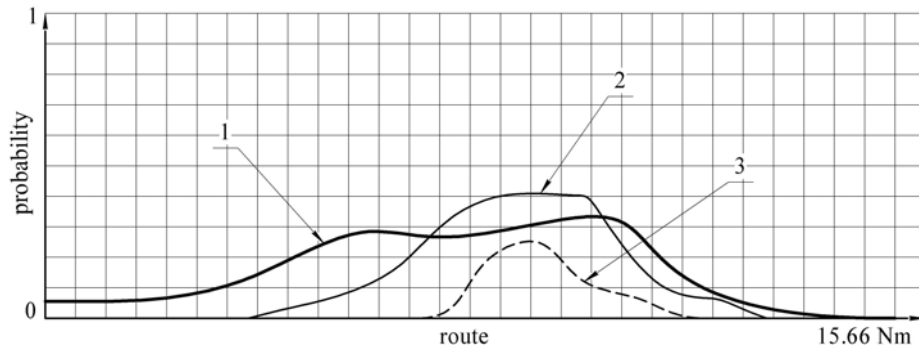


Fig. 9. Expediency use of VACA

The operation of VACA was also analyzed according to the levels of interaction with the navigation watch when passing a location. In this case, the VACA trigger levels are divided into three options (Fig. 10):

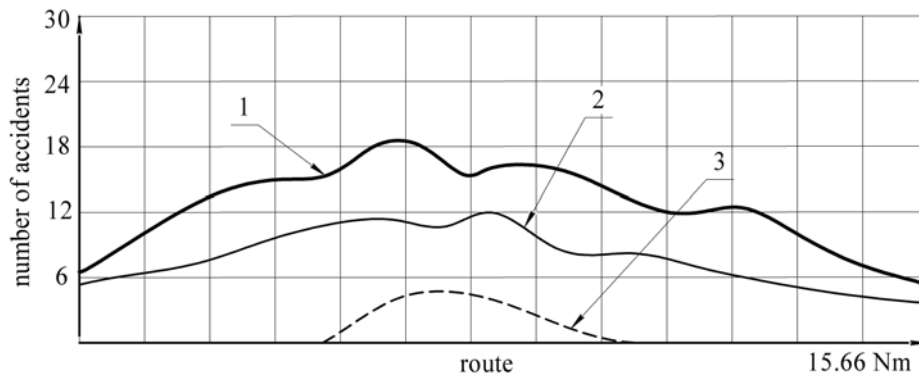


Fig. 10. Probability of operation of VACA

1 - VACA warns the navigational watch with a signal that the ship enters the zone of heightened attention, with no visible danger, but the probability is 20-30%.

2 - VACA warns the navigation watch with a signal that the vessel enters the zone of heightened attention, while informing the captain or chief officer that need to climb the bridge and take vessel control, the probability of an emergency is from 31 to 50%.

3 - VACA will switch the vessel handling to fully automated control mode, calling the captain and chief officer to the bridge with arise sound of vessel general alarm. At

the same time, the VACA urgently reduce main engine RPM till “Dead Slow Ahead” and determining the zones of zero probability of collision, performs an automated divergence maneuver.

In order to determine the effectiveness of the VACA application, a module was developed in the form of a three-dimensional graph (Fig. 11). The graph allows displaying the coverage of VACA during the execution of maneuvers on the ship’s passage. During the VACA operation, the function of plotting along the axes is activated: Y is the complexity of the route zone, X is the order of the zone, Z is the number of accidents.

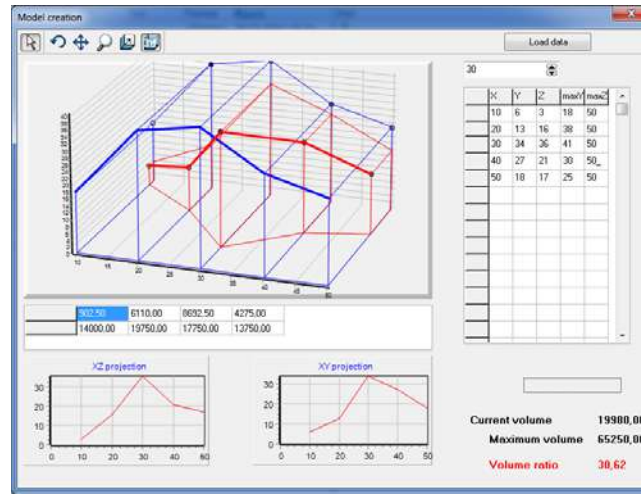


Fig. 11. Analysis of the effectiveness of the use of VACA

The graph shows that the three-dimensional figure in red reflects the complexity of the route, taking into account both geographic parameters and the characteristics of the experience of the navigation watch. The blue three-dimensional shape represents the precautionary response of VACA. In the area of the third zone, there is a small warning of a negative situation due to the human factor, which indicates the need to strengthen the navigation watch in this area of the vessel’s passage. The total reserve of VACA functionality over the time is 69%, which indicates its feasibility for use in the route planning and passage process. This graph allows us to perform an analysis of the likelihood of an emergency situation while keeping a navigation watch, as well as to prevent the human factor of the navigator in its early stages. The general scheme of interaction of the vessel automated course alteration system with the navigator is presented in Figure 12.

The base of navigator’s reactions to the stimulus accumulates over a long time. A selection of parameters that affects the loss of control of the situation can be interpreted as objects x_i on task classes $\kappa_i = \{0,1\}$. The objective of VACA is to reduce the likelihood of errors when using navigation hazard classification algorithms a_j .

A formal description of this process will be based on threshold classifiers, taking into account the specifics of VACA, [31]:

$$a_j(x_i)\kappa_i > 0 \Leftrightarrow \begin{cases} j > i, \kappa_i < 0, \\ j \leq i, \kappa_i > 0. \end{cases}$$

Then the condition $E_{(i)}$ for determining classification algorithms a_j giving error-free samples of object parameters will look like:

$$E_{(i)} = \{X : x_i \in \bar{X}, \forall \mu \in M \rightarrow \mu(X) = a_j : \begin{cases} j > i, \kappa_i < 0 \\ j \leq i, \kappa_i > 0 \end{cases}\}.$$

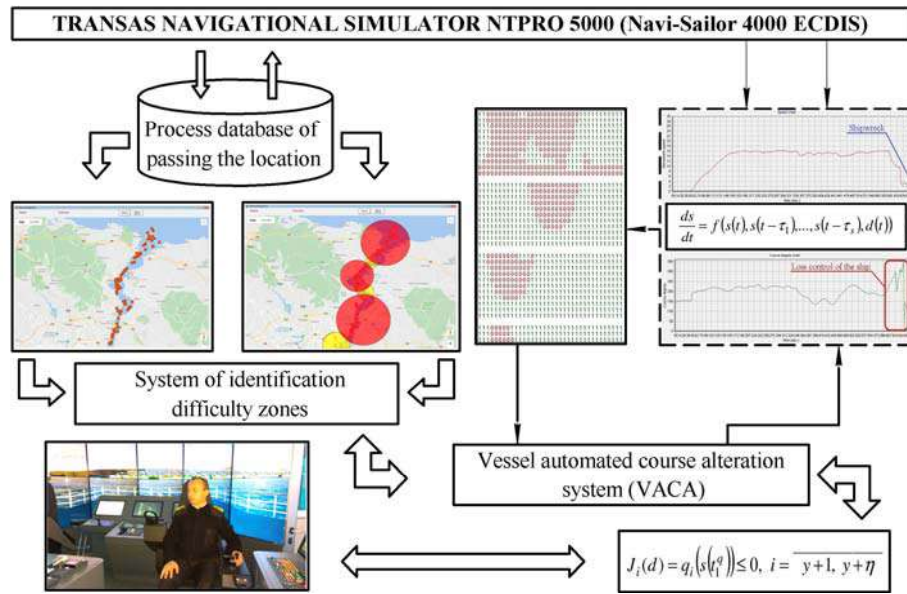


Fig. 12. The general scheme of interaction of the vessel automated course alteration system with the navigator

4 Discussion

The basic principles and research were discussed at the 10th International scientific and practical conference MINTT-2018 [32].

The introduction of software will allow quickly and efficiently process large amounts of data and highlight the classification signs of the negative manifestation of the human factor, both among cadets and experienced seafarers during traineeships and advanced training in accordance with the standards of IMO and STCW 78/85. It will also allow developing the professional competence of the cadets with the use of innovative learning and psychological technologies [33-36]. In the future, further research is planned to develop software, in the form of an expert system that determines deviations from a given course during the passage, as well as inadequate reactions when performing classical maneuvers when vessels diverge in constrained areas.

5 Conclusion

Analysis of the effectiveness of the developed models and software gives grounds to assert that the approaches described in the article allow increasing the safety of navigation in restricted navigation areas, reducing the periods of the negative influence of the human factor due to the transition to automated vessel control.

- Software tools have been developed that allow identifying on the cartographic panel the zones with the highest risk of accidents by analyzing the passage of locations on the NTPRO 5000, as part of an expert disaster prevention system.
- A module for automated analysis of the ECDIS database has been developed, which allows determining geolocation on Google maps, as well as periods of loss of control over the management of the vessel due to persistent negative reactions of the navigator during difficult navigation.
- Algorithmic software has been developed for automated recognition of the critical number of target vessels by AIS means for switching to autopilot in areas with the highest risk of accidents.
- The analysis of the performance of VACA in maritime transport was carried out by statistical analysis of simulator training data for cadets of the Kherson State Maritime Academy (Ukraine), confirming the appropriateness of the use of the designated approaches and software.

6 Acknowledgments

The work is carried out within the framework of “Development of software for improving the quality of functioning of dynamic positioning systems of ships” (state registration number 0119U100948), of navigation and ECDIS departments of Kherson State Maritime Academy Navigation Faculty (scientific adviser: Ph.D. Associate Professor, Deputy Rector for scientific and pedagogical work, Kherson State Maritime Academy, Ukraine, Ben A.P.).

References

1. Ming-Cheng, T.: Multi-target collision avoidance route planning under an ECDIS framework. *Ocean Engineering*. 121, 268-278, (2016). doi:10.1016/j.oceaneng.2016.05.040
2. Zinchenko, S.M., Nosov, P.S., Mateychuk, V.M., Mamenko, P.P., Grosheva, O.O.: Automatic collision avoidance with multiple targets, including maneuvering ones. *Radio Electronics, Computer Science, Control*. 211-221, (2019). doi:10.15588/1607-3274-2019-4-20
3. Puisa, R., Lin, L., Bolbot, V., Vassalos, D.: Unravelling causal factors of maritime incidents and accidents. *Safety Science*. 110 (a), 124-141, (2018). doi:10.1016/j.ssci.2018.08.001
4. Zinchenko, S., Nosov, P., Mateichuk, V. Mamenko, P., Popovych, I. Grosheva, O.: Automatic collision avoidance system with many targets, including maneuvering ones. *Bulletin of University of Karaganda. Technical Physics*. 96, 69-79, (2019). doi:0.31489/2019Ph4/69-79

5. Uğurlu, Ö., Yıldız, S., Loughney, S., Wang, J.: Modified human factor analysis and classification system for passenger vessel accidents (HFACS - PV). *Ocean Engineering*, 161, 47-61, (2018). doi:10.1016/j.oceaneng.2018.04.086
6. Akyuz, Emre.: Quantitative human error assessment during abandon ship procedures in maritime transportation. *Ocean Engineering*. 120, 21-29, (2016). doi:10.1016/j.oceaneng.2016.05.017
7. Longo, F., Chiurco, A., Musmanno, R., Nicoletti, L.: Operative and procedural cooperative training in marine ports. *Journal of Computational Science*. 10, 97-107, (2015). doi:10.1016/j.jocs.2014.10.002
8. Tvedt, S., Oltedal, H., Batalden, B.M., Oliveira, M.: Way-finding on-board training for maritime vessels. *Entertainment Computing*. 26, 30-40, (2018). doi:10.1016/j.entcom.2018.01.002
9. Ventikos, N.P., Papanikolaou, A.D., Louzis, K., Koimtzoglou, A.: Statistical analysis and critical review of navigational accidents in adverse weather conditions. *Ocean Engineering*. 163, 502-517, (2018). doi:10.1016/j.oceaneng.2018.06.001
10. Xi, Y.T., Yang, Z.L., Fang, Q.G., Chen, W.J., Wang, J.: A new hybrid approach to human error probability quantification—applications in maritime operations. *Ocean Engineering*. 138, 45-54, (2017). doi:10.1016/j.oceaneng.2017.04.018
11. Pazouki, K., Forbes, N., Norman, R. A., Woodward, M.D.: Investigation on the impact of human-automation interaction in maritime operations. *Ocean Engineering*. 153, 297-304, (2018). doi:10.1016/j.oceaneng.2018.01.103
12. Lee, J., Chung, C.: A new methodology for accident analysis with human and system interaction based on FRAM: Case studies in maritime domain. *Safety Science*. 109, 57-66, (2018). doi:10.1016/j.ssci.2018.05.011
13. Popovych, I.S., Cherniavskiy, V.V., Dudchenko, S.V., Zinchenko, S.M., Nosov, P.S., Yevdokimova, O.O., et. al. (2020). Experimental Research of Effective “The Ship’s Captain and the Pilot” Interaction Formation by Means of Training Technologies. *Revista ESPACIOS*, 41(11). Page 30
14. Popovych I.S.; Blynova, O.Ye., Aleksieieva M.I., Nosov P.S., Zavatska N.Ye. & Smyrnova O.O. (2019). Research of Relationship between the Social Expectations and Professional Training of Lyceum Students studying in the Field of Shipbuilding. *Revista ESPACIOS*, 40(33). Page 21
15. Seabrooke, W., Hui, E.C.M., Lam, W.H.K., Wong, G.K.C.: Forecasting cargo growth and regional role of the port of Hong Kong. *Cities*. 20, 51-64, (2003). doi:10.1016/S0264-2751(02)00097-5
16. Szlapczynski, R., Krata, P.: Determining and visualizing safe motion parameters of a ship navigating in severe weather conditions. *Ocean Engineering*. 158, 263-274, (2018). doi:10.1016/j.oceaneng.2018.03.092
17. Shigunov, V., Guo, B., Reddy, D.N., Lalovic, I.: Maneuverability in adverse conditions: Case studies. *Ocean Engineering*. 179, 371-386, (2019). doi:10.1016/j.oceaneng.2019.02.026
18. Yuanzhi Zhang, Erjia Ge.: Temporal scaling behavior of sea-level change in Hong Kong — Multifractal temporally weighted detrended fluctuation analysis. *Global and Planetary Change*. 100, 362-370, (2013). doi:10.1016/j.gloplacha.2012.11.012
19. Chen, J., Li, K.X., Xiang, L., Haobo, L.: The development of ship registration policy in China: Response to flags of convenience. *Marine Policy*. 83, 22-28, (2017). doi:10.1016/j.marpol.2017.05.020
20. Nosov, P.S., Ben, A.P., Mateychuk, V.N., Safonov, M.S. Identification of “Human error” negative manifestation in maritime transport. *Radio Electronics, Computer Science, Control*. Zaporizhzhia National Technical University. 47, 204-213, (2018). doi:10.15588/1607-3274-2018-4-20

21. Nosov, P., Ben, A., Safonova, A., Palamarchuk, I.: Approaches going to determination periods of the human factor of navigators during supernumerary situations. *Radio Electronics, Computer Science, Control.* 49, 140-150, (2019). doi:10.15588/1607-3274-2019-2-15
22. Buldaev, A.S.: Control over oscillations in delayed systems for modeling of diseases. *Math modeling.* 6, 10-21, (1991).
23. Zinchenko, S.N., Liashenko, V.G.: Optimal control of the vessel when approaching the mooring facility. *Materials of the X international scientific and practical conference on Modern information and innovations technology in transport, Kherson*, (2018).
24. Fujimoto, K.: Classes and truths in set theory. *Annals of Pure and Applied Logic.* 163, 1484-1523, (2012). doi:10.1016/j.apal.2011.12.006
25. Kastrisios, C., Pilikou, M.: Nautical cartography competences and their effect to the realization of a worldwide Electronic Navigational Charts database, the performance of ECDIS and the fulfilment of IMO chart carriage requirements. *Marine Policy.* 75, 29-37, (2017). doi:10.1016/j.marpol.2016.10.007
26. Pengfei Chen, Yamin Huang, Junmin Mou, P.H.A.J.M. van Gelder: Ship collision candidate detection method: A velocity obstacle approach. *Ocean Engineering.* 170, 186-198, (2018). doi:10.1016/j.oceaneng.2018.10.023
27. Liu Yan, Frangopol Dan M.: Probabilistic risk, sustainability, and utility associated with ship grounding hazard. *Ocean Engineering.* 154, 311-321, (2018). doi:10.1016/j.oceaneng.2018.01.101
28. Maling, D.H.: The Transverse Mercator projection. *Coordinate Systems and Map Projections.* Pergamon. 336-363, (1992). doi:10.1016/B978-0-08-037233-4.50021-2
29. Bole, A., Wall, A., Norris, A.: *Navigation Techniques Using Radar and ARPA. Radar and ARPA Manual.* Butterworth-Heinemann. 371-405, (2014). doi:10.1016/B978-0-08-097752-2.00008-8
30. Zinchenko, S.M., Nosov, P.S., Mateichuk, V.M., Mamenko, P.P., Grosheva, O.O.: Automatic collision avoidance with many targets, including maneuvering ones. *The international scientific and practical conference dedicated to the memory of professors Fomin Y.Y. and Semenov V.S. (FS-2019), Odessa – Istanbul*, (2019).
31. Guz, I.S.: Constructive estimates of full moving control for threshold classification. *Mat. Biology and Bioinform.* 6, 173-189, (2011).
32. Nosov, P.S., Ben, A.P., Safonov, M.S.: Model construction of individual scenarios for the elimination of the human factor. *Materials of the X international scientific and practical conference on Modern information and innovations technology in transport, Kherson*, (2018).
33. Yurzhenko, A.Y., Popova, H.V.: Competency framework as an instrument to assess professional competency of future seafarers. *Proc. 15 th Int. Conf. ICTERI, Kherson*, (2019).
34. Popovych, I.S., & Blynova, O.Ye. (2019). The Structure, Variables and Interdependence of the Factors of Mental States of Expectations in Students' Academic and Professional Activities. *The New Educational Review.* 55, 293-306. doi:10.15804/tner.2019.55.1.24
35. Popovych, I.S.: *Psychology of Social Expectations of Personality: Methodology, Theory and Practice.* OLDI-PLUS, Kherson (2019).
36. Voloshynov, S.A. (2020). Mobile training technologies in the professional training for the prospective river and maritime transport specialists. *Information technology and training tools.* 75, 108-121. doi:10.33407/itlt.v75i1.2729

Optimization Calculation of Vortex Type Granulation Devices: Application of Software Products and Computer Modeling

Artem Artyukhov¹[0000-0003-1112-6891] and Jan Krmela²[0000-0001-9767-9870]

¹Sumy State University, 2 Rymskogo-Korsakova Street, Sumy 40007, Ukraine
a.artyukhov@pohnp.sumdu.edu.ua

²Alexander Dubcek University of Trencin, 491/30 I. Krasku, Puchov 02001, Slovak Republic
jan.krmela@fpt.tnuni.sk

Abstract. The paper studies the hydrodynamic conditions of gas flow motion and vortex granulator workspace design optimization. A comprehensive approach for determination of the hydrodynamic characteristics of a vortex gas flow and their visualization is proposed. The mathematical approach, based on Reynolds equations for turbulent flows solution, is presented. The mathematical model equations solution with the definition of gas flow velocity components in any point on the radius and height of vortex granulator, is obtained. The obtained results form the basis of original algorithm for calculating of vortex granulator hydrodynamic calculation and its basic dimensions optimization selection.

Keywords: Software, Modeling, Vortex Granulator, Hydrodynamics, Optimization

Key Terms: Development, Software Engineering Process, Research

1 Introduction

Flows of single-phase and multiphase medium play a key role in the working process of many contemporary engineering devices. The design of these devices for the required operating parameters is impossible without a reliable prediction of the characteristics of these flows. Since many modern engineering devices are expensive and time consuming to manufacture, physical modeling with the experimental determination of the parameters of their work in different modes, as a rule, requires a lot of time and cost. In addition, because of limited possibilities of modern experimental sensors and measuring instruments, experimental observations do not provide a complete picture of the phenomenon being investigated. Because of nature of these environments, flow of liquids and gases occur often very in complex manner to form nonstationary effects, “dead” zones and vortex structures [1,2]. The situation is further complicated by the presence of heat transfer, in considering flows mixture of several substances, flows with free surfaces, weighted particles in stream, flows with cavitation, boiling, condensation, combustion, chemical reactions [3-7]. These factors lead to growing interest of mathematical modeling tools for flows of liquids and gases, which allow to

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

predict the flow characteristics and parameters of the devices at design stages and stages of manufacturing in the metal.

Software can adequately simulate the complex physical effects taking place in flows motion in vortex granulator, and perform calculation of flows within the reasonable time [8,9]. They provide user convenient tools of data preparation and analysis of calculations results, and are a powerful tool for accurately predict the characteristics of flow of hydraulic parts at design stage, saving the resources to conduct a physical experiment.

Methods CFD (Computational Fluid Dynamics) suggest calculation of flows of liquids and gases through the numerical solution of equations of Navier-Stokes and continuity, describing the most general case of this medium motion (for turbulent flows - Reynolds equations).

The purpose of work is creation of vortex granulator optimization calculation algorithm. The author's software and calculation data visualization using software for calculation flows hydrodynamics should be created on base of the suggested algorithm.

As a working hypothesis according to the possibility of controlling the trajectory of motion and the residence time of the dispersed phase in the working space of the granulator due to the creation of mechanisms of its directed movement. A joint solution of the fundamental equations of the hydrodynamics of flow motion and kinetics of changes in the temperature and humidity characteristics of the interacting flows, as well as the growth rate of the granules, will allow inventing a rational design of the working space, the optimal flow of coolant and its temperature and humidity characteristics in the vortex granulator. The calculation is carried out according to the criterion of optimization "minimum" hydrodynamic "residence time of the dispersed phase in the working space of the device." "Hydrodynamic" time should be equal to "kinetic" - the time during which the temperature and humidity characteristics of the dispersed phase must acquire a normative value.

2 Theoretical basics

Given that, in practice interest is usually deals with not instantaneous, but with the average in time velocity value, for mathematical description of turbulent swirling motion of gas flow Reynolds equation as modification of Navier-Stokes equations is used [10]

$$\frac{\partial}{\partial t}(\rho \bar{V}_i) + \frac{\partial}{\partial q_j}(\rho \bar{V}_i \bar{V}_j) + \frac{\partial}{\partial q_j}(\rho \bar{V}_i' \bar{V}_j') = -\frac{\partial p}{\partial q_i} + \frac{\partial}{\partial q_j} \left[\mu \left(\frac{\partial \bar{V}_i}{\partial q_j} + \frac{\partial \bar{V}_j}{\partial q_i} \right) \right] + f_i \quad (1)$$

where $\bar{V}$ - average in time values of velocity; $\bar{V}'$ - components of velocity pulsations; μ - coefficient of turbulent viscosity; t - time; ρ - gas density; p - pressure; f_i - element, characterizing the mass forces effect; q_j - coordinate axes (in the case of hydrodynamic modeling in granulator working volume as it is shown above, it is advisable to use curvilinear coordinate system), $i, j = 1 \dots 3$; for cylindrical coordinate system (fig. 1) code "1" - axial direction (z), code "2" - radial direction (r), code "3" - circular direction (φ) (figs 1,2).

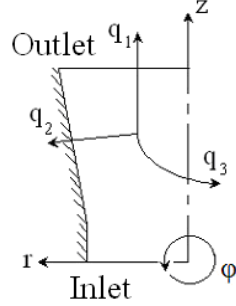


Fig. 1. Scheme of vortex granulator workspace and coordinate system.

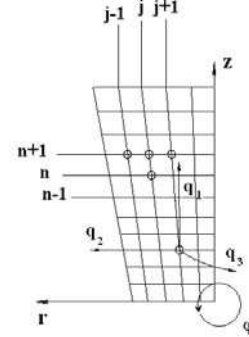


Fig. 2. Construction of calculation grid.

The Reynolds equations system is supplemented with flow continuity equation

$$\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial q_j} (\rho V_j) = 0. \quad (2)$$

The main advantage of description and problem solving hydrodynamics method, based on numerical solution of Reynolds complete equations - accuracy and versatility.

For Reynolds equations (1) solving Boussinesq hypothesis [10] is used. According to this hypothesis, the members with velocity pulsations $(\rho \overline{V_i' V_j'})$ in equation (3) are associated with the averaged flow characteristics in such relation:

$$\rho \overline{V_i' V_j'} = -\mu \left(\frac{\partial \overline{V_i}}{\partial q_j} + \frac{\partial \overline{V_j}}{\partial q_i} \right) + \frac{2}{3} \rho \delta_{ij} k, \quad (3)$$

where $k = 0,5 (\overline{V_j' V_j'})$ – turbulence kinetic energy, $\delta_{ij} = 1$ when $i = j$, $\delta_{ij} = 0$ when $i \neq j$.

The Reynolds system of equations is elliptic. It is used to calculate the trends in those cases where flow characteristics at arbitrary point area depend on the structure of flow both above and downstream, i.e. when the dominant direction of the fluid is absent or weakly expressed. Ellipticity system of equations means to address it is necessary to set the boundary conditions for all variables in all the borders of calculation area.

When axisymmetrical flow modeling equation of motion (1) and continuity of flow (2) is significantly simplified. For curved (cylindrical) coordinate system are as follows (with the introduction of equation (3) Reynolds number $Re = V_0 D / \nu$ where characteristic parameters D – diameter of input cross section calculation area; V_0 – average rate velocity in the input section; ν – kinematic viscosity):

– Reynolds equations projected on the axial direction q_1 :

$$\frac{V_1}{H_1} \frac{\partial V_1}{\partial q_1} + \frac{V_2}{H_2} \frac{\partial V_1}{\partial q_2} - \frac{V_2^2}{H_1 H_2} \frac{\partial H_2}{\partial q_1} - \frac{V_3^2}{H_1 H_3} \frac{\partial H_3}{\partial q_1} + \frac{1}{H_2} \frac{\partial (\overline{V_1' V_2'})}{\partial q_2} + \frac{(\overline{V_1' V_2'})}{H_1^2 H_2 H_3} \frac{\partial (H_1^2 H_3)}{\partial q_2} +$$

$$\begin{aligned}
& + \frac{1}{H_1} \frac{\partial \left(\overline{V'_1 V'_1} \right)}{\partial q_1} + \frac{\left(\overline{V'_1 V'_1} \right)}{H_1^2 H_2 H_3} \frac{\partial (H_1 H_2 H_3)}{\partial q_1} - \frac{\left(\overline{V'_2 V'_2} \right)}{H_1 H_2} \frac{\partial H_2}{\partial q_1} - \frac{\left(\overline{V'_3 V'_3} \right)}{H_1 H_3} \frac{\partial H_3}{\partial q_1} = - \frac{1}{H_1} \frac{\partial p}{\partial q_1} + \\
& + \frac{1}{\text{Re}} \left(\frac{1}{H_1^2} \frac{\partial^2 V_1}{\partial q_1^2} + \frac{1}{H_2^2} \frac{\partial^2 V_1}{\partial q_2^2} + \frac{1}{H_1 H_2 H_3} \frac{\partial V_1}{\partial q_1} \frac{\partial (H_2 H_3 / H_1)}{\partial q_1} + \frac{1}{H_1 H_2 H_3} \frac{\partial V_1}{\partial q_2} \frac{\partial (H_1 H_3 / H_2)}{\partial q_2} + \right. \\
& \quad \left. - \frac{2}{H_1 H_2^2} \frac{\partial H_2}{\partial q_1} \frac{\partial V_2}{\partial q_2} + \frac{V_1}{H_1} \frac{\partial}{\partial q_1} \left(\frac{1}{H_1 H_2 H_3} \frac{\partial (H_2 H_3)}{\partial q_1} \right) + \right. \\
& \quad \left. + \frac{V_2}{H_1} \frac{\partial}{\partial q_1} \left(\frac{1}{H_1 H_2 H_3} \frac{\partial (H_1 H_3)}{\partial q_2} \right) - \frac{V_2}{H_2 H_3} \frac{\partial}{\partial q_2} \left(\frac{H_3}{H_1 H_2} \frac{\partial H_2}{\partial q_1} \right) \right); \tag{4}
\end{aligned}$$

– Reynolds equations projected on the radial direction q_2 :

$$\begin{aligned}
& \frac{V_1}{H_1} \frac{\partial V_2}{\partial q_1} + \frac{V_2}{H_2} \frac{\partial V_2}{\partial q_2} - \frac{V_1 V_2}{H_1 H_2} \frac{\partial H_2}{\partial q_1} - \frac{V_3^2}{H_2 H_3} \frac{\partial H_3}{\partial q_2} + \frac{1}{H_1} \frac{\partial \left(\overline{V'_1 V'_2} \right)}{\partial q_1} + \frac{\left(\overline{V'_1 V'_2} \right)}{H_1 H_2^2 H_3} \frac{\partial (H_2^2 H_3)}{\partial q_1} + \\
& \quad + \frac{1}{H_2} \frac{\partial \left(\overline{V'_2 V'_2} \right)}{\partial q_2} + \frac{\left(\overline{V'_2 V'_2} \right)}{H_1 H_2^2 H_3} \frac{\partial (H_1 H_2 H_3)}{\partial q_2} - \frac{\left(\overline{V'_3 V'_3} \right)}{H_2 H_3} \frac{\partial H_3}{\partial q_2} = - \frac{1}{H_2} \frac{\partial p}{\partial q_2} + \\
& \quad + \frac{1}{\text{Re}} \left(\frac{1}{H_1^2} \frac{\partial^2 V_2}{\partial q_1^2} + \frac{1}{H_2^2} \frac{\partial^2 V_2}{\partial q_2^2} + \frac{1}{H_1 H_2 H_3} \frac{\partial V_2}{\partial q_1} \frac{\partial (H_2 H_3 / H_1)}{\partial q_1} + \frac{1}{H_1 H_2 H_3} \frac{\partial V_2}{\partial q_2} \frac{\partial (H_1 H_3 / H_2)}{\partial q_2} + \right. \\
& \quad \left. + \frac{2}{H_1 H_2^2} \frac{\partial H_2}{\partial q_1} \frac{\partial V_1}{\partial q_2} + \frac{V_2}{H_2} \frac{\partial}{\partial q_2} \left(\frac{1}{H_1 H_2 H_3} \frac{\partial (H_1 H_3)}{\partial q_2} \right) \right); \tag{5}
\end{aligned}$$

– Reynolds equations projected on the circular direction q_3 :

$$\begin{aligned}
& \frac{V_1}{H_1} \frac{\partial V_3}{\partial q_1} + \frac{V_2}{H_2} \frac{\partial V_3}{\partial q_2} + \frac{V_1 V_3}{H_1 H_3} \frac{\partial H_3}{\partial q_1} + \frac{V_2 V_3}{H_2 H_3} \frac{\partial H_3}{\partial q_2} + \frac{1}{H_2} \frac{\partial \left(\overline{V'_2 V'_3} \right)}{\partial q_2} + \frac{\left(\overline{V'_2 V'_3} \right)}{H_1 H_2 H_3^2} \frac{\partial (H_1 H_3^2)}{\partial q_2} + \\
& \quad + \frac{1}{H_1} \frac{\partial \left(\overline{V'_1 V'_3} \right)}{\partial q_1} + \frac{\left(\overline{V'_1 V'_3} \right)}{H_1 H_2 H_3^2} \frac{\partial (H_2 H_3^2)}{\partial q_1} = \frac{1}{\text{Re}} \left(\frac{1}{H_1^2} \frac{\partial^2 V_3}{\partial q_1^2} + \frac{1}{H_2^2} \frac{\partial^2 V_3}{\partial q_2^2} + \right. \\
& \quad \left. + \frac{1}{H_1 H_2 H_3} \frac{\partial V_3}{\partial q_1} \frac{\partial (H_2 H_3 / H_1)}{\partial q_1} + \frac{1}{H_1 H_2 H_3} \frac{\partial V_3}{\partial q_2} \frac{\partial (H_1 H_3 / H_2)}{\partial q_2} + \frac{V_3}{H_1 H_2} \frac{\partial}{\partial q_2} \left(\frac{H_1}{H_2 H_3} \frac{\partial H_3}{\partial q_2} \right) \right); \tag{6}
\end{aligned}$$

– continuity equation:

$$\frac{1}{H_1 H_2 H_3} \left(V_1 \frac{\partial (H_2 H_3)}{\partial q_1} + V_2 \frac{\partial (H_3 H_1)}{\partial q_2} \right) + \frac{1}{H_1} \frac{\partial V_1}{\partial q_1} + \frac{1}{H_2} \frac{\partial V_2}{\partial q_2} = 0, \tag{7}$$

where H_1, H_2, H_3 – Lamé coefficient [10].

Further simplification of system of equations (4) - (7) for simulating the vortex flow of gas phase in the workspace granulator is possible using the following assumptions [11-13]:

- expected the presence of dominant flow direction along which the axial component of gas flow velocity is everywhere positive and far exceeds the radial;
 - gas flow velocity component in the axial direction varies considerably slower than in the radial;
 - velocity and pressure values in every elementary volume of gas flow depend only on the conditions downstream and do not depend on the conditions upstream.
- These assumptions allow to conduct analysis of components in equations (4) - (7) and discard those, that provide significant impact on the result of the calculation. After accounting assumptions for axially symmetric gas flow equation (4) - (7) be written as

$$\begin{aligned} \frac{V_1}{H_1} \frac{\partial V_1}{\partial q_1} + \frac{V_2}{H_2} \frac{\partial V_1}{\partial q_2} - \frac{V_3^2}{H_1 H_3} \frac{\partial H_3}{\partial q_1} + \frac{1}{H_2} \frac{\partial \left(\overline{V_1' V_2'} \right)}{\partial q_2} + \frac{\left(\overline{V_1' V_2'} \right)}{H_1^2 H_2 H_3} \frac{\partial \left(H_1^2 H_3 \right)}{\partial q_2} = \\ = - \frac{1}{H_1} \frac{\partial p}{\partial q_1} + \frac{1}{\text{Re}} \left(\frac{1}{H_2^2} \frac{\partial^2 V_1}{\partial q_2^2} + \frac{1}{H_1 H_2 H_3} \frac{\partial V_1}{\partial q_2} \frac{\partial \left(H_1 H_3 / H_2 \right)}{\partial q_2} \right); \end{aligned} \quad (8)$$

$$\frac{\partial p_r}{\partial q_2} = \frac{V_3^2}{H_3} \frac{\partial H_3}{\partial q_2}; \quad (9)$$

$$\begin{aligned} \frac{V_1}{H_1} \frac{\partial V_3}{\partial q_1} + \frac{V_2}{H_2} \frac{\partial V_3}{\partial q_2} + \frac{V_1 V_3}{H_1 H_3} \frac{\partial H_3}{\partial q_1} + \frac{V_2 V_3}{H_2 H_3} \frac{\partial H_3}{\partial q_2} + \frac{1}{H_2} \frac{\partial \left(\overline{V_2' V_3'} \right)}{\partial q_2} + \frac{\left(\overline{V_2' V_3'} \right)}{H_1 H_2 H_3^2} \frac{\partial \left(H_1 H_3^2 \right)}{\partial q_2} = \\ = \frac{1}{\text{Re}} \left(\frac{1}{H_2^2} \frac{\partial^2 V_3}{\partial q_2^2} + \frac{1}{H_1 H_2 H_3} \frac{\partial V_3}{\partial q_2} \frac{\partial \left(H_1 H_3 / H_2 \right)}{\partial q_2} + \frac{V_3}{H_1 H_2} \frac{\partial}{\partial q_2} \left(\frac{H_1}{H_2 H_3} \frac{\partial H_3}{\partial q_2} \right) \right); \end{aligned} \quad (10)$$

$$\frac{1}{H_1 H_2 H_3} \left(V_1 \frac{\partial \left(H_2 H_3 \right)}{\partial q_1} + V_2 \frac{\partial \left(H_3 H_1 \right)}{\partial q_2} \right) + \frac{1}{H_1} \frac{\partial V_1}{\partial q_1} + \frac{1}{H_2} \frac{\partial V_2}{\partial q_2} = 0. \quad (11)$$

This system of equations is closed equation sustainability costs:

$$\int_0^{Q_2} V_1 H_2 H_3 dq_2 = \text{const}, \quad (12)$$

where Q_2 - coordinate q_2 on the wall of working volume of vortex granulator.

Obtained system of equations (8) - (11) has a parabolic character, and its decision based on the method proposed by Patankar and Spalding [14] and realized in SIMPLE (Semi-Implicit Method for Pressure-Linked Equations) procedure and its modifications.

Numerical solutions of mathematical models equations is performed in one marching passage from the working volume input cross section to output using finite volumes method with the elements of finite-element approach. Before calculating the estimated grid is constructed (fig. 2), and unknown values of velocity and pressure are found in the nodes of this grid.

3 Software description

The program «Conical channel»[©] is designed to calculate axial symmetric gas flows in channels, including the swirling flows in the diffuser [15,16].

Initial data for calculation are geometric configuration of channel properties and parameters of gas and parameters of flow in the input cross-section. The calculation objective is to determine the fields of velocity and pressure in full range of calculated area and, consequently, the determination of energy losses between the input and output cross sections of channel. The program displays the calculation data as graphic dependencies.

Program's work starts with opening or creating new file of calculated area configuration. Example of calculation area configuration is shown in fig. 3.

After selecting the necessary parameter for calculation program builds estimated grid (fig. 2). For the analysis of motion velocity components it is necessary to define the nature of dependencies graphical components and match them within a comparative graph as it is shown in figs 4-6.

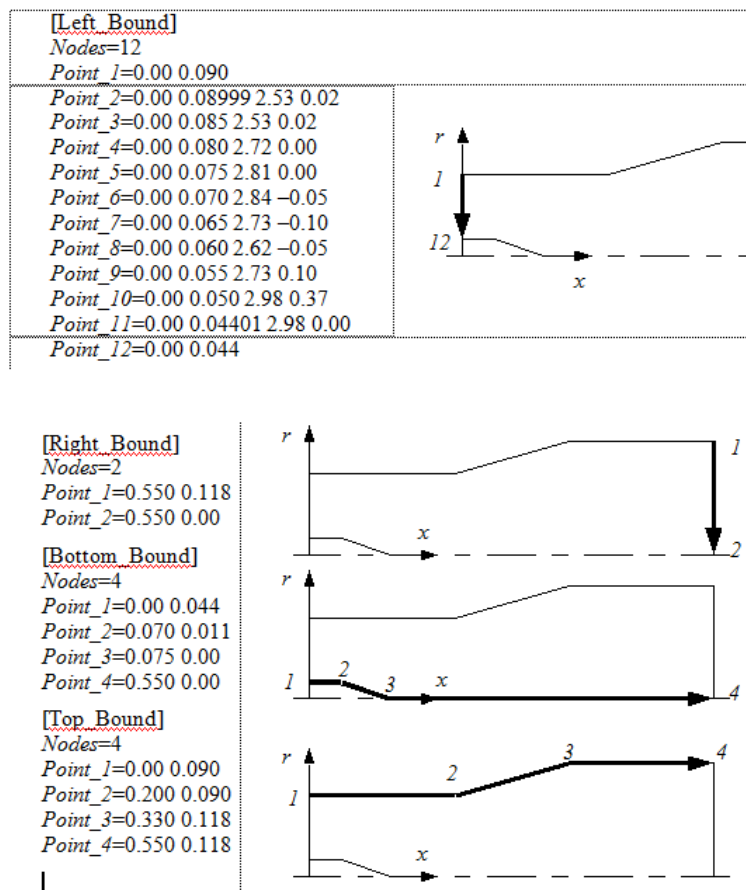


Fig. 3. Example of calculation area configuration.

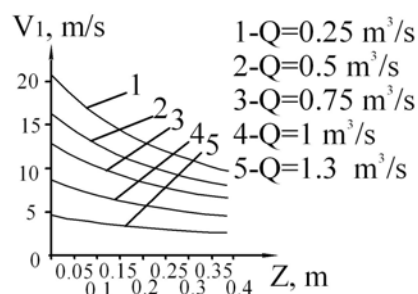


Fig.4. Calculated change of the longitudinal velocity of the gas flow.

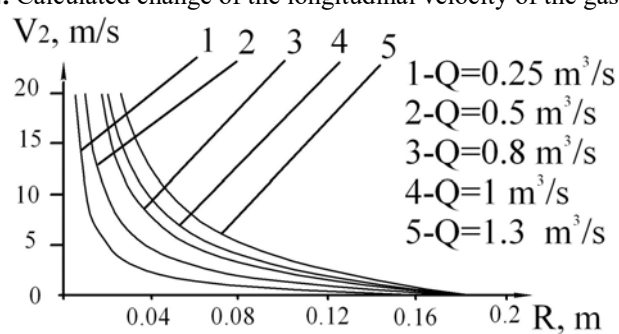


Fig.5. Calculated change of the radial velocity of the gas flow.

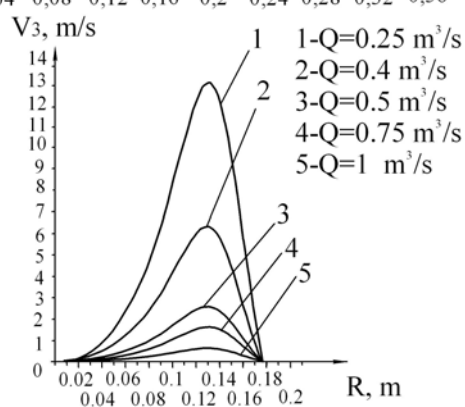
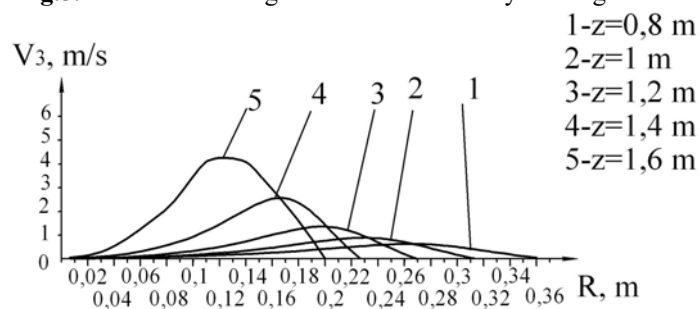


Fig.6. Calculated change of the circumferential velocity of the gas flow.

4 Visualization of results and discussion

For research results visualization in this study we have used complex ANSYS CFX [17], which allows to export obtained using «Conical channel»[©] results. The main results of visualization are shown in figs 7-11.

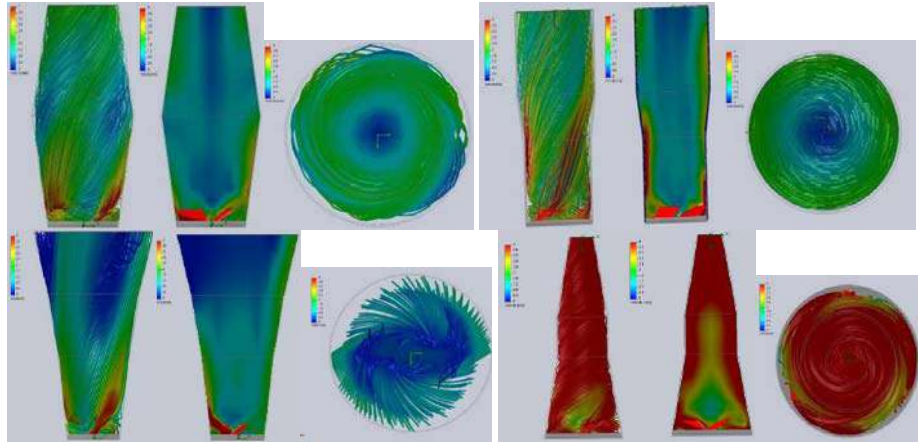


Fig.7. Filling the gas flow velocity field when installing of gas-distributing unit with blades and various configuration of granulator workspace with blades inclination angle $\alpha = 30^\circ$ and their number $n = 3$.

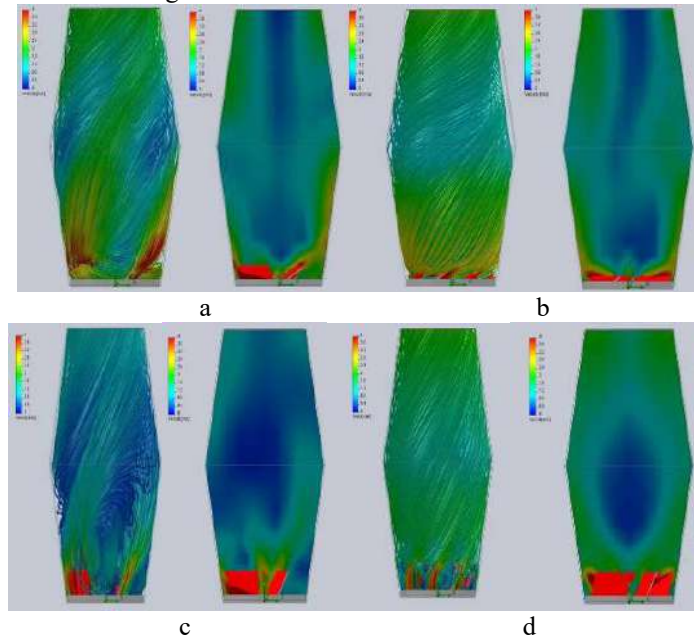


Fig. 8. Filling the gas flow velocity field when installing of gas-distributing unit with blades: a – $\alpha=30^\circ$, $n=3$; b – $\alpha=30^\circ$, $n = 8$; c – $\alpha =60^\circ$, $n =3$; d – $\alpha=60^\circ$, $n=8$.

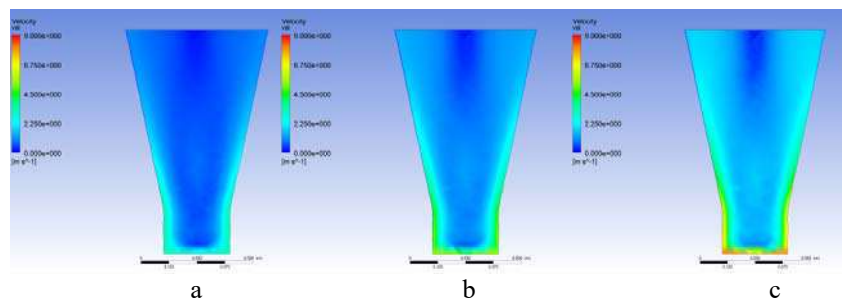


Fig. 9. The impact of granulator workspace design and gas flow velocity components on the gas flow total velocity and gas movement trajectory (working space "diffuser" with cylindrical insertion in the lower part), $d=300$ mm; $l_c=200$ mm; $L=1000$ mm, $\varphi=12^\circ$: a - $V_r=V_z=1$ m/s; $V_\varphi=3$ m/s; b - $V_r=V_z=1$ m/s; $V_\varphi=5,2$ m/s; c - $V_r=V_z=1$ m/s; $V_\varphi=8$ m/s.

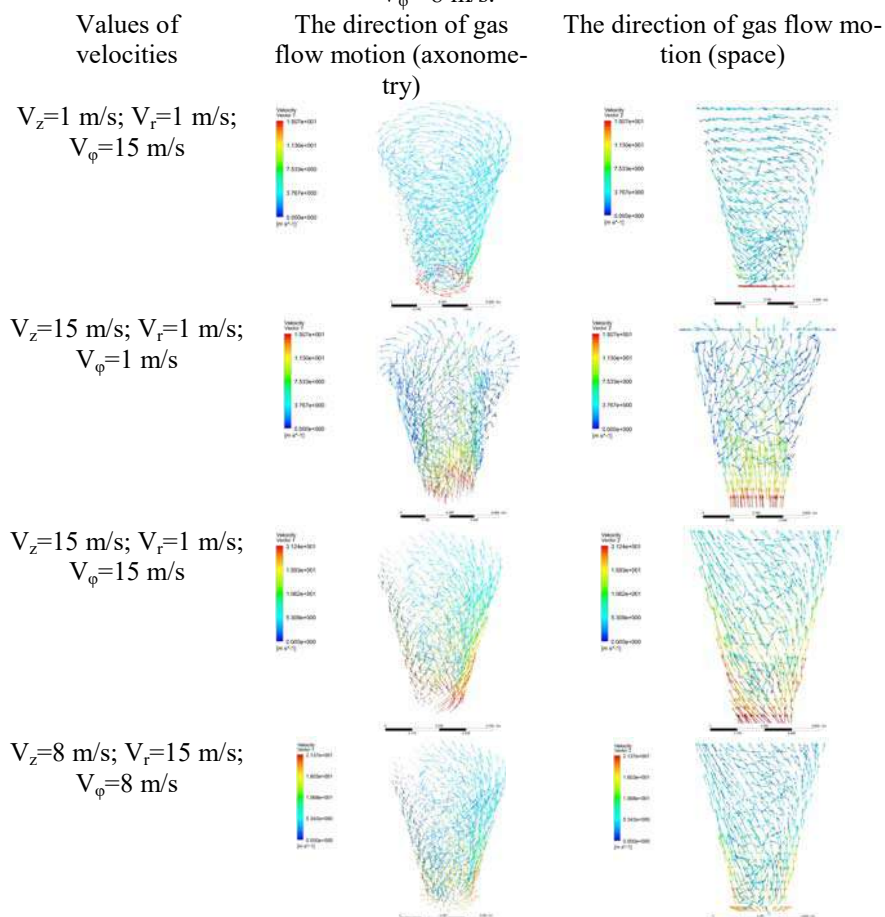


Fig.10. Gas flow movement vectors for different values of speed components.

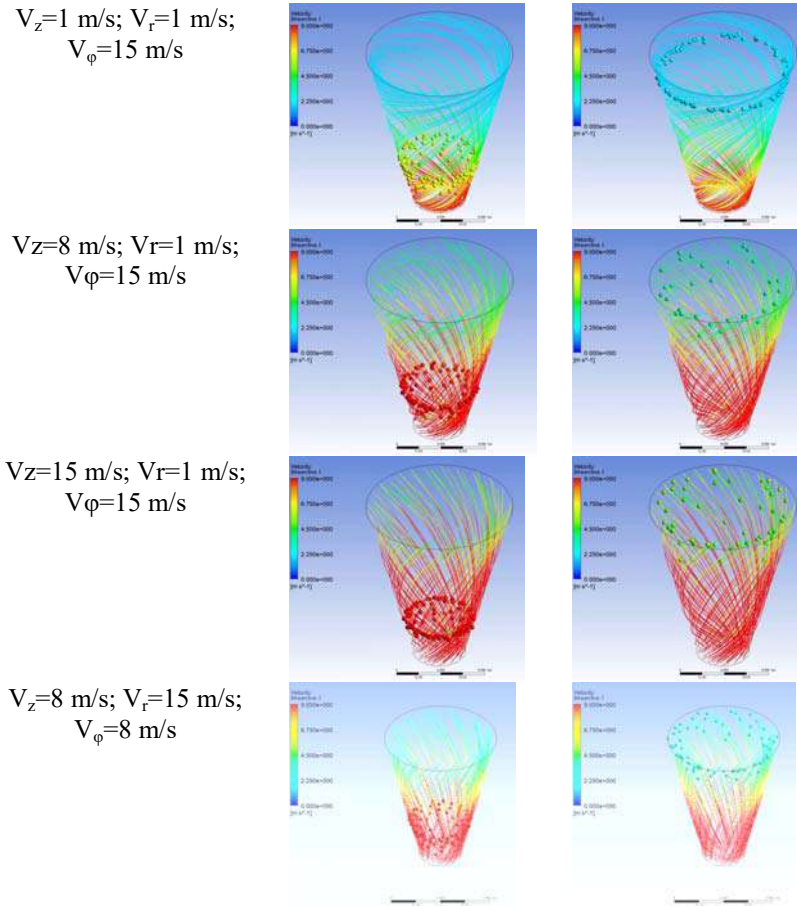


Fig.11. Gas flow movement trajectory for different values of speed components.

Analysis of graphic dependences of gas flow rate velocity under various conditions has identified monotonous nature of rate velocity reduction under all calculations conducting conditions.

Comparison of theoretical calculation for different cross sections of working space and different opening cone angles shows that the area of maximum circular speed for arbitrarily chosen section is $r = 0,66-0,72 R$.

Analysis of graphical dependencies shows the reduction zone of circular velocity growth by increasing the gas flow rate and narrowing the range of maximum speed. This velocity peak remains constant at geometric place in the working space of vortex granulator.

The value of circular velocity with increasing the gas flow rate increases faster. At the same time, with the increasing gas flow rate in workspace space of vortex granulator the angular velocity slow growth zone is growing. Thus, the vortex is moving toward solid wall and has a greater intensity.

Analysis of dependences of circular velocity changes under different conditions identified pattern of velocity distribution by the radius of device and its quality change according to the set of parameters.

Given the results of analysis of circular velocity peak of gas flow in cross section of device workspace for granular product from the melt (solution) device (or group of devices) for spraying is advisable to locate within the working space vortex granulator in specified range of current radius at random selected height and cone opening angle with the advice, received the results of experiment [2] (optimal height of location sprayer is defined to be $h = 0,6-0,8H$).

Continuously increasing the circular velocity of gas flow before reaching peak and reduction after passing its peak had obtained confirmation by experiment and defined range of gas distribution unit operation with maximum efficiency. In the area of minimum circular velocity the granules motion intensification and "dead" zones prevention, that is possible in the range up to $r = 0,25R$, is achieved by reasonable selection of gas-distributing unit with additional elements for flow swirling.

The value of gas flow radial velocity, received by the results of analytical solution, and experimental data have high degree of compliance. There is mutual rejection numerical value of speed and overall graphic image of diagrams of full speed radial component of gas flow has the same character. It is mentioned even radial velocity decreasing from the center (core) of gas flow to the periphery.

According to the analysis the influence of conical workspace opening angle on the possibility of reverse vortex in center of weighted layer and place a geometric location of individual elements weighted layer connection place and the transition region combined weighted layer connection was identified. With the increasing of cone opening angle (up to $10-13^\circ$) zone of reverse vortex has a constant value and is determined within $r = (0,15-0,25) R$. This situation is observed before $h=(0,3-0,4) H$. Recirculation zone is located in the inlet of diffuser.

In general in the following functional dependencies for each component of gas flow velocity it is observed is the same quality pattern. For rate and radial velocity components of gas flow for the initial conditions and changing set of geometrical and technological conditions does not cause such significant change of quantitative distribution value along the radius and height for both angular velocity component gas stream. Based on the analysis of hydrodynamic properties of gas flow due to certain variables the rational selection of geometry workspace vortex granulator becomes possible (opening angle α and height z).

Analysis of simulation using visualization software ANSYS CFX has showed:

- depending on gas flow velocity component value trajectory of its movement has a different configuration with the predominance of one or another direction;
- when the axial velocity component predominance the area of gas flow movement is narrowed;
- when velocity circular component predominance we observe an increasing zone of gas flow vortex motion in height;
- when the velocity radial component predominance gas flow movement to the wall of vortex granulator is made more intense;
- under certain conditions the intensity of gas flow initial twisting does not affect the trajectory of its movement, and affects only the value of gas flow resulting velocity;

- velocity radial component is characterized by maximum value on the axis of the device.

5 Conclusions

Simulation results allowed us to develop an algorithm for vortex granulator hydrodynamic calculating, which is represented on fig 12.

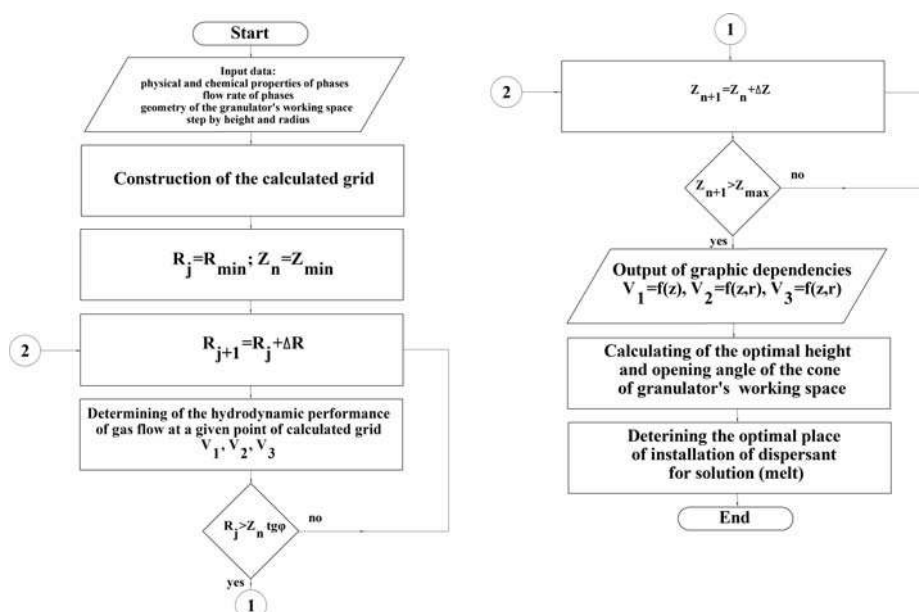


Fig. 12. An algorithm for vortex granulator hydrodynamic calculating

Example of granulator, designed based on hydrodynamic calculation (porous ammonium nitrate production, final product performance 3000 kg / day, commodity fraction - granules 2-3 mm) is shown in figs 13, 14.

The review of the modern software products (simulation modeling systems) used in the calculations for the chemical production plants (Aspen HYSYS®, PRO/II, Gas-CondOil, ChemCAD, DESIGN II, SIGMAFINE, gPROMS ModelBuilder, GIBBS and others) [18-25] shows that such software is mainly implemented for the technological calculation of the specified construction and under certain technological parameters of the chemical plants. The optimization calculation of the technological process requires a multiple initial provisioning; a quantity of such provisions for the optimization calculations may be numerous. The simulation software used for the hydrodynamic and heat-mass transfer processes in the chemical production (ANSYS Fluent, COSMOS FlowWorks, FlowVision, Flow-3D CFD, Ansys CFX, COMSOL Multiphysics and others) [26-31] enable to model only one construction variant of the installation and, thus, it reduces the optimization calculation performance. In practice, we use various models of the numerical solution of the classical equation of hydrodynamics and heat-mass transfer for turbulent flows.

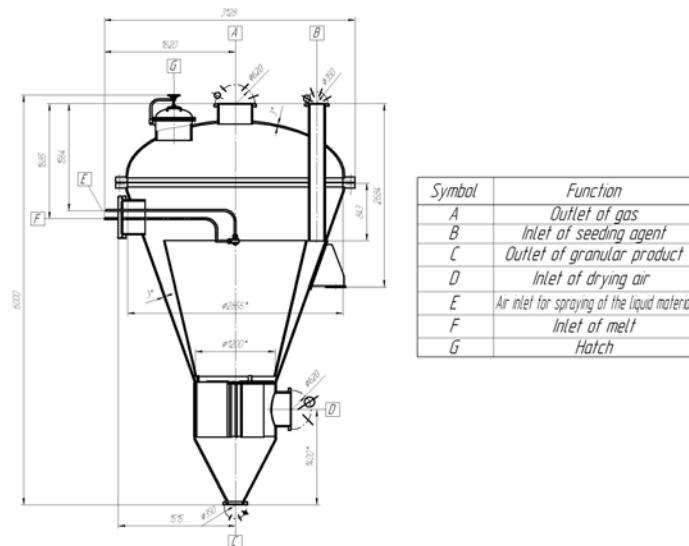


Fig.13. The industrial design of vortex granulator.

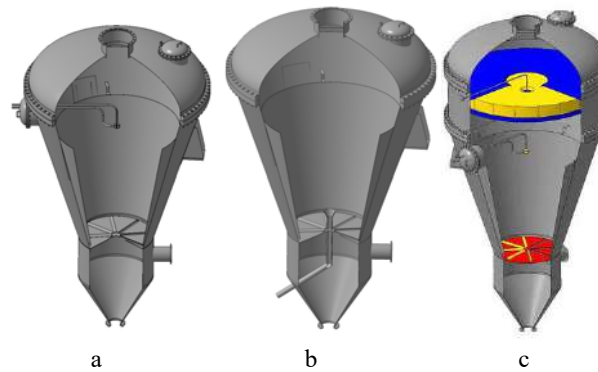


Fig. 14. Vortex granulator designs: a – with spraying of the melt; b – with a previous wetting of granules and simple inner case; c – with separating device for exhaust gases cleaning

Creating a comprehensive algorithm for determining the hydrodynamic characteristics of vortex flows (the author's software product and Ansys CFX) allows us to achieve high accuracy results and choose the optimal granulator configuration following the optimization criterion. This approach is more effective than using only the software products which are mentioned above.

6 Acknowledgments

This work was carried out under the project «Small-scale energy-saving modules with the use of multifunctional devices with intensive hydrodynamics for the production, modification and encapsulation of granules», project No. 0119U100834. While pre-

paring the article it was used information about the original designs of vortex granulators (patents Nos 99023 Ukraine IPC (2012.01) B01J2/16, Ukraine IPC 110992 B01J 2/00, B01J 2/16, 111406 Ukraine IPC B01J 2/04, B01J 2/16 etc.) and author's software product (certificate of authorship No 67472).

7 References

1. Artyukhov, A.E., Sklabinskyi, V.I.: Experimental and industrial implementation of porous ammonium nitrate producing process in vortex granulators. *Naukovyi Visnyk Natsionalnoho Hirnychoho Universytetu*, 6, 42-48 (2013).
2. Artyukhov, A.E., Fursa, A.S., Moskalenko, K.V.: Classification and separation of granules in vortex granulators. *Chemical and Petroleum Engineering*, 51, Nos 5–6, 311–318 (2015). <https://doi.org/10.1007/s10556-015-0044-x>
3. Shi, De-Pan, Luo, Zheng-Hong Guo, An-Yi.: Numerical Simulation of the Gas–Solid Flow in Fluidized-Bed Polymerization Reactors. *Industrial & Engineering Chemistry Research*, 49 (9), 4070–4079 (2010). <https://doi.org/10.1021/ie901424g>
4. Pandaba, P., Sukanta, K.D.: Numerical Simulation for Hydrodynamic Analysis and Pressure Drop Prediction in Horizontal Gas-Solid Flows. *Particulate Science and Technology Volume*, 32(1), 94-103 (2014). <https://doi.org/10.1080/02726351.2013.829543>
5. Feldmann, F., Hagemann, B., Ganzer, L., Panfilov, M.: Numerical simulation of hydrodynamic and gas mixing processes in underground hydrogen storages. *Environmental Earth Sciences*, 75, 1165-1172(2016). <https://doi.org/10.1007/s12665-016-5948-z>
6. Artyukhov, A.E.: Optimization of mass transfer separation elements of columnar equipment for natural gas preparation. *Chemical and Petroleum Engineering*, 49, Nos. 11-12, 736-740 (2014). <https://doi.org/10.1007/s10556-014-9827-8>
7. Prokopov, M.G., Levchenko, D.A., Artyukhov, A.E.: Investigation of liquid-steam stream compressor. *Applied Mechanics and Materials*, 630, 109-116 (2014). <https://doi.org/10.4028/www.scientific.net/AMM.630.109>
8. Tenneti, S., Subramaniam, S.: Particle-resolved direct numerical simulation for gas-solid flow model development. *Annual Review of Fluid Mechanics*, 46, 199-230 (2014). <https://doi.org/10.1146/annurev-fluid-010313-141344>
9. Yang, J., Wang, Q., Zeng, M., Nakayama, A.: Computational study of forced convective heat transfer in structured packed beds with spherical or ellipsoidal particles. *Chemical Engineering Science*, 65(2), 726-738 (2010). <https://doi.org/10.1016/j.ces.2009.09.026>
10. Sinaiski, E.G.: *Hydromechanics: theory and fundamentals*. Weinheim: WILEY-VCH Verlag GmbH & Co. KGaA (2010).
11. Artyukhov, A.E., Fursa, A.S., Moskalenko, K.V.: Classification and separation of granules in vortex granulators. *Chemical and Petroleum Engineering*, 51, Nos 5–6, 311–318 (2015). <https://doi.org/10.1007/s10556-015-0044-x>
12. Artyukhov, A.E., Sklabinskyi, V.I.: 3D Nanostructured Porous Layer of Ammonium Nitrate: Influence of the Moisturizing Method on the Layer's

Structure. Journal of Nano- and Electronic Physics, 8, No 4(1), 04051-1–04051-4(2016). [https://doi.org/10.21272/jnep.8\(4\(1\)\).04051](https://doi.org/10.21272/jnep.8(4(1)).04051)

13. Artyukhov, A.E., Sklabinskyi, V.I.: Investigation of the temperature field of coolant in the installations for obtaining 3D nanostructured porous surface layer on the granules of ammonium nitrate. Journal of Nano- and Electronic Physics, 9(1), 01015–01015-4 (2017). [https://doi.org/10.21272/jnep.9\(1\).01015](https://doi.org/10.21272/jnep.9(1).01015).
14. Patankar, S.V.: Numerical heat transfer and fluid flow. Washington, DC: Hemisphere Publishing Corp (1980).
15. Artyukhov, A., Sklabinskyi, V.: Theoretical analysis of granules movement hydrodynamics in the vortex granulators of ammonium nitrate and carbamide production. Chemistry & chemical technology, 9 No 2, 175-180 (2015). <https://doi.org/10.23939/chcht09.02.175>
16. Artyukhov, A., Sklabinskyi, V.: Hydrodynamics of gas flow in small-sized vortex granulators in the production of nitrogen fertilizers. Chemistry & Chemical Technology, 9 No 3, 337-342 (2015). <https://doi.org/10.23939/chcht09.03.337>
17. Computational Fluid Dynamics: ANSYS CFX and FLUENT CFD Software [Electronic resource]. Access mode: <https://caei.com/ansys-software-support/ansys-software/computational-fluid-dynamics-ansys-cfx-and-fluent-cfd-software>.
18. Aspen HYSYS® [Electronic resource]. Access mode: <http://www.aspentech.com/products/aspen-hysys/>
19. SimSci PRO/II [Electronic resource]. Access mode: <http://software.schneider-electric.com/products/simsci/design/pro-ii/>
20. GasCondOil [Electronic resource]. Access mode: <http://thermogas.kiev.ua/Products.htm>
21. ChemCAD [Electronic resource]. Access mode: <http://www.chemstations.com>
22. DESIGN II [Electronic resource]. Access mode: <https://www.winsim.com/design.html>
23. SIGMAFINE [Electronic resource]. Access mode: <https://sigmafine.pimsoftinc.com>
24. gPROMS ModelBuilder [Electronic resource]. Access mode: <https://www.psenterprise.com/products/gproms>
25. GIBBS [Electronic resource]. Access mode: <http://www.gibbsim.ru>
26. ANSYS Fluent [Electronic resource]. Access mode: <http://www.ansys.com/Products/Fluids/ANSYS-Fluent>
27. COSMOS FlowWorks [Electronic resource]. Access mode: <http://www.solidworks.com/sw/products/simulation/flow-simulation.htm>
28. FlowVision [Electronic resource]. Access mode: <http://tesis.com.ru>
29. Flow-3D CFD [Electronic resource]. Access mode: <https://www.flow3d.com>
30. Ansys CFX [Electronic resource]. Access mode: <http://www.ansys.com/products/fluids/ansys-cfx>
31. COMSOL Multiphysics [Electronic resource]. Access mode: <https://www.comsol.com>

Complex Technical System Condition Diagnostics and Prediction Computerization

Nicolay Rudnichenko¹ [0000-0002-7343-8076], Vladimir Vychuzhanin¹ [0000-0002-5244-5808],
Andrii Polyvianchuk² [0000-0002-9966-1938], Vasyl Mateichyk³ [0000-0002-3683-7246]

¹Odessa National Polytechnic University, Odessa, Ukraine
nickolay.rud@gmail.com, vint532@yandex.ua

²Beketov National University of Urban Economy in Kharkiv, Kharkiv, Ukraine
ap3@ukr.net

³National Transport University, Kyiv, Ukraine
wmate@ukr.net

Abstract. Based on the analysis of literary sources, the article sets the goal of forming a methodological application of information technology in diagnostics, as well as in predicting the state of complex technical systems. It is advisable to carry out a diagnostic assessment of the system failures risk based on modeling their components interaction. To achieve this goal, an informational cognitive model has been developed that allows for diagnosis to assess complex technical systems components failure risk. In order to provide a search for the failures causes of a complex technical system diagnosed subsystems components, a decision support model has been developed and researched. Using the developed informational cognitive model for diagnosing complex technical systems, the method and decision support model allows us to: diagnose the risk values of system component failures when information about component failures is received; to predict system components failure risk value in order to select a strategy for their recovery; support decision making when searching for the causes of system component failures. The developed methods and models, the proposed solutions for informatization of diagnostics and prediction of the complex system technical condition provide flexibility and adaptability.

Keywords: complex technical system, diagnostics, simulation, cognitive model, decision support

1 Introduction

Designed complex technical systems (CTS) are characterized by multicomponent structural and functional complexity. The increasing complexity of CTS requires the development of new methods to ensure such systems reliability. Reliable CTS operation isn't possible without diagnostic tools and predicting the technical condition of systems. Diagnosis and prediction CTS's state with an assessment components failure risk needs information support based on modern advances in information technology.

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

2 Description of Problem

Currently, it remains relevant to develop new methods based on information technology applications in the CTS diagnosis with an assessment of their components failure risk [1-3]. A widely used method for predicting system's technical condition based on modeling using time series CTS parameters characterizing [4]. Advantages and disadvantages of such methods are described in several publications [5,6]. When forecasting based on retrospective data, machine learning systems based on artificial neural networks are used [7]. For monitoring, diagnosing the causes of CTS failures, forecasting, deep learning methods are also used [8-10]. Intelligent methods used in CTS diagnostics and prediction systems include evolutionary programming. Modifications to the Bayesian approach and Bayesian networks can be used to predict CTS accidents and failures occurrence. In [11], a dynamic Bayesian trust network is presented, which allows predicting the values of failure probabilities and searching for defects and malfunctions in decision support systems.

At present, simulation modeling (SM) is widely used, which makes it possible to experiment with the analytic-probabilistic model, exploring various situations and simplifying the decision-making process. Simulation software in particular, SM products, such as Arena, AutoMod, AnyLogic, Extend, GPSS World and others, contributes to the widespread are widely used for research tasks. However, software tools facilitate the process of diagnosing and predicting CTS components failures risk, but do not facilitate the solution of the time-consuming task of collecting initial information, its interpretation, formalization, and an adequate ratio with a specific CTS. A promising SM method for studying systems reliability during their transitions between different state variants is cognitive simulation (CIM) based on models in the form of oriented graphs that reflect the CTS components interaction [12-17].

The analysis CTS cognitive models for both diagnosis and prediction system components failures risk showed the need to develop an informational cognitive model for diagnosing CTS components failures risk as a whole, a forecasting model and decision support when searching for the causes diagnosed systems components failures. Prospects for the further development of CIM methods for diagnosing and predicting CTS components failures risk with interpretation, formalization and an adequate ratio of incoming information about the technical condition of the system. In order to interpret, formalize and adequately correlate the incoming information during the study of the reliability of systems during their transitions between different state options, further improvement of the software for diagnostics and prediction CTS components failures risk is necessary. It is necessary to develop methods and models, new solutions for informatization of diagnostics and predicting the technical condition of a complex system to ensure: flexibility - the ability to use methods at any level to assess CTS subsystems components failure risk with their various configurations; adaptability - methods must have the ability to adapt when changing the configuration of CTS subsystems.

3 Information software for the diagnostics, forecasting and decision support CTS subsystem failures risk assessment

3.1 System concept

The concept of the CTS risk failures assessing in emergency scenarios is based on combining heterogeneous CTS components into a single model [18-20]. The model should provide an CTS failures risk assessment taking into account the interconnect- edness and interaction of their components in regard to significance and criticality for the entire system functioning as a whole, and also ensure the identification of struc- tural threats and vulnerabilities in the CTS.

The cognitive model can be represented as a functional graph

$$F_{gr}(G, X, F, Q) \quad (1)$$

where $G = \langle V, T, E \rangle$, G - sign oriented graph; $V = \{v_i\}$, $i=1,2,\dots,k$ - cognitive map vertices; $E = \{e_{ij}\}$ - many arcs connecting the vertices v_i and v_j ; T - time; $X = \{x_i\}$ - vertex parameters; $F = f\{v_i, v_j, e_{ij}\}$ - connection function between vertices; Q - vertex parameter space.

As a measure of damage to an undesirable event, it is proposed to determine the structural damage of components and inter-component communications (IC) in accor- dance with the method for assessing the CTS structural failure risk. Performing diag- nostics to assess the components failures risk and the IC of the ICE with the subsys- tems, failures probabilities for the system's components and IC are preliminarily de- termined. For this, statistical data obtained for a fixed time, containing information on the number of component and IC outages, is used. The components and IC failure probabilities CTS sub-systems are determined as

$$P_{v_i} = \frac{n_{v_i}}{\tau}, \quad P_{a_i} = \frac{n_{a_i}}{\tau}, \quad (2)$$

where p_{v_i} - the i -th component failure probability; p_{a_i} - the i -th IC failure probability; n_{v_i} - the i -th component number of failures; n_{a_i} - the i -th IC number of failures ; $\tau = 10^6$ ч. - statistical testing period.

3.2 Modelling CTS diagnosis software development

When implementing CIM, it is proposed to use developed application software for modeling is based on the client-server architecture. A striking modeling impulse on the system effect was applied for the CTS diagnosis. For this, the distribution of the operating system Debian GNU / Linux 8.0 (stable) is used. Python was used as the programming language. Data on the CTS components is hosted in the NoSQL Mon- goDB. Data exchange between the client and server side is carried out using the Rest- full API. The initial data of the models are presented in JSON format. Automation of the system was carried out on the basis of GNU make tools. Analysis of the results

was carried out by Calc Libre Office means. The diagrams of use cases are used to determine the general boundaries and context of the simulated domain at the developing software initial stages for assessing the risk of CTS failures, damage from loss of system operability and formulating general requirements for its behavior (Fig. 1). Software use cases diagram allows us to develop project main entities models by building class diagrams. A software class diagram fragment is shown in Fig. 2.

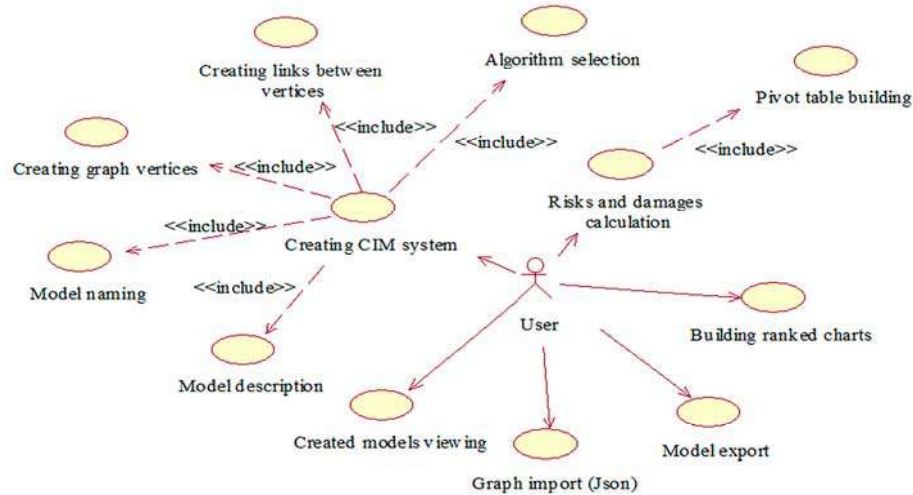


Fig. 1. Diagnosis software use case diagram

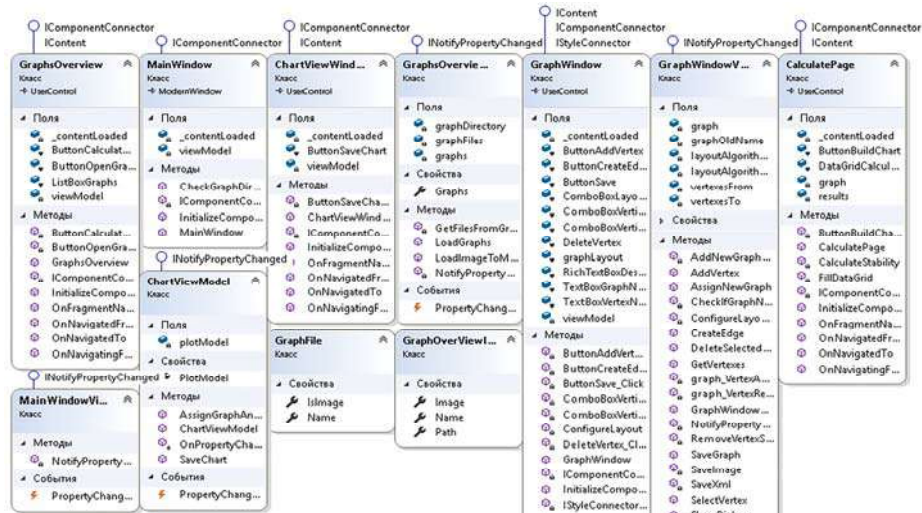


Fig. 2. Diagnosis software class diagram

The classes GraphsOverview, MainWindows, ChartViewWindow, GraphsOverviewView, GraphWindow, GraphWindowView, CalculatePage, MainWindowsView, and ChartViewModel implement a number of interfaces to provide the

necessary visualization functions for the processed and calculated data. To implement the indicated logic, the following interfaces for CIM were used and involved in the form of a graph: IComponentConnector (to ensure the connection of element objects among themselves); IContent (for displaying and implementing functions of dynamic linking and drawing objects of a working graphic container on the corresponding form); INotifyPropertyChanged (for handling events created graphic objects properties changing); IStyleConnector (to change the type of communication between components). The features of the system's physical representation are described in the relationship between system basic components formalization. To do this, use the component diagram (Fig. 3) allows us to determine system architecture, taking into account the relationship between software components, which may be the source, binary and executable code. The main module (MainApp) carries out the functions of calling the appropriate modules to process user requests for: constructing a graph model (GraphBuilder) based on the use of external dependencies Graph # and Oxy-Plot, as well as the WPF GraphLayout module to ensure the interactive visualization container operation with received models calculation system failures damage and risk; viewing the results in tabular form to evaluate their values; construction and viewing a graph showing calculations results in a ranked form..

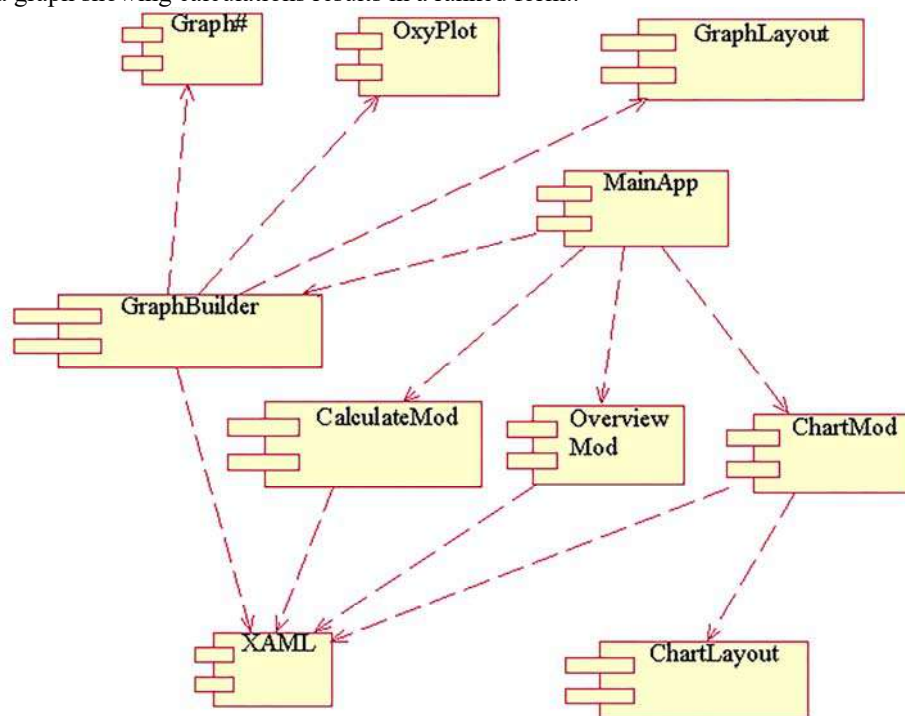


Fig. 3. Diagnosis software component diagram

As graphic libraries connected in external dependencies mode, Graph # and Oxy-Plot were used. Graph # library for graph visualization, containing some layout algo-

rithms, as well as for controlling GraphLayout WPF applications. Supported build and delete algorithms: Fruchterman - Reingold; Kamada - Kawai; ISOM LinLog; simple tree layout; Sugiyama; Force-Scan allocation algorithm. In order to simulate the interaction of objects in the designed software over time, as well as messages exchange between them, a sequence diagram (Fig. 4) is used. Each of the above forms (except the main one) is a separate fragment dynamically loaded in a single space in the new tab form (TabPane object), located at the main form. The basic component of the graph structure construction method is the Sugiyama algorithm.

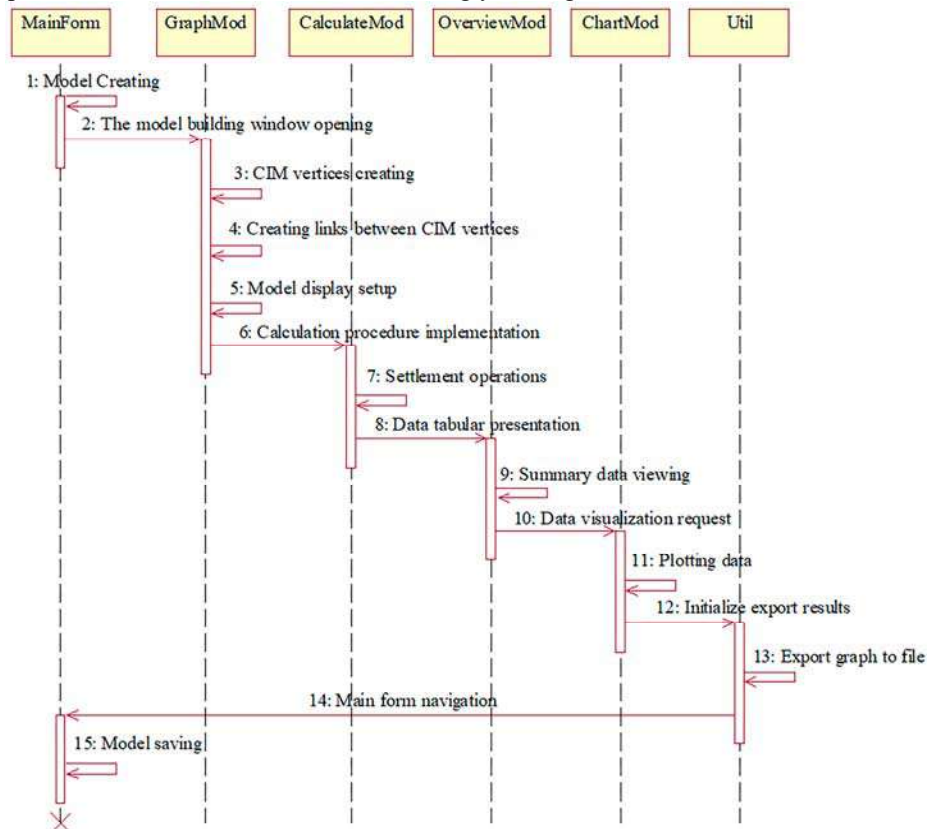


Fig. 4. Diagnosis software flow chart

In order to describe the CTS diagnosis software functionality, we denote the key classes that implement created software's business logic:

1. Calculations class CalculatePage: UserControl, IContent, IComponentConnector. It interprets and uses the obtained results of constructing a system model in the graph form to assess damage and the failures risk.
2. A class for constructing and displaying a ranked graph of calculated failure risk values ChartViewModel: INotifyPropertyChanged.
3. The CIM construction class in the graph form GraphWindowViewModel: INotifyPropertyChanged.

The system resulting graph model display class Graphs Overview: UserControl, IContent, IComponentConnector. The developed CTS diagnosis software used for research allows the user to:

- create a CTS CIM in the graph form, with support for the model name functions, setting a brief description, creating a new vertex and its image on the panel, creating a connection between the vertices, choosing an algorithm for positioning and displaying the structure in a graphic container;
- view the structure of the previously created CIM in the form of a graph with the vertices and edges total number display, supporting the loading operation into the program workspace;
- import a graph in *.json format for its visualization in the system;
- export CIM as a graph to a separate graphic file in *.jpg format;
- calculate the damage from failures values and the simulated CTS components failures risks and display the results in a summary table;
- build graphs obtained results visualization in a ranked form.

CTS subsystems are a dynamic structure, because their components have a different wear degree and change their characteristics at different speeds. All this leads to the following requirements for the decision-making support method (DSM) used [11,20] when searching diagnosed CTS subsystems components failures causes: flexibility - the ability to use the method at any level to assess the ICE components failure risk with their various configurations; adaptability - the method should have the ability to adapt to changes in the CTS subsystems configuration. To support decision-making on failure ICE subsystems risk assessments as well as when searching for failed system components, a method based on dynamic Bayesian trust networks (DBTN) is used [21]. The method has several advantages over other methods in assessing the likelihood of a CTS failure-free operation. The use of DBTN allows us to: assess the risk of CTS failures upon receipt of a new information about element failures; to predict the value of the risk of CTS failures in order to choose a system recovery strategy; support decision making when searching for the cause of component failure and the CTS as a whole. The proposed approach, regardless of the structure of the CTS, is based on the fact that the objective function of assessing the health of the objective function CTS components efficiency assessing through DBTN is

$$F(P_b) = \{G, M\}, \quad (3)$$

where G – acyclic directed network graph; M – ICE subsystems DBTN components.

The graph vertices are CTS's subsystems, which, taking into account the hierarchy, are determined as

$$v = \{v_i^j | \overline{1, n}, j = \overline{1, m}\}, \quad (4)$$

where v – CTS element name; i – network block number, n – number of blocks in the network; j – network level number; m – levels number in the network.

Based on the component location in the DBTN structure, two types of graph elements (vertices) are possible: parents ($v_i^j = \text{parent}(v_i^{j+1})$) variety vertex v_i^j , ver-

tex v_i^{j+1} , providing parent connections v_i^j to child element v_i^{j+1} ; child element ($v_i^{j+1} = \text{children}(v_i^j)$) variety vertex v_i^{j+1} , vertex - v_i^j , providing relationships between a child v_i^{j+1} and parent v_i^j .

CTS subsystems diagnosed DSM model implementation is carried out in accordance with the decision support algorithm for assessing CTS subsystems failure risk and consists in constructing the CTS DBTN using the following databases: design and regulatory documentation; expert evaluations of decision support for typical failure risk scenarios; decision support criteria; statistics of diagnosed data on the static and dynamic CTS subsystems elements characteristics.

A data sample is formed for a specific scenario CTS subsystems components data failure risk occurrence during analyzing procedure. Data is then interpreted and processed using blocks for acquiring knowledge, supporting decision making, and managing rules. As a result, they are replenished with new bases and knowledge data, data and rules, which then enter the analysis decisions unit.

4 Experiments and results analysis

4.1 Developed simulation model research

An automobile internal combustion engine system was chosen as an example of the diagnostic method practical implementation for assessing the CTS subsystems failure risk. Automobile internal combustion engine (ICE) as a directed graph diagram with subsystems is shown in Fig. 5.

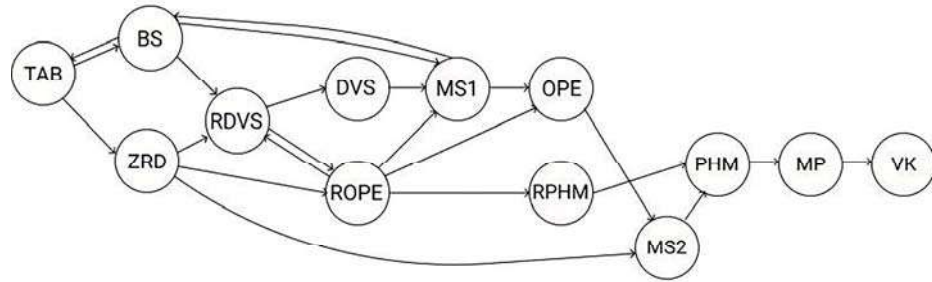


Fig. 5. Internal combustion engine directed graph diagram

The scheme of an internal combustion engine with subsystems (Fig. 5) consists of: TAB – traction battery; DVS – internal combustion engine; ZRD – drive mode dial; BS – unit for summing voltages and powers; OPE – reversible energy converter; PHM – speed and torque converter; MP – mechanical transmission; VK – driving wheels; MSI – clutch between DVS and OPE shafts; MS2 – clutch between OPE and MP shafts; ROPE – OPE knob; RPHM – regulator PHM; RDVS – DVS controller.

CIM modeling results let us to evaluate failures risk for all internal combustion engine's elements with subsystems and rank the calculation results (Fig. 6). From the internal combustion engine elements structural damages results it follows that the most critical elements are the elements TAB, BS and RDVS. This is due to the high values of their structural damage (1.0, 0.85 and 0.75). Less critical elements are elements MP, VK, namely mechanical transmission, drive wheels; having slightly lower structural damage values (0.15 and 0.05).

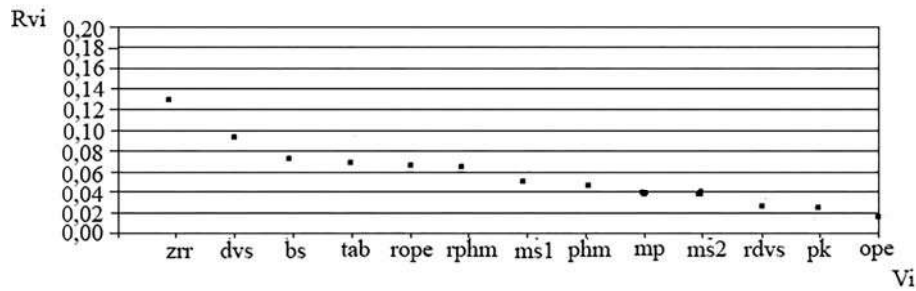


Fig. 6. Structural risk results ranking of CTS elements with sub-systems

It also follows that the engine itself belongs to the most ICE vulnerable elements with subsystems, based on the obtained values of the element failure risk (0,089). Less vulnerable ICE elements with subsystems include the OPE regulator (0.02).

The construction and study of the CTS subsystems component failure risk assessment DBTN was carried out using the GiNIe software product [22].

4.2 Decision support method CTS subsystems failure risk assessing research

Simulating ICE subsystems DBTN (Fig. 7) for different values of the subsystem failure probability (risk) at the CTS model input the risk values for 20,000 hours of ICE operation was determined (Fig. 8).

Research results let us define that with an increase subsystem's failure risk at the CTS model input from 0.09 to 0.26, the daughter CTS failures risk values increase notably.

Simulation model for assessing the failure risk during the ICE subsystems operation study showed that even a relatively small number of the subsystems considered components generate a large possible scenarios number and options leads to extreme situation when any component might be damaged. When models are supplemented with indicators of real subsystems criticality and spatial arrangement, models scale increases several times. The studied subsystems scale enlargement leads to a further increase in sub-systems emergency conditions.

The practical implementation of the proposed method for diagnosing the risk of failure of ICE subsystems can easily apply to any CTS structure, which has any complexity degree, any relationships between CTS components and subsystems.

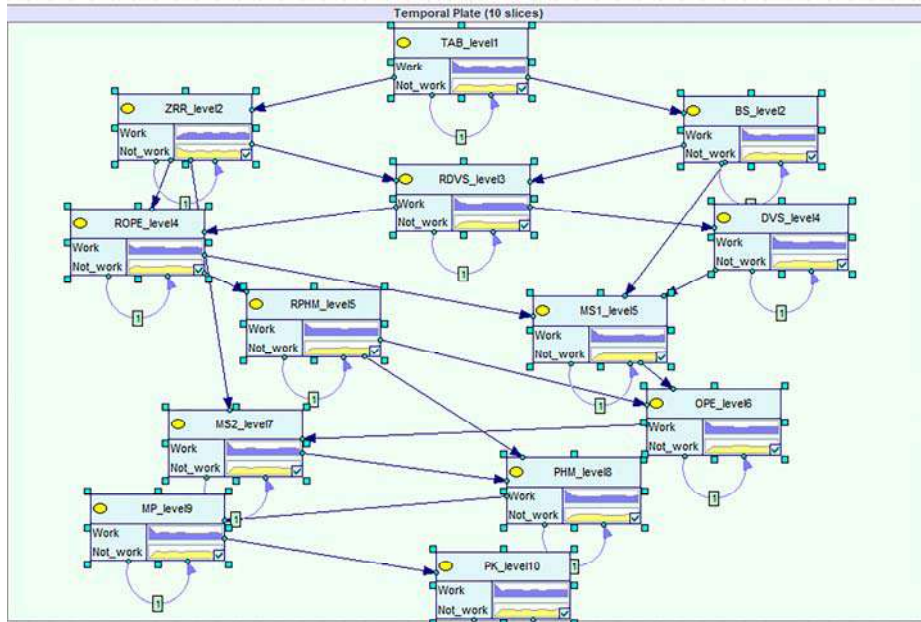


Fig. 7. DBTN for ICE subsystems in the GeNIe environment when determining CTS subsystem failures risk (subsystem failure risk at the model 0.26 input)

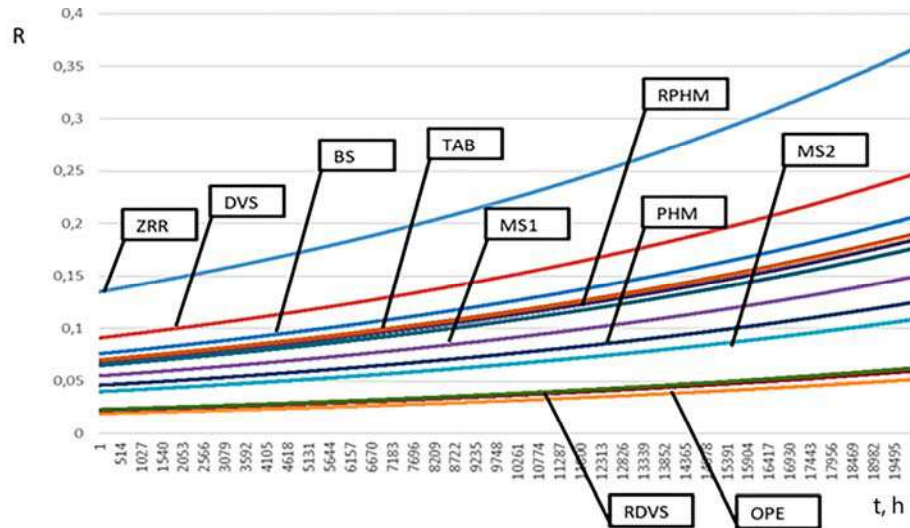


Fig. 8. ICE subsystems failure risk with the subsystem operability loss probability at the system input 0.26

4.3 Practical implementation and research of the decision support forecasting method and model in CTS subsystems finding failures causes

Practical implementation and research of the decision support forecasting method and model in CTS subsystems finding failures causes are provided on ICE example. When conducting studies diagnosed ICE subsystems DSM model each subsystem influence degree on the probability of performance loss the both subsystems and CTS failures risk as a whole was determined.

From a retrospective research results analysis the subsystems were installed to the greatest extent, affecting the overall system performance.

Emergency situations studying and CTS events analysis has the main goal to determine the cause of the system performance loss. From the research results it follows that the maximum failure risk during the subsystems operation is 20,000 hours for the ZRR subsystem. The ZRR subsystem is interdependent in operation from other ICE subsystems. Therefore, it is necessary to check the subsystem in order to find the failure cause. To identify the possible cause of the ZRR failure, relevant research processes were carried out using the search scheme for the ZRR subsystem failure cause, shown in Fig. 9.

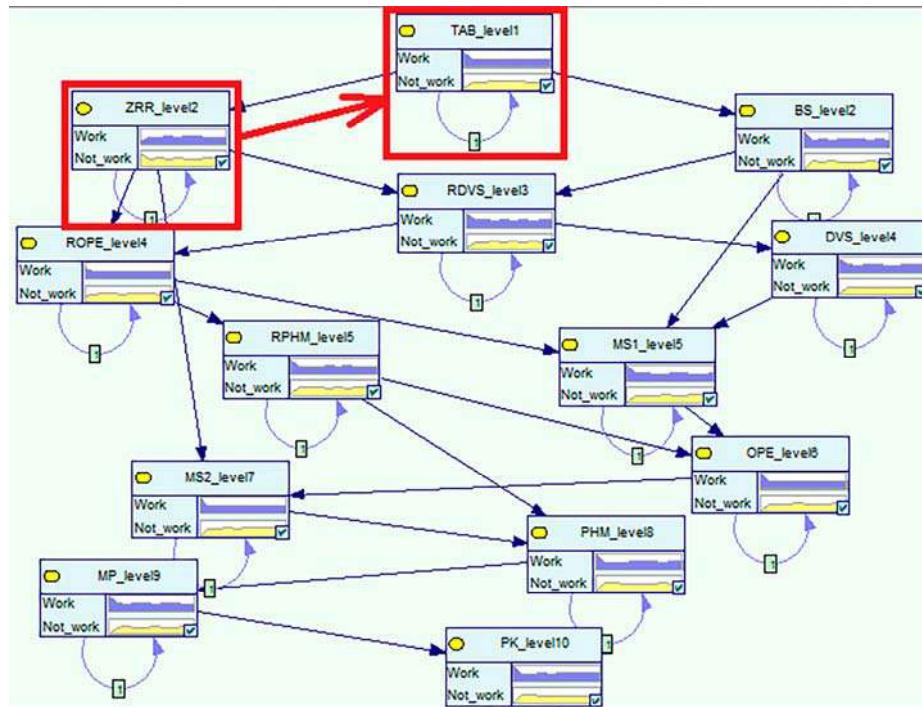


Fig. 9. ZRR subsystem cause failure search scheme

The ZRR subsystem failure cause search was performed in accordance with the algorithm shown in Fig. 10.

The purpose of the use of DBTN in assessing both the performance loss probability and the subsystems failure risk is an a posteriori conclusion. It consists in the fact that upon information receipt about subsystems failures, a priori failures probability (risk) that is incompatible with the evidence and is equal to zero.

A priori data are listed and form in a posteriori failures probability (risk) estimate, which is a priori data for processing new information. The posterior conclusion is based on data analysis procedures resulting from the DBTN usage.

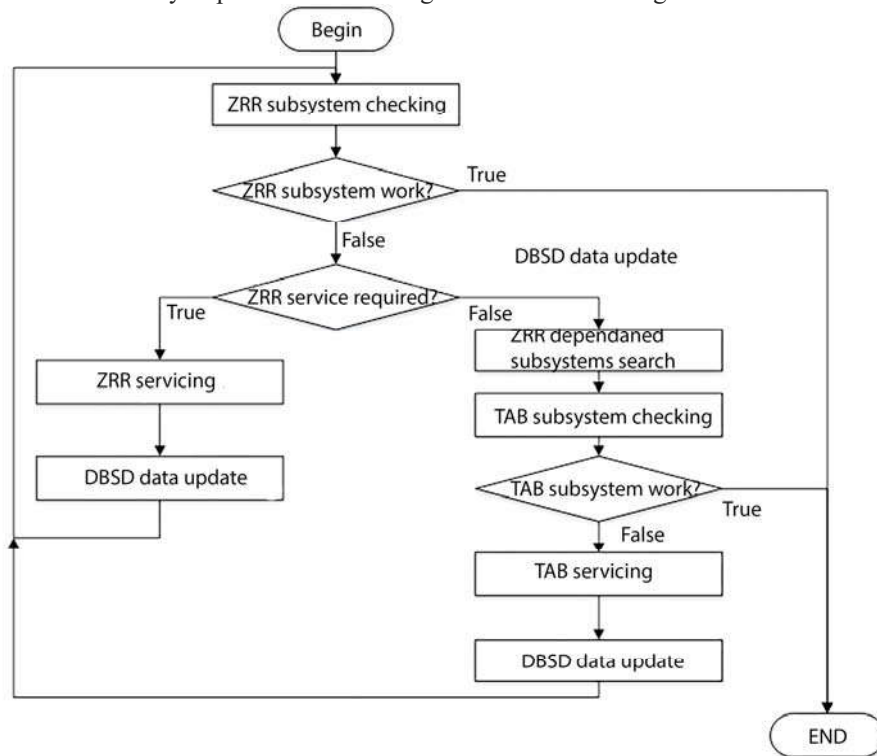


Fig. 10. ZRR subsystem failure cause search algorithm

When this approach is implemented, studies based on a priori and posterior data modeling have determined CTS subsystems for various time intervals that have the vastly impact on the CTS performance. It has been established that BS and ZRR belong to such subsystems (Fig. 11, 12).

The performed researches allow us to obtain algorithmic and methodological support for making informed decisions at the stage CTS subsystems operation in any complexity. The used troubleshooting algorithm in CTS provides: finding technically critical subsystems at all system levels, which maintenance must be performed immediately; troubleshooting time optimization.

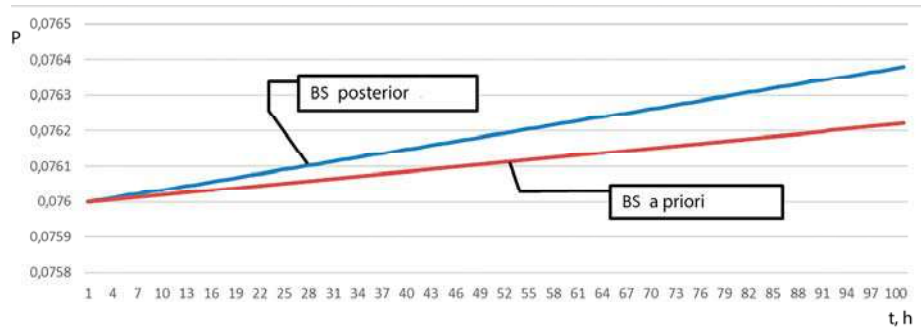


Fig. 11. A priori and a posteriori probability BS performance loss upon receipt of information about failures in ICE subsystems

Developed models research results application for the CTS emergency situations retrospective analysis purpose allows us to solve the determining their causes problem.

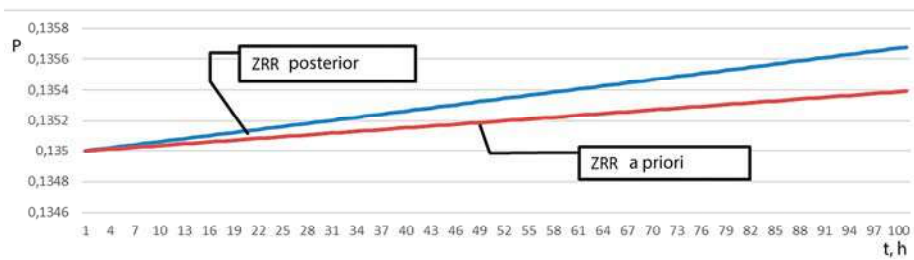


Fig. 12. A priori and a posteriori ZRR performance loss probability upon receipt of information about failures in ICE subsystems

This becomes especially relevant when the accident analysis the identification of its root cause and appropriate measures adoption eliminates or reduces the adverse events recurrence likelihood, which means that it will fulfill the task and increase the CTS components operation reliability.

5 Conclusion

Using the developed methods and models of diagnostics, forecasting and decision support when searching CTS diagnosed subsystems components failures causes allows you to: diagnose the system components risk failures values upon receipt of information about failures in subsystems; to predict the risk of system components failure in order to select a strategy for their recovery; support decision making when searching for the causes of system components failures; increase the efficiency of the

operation of the CTS as a result of subsystems elements prone to failure early identification.

The application of the developed software for the CTS diagnostic processes with an assessment system failures risk makes it possible to identify the least efficient components and inter-component communications, the functioning of which significantly affects the operability and reliability of the entire system.

The developed methods and models, the proposed solutions for informatization of diagnostics, as well as forecasting the technical condition of a complex system, unlike existing approaches, provide: flexibility - the ability to use methods and models at any level to assess CTS subsystems components failure risk with their various configurations; adaptability - methods and models have the ability to adapt when changing the configuration CTS subsystems.

Reference

1. Vychuzhanin, V.V., Rudnichenko N.D.: *Metody informatsionnykh tekhnologiy v diagnostike sostoyaniya slozhnykh tekhnicheskikh sistem*. Monografiya, Odessa (2019).
2. Tsvetkov, V.YA.: *Slozhnyye tekhnicheskiye sistemy*. Obrazovatel'nyye resursy i tekhnologii, №3(20), pp. 86-92 (2017).
3. Vychuzhanin, V.V., Rudnichenko N.D.: Assessment of risks structurally and functionally complex technical systems. *Eastern-European Journal of Enterprise Technologies*, Vol.1, №2, pp. 18-22 (2014).
4. Shevtsov, YU.D. Dudnik, L.N., Aref'yeva, S.A., Fadeyev, Ye.D.: Prognozirovaniye parametrov tekhnicheskogo sostoyaniya dvigateley energeticheskikh ustanovok. *Politehmicheskiiy setevoy elektronnyy nauchnyy zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta*, № 132, pp. 508-517 (2017).
5. Bukreyev, V.G., Kolesnikova, S.I., Yankovskaya A.Ye.: *Vyyavleniye zakonomernostey vo vremennykh ryadakh v zadachakh raspoznavaniya sostoyaniy dinamicheskikh ob'yektov*: monografiya., Tomsk (2010).
6. Vychuzhanin, V.: Evaluation model of marine engineering systems survivability. *NUS Journal. Electronic Edition*, №3, pp. 62-67 (2012).
7. Shin, S.J., Meilanitasari, P.: Developing a big data analytics platform for manufacturing systems: architecture, method, and implementation. *International Journal of Advanced Manufacturing Technology*, pp.1-42 (2018).
8. Wang, J., Ma, Y., Zhang, L., Gao R.X.: Deep learning for smart manufacturing: Methods and applications. *Journal of Manufacturing Systems*, Vol. 48, pp. 144-156 (2018).
9. Park, J.K., Kwon, B.K., Park, J.H., Kang D.J.: Machine learning-based imaging system for surface defect inspection. *International Journal of Precision Engineering and Manufacturing-Green Technologies*, Vol. 3, No. 3, pp. 303-310 (2016).
10. Zhao, R., Yan, Z., Chen, Z, Mao, K., Wang, P.: Deep learning and its applications to machine health monitoring: a survey: <https://arxiv.org/pdf/1612.07640.pdf>.
11. Vychuzhanin, V. Shibaeva, N.: Informatization of prognostication of risk structurally of the difficult technical systems by means of models of the bayes networks of trust. *Informatics and mathematical methods in simulation*, Vol. 6, №. 3, pp. 205-214 (2016).
12. Rudnichenko, N. Vichuzhanin, V.: Information cognitive model of technological interdependence of marine technical recourses. *Informatics and mathematical methods in simulation*, Vol.3, №3, pp. 240-247 (2013).

13. Zenkin, E., Vychuzhanin, V. Rudnichenko, N.: Cognitive Model of the Internal Combustion Engine. SAE International, 2018-01-1738, pp. 1-10 (2018).
14. Frank, A. Bittner, S., Raubal, M.: Spatial and cognitive simulation with multi-agent systems. *Spatial Information Theory*, pp. 124–139 (2001).
15. Buss, M., Beetz, M. Wollherr D.: CoTeSys – cognition for technical systems. in *Proceedings of the 4th COE Workshop on Human Adaptive Mechatronics (HAM)* (2007).
16. Schwenk, C.R.: The cognitive perspective on strategic decision making. *Journal of Management Studies*, Vol.25, №1, pp. 41–55 (2007).
17. Vose, D.: Risk analysis: a quantitative guide. John Wiley Sons (2008).
18. Vychuzhanin, V., Rudnichenko, N., Shibaeva, N., Kononov, S.: Devising a method for the estimation and prediction of technical condition of ship complex systems. *Eastern-European Journal of Enterprise Technologies*, Vol. 84, № 6/9, pp. 4 -11 (2016).
19. Beetz, M. Buss, M., Wollherr D.: Cognitive technical systems - what is the role of artificial intelligence. *Advances in Artificial Intelligence, 30th Annual German Conference on AI*, pp. 19 – 42 (2007).
20. Vichuzhanin, V., Rudnichenko, N.: Development of the neural network model for prediction failure risk's of complex technical systems components. *Informatics and mathematical methods in simulation*, Vol. 6, №4, pp. 333-338 (2016).
21. Dorozhko, I. V., Tarasov, A. G., Baranovskiy A. M.: Otsenka nadezhnosti strukturno slozhnykh tekhnicheskikh kompleksov s pomoshch'yu modeley bayyesovskikh setey doveriya v srede GeNIe. *Intellectual Technologies on Transport*, № 3, pp. 36-45 (2015).
22. Pohorletskiy, D., Gritsuk I., Zenkin, E., Bulgakov, N., Golovan, A.: The Complex Application of Monitoring and Express Diagnosing for Searching Failures on Common Rail System Units. SAE International, 2018-01-1773, pp. 1-10 (2018).

Smart Crossover Mechanism for Parallel Neuroevolution Method of Medical Diagnostic Models Synthesis

Serhii Leoshchenko^{1[0000-0001-5099-5518]}, Sergey Subbotin^{2[0000-0001-5814-8268]}, Andrii Oli-
inyk^{3[0000-0002-6740-6078]}, Viktor Lytvyn^{4[0000-0003-4061-4755]}
and Matviy Ilyashenko^{5[0000-0003-4624-4687]}

^{1,2,3,4} Dept. of Software Tools, National University "Zaporizhzhia Polytechnic", 69063 Za-
porizhzhia, Ukraine

⁵Dept. of Computer Systems and networks, National University "Zaporizhzhia Polytechnic",
69063 Zaporizhzhia, Ukraine

¹sergleo.zntu@gmail.com ²subbotin@zntu.edu.ua
³olejnikaa@gmail.com ⁴lytvynviktor.a@gmail.com
⁵matviy.ilyashenko@gmail.com

Abstract. Information technologies significantly expand the capabilities of modern medicine. Using of artificial neural networks is becoming particularly promising. They are actively used for diagnostics based on patient data. However, the problem of network synthesis with a satisfactory topology and accurate diagnosis remains important. Neuroevolution methods allow to solve this problem without much involvement of an expert. Moreover, these methods make it possible to effectively use the parallel power of modern computer systems. On the other hand, parallelization raises a number of new problems. The paper suggests mechanisms for improving the crossover operator. The proposed solution allows you to reduce resource consumption and improve the synthesis process.

Keywords: Medical Diagnosis, Forecasting, Neuroevolution, Synthesis, Adaptive Mechanism, Genetic Algorithm, Parallel Genetic Algorithm, Crossover.

1 Introduction

The human factor often causes a number of problems. This also applies, of course, to medicine. A doctor's mistake can mean the loss of a patient's health or even their life, and doctors make mistakes not infrequently. Even the highest-level professional can make mistakes, because the specialist can be tired, irritated, concentrating on the problem worse than usual [1-3].

In this case, information technologies can come to the rescue. For example, the IBM Watson cognitive system [4-7] copes with work in the medical field at a fairly high level (oncology, reading x-rays, etc.) [4], [5]. But there are other solutions proposed by independent researchers. A number of scientists regularly present successful results of using artificial intelligence systems in medical diagnostics [8-11].

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

Artificial neural networks (ANN) have come into practice wherever it is necessary to solve problems of forecasting, classification or management. Problems of medical diagnostics are a sub-division of problems classification and forecasting of the object's condition. The following reasons determine the impressive success of the use of ANNs [10-13]:

- a huge number of opportunities [12], [13]. ANNs are a powerful modeling method that allows reproducing extremely complex dependencies. In particular, neural networks naturally are nonlinear. In problems where linear application is unsatisfactory (most medical diagnostics problems), linear models do not work well. In addition, neural networks cope with modeling linear dependencies in the case of a large number of variables;
- learning on examples (data about the object) [14], [15]. Neural networks are trained using examples. Initially, representative data is selected, and then a training algorithm is run that automatically perceives the data structure.

On the other hand, a user who trains a neural network needs a certain set of knowledge about how to select and prepare data, select the appropriate network architecture, and interpret the results [14-20]. Therefore, it can be assumed that the probability of error passes from the expert in knowledge of the problem area to the expert in the design of the ANN [16], [18].

Neuroevolution methods of ANN synthesis are based on an evolutionary approach [21]. These methods can simultaneously configure the network structure and weights. This allows getting ready-made neural network solutions with only data from the training sample, and in most cases does not require the user to have deep knowledge of the ANN theory [22], [23].

Advantages of using neuroevolution methods [24-27]:

- a wide variety of resulting topologies – possible solutions with "non-standard" ANNs structures;
- adaptivity;
- universality;
- the use does not require a deep knowledge about the ANN;
- possibility of using a parallel approach.

The most well-established neuroevolution method is the genetic algorithm (GA).

Unlike other optimization technologies, GA contain a population of trial solutions that are competitively managed using defined operators [28-30]. GA is inherent in iterative training of a population of individuals. The capacity of GA increases with the use of distributed computing. Such algorithms are called parallel genetic algorithms. They are based on splitting a population into several separate subpopulations, each of which will be processed by GA independently of other subpopulations.

However, when it comes to parallelizing, it should be taken into account that the main disadvantage of GA is a constant desire for a population containing one local optimum, which leads to a constant decrease in the genetic diversity of the population, which reduces the ability of GA to search for a global optimum and/or adapt to chang-

ing parameters of target functions [31], [32]. Therefore, the task of improving the methods of evolutionary optimization based on increasing the adaptive characteristics of the method is urgent.

2 Literature review

As noted above, GA is one of the most acceptable ways to synthesize ANN. This is due to the fact that at the initial stage there is absolutely no information about the direction of movement in terms of setting the matrix weights. Under conditions of uncertainty, evolutionary methods, including GA, have the highest chances to achieve the necessary results.

At the same time, there is a significant disadvantage of using genetic algorithms—the multi – iterative nature of the methods, serious time costs. To solve such problems, parallelization can be used to branch out the execution of calculations between the cores of modern computer systems. However, when designing a parallel approach, you should consider the most common problems encountered:

- selecting or developing a strategy for interaction between the components of the algorithm;
- choosing the frequency of migration between populations;
- determination of migrating individuals and their number;
- determining the structure of the evolution of individual populations.

Consider the problems in more detail. The structure of a parallel system is an important factor in the performance of a parallel algorithm, since it determines how quickly (or how slowly) a good solution spreads to other populations. If the system is strongly connected, then good solutions will quickly spread to all streams and can quickly "saturate" the population [33]. On the other hand, if the network is loosely connected, solutions will spread more slowly and threads will be more isolated from each other. Further parallel development and recombination of different solutions can occur to obtain potentially better solutions.

A common trend in promising parallel genetic algorithms (PGA) [34], [35] is the using of static system structures that are defined before the algorithm is run and remain unchanged.

Another method of constructing a structure is to create a dynamic system [33-35]. In this case, the flow is not limited to links with a certain fixed number of threads; instead, migrants are directed to threads that meet a certain criterion. As a similar criterion, the degree of population diversity or the measure of genotypic distance between two populations (or the distance from a characteristic individual of the population, for example, a favorite) is taken. This structure requires mechanisms for tracking events in neighboring populations, and if an event occurred in one of the neighboring populations, then an event should be expected in the second population.

The frequency of migrations also has a big impact on the final decision [33], [34]. As it known, too frequent migrations lead to degeneration of populations, and rare

ones, on the contrary, to a decrease in convergence. Various methods are used to regulate the frequency of migration, which can be divided into two types: adaptive and event-based. In the first case, adaptation methods are used to adjust the migration frequency during the algorithm operation. In the second case, methods are used that determine the need for migration, that is, migration is performed only when an event occurs.

Selection mechanisms are used to select individuals for migration [36-40]. It is known that individual chromosomes may contain "good" fragments of the genetic code, but these parts may be in chromosomes that are poorly adapted. But at the same time, excluding similar ones can lead to premature convergence, or skipping the global optimum.

Using different strategies imposes the main restriction: the necessity of forming the same type of chromosome structure. But the effect that is possible with successful formation can be much greater than when using a single GA structure in all populations [37].

It is also worth noting that a large number of migrations and even dynamic exchange of intermediate information between threads requires additional overhead, which sometimes largely negates the achievement of parallel execution of calculations [36], [38].

Moreover, we should not forget that even the sequential version of GA imposes significant resource requirements (especially RAM), compared to gradient methods [39], [40]. This is because in most cases a population of neural networks is used. Hence, when parallelizing, keep in mind that the requirements will increase according to the number of threads involved.

Therefore, the use of additional mechanisms and the introduction of hybrid methods, at different stages of operation, will improve the performance of the PGA.

3 Parallel genetic algorithm with smart crossover

The paper proposes a parallel genetic method. Parallelization will occur in the following way. Initially, we will generate a population of individuals on the main core of the system, where each individual is a separate neural network:

$$P = \{NN_1, NN_2, \dots, NN_n\}, \quad (1)$$

where, P is population,

NN_n is neural network,

n is the size of population.

Also, at the initial stage, the main free parameters of the method will be set: the stop criteria, population size, and so on.

The next step is to divide the population into subpopulations and distribute subpopulations between the cores of the multi-core system.

The cores will perform the same sequence of actions with subpopulations: evaluating the genetic information of individuals, sorting and selecting the best individuals, crossing and selecting the new best individuals.

Then the selected best individuals are sent to the main core, where they are re-sorted and selected the best representatives. The new best individual is evaluated for the stop criterion. If the resulting individual is satisfied, it becomes a solution and the method ends. If the solution is not satisfactory, the best individuals from the subpopulations are crossed and a new population is obtained, which is re-distributed between the system cores.

Diversity is maintained in the process of evolution by the fact that each species (subpopulation) develops without exchanging genetic material with other species. This is an important aspect of this model. The exchange of genetic material between two different species usually produces non-viable offspring. In addition, mixing of genetic material can reduce the diversity of populations. This approach will also significantly reduce the share of overhead associated with the transfer of information between the system cores.

Now let's look at the mechanisms that allow optimizing the use of RAM, while maintaining the adaptive characteristics of the method. In the proposed method, it is recommended to use a smart crossover, which is based on uniform crossing and rank selection enhanced by criteria conditions.

Uniform crossover is one of the most effective recombination operators in standard GA [41], [42].

Uniform crossover is performed according to a randomly selected standard that specifies which genes should be inherited from the first parent (other genes are taken from the second parent) [42]. In other words the general rule of uniform crossover can be represented as follows:

$$\begin{aligned}
 Ind_3 &= \text{Crossover}(Ind_1, Ind_2, \text{DataofCros}); \\
 g_{Ind_3} &= \{g_1 = \text{Rand}(g_{1Ind_1}, g_{1Ind_2}), \\
 g_2 &= \text{Rand}(g_{2Ind_1}, g_{2Ind_2}), \dots, \\
 g_i &= \text{Rand}(g_{iInd_1}, g_{iInd_2})\}.
 \end{aligned} \tag{2}$$

It has long been known that setting the probability of passing a parent gene to a descendant in a uniform crossover can significantly increase its efficiency [35], [36], and also allows emulating other crossing operators (single-point, two-point). It is also known that the use of the uniform crossing operator makes it possible to apply the so-called multi-parent recombination, when more than two parents are used to generate one offspring. Despite this, in most studies, only two parents are used and the fixed probability of transmitting the gene is 0.5 [41].

Even crossing gives greater flexibility when combining rows, which is an important advantage when working with GA.

However, it should be noted that even crossing requires additional computing power. On the other hand, uniform crossover makes it possible to emulate the operation of simpler types of crossing, such as two-point crossover. Therefore, we will use a two-

point crossover to work on threads. This approach will allow you to implement this method in the future on computing systems running graphics processors (GPUs).

It is proposed to strengthen the ranking selection by introducing additional criteria that will help to track various characteristics of neural networks more subtly, namely: excessive memory usage and approximation properties of the neural network.

The general scheme of the method is shown in Fig. 1.

4 Experiment

4.1 Description of the experiment

A sample of Parkinson's Disease Classification Data Set will be used for the experiment [43].

The data used in this study were gathered from 188 patients with PD (107 men and 81 women) with ages ranging from 33 to 87 (65.1 ± 10.9). The data used in this study were gathered from 188 patients with PD (107 men and 81 women) with ages ranging from 33 to 87 (65.1 ± 10.9) at the Department of Neurology in Cerrahpasa Faculty of Medicine, Istanbul University.

The control group consists of 64 healthy individuals (23 men and 41 women) with ages varying between 41 and 82 (61.1 ± 8.9). During the data collection process, the microphone is set to 44.1 KHz and following the physician's examination, the sustained phonation of the vowel was collected from each subject with three repetitions. Table 1 shows the main characteristics of the data sample.

Table 1. Main characteristics of the Parkinson's Disease Classification Data Set

Criterion	Characteristic	Criterion	Characteristic
Data Set Characteristics	Multivariate	Number of Instances	756
Attribute Characteristics	Integer, Real	Number of Attributes	754

The following hardware and software have been used for experimental verification of the proposed parallel genetic method for ANN synthesis: the computing system of the Department of software tools of Zaporizhzhya national technical university (National university "Zaporizhzhia politechnic"), Zaporizhzhia: Xeon processor E5-2660 v4 (14 cores), RAM 4x16 GB DDR4, the programming model of Java threads.

The results of the proposed method will be compared with the results of the method considered in the previous works [44]. Old modification will be called PGM (Parallel genetic method) and new variant – PGM SC (Parallel genetic method with smart crossover). Note that when working, the size of the parent pool for even crossing will be equal to the number of system cores involved.

4.2 The results of the experiment

In the Fig. 2 is graph of the execution time (in minutes) of the proposed method on computer systems, which depends on the number of cores involved.

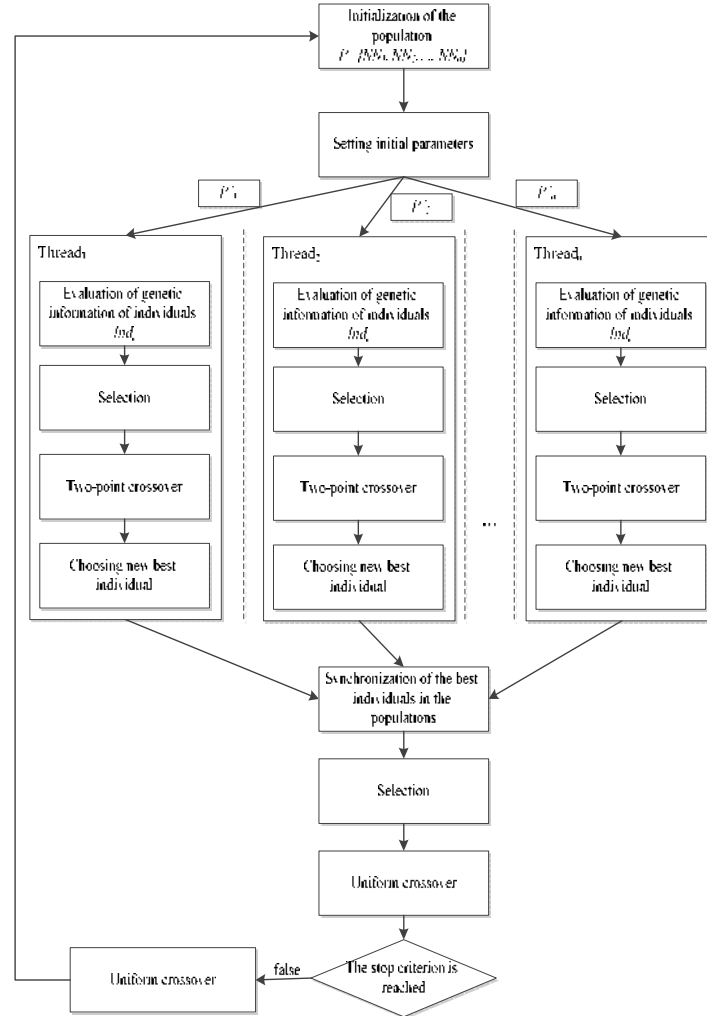


Fig. 1. Parallel genetic method for ANN synthesis with smart crossover

It can be seen from the graphs that the proposed method has an acceptable degree of parallelism and is effectively performed on MIMD parallel system. In addition, the processor in multi-core computer supports Turbo Boost technology [45–47], making the time of the method execution on the single core much less than on the core of which does not support this technology.

Fig. 3 shows graphs of changes in communication overhead. Since the new method offers the transfer of only the best individuals, this allows you to significantly reduce the transfer of excess information. This is explained by the fact that communication

overhead of the proposed method execution on computer systems is relatively small, and the number of parallel operations significantly exceeds the number of serial operations and synchronizations. In communication overhead, is understood the ratio of the time spent by the system for transfers and synchronization among cores to the time of target calculations on a given number of cores.

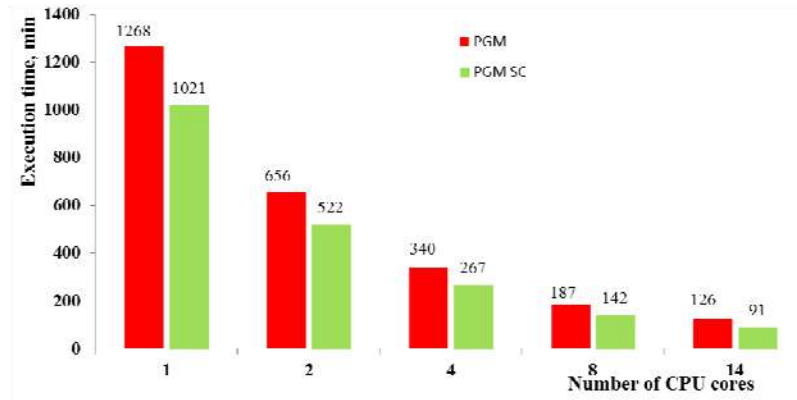


Fig. 2. Dependence the execution time of the proposed method to the number of involved cores of the computing system

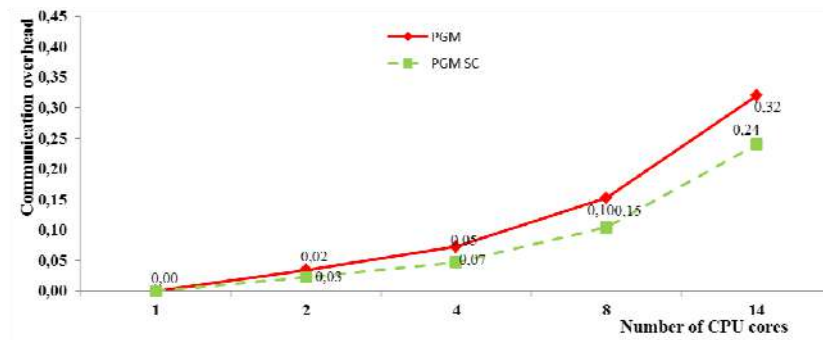


Fig. 3. Communication overhead performing the proposed method to the number of cores involved of the computing system

Fig. 4 shows graphs of speedup changes. Based on the fact that the communication overhead has decreased, the speedup of execution increases significantly.

The graph of efficiency of computer systems is presented in Fig. 5. It shows that the using of even 14 cores of computer systems for the implementation of the proposed method retains the efficiency at a relatively acceptable level and indicates the potential, if necessary and possibly, to use even more cores.

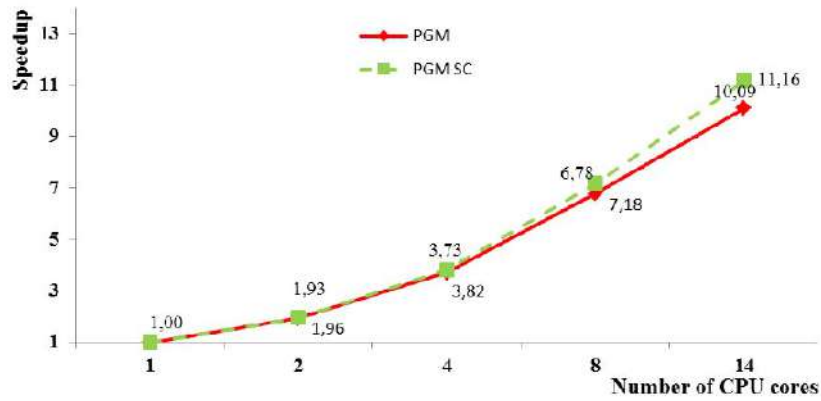


Fig. 4. The speedup graphics of calculations on computing system

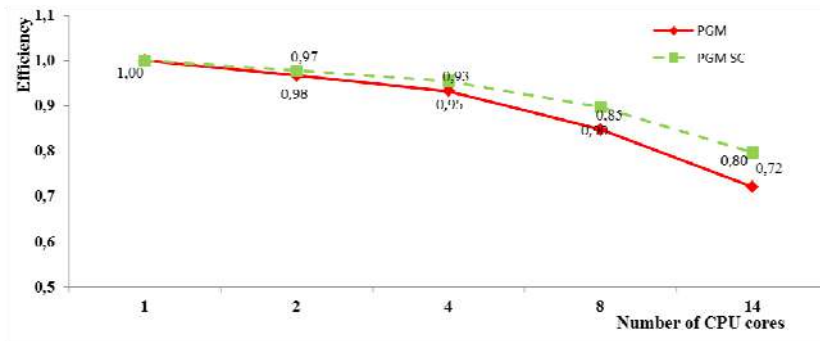


Fig. 5. The efficiency graph of computing systems when executing the proposed method

5 Conclusion

Based on the results of experiments, it can be argued that the proposed method can be used in the synthesis of neural network diagnostic models. The proposed smart crossover mechanism significantly optimizes the synthesis process by using an adjustable uniform crossover and additional criteria at the selection stage.

However, we are not talking about large-scale implementation of neural networks in hospitals around the world. The main problem in terms of the spread of these technologies is that neural networks are a kind of "black box". Specialists enter data and get a certain result. But the creators of such systems may not fully understand how this result was obtained, what algorithms and in what sequence are involved. If neural networks could be made more transparent, and the principle of their operation could be easily explained to medical practitioners, then the rate of spread of this technology would be much higher.

Acknowledgment

The work is supported by the state budget scientific research project of National University "Zaporizhzhia Polytechnic" "Intelligent methods and software for diagnostics and non-destructive quality control of military and civilian applications" (state registration number 0119U100360).

References

1. Einarson, T.R., Acs, A., Ludwig, C., Panton, U.H.: Prevalence of cardiovascular disease in type 2 diabetes: a systematic literature review of scientific evidence from across the world in 2007–2017. *Cardiovasc Diabetol* 17 (1), 1–19 (2018). <https://doi.org/10.1186/s12933-018-0728-6>
2. Petrie, J.R., Guzik, T.J., Touyz, R.M.: Diabetes, Hypertension, and Cardiovascular Disease: Clinical Insights and Vascular Mechanisms. *The Canadian journal of cardiology* 34(5), 575–584 (2017). <https://doi.org/10.1016/j.cjca.2017.12.005>
3. Jungen, C., Scherschel, K., Eickholt, C., Kuklik, P., Klatt, N.: Disruption of cardiac cholinergic neurons enhances susceptibility to ventricular arrhythmias. *Nature Communications* 8 (2017). <https://doi.org/10.1038/ncomms14155>
4. Bakkar, N., Kovalik, T., Lorenzini, I. et al.: Artificial intelligence in neurodegenerative disease research: use of IBM Watson to identify additional RNA-binding proteins altered in amyotrophic lateral sclerosis. *Acta Neuropathol* 135, 227–247 (2018). <https://doi.org/10.1007/s00401-017-1785-8>
5. Gomes, C., Dietterich, T., Barrett, C. et al.: Computational sustainability: computing for a better world and a sustainable future. *Communications of The ACM* 62 (9), 56–65 (2019). <https://doi.org/10.1145/3339399>
6. Kolpakova, T., Oliinyk, A., Lovkin, V.: Improved method of group decision making in expert systems based on competitive agents selection. *UKRCON: IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON)*, 939–943. KPI, Kyiv (2017). <https://doi.org/10.1109/UKRCON.2017.8100388>
7. IBM Watson Homepage, <https://www.ibm.com/watson>
8. Pranav, R., Anirudh, J., Anuj, P. et al.: CheXpedition: Investigating Generalization Challenges for Translation of Chest X-Ray Algorithms to the Clinical Setting (2020), <https://arxiv.org/pdf/2002.11379.pdf>
9. Artificial intelligence rivals radiologists in screening X-rays for certain diseases, <https://med.stanford.edu/news/all-news/2018/11/ai-outperformed-radiologists-in-screening-x-rays-for-certain-diseases.html>
10. Stanford researchers develop artificial intelligence tool to help detect brain aneurysms, <https://news.stanford.edu/2019/06/07/ai-tool-helps-radiologists-detect-brain-aneurysms/>
11. Leoshchenko S, Oliinyk A., Subbotin, S, Zaiko T: Using modern architectures of recurrent neural networks for technical diagnosis of complex systems. In: *Proceedings of the 2018 international scientific-practical conference problems of infocommunications. science and technology (PIC S&T)*, 411–416. IEEE, Kharkov (2018). <https://doi.org/10.1109/infocommst.2018.8632015>
12. Schmidt, J., Marques, M.R.G., Botti, S. et al.: Recent advances and applications of machine learning in solid-state materials science. *npj Comput Mater* 5, 83 (2019). <https://doi.org/10.1038/s41524-019-0221-0>

13. Bradley, J.R., Holan, S.H., Wikle, C.K.: Multivariate spatio-temporal models for high-dimensional areal data with application to Longitudinal Employer-Household Dynamics. *The Annals of Applied Statistics*, 9(4), 1761-1791 (2015).
14. A Tour of Machine Learning Algorithms, <https://machinelearningmastery.com/a-tour-of-machine-learning-algorithms/>
15. Sabater-Mir, J., Torra, V., Aguiló, I., González-Hidalgo, M.: *Artificial Intelligence Research and Development*. IOS Press, Amsterdam (2019).
16. Scher, S., Messori, G.: Generalization properties of feed-forward neural networks trained on Lorenz systems. *Nonlinear Processes in Geophysics* 26, 381–399 (2019).
17. How to Scale Data for Long Short-Term Memory Networks in Python, <https://machinelearningmastery.com/how-to-scale-data-for-long-short-term-memory-networks-in-python/>
18. Learning process of a neural network, <https://towardsdatascience.com/how-do-artificial-neural-networks-learn-773e46399fc7>
19. Oliinyk, A.O., Zayko, T.A., Subbotin, S.O.: Synthesis of Neuro-Fuzzy Networks on the Basis of Association Rules. *Cybern Syst Anal* 50, 348–357 (2014). <https://doi.org/10.1007/s10559-014-9623-7>
20. Oliinyk, A.O., Skrupsky, S.Y., Subbotin, S.A.: Using parallel random search to train fuzzy neural networks. *Aut. Control Comp. Sci.* 48, 313–323 (2014). <https://doi.org/10.3103/S0146411614060078>
21. Stanley, K.O., Clune, J., Lehman, J. et al.: Designing neural networks through neuroevolution. *Nature Machine Intelligence* 1, 24–35 (2019). <https://doi.org/10.1038/s42256-018-0006-z>.
22. Baldominos, A., Saez, Y. & Isasi, P.: On the automated, evolutionary design of neural networks: past, present, and future. *Neural Comput & Applic* 32, 519–545 (2020). <https://doi.org/10.1007/s00521-019-04160-6>
23. Gaier, A., Asteroth, A., Mouret, J.-B.: Data-efficient Neuroevolution with Kernel-Based Surrogate Models. *GECCO '18: Proceedings of the Genetic and Evolutionary Computation Conference*, 85–92 (2018). 10.1145/3205455.3205510.
24. Albrigtsen, S.I., Imenes A.: *Neuroevolution of Actively Controlled Virtual Characters*. University of Agder, Kristiansand & Grimstad (2017).
25. Bohrer, J., Grisci, B., Dorn, M.: *Neuroevolution of Neural Network Architectures Using CoDeepNEAT and Keras* (2020), <https://arxiv.org/abs/2002.04634>
26. Phillips B.: *Deep Neuroevolution: Genetic Algorithms are a Competitive Alternative for Training Deep Neural Networks for Reinforcement Learning – Paper Summary*, <https://towardsdatascience.com/deep-neuroevolution-genetic-algorithms-are-a-competitive-alternative-for-training-deep-neural-822bfe3291f5>
27. Gradient descent vs. neuroevolution, <https://towardsdatascience.com/gradient-descent-vs-neuroevolution-f907dace010f>
28. García-Martínez C., Rodríguez F.J., Lozano M. (2018) Genetic Algorithms. In: Martí R., Pardalos P., Resende M. (eds) *Handbook of Heuristics*. Springer, Cham
29. Lee, W.-P., Hsiao, Y.-T., Hwang, W.-C.: Designing a parallel evolutionary algorithm for inferring gene networks on the cloud computing environment. *BMC Systems Biology* 8(1), 23-28 (2014). <https://doi.org/10.1186/1752-0509-8-5>.
30. Gonçalves, I., Gomes, A., Henggeler, C.: Optimizing residential energy resources with an improved multi-objective genetic algorithm based on greedy mutations. *GECCO '18: Proceedings of the Genetic and Evolutionary Computation Conference*, 1246–1253 (2018). <https://doi.org/10.1145/3205455.3205616>

31. Zhang, F., Mei, Y., Zhang, M.: A two-stage genetic programming hyper-heuristic approach with feature selection for dynamic flexible job shop scheduling. GECCO '19: Proceedings of the Genetic and Evolutionary Computation Conference, 347–355 (2019). <https://doi.org/10.1145/3321707.3321790>
32. Stork, J., Eiben, A., Bartz-Beielstein, T.: A new Taxonomy of Continuous Global Optimization Algorithms (2018), <https://arxiv.org/pdf/1808.08818.pdf>
33. Potuzak, T.: Optimization of a genetic algorithm for road traffic network division using a distributed/parallel genetic algorithm. 2016 9th International Conference on Human System Interactions (HSI), 2016, pp. 21–27. <https://doi.org/10.1109/HSI.2016.7529603>
34. Hoseini Alinodehi, S.P., Moshfe, S., Saber Zaeimian, M., et al.: High-Speed General Purpose Genetic Algorithm Processor. IEEE Transactions on Cybernetics 46(7), 1551–1565, (2016). <https://doi.org/10.1109/TCYB.2015.2451595>
35. Hou, N., He, F., Zhou, Y., Chen, Y., Yan, X.: A Parallel Genetic Algorithm With Dispersion Correction for HW/SW Partitioning on Multi-Core CPU and Many-Core GPU. IEEE Access 6, 883–898 (2018). <https://doi.org/10.1109/ACCESS.2017.2776295>
36. Varun Kumar, S.G., Panneerselvam, Dr.R.: A Study of Crossover Operators for Genetic Algorithms to Solve VRP and its Variants and New Sinusoidal Motion Crossover Operator. International Journal of Computational Intelligence Research 13(7), 1717–1733 (2017).
37. Hassanat, A., Alkafaween, E.: On Enhancing Genetic Algorithms Using New Crossovers, <https://arxiv.org/ftp/arxiv/papers/1801/1801.02335.pdf>
38. Umbarkar, Dr.A., Sheth, P.: Crossover operators in genetic algorithms: a review. ICTACT Journal on Soft Computing 6(1), 1083–1092 (2015). <https://doi.org/10.21917/ijsc.2015.0150>.
39. Oliinyk, A.O., Skrupsky, S.Y., Subbotin, S.A.: Experimental investigation with analyzing the training method complexity of neuro-fuzzy networks based on parallel random search. Aut. Control Comp. Sci. 49, 11–20 (2015). <https://doi.org/10.3103/S0146411615010071>
40. Leoshchenko, S., Oliinyk, A., Skrupsky, S., Subbotin, S., Zaiko T.: Parallel Method of Neural Network Synthesis Based on a Modified Genetic Algorithm Application. Workshop Proceedings of the 8th International Conference on “Mathematics. Information Technologies. Education”, MoMLet&DS-2019, 11–23. Lviv Polytechnic National University, Lviv (2019). <https://dblp.org/rec/conf/momlet/LeoshchenkoOSSZ19>
41. Baeck, T., Fogel, D.B., Michalewicz, Z.: Evolutionary Computation 1: Basic Algorithms and Operators. CRC Press, Boca Raton (2000).
42. Das, A.K., Pratihari, D.K. A directional crossover (DX) operator for real parameter optimization using genetic algorithm. Appl Intell 49, 1841–1865 (2019). <https://doi.org/10.1007/s10489-018-1364-2>
43. Parkinson's Disease Classification Data Set, <https://archive.ics.uci.edu/ml/datasets/Parkinson%27s+Disease+Classification>
44. Leoshchenko, S., Oliinyk, A., Skrupsky, S., Subbotin, S., Lytvyn V.: Parallel Genetic Method for the Synthesis of Recurrent Neural Networks for Using in Medicine. In: Proceedings of the Second International Workshop on Computer Modeling and Intelligent Systems (CMIS-2019), pp. 1–17 (2019). <https://dblp.org/rec/conf/cmis/LeoshchenkoOSS19>
45. Intel Turbo Boost Technology 2.0, <https://www.intel.com/content/www/us/en/architecture-and-technology/turbo-boost/turbo-boost-technology.html>
46. Oliinyk, A., Subbotin, S., Lovkin, V., Leoshchenko, S., Zaiko, T. Feature Selection Based on Parallel Stochastic Computing. 13th International Scientific and Technical Conference

- on Computer Sciences and Information Technologies (CSIT), 347-351. IEEE, Lviv (2018). <https://doi.org/10.1109/STC-CSIT.2018.8526729>.
47. Shkarupylo, V., Skrupsky, S., Oliinyk, A., Kolpakova, T.: Development of stratified approach to software defined networks simulation. Eastern-European Journal of Enterprise Technologies, 5. 67-73 (2017). <https://doi.org/10.15587/1729-4061.2017.110142>.

Monitoring System for Tests of the Mg Implants

Galyna Tabunshchik¹[0000-0003-1429-5180], Vadym Shalomeev¹[0000-0002-6091-837X], Peter Arras²[0000-0002-9625-9054]

¹ National University “Zaporizhzhia Polytechnic”, Zaporizhzhia, Ukraine

² KU Leuven, Campus De Nayer, Belgium

galina.tabunshchik@gmail.com, shalomeev@radiocom.net.ua,
peter.arras@kuleuven.be

Abstract –In the article authors are providing requirements analysis for the informational systems for medical trials. The author considered types of the experiments which could be conducted remotely with the Mg implants and designed an experiment for the in vivo tests of the Mg stem. In the paper the authors present three level architecture for the monitoring system the corrosion and solubility of Mg implants and model for data integration and transmission. Also suggested the model for the data integration and transition in to repository.

Keywords —critical systems, implants, monitoring, data analysis, hardware and software systems, robust models.

1 Introduction

Implementation of IoT-technology in the research process allows users to collect, process and visualise data. Especially in critical fields as for example the testing of new materials for artificial implants, an IoT approach can present new opportunities for speeding up and get more accurate data.

Internet of medical things (IoMT), a connected infrastructure of medical devices, software applications and health systems and services, transforming greatly the sector of medical trials improving the quality of data and its interoperability[1]. It enables machine-to-machine interaction and real-time data streaming between an almost infinite range of medical device and involve multiple participants, such as doctors and patients, hospitals, and research institutions in multi.

New regulations, digitisation, data analytics, artificial intelligence, automation and the development of value-based health care represent some of the numerous challenges as well as opportunities facing the medical technology industry [2]. And one of the biggest challenge is interoperability, including complying with various national and international standards and protocols around the exchange and use of data.

Real-time monitoring systems could be implemented as for in-vivo and for in-vitro test systems and complement to the existing approaches in both stages. Within the international Erasmus+ KA2 project BIOART (Innovative Multidisciplinary Curricu-

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

lum in Artificial Implants for Bio-Engineering BSc/MSc Degrees [3] in NU “Zaporizhzhia Polytechnic” are developed new Mg alloys for implants [4]. To test these new materials it is required to create a new robust testing system, which could provide researchers with prompt and reliable information on their material tests. For development of the requested monitoring system there is required integration of the biomedical data, material science data and IoT.

2 Investigation of the in-vivo and in-vitro real-time monitoring approaches

Mg has a particular material property which makes it suitable for using as an implant material: it is soluble in a watery environment. If the solubility rate of the material can be controlled it opens up a range of possibilities in the implant-surgery as it would avoid a second surgery for taking the temporary implants out again, making in more comfortable for the patient, and giving an economic advantage in avoiding extra medical costs.

The aim of the current research is to accurately determine the solubility rate of the material and to match it with other retrieved data. The rate is measured by putting stems of Mg in water and monitoring its state under pressure. After a predetermined time the stems are taken out and their characteristics (weight and corrosion state) are measured. The stems should dissolve in a controlled concrete time. The continuous real-time measurement data of the this process should be collected at a server and then used for decision making in in-vivo tests. In in-vivo test the state of the implant is monitored by making X-rays or tomography. The base hypothesis is that the rate of dissolving is the same in-vitro as in-vivo. By measuring and comparing a prediction of the dissolving time will be possible.

The combination of a test-method(s) and appropriate techniques for assessing solubility in-vitro and validating it in-vivo is the basis for the research. Underneath there is an overview of the different state-of-the-art solubility tests and currently used measuring techniques.

For the actual monitoring during the in-vitro test an EIS-technique is used for reasons of efficient measuring and digitalizing of signals over a longer period of time.

In the work [5,5] there is provided an overview solubility test-methods for Mg alloys:

- Short term immersion test (hydrogen evolution) in $\text{NHCO}_3/\text{CO}_2$ – buffered simulated body fluid (SBF) at 37°C – up to 14 days.
- Electrochemical impedance spectroscopy and immersion tests in DMEM supplemented by 10% (v/v) fetal bovine serum (FBS) pre-conditioned at 37°C up to 28 days.
- Long-term immersion test (up to 365 days) in Dulbecco’s Modified Eagle Medium (DMEM) supplemented with 10 vol.% fetal bovine serum buffered with 4-(2-

hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) at 37°C and refreshed periodically.

Assessment techniques for in-vivo corrosion:

- SEM (Scanning Electron Microscopy) and WDS (Wavelength Dispersive Spectroscopy) for the morphological imaging of retrieved implants and on-the-spot semi-quantitative chemical composition analysis.
- FTIR (Fourier-transform infrared spectroscopy (FTIR) is a technique used to obtain an infrared spectrum of absorption or emission of a solid, liquid or gas. An FTIR spectrometer simultaneously collects high-spectral-resolution data over a wide spectral range. This confers a significant advantage over a dispersive spectrometer, which measures intensity over a narrow range of wavelengths at a time.
- Histology: also known as microscopic anatomy or microanatomy, is the branch of biology which studies the microscopic anatomy of biological tissues. Histology is the microscopic counterpart to gross anatomy, which looks at larger structures visible without a microscope.
- USG: Medical ultrasound (also known as diagnostic sonography or ultrasonography) is a diagnostic imaging technique.
- X-ray radiography - non-invasive hard tissue imaging, produce 2D image of gross tissue (combination of hard and soft tissue), image nearly all biomaterials, similar to USG, qualitative analysis by visual observation of radiographs, quantitative analysis by means of image processing software
- μ CT - Enhanced capabilities of X-ray radiography with possibility to generate a complete 3D image, as a product of stacking discontinuous 2D slices, of an implant and its surrounding tissue
- MRI - non-invasive medical diagnostic tool for real-time soft and hard tissue imaging

Real-time in-vitro corrosion monitoring of absorbable metal implants [6]:

- Electromechanical-based monitoring systems - quantitatively measure corrosion rate and electrochemical properties of corroded surface in short-time (accelerated process);
- Electrochemical impedance spectroscopy (EIS): EIS is a highly sensitive characterization technique used to establish the electrical response of chemical systems in a nondestructive manner. EIS systems characterize the time response of chemical systems using low amplitude alternating current (AC) voltages over a range of frequencies. Using an electrode setup consisting of a working, reference, and counter electrodes a known voltage is passed from the working electrode through an electrolytic solution and into the counter electrode. Quantitative measurements are produced by the EIS and enable the evaluation of small scale chemical mechanisms at the electrode interface and within the electrolytic solution. Therefore, EIS is useful in determining a wide range of dielectric and electrical properties of components in research fields studying batteries, corrosion, etc.

- Sensor-based monitoring systems - accurately measure certain corrosion parameters using specific sensors, e.g., H_2 sensor, Mg ion sensor, possible use for real-time and continuous in-vivo corrosion monitoring as the sensor is connected to a data acquisition device.
- Micro-dialysis-based monitoring system: Micro dialysis is a minimally-invasive sampling technique that is used for continuous measurement of free, unbound analyte concentrations in the extracellular fluid of virtually any tissue. Analytes may include endogenous molecules (e.g. neurotransmitter, hormones, glucose, etc.) to assess their biochemical functions in the body, or exogenous compounds (e.g. pharmaceuticals) to determine their distribution within the body or fluid.

3 Design of the in vivo-experiment

Advances in microelectronics, in particular, such as miniaturization, computing power increase per given area of silicon, reduced power consumption, increased efficiency of radio frequency telecommunications and monolithic system-on-chip (SoC) integration are leading to data acquisition equipment which is becoming increasingly portable and unobtrusive, permitting uninterrupted real-time monitoring in a variety of scenarios [7].

For the setup of the experiment (Fig.1) there was chosen an immersion test which resembles body conditions. The Mg stems are put in a row in a plastic tube which is pressurized with a saline solution by a pump. The saline solution is kept at a constant temperature of 36.6°C to mimic body conditions.

The saline solution dissolves the Mg from the stem and as such will change its electric properties. The impedance change of the solution is measured in real-time and logged. After each test period a stem is removed and measured for weight and visual (microscopic) inspection to assess the rate of dissolving. The other stems are kept in the saline solution for the next test.

The pressure and flow of the saline will be controlled to assess the influence of the different parameters. The H_2 which is set free from the reaction $\text{Mg} + \text{H}_2\text{O} \Rightarrow \text{MgO} + \text{H}_2$ is collected in a gas reservoir to measure the dissolving rate.

The control system is managing the temperature of the saline, pump pressure, periodic pictures of the stems, data from the sensors. It is developed on the Xilinx platform [9], which makes its flexible for integration with the user application and verification of the system functioning.

Web-interface is providing access to the experiments measurements in the real time, allowing user to change the conditions of the experiments and manually control measurements.

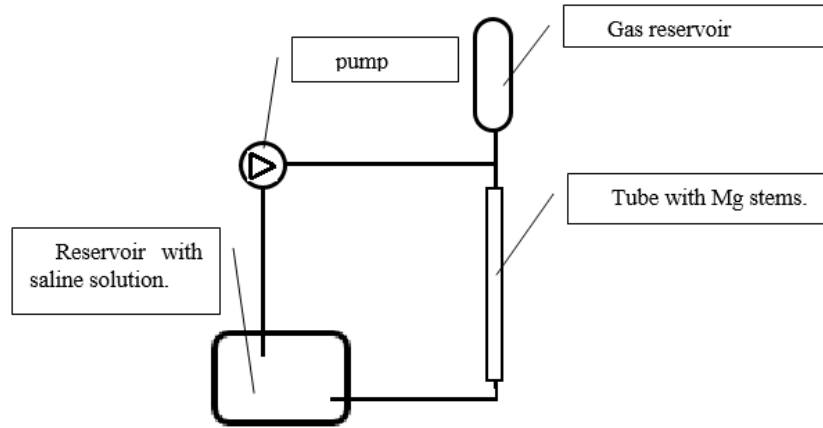


Fig. 1. Experimental setup for immersion test on Mg -stems

4 Monitoring system for Mg implants

A feature of modern monitoring is the extensive use of various instruments, sensors and communication infrastructures capable of transmitting and processing data in real-time. At the same time, a data collection system with low energy costs and the ability to simultaneously service a large number of IoT devices is necessary for efficient collection and processing of data on end nodes of information systems based on IoT.

For the monitoring system we will consider a three layer architecture [10] which contains device level, server level and application level (Fig.2).

At device level data is collected from different sources, but in general it could be divided into two types: sensor-information and images. Both types can be collected at in-vitro tests. For in-vivo tests mostly only imaging data (X-ray, tomography, scans..) can be used, and in the case of surgical extraction of implants other data (loss of material...) can be measured too.

At sever level, the technique to classify real-time data from several different sources involves the following eight steps: data normalization; prediction future points; analysis of residuals; probability calculation; conflicts definition; data fusion; classification; estimation of classification accuracy.

At application/end user level tools are needed for the collection, the analysis, and the representation of results of the experiments.

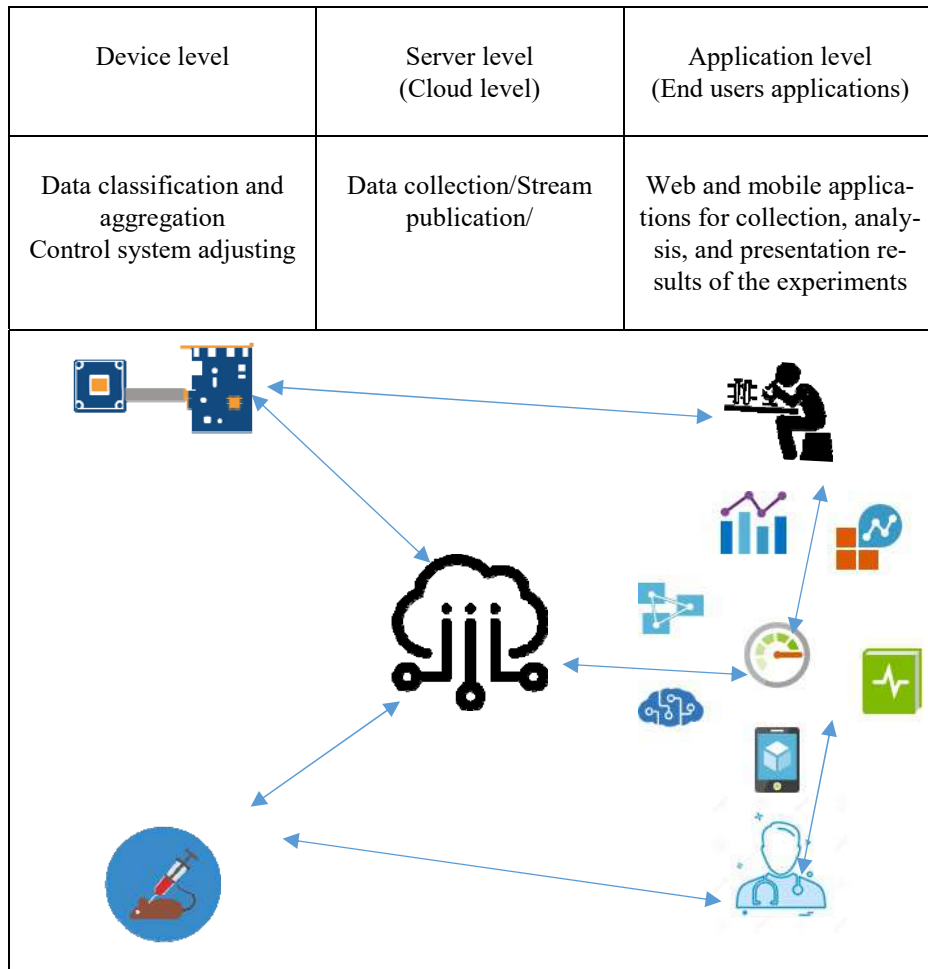


Fig. 2. Architecture of the monitoring system

Data which is received from sensors during the tests is also used for the manipulation of the test control system.

Data resulting from in-vivo tests should be used later for the further in-vitro tests. There should be possibility for continuous update of all the information and the matching of the different types of data.

For the testing of the Mg implants a monitoring system is developed, which is described in Fig.3. We are implementing a single model of combining all data which makes it possible to develop a high quality system for supporting medical trial decisions. In the system we integrate medical, biomedical, physical and clinical data .

Prediction model for implant corrosion		
In-Vitro data	In-Vivo Data	
Experimental data and results	Real time data for implant corrosion Clinical data	
Experiments data repository 	Clinical repository 	
Data Integration		
Predictive Models	Data-driving trials	Diagnostic alerts

Fig. 3. Integration model for the trial data

Data integration model will be considered as a multidimensional cube:

$$M_{data} = H(D, A), \quad (1)$$

where $D = \{d_1, d_2, \dots, d_i\}$ is set of dimensions (d_1 – "set of material properties of the sample", d_2 – "set of trial conditions", d_3 – "source of information", d_4 – "time dimension"), $A = A_{d1} \cup A_{d2} \cup \dots \cup A_{di}$ – set of attributes:

$$A_{di} = \{a_1^i, a_2^i, \dots, a_g^i\}, \quad (2)$$

where $i = 1, 2, 3$ – set of attributes of dimension di .

Each cell of the hypercube of data corresponds to the only possible set of measurement attributes containing unique experimental data for set of measurements in concrete moment of time.

As developed test system is system with limited resources we need periodical data transfer to the cloud, and for the data transfer period from the testing system to the cloud we will use [11]:

$$\eta' = (1 + \omega) \left(\tau B + \sigma C \right) \left(\frac{\tau}{S} - \frac{1}{v_{net}} \right), \quad (3)$$

where ω – safety factor; B – volume of the modified data within specified period; C – volume of the recorded data or Δ period; τ – the predicted device operating time; σ – predicted volume of the recording data; v_{net} – mean speed of data transmission S – available volume.

For realisation of the interoperability repositories should comply following requirements:

- (1) it should be a digital system able to manage N-dimensional hyper-cubes and allow data analytics;
- (2) it should be accessible and interoperable at the software level from the external channels;
- (4) it should have managing mechanism;
- (5) the realisation of the data exchange should be organised regarding to the health information technology standards [12] to have possibility for the further in-vivo monitoring.

The main regulation standards for file format are xml and json, for electronic health information architecture is ISO 18308:2011 [14], for messaging of healthcare data are HL7 v2.x [15], v3, FHIR [16], openEHR [13], for clinical documents (Clinical Document Architecture – CDA), for medical imaging and communication (DICOM) [17], and for patient health summaries (Continuity of Care Record – CCR) [15].

5 Conclusion

In the paper authors suggest an architecture of the real-time monitoring system which allows to check the state of the artificial implants based on Mg alloys.

Starting from a predictive model on the dissolving rate of the Mg in the human body, data from in-vitro tests and in-vivo checks is combined to validate and update the predictive model.

The architecture of the monitoring systems contains aggregation data from different sources, processing, automated control of the testing equipment, and decision support system for the researcher.

The setup for in-vitro experiments was constructed, which could be used as the remote experiments for research and education purposes.

The aim is to have a robust and validated prediction for the dissolving of the Mg-implants in the body.

6 Acknowledgement

This work is carried out partly with the support of Erasmus + KA2 project 586114-EPP-1-2017-1-ES-EPPKA2-CBHE-JP «Innovative Multidisciplinary Curriculum in Artificial Implants for Bio-Engineering BSc/MSc Degrees» [BIOART].

7 References

1. Tian S., Yang W., Le Grange J. M., Wang P., Huang W., Ye Zh, Smart healthcare: making medical care more intelligent, *Global Health Journal*, Volume 3, Issue 3, 2019, pp 62-65. (2019)
2. Medtech and the Internet of Medical Things. How connected medical devices are transforming health care, Deloitte NWE LLP, 56p. (2018),
3. BIOART. <https://bioart.iucc.ac.il/>
4. Shalomeev V., Aikin N., Chorniy V., Naumik V. Design and examination of the new bio-soluble casting alloy of the system mg–zr–nd for osteosynthesis Vol 1, No 12 (97) pp. 40-48. (2019) doi:10.15587/1729-4061.2019.157495
5. Esmaily M., Svensson J.E., Fajardo S., Birbilis N., Frankel G.S., Virtanen S. Fundamentals and advances in magnesium alloy corrosion, *Progress in Materials Science*, vol. 89, pp. 92-193. (2017)
6. Ulum M., Caesarendra W., Alavi R., Hermawan H.) In-Vivo Corrosion Characterization and Assessment of Absorbable Metal Implants., *Coatings* , vol. 9, pp. 282-391. (2019)
7. Astaras, A., Bamidis, P. D., Kourtidou-Papadeli, C., & Maglaveras, N. Biomedical real-time monitoring in restricted and safety-critical environments. *Hippokratia*, 12 Suppl 1(Suppl 1), 10–14.(2008)
8. Ren, L.; Yu, K.; Tan, Y. Monitoring and Assessing the Degradation Rate of Magnesium-Based Artificial Bone In Vitro Using a Wireless Magnetoelastic Sensor. *Sensors* , 18, 3066. (2018) doi:10.3390/s18093066
9. PYNQ: python productivity for ZYNQ. <http://www.pynq.io/>
10. Slideshare.net. Dynamic Data Analytics for the Internet of Things: Challenges and Opp. <https://www.slideshare.net/PayamBarnaghi/dynamic-data-analytics-for-theinternet-of-things-challenges-and-opportunities/8>
11. Tabunshchyk G. V., Kapliienko T. I., Shytikova O. V. Verification model for the systems with limited resources, *Radio Electronics, Computer Science, Control No 4*, pp. 162 – 167. (2017)
12. Schulz S., Stegwee R., Chronaki C. Standards in Healthcare Data. In: Kubben P., Dumontier M., Dekker A. (eds) *Fundamentals of Clinical Data Science*. Springer, Cham. (2019)
13. Open industry specifications, models and software for e-health. https://www.openehr.org/about/what_is_openehr
14. ISO 18308:2011 Health informatics — Requirements for an electronic health record architecture, 25p. (2011)
15. Introduction to HL7 Standards. <https://www.hl7.org/implement/standards/>
16. HL7 Standards - Section 1c: FHIR®–Fast Health Interop Resources. https://www.hl7.org/implement/standards/product_section.cfm?section=12
17. DICOM® (Digital Imaging and Communications in Medicine). <https://www.dicomstandard.org/>

The Monitoring System Based on a Multi-Agent Approach for Moving Objects Positioning in Wireless Networks

Oleksii Tohoiev¹[0000-0003-3465-7767], Ivan Burlachenko¹[0000-0001-5088-6709],
Iryna Zhuravska¹[0000-0002-8102-9854], Volodymyr Savinov¹[0000-0002-0862-5879]

¹ Petro Mohyla Black Sea National University, 68 Desantnykiv str., 10, Mykolaiv, 54003,
Ukraine
iryna.zhuravska@chmnu.edu.ua

Abstract. The problem of mobile device positioning system deployment is considered. Existing mobile positioning techniques that were intended for the moving object itself will not get on the Wi-Fi sniffer. A radio frequency propagation model method based on a multi-agent approach is proposed. Also, the router, host sniffer and phone can be in different network segments. This makes it possible to track any device that can access the Internet via Wi-Fi or Ethernet, like a mobile device in this approach, and to increase performance by 2% while reducing the number of calculations in the model. The sniffing software that implements the proposed methods is developed. It can be argued that the considered task is a supervised learning task with significantly accelerating the learning process.

Keywords: positioning for mobile devices, radiofrequency propagation, Wi-Fi N, hot spot mode, multi-agent analysis, sniffing software, accelerating the learning process

1 Introduction

The core concept of the Wi-Fi wireless network is the presence of the access point (AP), which connects to the terrestrial network infrastructure (Internet Service Provider) and transmits the radio signal. Usually, the AP consists of a receiver, a transmitter, a cable network interface, and firmware to quickly configure. A space area of 50–100 meters radius (called a hotspot or Wi-Fi zone) is formed around the AP, within which you can use a wireless network. The transmission distance depends on the transmitter power (which is programmatically configured in some equipment models), the presence and characteristics of the interference, the type of antenna. Today, the 802.11n standard is widely used, which provides data rates of up to 320 Mbps [1].

Wi-Fi positioning techniques can be divided into two main groups. One is based on the Cartographic Catalog (CC) [2] of cartography and the other is based on modeling of radio wave propagation (RF model) [3]. The RF model determines the

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

relationship between signal strength and distance. Determining the distance between known points and the mobile device allows using trilateration algorithms [4].

The purpose of the article is to study the operation of location algorithms' efficiency, and to develop a high-precision location algorithm of moving hosts in wireless networks.

2 Formal problem statement

Using RADAR technology, the mobile device uses the CC card of the required space [5]. The original CC map is formed from the coordinates, the CC measurements and the location of the mobile device [6]. The CC card can be based on calculations as well as physical measurements. The strength of the signal from each AP is compared with the corresponding indicators in the database, and the appropriate location is assumed. Transparent location fingerprinting uses a map of baselines. The reference point is a sequence of pairs (ss_j, c_j) . Where ss_j is the set of signal strength measurements and c_j is the corresponding physical coordinates. The agreement of a new set of measurements of the signal strength ss is carried out by choosing k basis points closest to the obtained measurements, the average weight c_f which calculated using the formula (1):

$$c_f = \frac{\sum_{j=1}^k \frac{1}{d(ss_j, ss) + \varepsilon} c_j}{\sum_{j=1}^k \frac{1}{d(ss_j, ss) + \varepsilon}}, \quad (1)$$

where $d(ss_j, ss)$ is the Euclidean distance between two triples of APs and ε is a constant. The average error of this method is 1.78 m, but the maximum error can be up to 10 m.

We propose solving the problem of improving moving hosts' positioning in wireless network accuracy via the identification performed by a neuron network, depending on the types of tasks performed by devices.

3 Literature review

3.1 Technologies for positioning a mobile device based on data mining

The location of the object can be determined both in global and local coordinates. To date, there are widespread global positioning systems GPS, GLONASS, Galileo, Bei-Dou [7].

The main advantages of these technologies are its large coverage and high enough accuracy for determining outdoor locations. The main disadvantages of such systems include the dependence on weather conditions and the inability to use indoors with a

lot of noise both from various equipment inside the building and from the building itself [8].

For positioning in local coordinates (to determine the movement of customers, intra-warehouse logistics, etc.), equipment with built-in Wi-Fi, RFID, Bluetooth modules, as well as accelerometers, compasses, smartphones antennas are used [9, 10]. However, data from these modules are hampered for both registration and analysis.

Probabilistic technologies use the probability distribution of signal strength for each reference point with indicators above the average. For example, the Ekahau Positioning Engine (EPE) uses two evaluation functions to coordinate measurements with a database using the formula (2).

$$C_{xy}(S_i) = \begin{cases} 0, & \text{if } r(\theta) + re(\theta) \leq d(S_i, P) \\ f(\theta)e^{-\lambda\alpha\beta}, & \text{if } r(\theta) - re(\theta) < d(S_i, P) < r(\theta) + re(\theta), \text{ and } \theta \in [0, 2\pi] \\ f(\theta), & \text{if } r(\theta) - re(\theta) \geq d(S_i, P), \text{ and } \theta \in [0, 2\pi] \end{cases} \quad (2)$$

The first function is calculated by the kernel method, and the second function is calculated by the histogram method [11].

The HORUS system is based on Bayesian logic [12]. Its main feature is data grouping, which reduces the computational cost and time required to determine the location of a moving device connected to AP. This approach is also probabilistic. Each reference point contains a sample of 240 measurements. Samples are stored in the form of histograms, each of which combines all APs into a joint distribution.

It should be noted that the need for multiple measurements during continuous monitoring of users to determine their positions leads to high power consumption and, ultimately, to a significant reduction in the operating time of a mobile device without recharging [13, 14]. In this case, it is advisable to redistribute the tasks of continuous monitoring for execution on slower cores of a mobile device with low energy consumption [15].

3.2 Positioning based on modeling the propagation of radio waves

The purpose of this simulation is to express the mathematical relationship between the distance from the receiver to the transmitter and the signal strength. The mathematical expression is obtained from the polynomial regression of the third order. The main advantage of this technology is in positioning speed. In addition, an important point is that APs are stationary with known coordinates. An actual problem is the development of methods for determining the location of a user based on a combination of characteristics of signals from APs.

However, regression requires a large amount of accurate information about signal strength over a fairly long time. This technique provides positioning accuracy from 1 m to 3 m. An integral quadratic quality criterion is used to evaluate the effectiveness of positioning technologies [4].

Several dozen measurements are required to determine the relationship between distance and signal strength. It follows that this model is not fully dynamic.

Studies show that trilateration-based systems are characterized by lower location errors compared to the Bayesian approach [16]. However, it can be seen from the

presented results that on average in 10% of cases a mobile object will be unprotected when a user works on a corporate network [17].

To monitor numerous objects authorized in a wireless corporate network, a multi-agent approach is often used [18, 19]. However, in this case, it is important to take into account the multipath nature of the radio channel in calculating the characteristics of the signal during direct communication of moving objects and communication through access points [20].

In the last decade, the localization of Wi-Fi-based objects has become one of the most popular solutions and is considered the most promising for the study of the raised issues in both scientific and industrial communities [21].

4 The modified method of a moving object positioning

4.1 Wi-Fi hot spot modes

The Wi-Fi net diagram contains at least one AP and can be easily scaled.

It is also possible to connect two clients in Ad-hoc mode when the AP is not in use and clients are connected to network adapters directly. The AP transmits its network identifier (SSID) via special signal packets at a rate of 0.1 Mbps every 100 ms. Therefore, 0.1 Mbps is the lowest data rate for Wi-Fi. Knowing the SSID of the network, the client can determine whether a connection to this AP is possible. When two hotspots with identical SSIDs are in range, the receiver can choose between them based on the signal level data. The Wi-Fi standard gives the customer full freedom in choosing the criteria for connection.

However, the standard does not describe all aspects of wireless LAN construction. Therefore, each manufacturer of the equipment solves this problem in its own way, applying the approaches that it considers to be the best from one point or another. Therefore, there is a need to classify ways to build wireless LANs.

By the pooling of APs into a single system, it's possible to distinguish:

- autonomous APs (standalone, decentralized, smart);
- manageable APs (controller-based, centralized);
- uncontrolled but not autonomous (cloud-based).

By the way of radio channels organization and management, wireless LANs can be distinguished:

- with static radio channel settings;
- with the dynamic (adaptive) tuning of radio channels;
- with the multi-layered structure of radio channels.

4.2 The method of multi-agent analysis

To view the specific traffic data that the agent $A_{m(i)}^s$ sends to the network, you will need more agents $A_{m(i)}^{sn}$ with a special sniffing role. Sometimes multi-agent monitoring software can be installed on the agent's host H_{RoA} (3):

$$RoA = \{ j \in RoA \mid A_{m(i)}^j \}, \quad (3)$$

where RoA is a set of agents' behaviors. This is the most reliable way, but it is not always available. If the H_{RoA} has a wireless network connection, then you can use additional agents' devices $m(i)$. That means, connect it via Wi-Fi to a router or an agent's host with a sniffing role, or if there is no access to a router with Wi-Fi, connect it to an unprotected Wi-Fi network and analyze all traffic's data from another movable agents' hosts (4):

$$RoA_B = \{m(i - k_{lb}), \dots m(i), \dots m(i + k_{rb}) \mid (k_{lb}, k_{rb}) \in B\}, \quad (4)$$

Consider the most typical case a corporate network controlled by a regular Wi-Fi AP. Most corporate APs are running widespread OS, which can be accessed by agents of the multi-agent monitoring system (MAMS) without problems via Telnet protocol.

This experiment will investigate the process of the traffic's redirection from movable agents (or from any other host on the network) to an agent with a sniffing role to analyze the H_{RoA} under monitoring (Fig. 1, a).

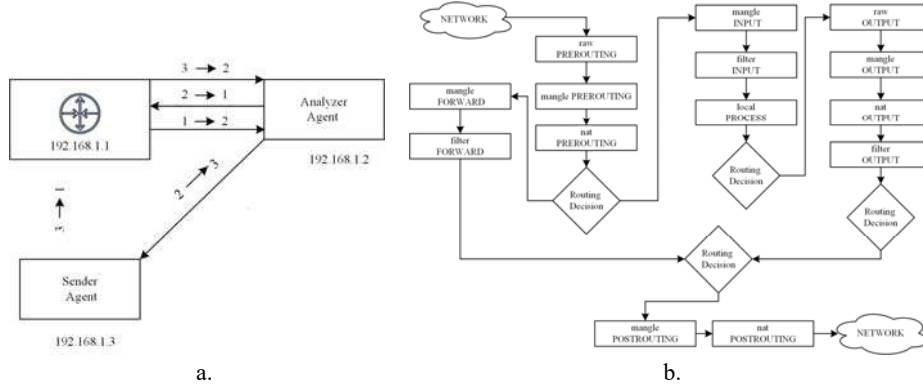


Fig. 1. Dataflow between movable agents at the subnet (a) and *iptables* routing diagram (b)

A network router as an agent was configured on the Linux base. The agent had a setup API available onboard and had network IP address 192.168.1.1. An Analyzer Agent includes the *iptables* and *iproute2* utilities and the *tcpdump* sniffer, with which you can solve almost all network tasks. An Analyzer Agent has an IP address: 192.168.1.2. The device H_{RoA} under investigation is a Sender Agent which connected to the network and has IP address 192.168.1.3.

The research plan is using *iproute2*, we create a separate routing table on the router and a rule that will use it for all traffic from the moving agents. We indicate in the table the default gateway – the host with the sniffing agent's role. Two rules with the

same priority are not allowed. First, you need to check if there is another rule with the same priority: *ip rule list*.

It is advisable to choose a smaller number for priority – the lower the value, the higher the priority. The first four were busy. You can select any unused table (*\$TBL*).

Whether the table is used can be viewed using the *ip route list table \$TBL* command. A host with a sniffing agent's role should redirect all the traffic under investigation to the Network. NAT can handle this easily. Further, this traffic will easily go through the same router and go to its destination – because now IP packets have a different sender address and it won't get into our routing table. That would seem to be all. You can run an Analyzer Agent on the H_{RoA} and filter traffic by IP address 192.168.1.3.

But this scheme has a significant drawback: the traffic that was intended for the router agent itself will not get on the sniffer. To solve this problem the Sniffing Agent role was used (Fig. 2). Judging by the scheme from the *iptables* by MAMS, so that as a result of routing this traffic is redirected (to the right branch), you need to change its destination address in *PREROTING*: 192.168.1.3.

It should be noted that on the sniffer agent at the H_{RoA} host, traffic will fall under masquerading (*MASQUERADE*), that is, the source address will also change for packets. As a result, packets should move according to the following scheme in Fig. 2, b.

However, this option will not work if the H_{RoA} hosts are in the same segment, and the router acts as a network bridge. Then, according to the routing table on the Analyzer Agent's host, the reverse traffic will be sent directly to the recipient, and the recipient's OS should ignore it, because the sender's address will not be the one with which the connection is established.

The proposed way out was implemented in the RoA_B module of MAMS. According to k_{lb}, k_{rb} indices of range with agents' roles RoA_B redirects these packets to the router again, similar to how the redirection to the sniffer was done.

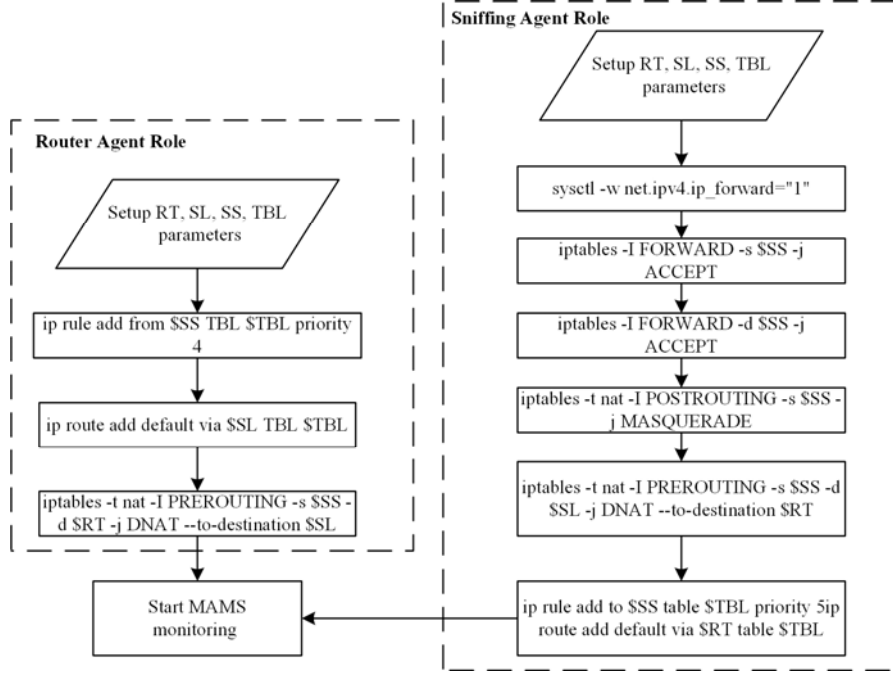


Fig. 2. Multi-agent Monitoring System (MAMS) for the network traffic analysis

When MAMS receiving this problematic traffic, the H_{RoA} with Router Agent Role, using the state-determining mechanism (*nf_conntrack* module), recognizes the connections changed with DNAT in it and replaces the packet receiver address with the original one. Next, the packets are sent to the recipient as if nothing had happened.

5 Experiments and results

The previous methodology that allows analyzing network properties of all networks' agents was described. So that gives the ability to investigate the problem of automatically classifying agents into many categories during establishing multiple connections in different networks' segments (Fig. 3).

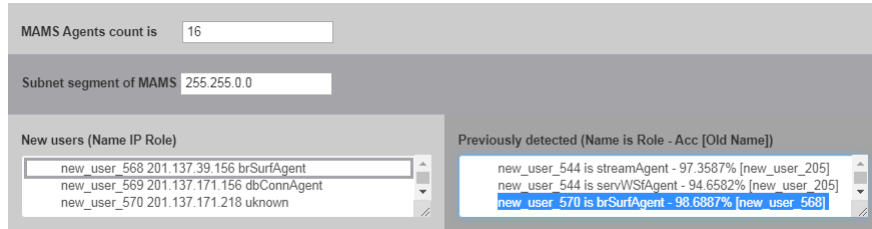


Fig. 3. The user interface of the proposed MAMS during the identification new user's role and network's positioning

As the source data for machine learning were provided a database containing ~10 thousand network packages from MAMS, divided into 8 categories according to the network's activities of agents which were positioned up to 256 network segments. Payloads content of network packages is divided into categories, the names of which were hidden intentionally, because this data cannot be shown, rather unevenly.

Here are features which an agent gets from a payload of the H_{RoA} devices (5):

$$A_H^{PL} = \{A_{DA}, A_{DO}, A_{IDs}, A_{DAI}, A_{DSD}, A_{NC}, A_{CD}\}, \quad (5)$$

A_{DA} describes device attributes: operating system information, hardware and software versions, battery level, signal strength, free storage space, browser type, names and types of applications and files, and plugins. A_{DO} is device operation: information about the device's operation and behavior during use. A_{IDs} describes IDs: the individual IDs, device IDs, and other IDs of the games, apps, or accounts you use, and the IDs of the devices or other unique software IDs for a single device or account. A_{DAI} are device alerts: Bluetooth alerts, as well as information on close Wi-Fi hotspots, beacons, and mobile towers. The feature A_{DSD} is used for storing device settings data. The information you allow us to retrieve through your device settings, such as access to GPS data, cameras, or snapshots. A_{NC} is network and connectivity: information about your network operator or ISP, language, time zone, mobile phone number, IP address, connection speed, and, in some cases, information about other networks H_{RoA} nearby. A_{CD} is cookie data: the cookie data stored on H_{RoA} , including cookie IDs and cookie settings.

5.1 Classification of agent category according to subnet position using the LSTM model

After the supervised training for $A_{m(i)}^{sn}$ agents is defined, its solution can be share between $A_{m(i)}^{sn}$ on the different H_{RoA} . This task for other agents with similar roles is currently being solved by the following algorithm in Fig. 4.

The essence of the task of classifying an agent's category classification when agents appear in another network segment after its preliminary preparation and cleaning is to compile a dictionary of all the features available in A_H^{PL} payload texts, replace each category with a number – a unique word number in the dictionary, align the length of each category and network segments to the needed size (usually the number of elements in the longest feature's vector), and then any classification algorithm can be applied to the data presented in this form. In our case, the raw data was presented in the *MariaDB* database management system (DBMS). These data were successfully uploaded to the *Pandas DataFrame*.

training algorithm will be recorded in the variables *payP1_test*, *payP2_test*, *payP13_test*. This stage takes an input array with descriptions of agent's features strings, an array with a numerical designation of the $A_{m(i)}^s$ agent's category labels and the variable needed to vary the amount of data for the test and verification. In our experimental investigation, 90% of the data for training and 10% for checking the result were used. For varying portions, agents' features mutation behavior dispatcher was implemented. The number of the categories of the agent's behavior in the dictionary we are dealing with is 8 and the number of subnet addresses is up to 256.

To solve the categorization problem the agent logic using Keras model (Fig. 5) was proposed. To define the position of moving agent could be used only 1700, which is approximately 17% of the entire dataset of 10 thousand frames with a payload.

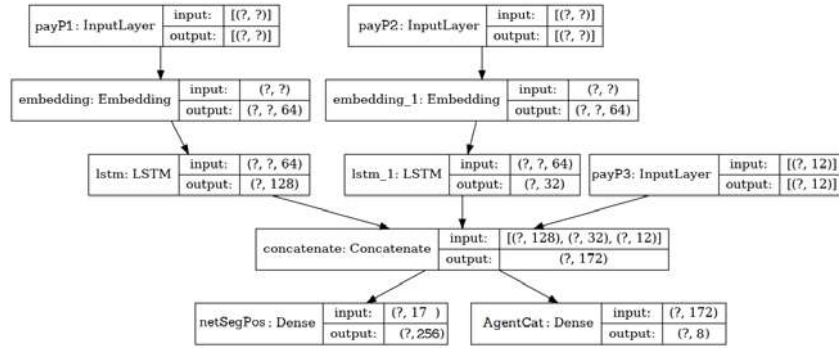


Fig. 5. Directed acyclic graph (DAG) of layers for the proposed agent's behavior identification based on the subnet addresses (*netSegPos*) and category (*AgentCat*)

Usage of mutation behavior dispatcher as part of MAMS significantly reduces the number of calculations and significantly accelerates the learning process. Based on these considerations, training and test data were redefined. Cross-validation data is used to investigate whether your model over-fits the data.

To find the categorization problem solution, Long Short-Term Memory (LSTM) models were included as a layer of DAG. On our data, this model showed 98.7% accuracy in classification (Fig. 6).

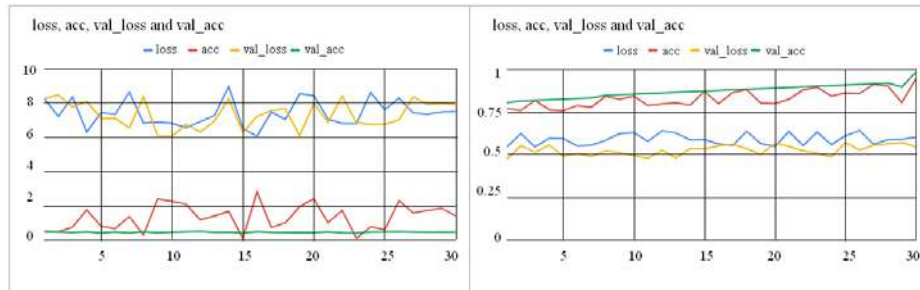


Fig. 6. The accuracy plot of the agent's classification models

We trained the resulting model by writing the learning history into the history store with the predefined batch size that was 32 and 30 epochs. After the model was trained the proposed model accuracy has been validated.

6 Conclusion

Both *iptables*, *iproute2* agent behavior packages must be installed on both the router and agents' hosts, and agents with traffic filtering role and advanced routing support must be included in the kernel. Independent MAMS configuration of the router's kernel has decrease packets lost on 14% during the agent's data exchange and improve performance time by 70 ms faster.

Agents with router-, hostsniffer- and sender roles can be in different network segments. The main condition is that they are connected, and all traffic from the agents passed through the MAMS hosts.

Any other device that can access the Internet via Wi-Fi or Ethernet can act as an agent. Remember that even Wi-Fi protected networks can be unsafe if you are not sure that the AP is completely protected from unauthorized persons. If this is important, use data transfer protocols that support encryption: HTTPS, XMPP.

Of course, the application of this method to a computer with a full-fledged OS is easily monitored using *traceroute* (*tracert* on Windows), or simply because of a decrease in TTL packets, which is noticeable when the router pings.

The proposed MAMS provides the result of the correct definition of the category in 98.7% of cases using LSTM during of agent's behavior analysis.

References

1. Colbach, G.: The WiFi Networking Book: WLAN Standards: IEEE 802.11bgn, 802.11n, 802.11ac and 802.11ax. Independently Published (2019)
2. Longley, P.A., Goodchild, M.F., Maguire, D.J., Rhind D.W.: Geographic Information Science and Systems. John Wiley & Sons (2015)
3. Radio Propagation and Antennas: A non-mathematical treatment of radio and antennas. Cerwin S. AuthorHouse (2019)
4. Pantiukhin, A.R., Belyaev, A.S.: System of determination of objects location inside of premises. *Int. Res. Jnl.* 10 (64): 81–84 (2017). doi: <https://doi.org/10.23670/IRJ.2017.64.012>
5. Bahl, V.: RADAR: An in-building RF-based user location and tracking system. In: Proc. of the 22nd Annual Int. Conf. on Mobile Computing and Networking (MobiCom 2016). <https://www.youtube.com/watch?v=XI5-t5kKENk>
6. Field-Map is a product of IFER – Monitoring and Mapping Solutions, Ltd. Praha, Czech Republic (2019). https://www.fieldmap.cz/download/FM_Catalogue_en.pdf
7. Santerre1, R., Pan, L., Cai, C., Zhu, J.: Single Point Positioning using GPS, GLONASS and BeiDou Satellites. *Positioning*. 5: 107–114 (2014). doi: 10.4236/pos.2014.54013
8. Rumiankov, D., Zhuravska, I., Solobuto, L., Musiyenko, M.: Reduction of noise similar to solar interference in computer networks based on Power Line Communication. In: Proc. of the IEEE 8th Int. Conf. on Intell. Data Acquisition and Adv. Computing Syst.: Technol.

and Applic. (IDAACS'2017), Bucharest, Romania, 21–23 Sept. 2017, vol. 1, pp. 215–221 (2017). doi: 10.1109/IDAACS.2017.8095079

9. Krainyuk, Y., Davydenko, Y., Tomas, V.: Configurable Control Node for Wireless Sensor Network. In: Proc. of the IEEE 3rd Int. Conf. on Adv. Information and Communic. Technol. (AICT'2019), Lviv, Ukraine, 2–6 July 2019, pp. 258–262 (2019). doi: 10.1109/AIACT.2019.8847732
10. Dabove, P., Di Pietra, V., Lingua, A.M.: Positioning Techniques with Smartphone Technology: Performances and Methodologies in Outdoor and Indoor Scenarios. In: Smartphones from an Applied Research Perspective. Nawaz Mohamudally (2017). doi: 10.5772/intechopen.69679
11. Krzysztofik, W.J.: Radio Network Planning and Propagation Models for Urban and Indoor Wireless Communication Networks. In: Antennas and Wave Propagation (2018). doi: 10.5772/intechopen.75384
12. Youssef, M., Agrawala, A.: The Horus WLAN location determination system. In: Proc. of the 3rd Int. Conf. on Mobile Syst., Applic., and Services (MobiSys). Seattle, Washington, USA (2005). doi: 10.1145/1067170.1067193
13. Huber, D.: Background Positioning for Mobile Devices – Android vs. iPhone (2019), https://www.snet.tu-berlin.de/fileadmin/fg220/courses/WS1011/snet-project/background-positioning_huber.pdf
14. Obukhova, K., Zhuravska, I., Burenko, V.: Diagnostics of power consumption of a mobile device multi-core processor with detail of each core utilization. In: Proc. of the IEEE 15th Int. Conf. on Adv. Trends in Radioelectronics, Telecomm. and Computer Engin. (TCSET), Lviv, Ukraine, 25–29 Feb. 2020 (2020)
15. Zhuravska, I.M., Koretska, O.O., Musiyenko, M.P., Surtel, W., et al.: Self-powered information measuring wireless networks using the distribution of tasks within multicore processors. In: Proc. of SPIE: Photonics Applic. in Astr., Communic., Ind., and High Energy Physics Exper., Wilga, Poland, 28 May – 6 June 2017, vol. 10445, pp. 1–13 (2017). doi: 10.1117/12.2280965
16. Markin, D.O., Makeev, S.M.: Mobile device location system model based on Monte-Carlo method. *Izvestiya TulGU. Technical science*. 2: 150–165 (2016)
17. Burlachenko, I., Zhuravska, I., Davydenko, Ye., Savinov, V.: Vulnerabilities analysis and defense based on MAS method in fast dynamic wireless networks. In: Proc. of the Wireless Syst. within of the IEEE 4th Int. Conf. on Intell. Data Acquisition and Adv. Computing Syst. (IDAACS-SWS), Lviv, Ukraine, pp. 98–102 (2018). doi: 10.1109/IDAACS-SWS.2018.8525692
18. Burlachenko, I., Zhuravska, I., Musiyenko, M.: Devising a method for the active coordination of video cameras in optical navigation based on multi-agent approach. *Eastern-European Jnl. of Enterprise Technol.* 1, 9 (85): 17–25 (2017). doi: 10.15587/1729-4061.2017.90863
19. Burlachenko, I.: Management of energy efficient distributed computer systems with self-contained remote modules using multi-agent system. In: Proc. of the IEEE 35th Int. Conf. on Electronics and Nanotechnology (ELNANO), Kyiv, Ukraine, 1: 512–514 (2015). doi: 10.1109/ELNANO.2015.7146940
20. Rida, J.F.A.: Improvement for performance radio frequency in wireless communication based on impulse signal. *Indonesian Jnl. of Electr. Engin. and Computer Sci.* 18 (2): 903–916 (May 2020). doi: 10.11591/ijeecs.v18.i2.pp903-916
21. Wu, C., Yang, Z., Liu, Yu.: *Wireless Indoor Localization: A Crowdsourcing Approach*. Springer (2018)

Parametric synthesis of a dynamic object control system with nonlinear characteristics

Oleg Nikonov^[0000-0002-8878-4318], Igor Kyrychenko^[0000-0002-2128-3500],

Vladyslav Shuliakov^[0000-0002-6654-6749], Fastovec Valentyna^[0000-0002-8428-747X]

Kharkiv National Automobile and Highway University,
Str. Yaroslava Mudrogo 25, Kharkiv, 61002, Ukraine
nikonov.oj@gmail.com, igk160450@gmail.com, jason07@ukr.net,
tinafast2013@gmail.com

Abstract. The aim of the research is to solve the problem of using the method of factor experiment for the problem of parametric synthesis of a closed dynamic object control system. The method of factor experiment allows, on the basis of modeling the behavior of the system in a random external environment, to select its parameters that satisfy the requirements of the minimum of the regression functional. A quality functional that realize the requirements for a dynamic object control system and reduces the computational resource in the parametric synthesis of the system was proposed. The behavior of a closed control system was simulated considering the random external perturbations acting on the control object. The introduction of a nonlinear link with variable amplification factor widens the stability area of a closed dynamic object control system. The parametric synthesis of the closed dynamic object control system considering nonlinear characteristics on the basis of the factor experiment method was considered.

Keywords: Car, Parametric Synthesis, Quality Functional, Transient Process, Dynamic Object.

1 Introduction

Recently, the simulation methods become widespread use, in particular the method of factor experiment. This method allows, on the basis of modeling the behavior of the system in a random external environment, to select its parameters that satisfy the requirements of the minimum of the regression functional. Additive quality functional of the closed system can be used as a regression function. In this case, the behavior of a closed control system is simulated into account of the random external perturbations acting on the control object.

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

2 Formulation of the problem

The aim of the research is to solve the problem of applying the method of factor experiment to the problem of parametric synthesis of the dynamic object control system.

Due to continuous external disturbance $M_{ed}(t)$, the position of the dynamic object is constantly changing. Changing the position of a dynamic object $M_{ed}(t)$ is a random function. In this situation, random functions are also functions $\phi_c(t)$ and $\omega_c(t)$ (Fig. 1). The accuracy of stabilizing a dynamic object is higher, the smaller the area under the dynamic process curves.

3 Literature review

The methods of classical automatic control theory do not work for high-order mathematical models. If the order of the differential equations that make up the mathematical model of the control object is higher than five, then the problem of parametric synthesis can be solved only by the methods of modern control theory, in particular, the method of state-space, the methods of the theory of analytic design of optimal regulators [1-3], minimax methods, methods of Liapunov functions, methods of simulation modeling.

Minimax methods and methods of Liapunov functions, which were applied by E.E. Alexandrov to solve the problem of parametric synthesis of the control system give satisfactory results [4-6], because they provide for the use of information in the control algorithm only about those components of the vector of the state of the control object, the measurement of which is not difficult. But in the process of parametric synthesis of a digital control system, these methods involve the transition from differential equations of the mathematical model of the control object to equations, which introduces a certain error in the synthesis process.

In addition, the high order of the differential equations describing the perturbed motion of the control object results in considerable computational difficulties associated with limited computer memory when applying the minimax method and the Liapunov function method. However, if the objects contain non-analytic nonlinearities, then the above synthesis methods are inactive. Therefore, in recent years, the simulation methods become widespread use, in particular the method of factor experiment [7, 8].

4 Research Methodology

Therefore, the quantitative accuracy of the stabilization of the dynamic object relative to the base position can be estimated by the following functional

$$I_{\phi_c} = M \left[\int_0^T \phi_c^2(t) dt \right] \quad (1)$$

where M – is the symbol of mathematical expectation.

The quality of stabilization processes is estimated not only by the changing of position of the dynamic object, but also by the angular velocity of motion of the dynamic object relative to its base position [9, 10]. The quantitative characteristic of this movement may be functional

$$I_{\omega_c} = M \left[\int_0^T \omega_c^2(t) dt \right] \quad (2)$$

It was an ideal case, when closed dynamic object control system solutions, both functionalities (1) and (2) would be minimized. But this is not possible and the minimums of functionals (1) and (2) correspond to different values of the parameters of the controller k_ϕ and k_ω . Therefore, it is advisable to choose the parameters k_ϕ and k_ω in condition of a minimum of additive functionality

$$I = \beta_1^2 I_{\phi_c} + \beta_2^2 I_{\omega_c} = M \left[\int_0^T [\beta_1^2 \phi_c^2(t) + \beta_2^2 \omega_c^2(t)] dt \right], \quad (3)$$

where β_1 and β_2 – are the weighting factors to be selected.

But, many experiments [11] proved that the obtained values of the varied parameters, which give a minimum of functionals (1) and (2) are almost not different (within 5%). Considering this fact, as well as the fact that the output signal from the control unit is limited and does not allow the dynamic object to pass excessive speed, it is advisable to use functional (1) or its modifications for this particular system.

In general, a functional that evaluates the quality of the stochastic system looks like

$$I = M \left[\int_0^T f[e(t), x(t), y(t), t] dt \right], \quad (4)$$

where f – is the function of error, input and output signals, as well as time. Using different combinations of system and time variables, different quality estimates can be obtained.

To reduce the contribution of a significant initial error and to account for a future error, it is more appropriate to use the form functionality

$$I = M \left[\int_{t_1}^T |\Delta \phi_c(t)| dt \right], \quad (5)$$

and when $M_{ed}(t)=0$ functional (5) is transformed into $I = \int_{t_1}^T t |\Delta\phi_c(t)| dt$.

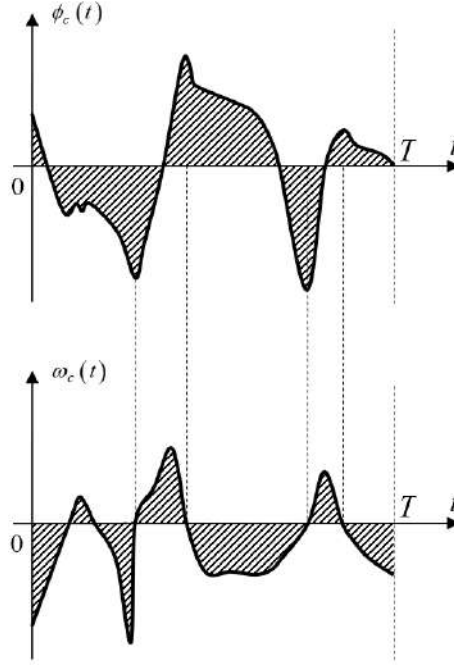


Fig. 1. Dynamic processes in a closed dynamic object control system

Choosing as variables parameters the coefficients k_ϕ and k_ω^* , and as the optimization parameters, the functional (5), we use the factor experiment theory [7, 8] to find the values of the variable parameters of the controller that give the least functional (5) for the dynamic object control system.

The points a, b correspond to the minimum of the objective function (5) without and taking into account $W_{H_1}(A_{H_1})$ respectively. Parameters for (5): $t_1 = 0.25$ s, $T = 5$ s. For the case without taking into account $W_{H_1}(A_{H_1})$ – $I = 0.005958$, $k_\phi^* = 199.5$, $k_\omega^* = 16.3$. For the case considered $W_{H_1}(A_{H_1})$ – $I = 8.186339$, $k_\phi^* = 323.2$, $k_\omega^* = 21.0$.

In Fig. 2 the transients processes of a closed dynamic object control system with the obtained values of the variable parameters of the controller k_ϕ^* and k_ω^* for the case without taking into account $W_{H_1}(A_{H_1})$ are presented. As you can see, transients processes are smooth without significant fluctuations. The amount of overshoot was 0.5 %.

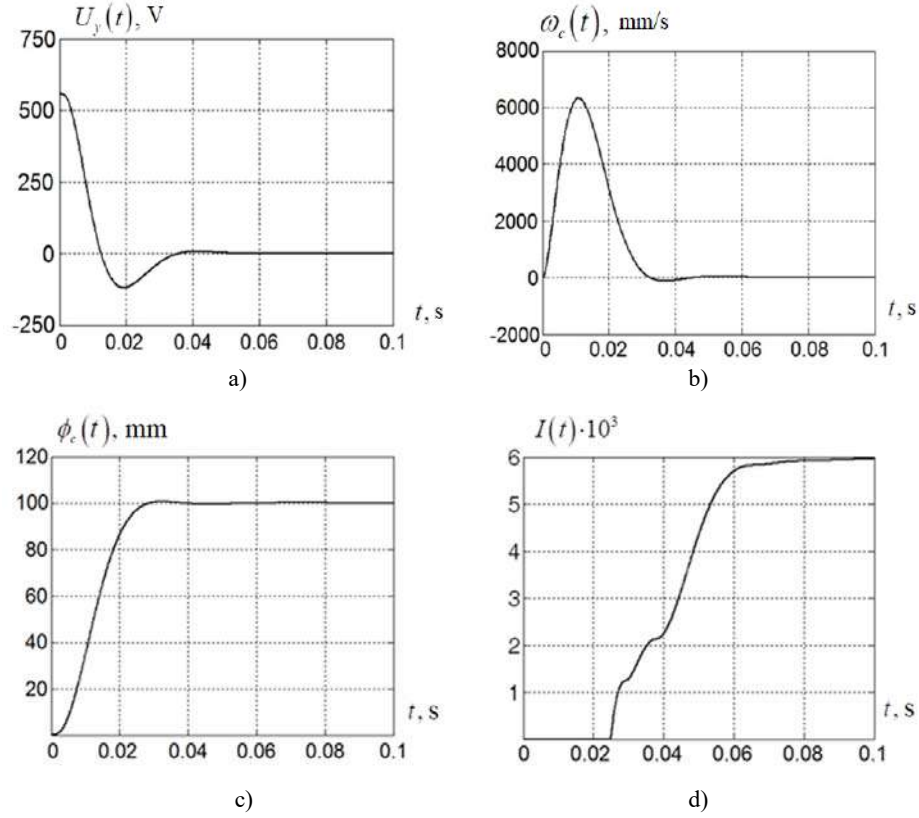


Fig. 2. Transients processes of a closed dynamic object control system when $k_{\phi}^* = 199.5$, $k_{\omega}^* = 16.3$: control voltage $U_y(t)$ (a), dynamic object speed $\omega_c(t)$ (b) and for the value of moving a dynamic object 100mm $\phi_c(t)$ (c) and the objective function $I(t)$ (d)

In Fig. 3 the transients processes of a closed dynamic object control system with the obtained values of the variable parameters of the controller k_{ϕ}^* and k_{ω}^* in the case of the given $W_{H_1}(A_{H_1})$ are presented. As you can see, transients processes are smooth without significant fluctuations. The amount of overshoot was 0.1 %.

A structural diagram of a dynamic object control system, in addition to a nonlinear link with a transfer function $W_{H_1}(A_{H_1})$ and characteristic, of the limited (saturation) zone type also contains a nonlinear link with a transfer function $W_{H_2}(A_{H_2})$ and characteristic with variable amplification factor.

In this case, this nonlinear link is introduced into the structural diagram in order to improve the quality of transients processes and increase the reliability of the system as a whole by extending the stability area.

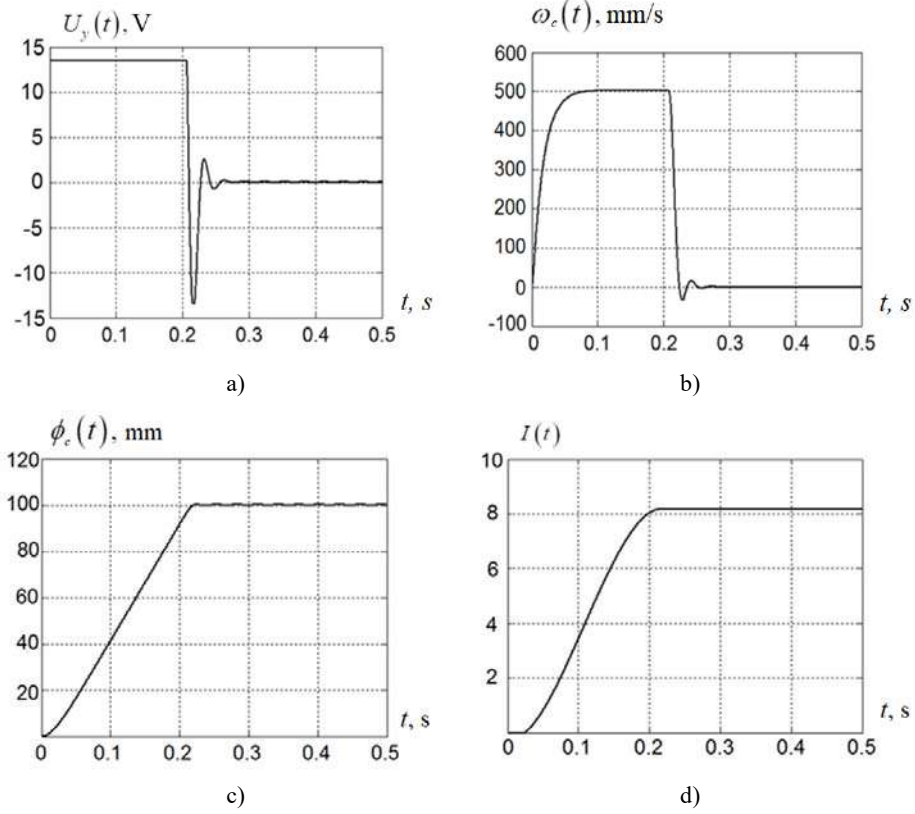


Fig. 3. Transients processes of a closed dynamic object control system when $k_\phi^* = 323.2$, $k_\omega^* = 21.0$: control voltage $U_y(t)$ (a), dynamic object speed $\omega_c(t)$ (b) and for the value of moving a dynamic object 100mm $\phi_c(t)$ (c) and the objective function $I(t)$ (d)

From the structural diagram we find the transfer function of the closed circuit with the nonlinear link with the transfer function $W_{H_2}(A_{H_2})$

$$W'_{3HH}(s) = \frac{W_{3H}(s) \cdot W_{H_2}(A_{H_2})}{1 + W_{3H}(s) \cdot W_{H_2}(A_{H_2}) \cdot k_\omega \cdot \frac{k_{gs}}{T_{\omega 1}^2 s^2 + T_{\omega 2} s + 1} \cdot s}. \quad (6)$$

We find the transfer function of the whole open system

$$W_{4HH}(s) = \frac{k_\phi \cdot W_{3H}(s) \cdot W_{H_2}(A_{H_2})}{1 + W_{3H}(s) \cdot W_{H_2}(A_{H_2}) \cdot k_\omega \cdot \frac{k_{gs}}{T_{\omega 1}^2 s^2 + T_{\omega 2} s + 1} \cdot s}$$

and a closed system

$$W'_{4HH} = \frac{W_{4HH}(s)}{1 + W_{4HH}(s)}. \quad (7)$$

We construct the stability domain of a closed dynamic object control system in the plane of variable parameters of the controller k_ϕ and k_ω , taking into account a nonlinear link such as a zone of limitation (saturation) and a nonlinear link with a variable amplification factor.

$$\begin{aligned} & T_y A_3 T_{\omega 1}^2 s^7 + (T_y A_3 T_{\omega 2} + T_y A_2 T_{\omega 1}^2 + A_3 T_{\omega 1}^2) s^6 + \\ & + (A_2 T_{\omega 1}^2 + T_y A_2 T_{\omega 2} + T_y A_3 + A_3 T_{\omega 2} + T_y A_1 T_{\omega 1}^2) s^5 + \\ & + (T_y T_{\omega 1}^2 + A_2 T_{\omega 2} + T_y A_2 + A_3 + T_y A_1 T_{\omega 2} + A_1 T_{\omega 1}^2) s^4 + \\ & + (T_{\omega 1}^2 + A_1 T_{\omega 2} + T_y A_1 + T_y T_{\omega 2} + A_2) s^3 + \\ & + (T_{\omega 2} + A_1 + T_y + k_\phi k W_{H_1}(A_{H_1}) W_{H_2}(A_{H_2}) k_y k'_2 T_{\omega 1}^2) s^2 + \\ & + (1 + k W_{H_1}(A_{H_1}) W_{H_2}(A_{H_2}) k_y k'_2 (k_{\phi 2} T_{\omega 2} + k_\omega k_{gs})) s + \\ & + k_\phi k W_{H_1}(A_{H_1}) W_{H_2}(A_{H_2}) k_y k'_2 = 0. \end{aligned} \quad (8)$$

From the transfer function of the closed system (7) we write the characteristic equation (8). In characteristic equation (8) we make a substitution $s = j\omega$, select the real and imaginary parts and equal them to zero. Obtained algebraic equations we solve referring to the parameters k_ϕ and k_ω :

$$k_\phi = \frac{\left(T_y T_{\omega 1}^2 + A_2 T_{\omega 2} + T_y A_2 + A_3 + T_y A_1 T_{\omega 2} + A_1 T_{\omega 1}^2\right) \omega^4}{k W_{H_1}(A_{H_1}) W_{H_2}(A_{H_2}) k_y k'_2 (\omega^2 T_{\omega 1}^2 - 1)} - \quad (9)$$

$$- \frac{\left(T_y A_3 T_{\omega 2} + T_y A_2 T_{\omega 1}^2 + A_3 T_{\omega 1}^2\right) \omega^6 + \left(T_{\omega 2} + A_1 + T_y\right) \omega^2}{k W_{H_1}(A_{H_1}) W_{H_2}(A_{H_2}) k_y k'_2 (\omega^2 T_{\omega 1}^2 - 1)};$$

$$k_\omega = \frac{\left(T_{\omega 1}^2 + A_1 T_{\omega 2} + T_y A_1 + T_y T_{\omega 2} + A_2\right) \omega^2 + T_y A_3 T_{\omega 1}^2 \omega^6}{k W_{H_1}(A_{H_1}) W_{H_2}(A_{H_2}) k_y k'_2 k_{gs}} -$$

$$- \frac{\left(A_2 T_{\omega 1}^2 + T_y A_2 T_{\omega 2} + T_y A_3 + A_3 T_{\omega 2} + T_y A_1 T_{\omega 1}^2\right) \omega^4}{k W_{H_1}(A_{H_1}) W_{H_2}(A_{H_2}) k_y k'_2 k_{gs}} - \quad (10)$$

$$- \frac{k_\phi k W_{H_1}(A_{H_1}) W_{H_2}(A_{H_2}) k_y k'_2 T_{\omega 2} + 1}{k W_{H_1}(A_{H_1}) W_{H_2}(A_{H_2}) k_y k'_2 k_{gs}}.$$

In Fig. 4 and 5 are shown the limits of stability area of the closed dynamic object control systems in the plane of variable parameters of the controller k_ϕ and k_ω taking into account a nonlinear variable with amplification factor, constructed using relations (9) and (10), where a, b, c – are the points of minimum of the objective function (5) without and with considering $W_{H_1}(A_{H_1})$, $W_{H_2}(A_{H_2})$ respectively.

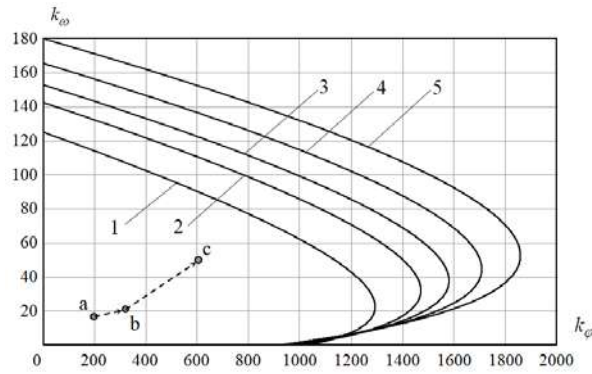


Fig. 4. Stability area of a closed dynamic object control system considering the nonlinear link with variable amplification factor at $A_{H_2}/b_{H_2} = 2 : 1 - k_{H_2} = k_{H_3}$; $2 - k_{H_2} = 0.8 k_{H_3}$; $3 - k_{H_2} = 0.7 k_{H_3}$; $4 - k_{H_2} = 0.6 k_{H_3}$; $5 - k_{H_2} = 0.5 k_{H_3}$

For the case with $W_{H_1}(A_{H_1})$ and $W_{H_2}(A_{H_2})$ based on the method of factor experiment we obtain: $I = 8.190428$, $k_\phi^* = 607.1$, $k_\omega^* = 49.9$. Parameters for (5): $t_1 = 0.25$ s, $T = 5$ s.

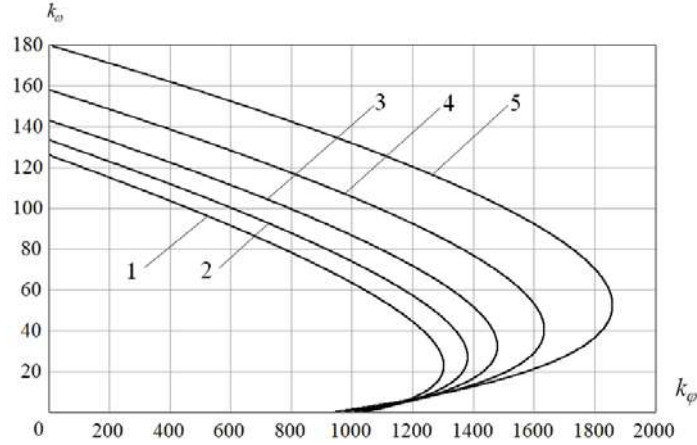


Fig. 5. Stability area of a closed dynamic object control system considering the nonlinear link with variable amplification factor at $k_{H_2} = 0.5 k_{H_3}$: 1 – $A_{H_2}/b_{H_2} = 100$; 2 – $A_{H_2}/b_{H_2} = 10$; 3 – $A_{H_2}/b_{H_2} = 5$; 4 – $A_{H_2}/b_{H_2} = 3$; 5 – $A_{H_2}/b_{H_2} = 2$

As you can see from Fig. 4, the introduction of a nonlinear link [12 - 14] with variable amplification factor widens the stability area of a closed dynamic object control system (with $k_{H_2} = k_{H_3}$ (curve 1), the stability area coincides completely with the stability of the linear system).

The optimal point c in this case moves to the area of high coefficients k_ϕ and k_ω . The above helps to improve the reliability and accuracy of a closed dynamic object control system.

As you can see from Fig. 5, when $A_{H_2}/b_{H_2} = 100$ the equation $W_{H_2}(A_{H_2}) \rightarrow k_{H_3}$, and the stability area approaches to the linear.

In Fig. 6 the transients processes of a closed dynamic object control system with the obtained values of the variable parameters of the controller k_ϕ^* and k_ω^* considering $W_{H_1}(A_{H_1})$ and $W_{H_2}(A_{H_2})$ are presented. The amount of overshoot was 0.1%.

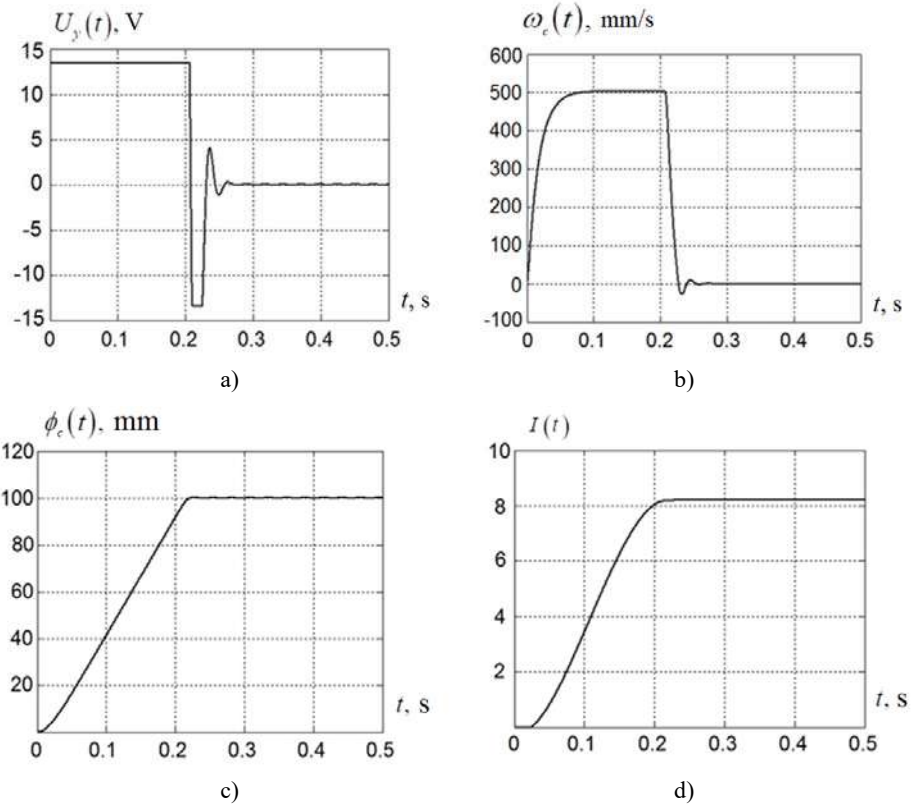


Fig. 6. Transients processes of a closed dynamic object control system when $k_\phi^* = 607.1$, $k_\omega^* = 49.9$: control voltage $U_y(t)$ (a), dynamic object speed $\omega_c(t)$ (b) and for the value of moving a dynamic object 100mm $\phi_c(t)$ (c) and the objective function $I(t)$ (d)

For a more positive result when using a variable structures need to use a more complicated law of the changing the structure of the system, but it can reduce the reliability of the system as a whole, due to the additional elements and connections.

5 Conclusions

The aim of the research to solve the problem of using the method of factor experiment for the problem of parametric synthesis of a closed dynamic object control system was considered. Quality functional that realize the requirements of a dynamic object control system and reduce the calculating resource of parametric system synthesis was proposed. Parametric synthesis of a closed dynamic object control system

with nonlinear characteristics on the basis of the factor experiment method was made by using three practical examples.

It was proved that the method of factor experiment allows, on the basis of modeling the behavior of the system in a random external environment, to select its parameters that satisfy the requirements of the minimum of the regression functional. The behavior of a closed control system was simulated considering the random external perturbations acting on the control object. It was defined that the introduction of a nonlinear link with variable amplification factor widens the stability area of a closed dynamic object control system.

Reference

1. Letov, A.: Flight dynamics and control. M, Science (1969) (in Russian)
2. Letov, A.: Mathematical theory of control processes. M, Science (1981) (in Russian)
3. Roitenberg, Y.: Automatic control. M, Science (1978) (in Russian)
4. Aleksandrov, E., Bogaenko, I., Kuznetsov, B.: Multichannel optimal control systems. K, Tekhnika (1995) (in Russian)
5. Aleksandrov, E., Kostenko, Y., Kuznetsov, B.: Optimization of multichannel control systems. Kharkiv, Osnova (1996) (in Russian)
6. Aleksandrov, E., Borisyuk, M., Kuznetsov, B.: Parametric optimization of multichannel automatic control systems. Kharkiv, Osnova (1995) (in Russian)
7. Adler, Y., Markova, E., Granovskii, Y.: Planning of the experiment in search of optimal conditions. M, Science (1976) (in Russian)
8. Alexandrov, E., Nikonov, O., Skvorchevsky, O.: Structural and parametric synthesis of the system of automatic control of the car brakes. Energy and resource saving, vol. 6, pp. 30-39 (2009) (in Ukrainian)
9. Xu, G., Ma, Z, Lu, F., Hou, P.: Kinematic Analysis of Hydraulic Excavator Working Device Based on DH Method. In: International Conference on Applied Mechanics, Mechanical and Materials Engineering, pp. 8 (2016). doi: 10.12783/dtmse/ammme2016/6857
10. Gurko, A., Kyrychenko, I., Yaryzhko, A.: Trajectories Planning and Simulation of a Backhoe Manipulator Movement. CMIS, pp. 771-785 (2019)
11. Xu, J., Yoon, H. S.: A Review on Mechanical and Hydraulic System Modeling of Excavator Manipulator System. Journal of Construction Engineering 2016, 9409370 (2016). doi: 10.1155/2016/9409370.
12. Uspensky, B., Avramov, K., Liubarskyi, B., Andrieiev, Y., Nikonov, O.: Nonlinear torsional vibrations of electromechanical coupling of diesel engine gear system and electric generator. Journal of Sound and Vibration, vol. 460, 114877 (2019) doi:10.1016/j.jsv.2019.114877
13. Uspensky, B., Avramov, K., Nikonov, O.: Nonlinear modes of piecewise linear systems forced vibrations close to superharmonic resonances. Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science, vol. 233, Issue 23-24, pp. 7489-7497 (2019) doi: 10.1177/0954406219869967
14. Gu, J., Ma, X. D., Ni, J. F., Sun, L. N.: Linear and nonlinear control of a robotic excavator. J. Cent. South Univ. 19, pp. 1823–1831 (2012). doi: 10.1007/s11771-012-1215-y

Intelligence Analysis Method of Automation Control System Archive Database for controlling Hot Blast Stove Block

Oleksiy Koyfman¹ [0000-0003-2075-7417], Oleksandr Simkin² [0000-0002-9939-7866]
Kseniia Serdiuk³ [0000-0002-4598-0562]

Pryazovskyi State Technical University, str. Universytets'ka 7, Mariupol, 87555, Ukraine,

¹koifman_o_o@pstu.edu, ²simkin@pstu.edu
³serdiuk.xen@gmail.com

Abstract. In automated process control systems the data from the archive database are used in most cases in a graphic form for analyzing the technological process history. Frequently, the controlling of a hot blast stove block occurs in the conditions of incomplete information. The method of data mining of the archive database was developed and implemented in the form of computer programs to obtain new interconnections between the main technological parameters of the blast air heating process in a hot blast stove block. The developed algorithm allows to determine the functional dependence of the base parameter on two other parameters with fixing the third parameter in a certain range. The obtained analytical dependencies allow us to make decisions about changes in control algorithms for a hot blast stove block.

Keywords: intelligence analysis, data mining, method, approximation, hot blast stove, HBS, flow-chart, computer application

1 Introduction

Nowadays, all the basic technological processes of metallurgical production are automated by Automated Control Systems (ACS) of various levels of complexity, in which the technological parameters measured by the sensors, the values of the calculated variables enter the visualization systems and are stored in current databases. This information is used both for process control and for informing personnel about the process condition. Current databases are stored forming an archive database of the technological process (ADB). Unfortunately, in most cases ADB information is used to display the process flow in a graph form and, if necessary, to evaluate the operation of the unit in the past.

The software of ACS includes programs that implement mathematical models – informational or controlling. During the operation of the technological facility, such

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

tuning coefficients of these programs must be adjusted due to aging or a change in the state of equipment and structural elements.

To solve this problem, some ACS are used data mining in recent years.

The goal of the research: the development and application of new methods of data mining to obtain new interconnections between the technological parameters of blast air heating in hot blast stove block (HBSB) in order to configure control algorithms for HBSB.

2 Formal problem statement

In many cases HBSB consists of three or four hot blast stoves (HBS). The block is designed to heat blast air (air enriched with oxygen up to 25-30%) to a temperature of 1000 - 1200 °C, which enters a blast furnace (BF).

HBS is a unit of discontinuous operation. The main two modes of operation are the checkerwork heating (on-gas period) and the blast heating (on-blast period). In on-gas period in the combustion chamber a fuel gas is burned (blast furnace gas (BFG) or a mixture of BFG and natural or coke gas). Then the hot combustion products heat the block refractory checkerwork, giving it the heat. In on-blast period, BF air passes through the hot checkerwork in the direction opposite to the combustion products, taking the heat from the checkerwork. In case of operation of HBSB, which includes 4 units, in sequential mode, three HBS operate in on-gas period, and one operates in on-blast period. As soon as one of HBS ends on-blast period, it is switched to the on-gas period, and the heated HBS is switched to on-blast period.

HBSB is controlled by an automation system that also includes a subsystem for maintaining a database of process parameters for each HBS. In modern ACS, there is a system for controlling the dome temperature, the fuel-air ratio and a monitoring system for a fuel gas flow, an air for combustion flow, a flue gases temperature, chemical analysis of the flue gases composition and fuel value.

In the studied system, there is a system for controlling the dome temperature and a system for monitoring the fuel gas flow and the temperature of the flue gases. Such number of parameters does not allow a clear understanding of process quality, i.e. in practice, ACS operates in the conditions of incomplete information.

Thus, the objective of the research aims at by using ADB that contains the above-mentioned minimum set of parameter values for the long-term operation of HBSB to find the interconnections between these parameters in order to their further use to correct the existing mathematical models of HBS operation and in block control algorithms.

3 Literature review

The article [1] describes various methods used in data analysis and approaches to the implementation of these methods.

Liu Y. et al. [2] use accumulated production data to compensate for the lack of technological information for the gas flow and conclude that the obtained sufficient

accuracy of reproducing the missing data provides powerful support for the operation of gas blast furnace facilities.

The quality and accuracy of the technological information stored in the current and archive databases are very important for the decision-making process in order to save resources. Unfortunately, due to design errors, instability and vulnerability of the Supervisory Control and Data Acquisition (SCADA), these anomalies usually exist in practice. This usually leads to a negative impact on the accuracy and stability of data-driven methodologies. The authors of [3] proposed two methods for searching for anomalies in the data presented in the current database.

Methodology, problems and solutions in the field of using multivariate polynomial regression are described in [4]. According to the authors, descriptive data analysis is a process of evaluating values based on a given data set, which can be achieved by regression analysis on this data set. In real time, data can include various data sets with different properties. Predictive (intelligent) Data Mining is the process of evaluating or predicting future values from an available set of values, uses the concept of regression to evaluate future values.

An example of the use of quadratic multiparametric polynomial regression is [5], where the authors performed an analysis to determine the relationships between the parameters (gas temperature, wall thickness) of heat exchange during drying with infrared convection. By analyzing the dispersion, the authors evaluate the order of influence of the significance of all factors on the temperature of hot air.

As indicated above, both the blast furnace itself and the hot blast stoves are units operating in conditions of incomplete information. For such objects, according to various authors, the only real opportunity for implementing management functions is data analysis.

The standard approach to the analysis of BF process is based on a detailed study of its individual phenomena (heat and mass transfer, gas and hydrodynamics, etc.) with the subsequent use of the identified patterns for determination of possible and desired trends in the development of processes and management. The authors of [6] proposed a thermochemical model of BF smelting, allowing the operator to apply corrective actions, based on an analysis of data from an archive database for 6 months. The calculation results obtained on the basis of the model were compared with real data.

The success of cluster data analysis primarily depends on the controllability of BF processes based on existing sets of measured parameters and control factors [7]. The uncertainty of the control of BF process is determined by uncontrolled parameters, which determine the spread of the values of the output parameters of the process. An analysis of the controllability of BF processes, carried out in this work on the basis of existing statistics of BF smelting, shows that setting specific values of control factors determine the productivity of BF with a certain accuracy. This fact justifies the management technique proposed in this paper. It is based on real data as applied to the calculation of the theoretical combustion temperature.

According to the authors of [8], measurements of technological parameters of BF smelting usually have non-Gaussian characteristics. This problem can be solved if we do not stick to the analytical solution, but use the empirical method to determine the control limit. In the general case, if the distribution of the measurement data is uni-

modal and the detection index value is constant for the data under normal conditions. To diagnose disturbances in the operation of BF, the authors of the article used multivariate statistical analysis, the purpose of which was to monitor disturbances associated with switching HBS to detect anomalies.

To monitor the process of cast iron production in BF, the authors of [9] use the Principal Component Analysis (PCA) method of analysis of the main components - a statistical procedure that orthogonally converts the initial n coordinates of the data set into a new set of n coordinates called the main components. The work shows that the distribution is not normal for the following reasons: peak-like interference caused by HBS switching; periodic loading of coke and tapping of cast iron, etc. Before detecting abnormalities, outliers are removed from the original database, the size of the training sample and the number of main components are selected.

In article [10] an approach is proposed to control BF process to increase the efficiency of process, based on modeling the operation of BF and on-line stabilization of parameters on the basis of the obtained mathematical models. Using machine learning methods, dependencies are determined to specify the controlled parameters at the output and coke speed. Intelligent data analysis for BF operation was performed to identify models of distribution and subsequent melting of materials. Based on the developed models an expert system was developed and implemented.

In practice, operators of BF make decisions on controlling the unit based on a visual assessment of the graphic profiles of key variables, and not through a complex calculation of the exact numerical values. Zhou, B. et al. [11] proposed a new qualitative trend analysis algorithm (QTA) based on piecewise linear polynomial approximation. Using the first and second derivatives, the rate of change and the direction of the trend are determined. The effectiveness of the algorithm is shown by testing on various models and real data of blast furnaces.

An analysis of the existing quality control systems performed by the author [12] showed that currently there are practically no systems that could be used for operational control of the dynamics of the quality of cast iron during the smelting process in real time. In this regard, the article presents a system for collecting and analyzing data on the controlled parameters of the BF process with the aim of real-time monitoring of the quality of cast iron during smelting. Data on the current values of the technological parameters are sent to a server from the controllers every 3 seconds and are smoothed using an exponential algorithm with the choice of the optimal parameter.

The work [13] describes the operation of a BF automation system based on the use of an archive database in the mode of an operator advisor. The author outlines a methodology for analyzing and optimizing the volume of the operational database in order to predict the course of BF smelting in real time and offers a method for analyzing data of the technological parameters, which allows you to obtain additional information about the smelting process. When designing the system, hierarchical clustering (the Lance-Williams method) was used, which allows them to be found in the process of performing the clustering procedure.

The main goal of [14] is to show the potential of modern intelligent technologies based on self-organizing neural networks for clustering operating modes and model-predictive control of the BF process. The method uses a prediction model of control

technology. Moreover, the construction of a model of the BF process includes data obtained in real time. The authors propose to use two approaches to clustering the effective values of parameters using the criteria for the efficiency of a blast furnace: using elliptical surfaces and a self-organizing Kohonen network.

A sufficient amount of research also applies to the use of modeling, solving optimization problems, multivariate statistical analysis, cluster analysis and data mining for the operation of automation systems for HBS.

For example, Nose, K. et al. [15] formulated and solved using modeling the non-linear optimization problem by using the generalized reduced gradient (GRG) method.

For a HBS control system the authors of [16] proposed a sequential procedure for improving process control based on multivariate methods of statistical analysis. In order to take advantage of a large amount of historical data, a combination of the hierarchical clustering method and statistical process control methods is used to detect and analyze key factors that significantly affect process performance. This technique consists of four successive stages: data collection and multivariate statistical analysis, hierarchical clustering and detection of the operating mode, the selection of dominant variables, a new operational benchmark and its verification.

The authors of [17] confirm that with intelligent data processing the Six Sigma Methodology increases the processing speed and the quality of the “gained” data.

4 Statement of Basic Materials

The procedure for obtaining the desired information from the archive database is shown in Fig. 1.

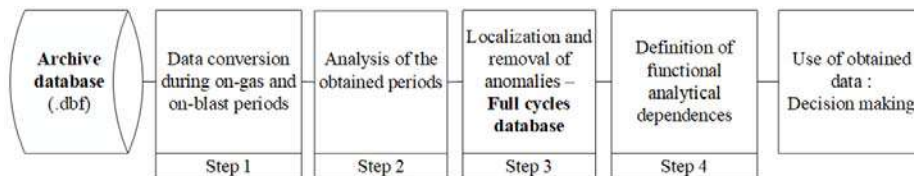


Fig. 1. Stages of information conversion

At one of the iron and steel plants of Ukraine, the values of 194 technological parameters of HBS in ACS for cast iron smelting are recorded in ADB (file with the extension *.dbf) with a frequency of 5 seconds, and the file is generated for every 24 hours and, accordingly, one file contains 3352320 records.

The ADB file structure is shown in Fig. 2, where Date and Time columns correspond to the date and time the parameter was written to the database, TagIndex is the index corresponding to the technological parameter, Value is the value of the technological parameter; Tagname – the name of the technological parameter in the database.

The studied ADB has a form that is inconvenient for its further analysis and use due to its structure (outdated recording format) and accordingly due to its size.

	A	B	C	D		A	B
1	Date	Time	TagIndex	Value	1	Tagname	TagIndex
19	26.02.2019	00:00:11	17	0,000	19	VN1\D	17
24	26.02.2019	00:00:11	22	1,000	24	VN1\N	22
26	26.02.2019	00:00:11	24	0,000	26	VN1\O	24
29	26.02.2019	00:00:11	27	23982,777	29	VN1\RSG	27
42	26.02.2019	00:00:11	40	122,300	42	VN1\TD	40
49	26.02.2019	00:00:11	47	1240,300	49	VN1\TKF	47
52	26.02.2019	00:00:11	50	0,000	52	VN2\D	50

a)

b)

Fig. 2. ADB structure: a) file with values, b) tag table

Therefore, firstly, it is necessary to transform the database into a form convenient for further analysis. The authors developed and implemented an ADB processing algorithm taking into account the following provisions:

1. The operation of each HBS will be described with the characteristics (values of technological parameters) of two periods – on-gas and on-blast including:

a) for on-gas period: time of the beginning and end of the period ($H_{G_{beg}}$, $M_{G_{beg}}$, $S_{G_{beg}}$, $H_{G_{end}}$, $M_{G_{end}}$, $S_{G_{end}}$); the duration of the period; the beginning and end values of the dome temperature ($TDG_{HBS1_{beg}}$, $TDG_{HBS1_{end}}$); the beginning and end values of the flue gases temperature ($TFG_{HBS1_{beg}}$, $TFG_{HBS1_{end}}$); the average integral value of the dome temperature ($TDG_{HBS1_{avg}}$); the average integral value of the flue gases temperature ($TFG_{HBS1_{avg}}$); dome heating rate; the rate of heating of the flue gases; the beginning and end consumption of the BFG ($FG_{HBS1_{beg}}$, $FG_{HBS1_{end}}$); volume of the consumed BFG (VG_{HBS1}); the average integral flow rate of the BFG during the period ($FG_{HBS1_{end}}$);

b) for on-blast period: time of the beginning and end of the period ($H_{B_{beg}}$, $M_{B_{beg}}$, $S_{B_{beg}}$, $H_{B_{end}}$, $M_{B_{end}}$, $S_{B_{end}}$); the duration of the period; the beginning and end values of the dome temperature ($TDB_{HBS1_{beg}}$, $TDB_{HBS1_{end}}$); the beginning and end values of the temperature of the blast air ($TBB_{HBS1_{beg}}$, $TBB_{HBS1_{end}}$); the average integral value of the dome temperature ($TDB_{HBS1_{avg}}$); the average integral value of the blast temperature at the inlet to HBS ($TBB_{HBS1_{avg}}$); the cooling rate of the dome temperature during the period.

2. Divide ADB into 8 parts (files), each of which will contain data on on-gas and on-blast periods for each HBS of the block.

Let us consider in details the proposed algorithm (Fig. 3) for converting the ADB into a form convenient for analysis of HBS #1 in particular.

When the program is initialized, a directory in which the ADB files are located and a directory in which the processing results will be written are selected.

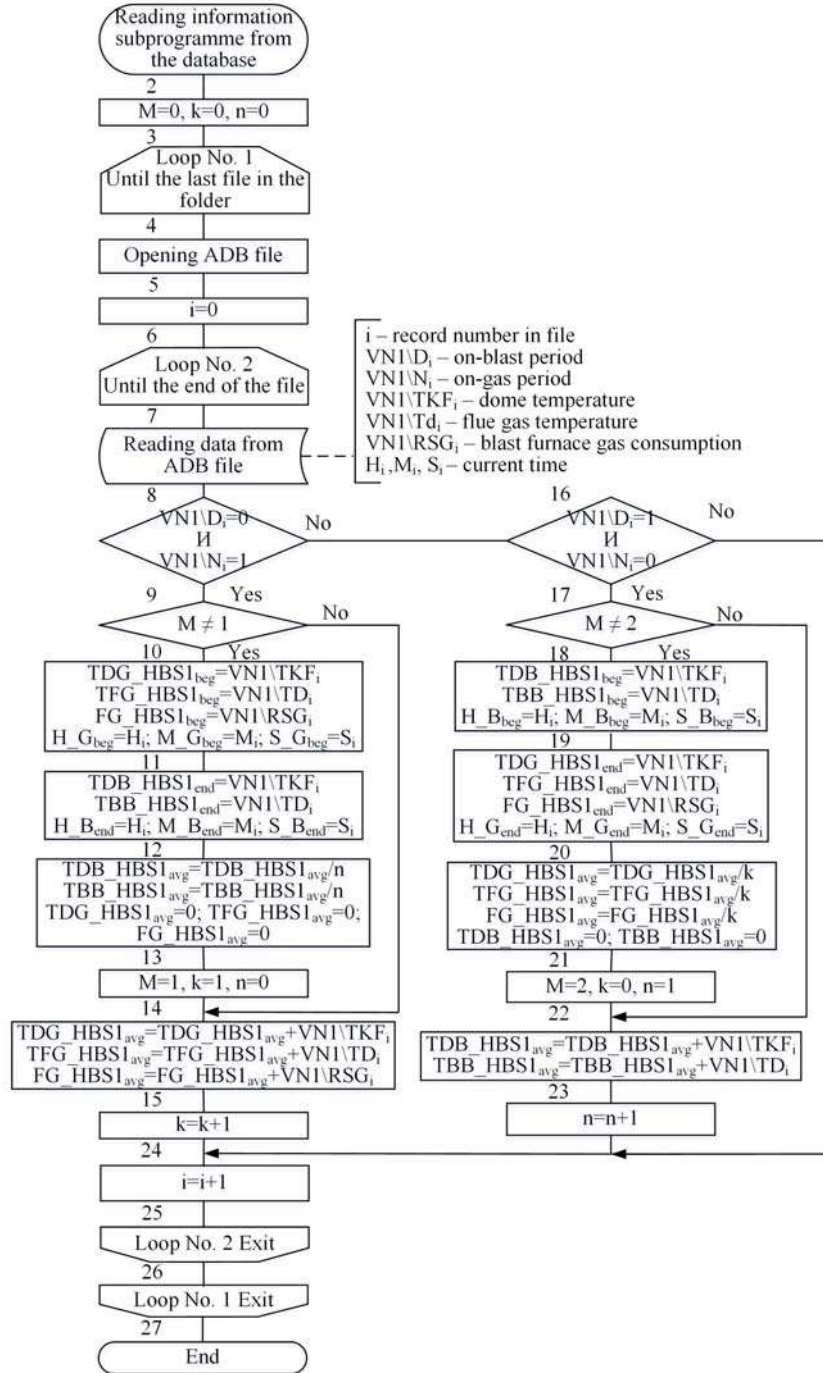


Fig. 3. Flow-chart of algorithm of ADB conversion in a form convenient for analysis

Step 1. Beginning of reading of information subprogramme from the database.

Step 2. Resetting variables: M – mode of operation of HBS, $M = 1$ – on-gas period, $M = 2$ – on-blast period, n – number of parameters values recorded over on-gas period; k – number of parameters values recorded over on-blast period.

Step 3. Loop No. 1 starting, it operates until all ADB files from the folder are processed.

Step 4. Opening the next ADB file.

Step 5. Resetting variable: i – number of records in a file.

Step 6. Loop No. 2 starting, it operates until the end of a file is reached.

Step 7. Reading data from ADB file using SQL-queries: i ; $VN1 \setminus D_i$ – on-blast period; $VN1 \setminus N_i$ – on-gas period; $VN1 \setminus TKF_i$ – dome temperature; $VN1 \setminus TD_i$ – flue gas temperature; $VN1 \setminus RSG_i$ – BFG flow.

Step 8. Mode checking: $VN1 \setminus D_i = 0$ AND $VN1 \setminus N_i = 1$? If YES – on-gas period, transition to step 9, if NOT – transition to step 16.

Step 9. Mode change checking: $M \neq 1$? If YES – mode changing from on-blast period to on-gas period, transition to step 10; if NOT – transition to step 14.

Step 10. Assigning initial values for on-gas period beginning: TDG_HBS1_{beg} , TFG_HBS1_{beg} , FG_HBS1_{beg} , H_G_{beg} , M_G_{beg} , S_G_{beg} .

Step 11: Assigning finite values for on-blast period end: TDB_HBS1_{end} , TBB_HBS1_{end} , H_B_{end} , M_B_{end} , S_B_{end} .

Step 12. Calculating for on-blast period: TDB_HBS1_{avg} , TBB_HBS1_{avg} .

Resetting for on-gas period: TDG_HBS1_{avg} , TFG_HBS1_{avg} , FG_HBS1_{avg} , VG_HBS1 .

Step 13: Assigning initial values: $M = 1$ – on-gas period; $k = 1$ – beginning of new on-gas period; $n = 0$ – end of on-blast period.

Step 14. Calculating: TDG_HBS1_{avg} , TFG_HBS1_{avg} , FG_HBS1_{avg} , VG_HBS1 .

Step 15. Incrementing number of parameters values $k = k + 1$.

Step 16. Mode checking: $VN1 \setminus D_i = 1$ AND $VN1 \setminus N_i = 0$? If YES – on-blast period, transition to step 17, if NOT – transition to step 24.

Step 17. Mode change checking: $M \neq 2$? If YES – mode changing from on-gas period to on-blast period, transition to step 18; if NOT – transition to step 22.

Step 18. Assigning initial values for on-blast period beginning: TDB_HBS1_{beg} , TBB_HBS1_{beg} , H_B_{beg} , M_B_{beg} , S_B_{beg} .

Step 19. Assigning finite values for on-gas period ending: TDG_HBS1_{end} , TFG_HBS1_{end} , FG_HBS1_{end} , H_G_{end} , M_G_{end} ; S_G_{end} .

Step 20. Calculating for on-gas period: TDG_HBS1_{avg} , TFG_HBS1_{avg} , FG_HBS1_{avg} .

Resetting for on-gas period: TDB_HBS1_{avg} , TBB_HBS1_{avg} .

Step 21. Assigning initial values: $M = 2$ – on-blast period; $k = 0$ – ending of new on-gas period; $n = 1$ – beginning of on-blast period.

Step 22. Calculating: TDB_HBS1_{avg} , TBB_HBS1_{avg} .

Step 23. Incrementing number of parameters values $n = n + 1$.

Step 24. Incrementing number of records $i = i + 1$.

Step 25. Exiting from the loop No. 2.

Step 26. Exiting from the loop No. 1.

Step 27. End.

The results of processing ADB are written to an Excel file (Fig. 4). Each row of eight tables represents one on-gas period or one on-blast period.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Period Start Date	Period End Date	Period Duration	Period Duration (Mins)	Beg Dome Temp	End Dome Temp	Avg Dome Temp	Dome Temp Change Rate	End BFG Flow Rate	Avg BFG Flow Rate	Vol BFG Flow Rate	Beg FlueG Temp	End FlueG Temp	FlueG Temp Change Rate
137	1.3.20 14:30:53	1.3.20 20:06:39	6:35:46	335,8	1206	1333	1297	23	44150	35240	163543	94	401	55
138	1.3.20 22:40:40	1.4.20 6:49:02	8:08:22	488,4	1242	1339	1292	12	43301	34291	212925	123	401	34
139	1.4.20 9:37:35	1.4.20 14:43:57	5:06:22	306,4	1222	1274	1248	10	41793	31169	93918	128	401	54

Fig. 4. ADB conversion results

The next second processing step is the analysis of the obtained periods and the comparison of on-gas periods with the corresponding subsequent on-blast periods.

The following ADB processing results were obtained, which included 7 months of operation of HBSB in on-gas periods / on-blast periods: for HBS1 696 / 761; for HBS2 773 / 763; for HBS3 306 / 718; HBS4 207 / 237. For the next 3 months: for HBS1 155 / 152; for HBS2 166 / 167; for HBS3 59 / 81; for HBS4 173 / 173. For example, inconsistencies in the number of periods can be in the case when the recording of the technological parameters values in the ADB is incorrect (absence or failure of the sensor) and there is a mismatch between the values of the tags responsible for the periods and these periods.

In addition the operating values of time of on-gas and on-blast periods are known from the operating map of HBS block, so periods that are known to be small or long in duration are initially removed.

Then a search is made for on-blast periods, which correspond to the current on-gas period. After processing the following number of full cycles was obtained: for HBS1 – 641, for HBS2 – 736, for HBS3 – 135, for HBS4 – 201. Thus, for example, for HBS1 there were 851 entries in ADB, and only 641 full cycles fell into the table, i.e. approximately 25% of records are filtered.

At the third processing step, localization and removal of anomalies of the obtained values for each of the columns is performed using descriptive statistics: calculation of the maximum and minimum values, average values, standard deviation (σ): elements that do not belong to the range $[-3\sigma; 3\sigma]$ are removed from the selection. It can be seen from Fig. 5 that the initial distribution of the parameter values is processed in such a way that the resulting distribution is close to normal.

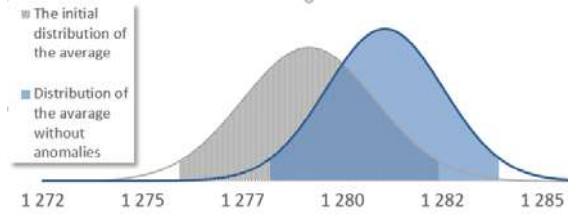


Fig. 5. Result of processing in steps 3

As a result, the final version of the database was obtained with a breakdown into full HBS operation cycles database (hereinafter referred to as FCDB).

In the fourth processing step we get the data from FCDB. At this stage, the opportunity is realized using FCDB and the algorithm developed by the authors (Fig.6) to determine the functional dependence of the base parameter on two other parameters with fixing the third parameter in a certain range.

Below is an algorithm that implements this possibility, using the example of obtaining the dependence of on-gas duration $\tau_{dur}(Z)$ on the average dome temperature $TDG\_HBS_{avg}(X)$ and the volume of consumed BFG $VG\_HBS(Y)$ within the range of BFG flow $FG\_HBS(F)$ from FG_HBS1_{min} to FG_HBS1_{max} (F_{min} to F_{max}):

$$Z = f(X, Y) \text{ within } F \in [F_{min}, F_{max}]. \quad (1)$$

Let us consider in details the proposed algorithm (Fig. 6):

Step 1. Beginning of the subprogram.

Step 2. Opening FCDB file.

Step 3. Reading k - the number of records in FCDB file.

Step 4. Reading data from FCDB file to X, Y, Z, F .

Step 5. Calculating the values of average, minimum, maximum, standard deviation for X and Y arrays: $X_{min}, X_{max}, X_{avg}, X_{ds}, Y_{min}, Y_{max}, Y_{avg}, Y_{ds}$.

Step 6. Input F_{min}, F_{max} .

Step 7. Input range values in which the X and Y data will be processed: $\sigma_{beg}, \sigma_{end}, \sigma_{step}$. By default: $\sigma_{beg} = -2, \sigma_{end} = 2, \sigma_{step} = 1$.

Step 8. Assigning initial values $\sigma = \sigma_{beg}, m = 0$.

Step 9. Loop No. 1 starting, it operates until $\sigma \leq \sigma_{end}$.

Step 10. Calculating values of $convX_m$ and $convY_m$ arrays.

Step 11. Incrementing σ and m .

Step 12. Exiting from the loop No. 2

Step 13. Resetting X_{rez} array.

Step 14. Loop No. 2 starting, it operates until $i \leq m-1$.

Step 15. Resetting Y_{rez} array elements to zero.

Step 16. Loop No. 3 starting, it operates until $j \leq m-1$.

Step 17. Resetting Z_{rez} array elements to zero; $n=0$.

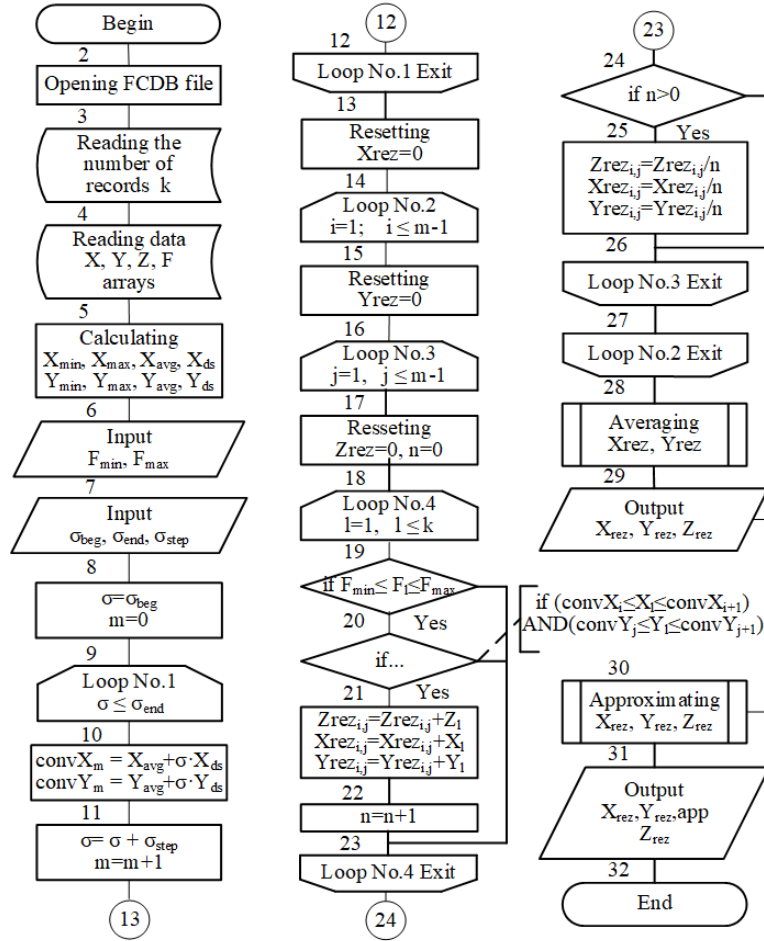


Fig. 6. Flow-chart of algorithm for determining the functional dependence

Step 18. Loop No. 4 starting, it operates until $l \leq k$.

Step 19. Checking condition: $F_{\min} \leq F_l \leq F_{\max}$? If YES – transition to step 20, if NOT – transition to step 23.

Step 20. Checking condition:

if $((\text{conv}X_i \leq X_l \leq \text{conv}X_{i+1}) \text{AND} (\text{conv}Y_j \leq Y_l \leq \text{conv}Y_{j+1}))$? If YES – transition to step 21, if NOT – transition to step 23.

Step 21. Calculating: $Z_{rez_{i,j}}$, $X_{rez_{i,j}}$, $Y_{rez_{i,j}}$.

Step 22. Incrementing n .

Step 23. Exiting from the loop No. 4

Step 24. Checking condition: if $n > 0$? If YES – transition to step 25, if NOT – transition to step 26.

Step 25. Calculating of average values: $Z_{rez_{i,j}}$, $X_{rez_{i,j}}$, $Y_{rez_{i,j}}$.

Step 26. Exiting from the loop No. 3
 Step 27. Exiting from the loop No. 2
 Step 28. Averaging X_{rez} and Y_{rez} values – getting a one-dimensional array from a two-dimensional array.
 Step 29. Output X_{rez} , Y_{rez} , Z_{rez} .
 Step 30. Regression analysis of the obtained results X_{rez} , Y_{rez} , Z_{rez} with multi-variate approximation.
 Step 31. Output X_{rez} , Y_{rez} , $appZ_{rez}$.
 Step 32. End.

5 Assessment of the research results

Based on the above algorithms, the following programs have been developed: for the implementation of the first and second stages (Fig. 1); for the implementation of the third and fourth stages (Fig. 1).

Below are the results of the programs.

To analyze the operation of HBS #4, the dependence (1) was obtained (Fig. 7) with the following samples: $[-3 \cdot \sigma; 3 \cdot \sigma]$, step σ (Fig. 7, a, b); $[-2,5 \cdot \sigma; 2,5 \cdot \sigma]$, step $0,5\sigma$ (Fig. 7, c, d, e); $[-2,5 \cdot \sigma; 2,5 \cdot \sigma]$, step $0,25\sigma$ (Fig. 7, f, g, h).

It can be seen from the graphs that when using a step of $0,5\sigma$ and especially $0,25\sigma$ (Fig. 7, c, f), the problem of the lack of input data in FCDB that simultaneously satisfy the set of specified parameters values (dome temperature and gas volume) arises. This is determined by the presence of zero values of heating time. In this case, the program replaces these points using an approximation based on nonzero values, which is visually seen by the presence of flat sections on the surface of the graphs (Fig. 7, d, g).

Thus using an interval equal to σ eliminates the problems that arise in the absence of values in the database (Fig. 7, a).

It should be noted that the most optimal sampling range is the range from -2σ to 2σ with an interval equal to σ .

The obtained results (Fig 7, a, d, g) in terms of the totality of values are unsuitable for subsequent use in control algorithms of HBS. Using the multi-parameter approximation, we obtain these dependencies in the form of analytical expressions ($R^2 > 0.8$), the graphs of which are shown in Fig. 7 (b, e, h). Examples of such expressions are given in Table 1.

Earlier [18], the authors developed and implemented a mathematical model of the thermal operation of HBS of BF, which was identified on the basis of ADB used in this research. For this purpose, the averaged basic operation parameters of each HBS of the block ([18], Fig. 4) were used. Thus, for example, for HBS4, the average dome temperature corresponds to 1275 °C, and the average on-gas period duration is 230 minutes, which fully corresponds to the calculated value in accordance with the analytical expression (Tab. 1).

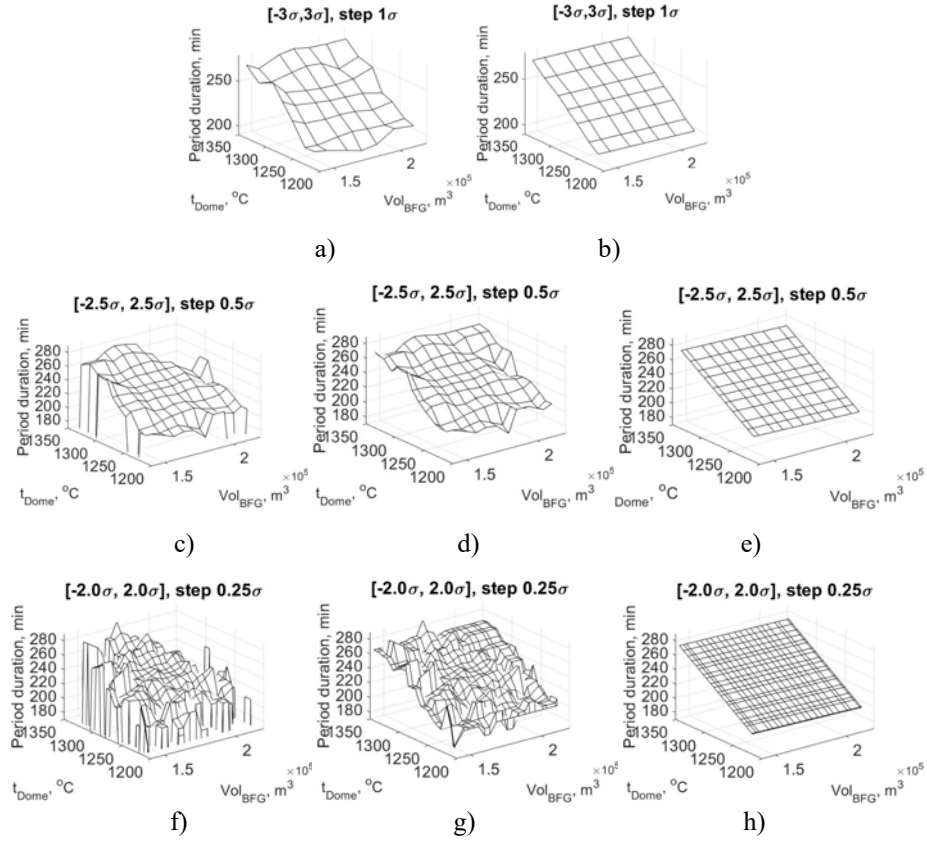


Fig. 7. The dependence (1) for HBS #4 for different range of σ

Table 1. Analytical expressions

Range	Analytical expressions	R^2
$[-3 \cdot \sigma; 3 \cdot \sigma]$, step σ	$\tau = -0.0001 \cdot V_{BFG} + 0.5 \cdot t_{Dome} - 471.3$	0.8939
$[-2.5 \cdot \sigma; 2.5 \cdot \sigma]$, step 0.5σ	$\tau = -0.0001 \cdot V_{BFG} + 0.6 \cdot t_{Dome} - 511.6$	0.8578
$[-2.5 \cdot \sigma; 2.5 \cdot \sigma]$, step 0.25σ	$\tau = -0.00005 \cdot V_{BFG} + 0.6 \cdot t_{Dome} - 534.3$	0.8281

Fig. 8 graphically presents the dependences of on-gas duration on the average dome temperature and the BFG volume for various average BFG flow for each HBS of the block obtained as a result of the research.

Such data presentation (Fig.8) makes it possible to change in real time the parameters of the mode map of the blast heating of HBSB in case of production need or significant changes in the production situation.

6 Conclusion

1. The methods of data mining and their use, including in control systems for blast furnace production, which allows one to extract new opportunities from a minimum set of values of technological parameters, have been analyzed. The authors concluded that there is no information on the use of data mining in automation systems of HBSB.

2. For the first time, a method and algorithm for processing an archive database which includes 4 stages is proposed.

3. The specified algorithm is implemented in the form of two software modules developed using Visual Studio 2019.

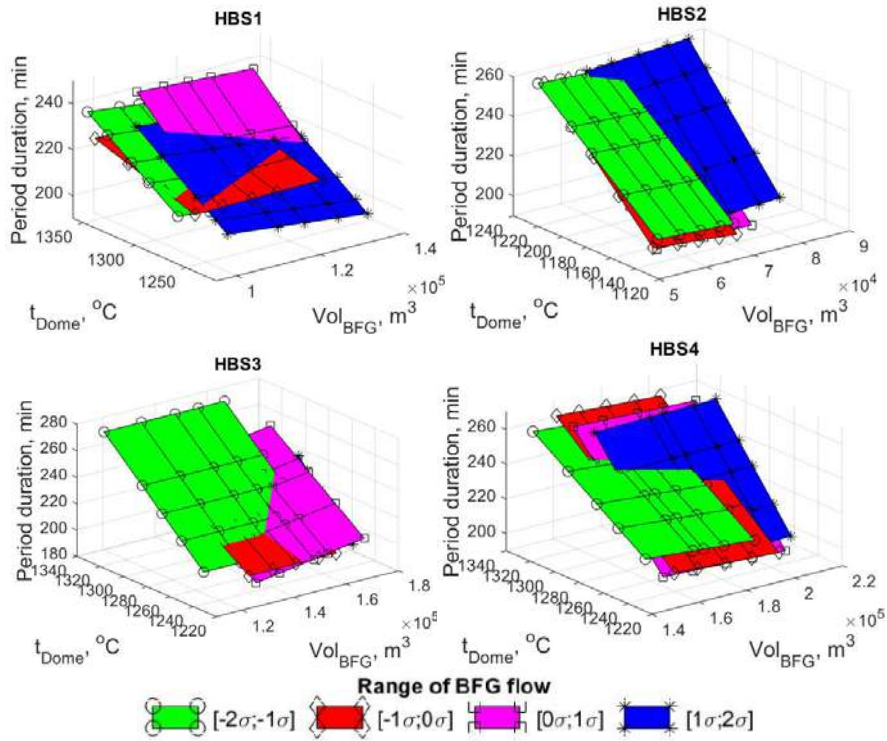


Fig. 8. Dependence of on-gas duration on the average dome temperature and the volume of consumed BFG within the range of BFG flow for each HBS of block

4. As a result of the research, analytical dependencies have been obtained, they allow to make decisions about changes in the heating flow chart and implement control algorithms for HBSB and each HBS of the block.

References

1. Qin, S. J.: Process data analytics in the era of big data. *AIChE Journal*, Vol. 60(9): pp. 3092-3100 (2014). doi: 10.1002/aic.14523
2. Liu, Y., Lv, Z., Wang, W.: An Improved Generalized-Trend-Diffusion-Based Data Imputation for Steel Industry. *Mathematical Problems in Engineering*, Vol. 2013: pp. 1-10 (2013). doi: 10.1155/2013/136241
3. Zhao, J., Liu, K., Wang, W., Liu, Y.: Adaptive fuzzy clustering based anomaly data detection in energy system of steel industry. *Information Sciences*, Vol. 259: pp. 335-345 (2014). doi: 10.1016/j.ins.2013.05.018
4. Sinha, P. Multivariate polynomial regression in data mining: methodology, problems and solutions. *International Journal of Scientific and Engineering Research*, Vol. 4(12): pp. 962-965 (2013).
5. Liu, C., Chen, S., Hai, Y., et al.: Experimental study and analysis on heat transfer hot air temperature of far infrared convection combined drying. In 2018 International Conference on Robots & Intelligent System (ICRIS): pp. 552-555. IEEE (2018). doi: 10.1109/ICRIS.2018.00142
6. Gasparini, V. M., de Castro, L. F. A., et. al.: Thermo-chemical model for blast furnace process control with the prediction of carbon consumption. *Journal of Materials Research and Technology*. Vol. 6(3): pp. 220-225 (2017). doi:10.1016/j.jmrt.2016.12.001
7. Kazarinov, L.S., Barbasova, T.A., Mohsen, H.M.: An efficient control method for furnace process based on cluster analysis. *Bulletin of the South Ural State University. Ser. Computer Technologies, Automatic Control, Radio Electronics*. Vol. 16 (2): pp. 164–169 (2016). (In Russian) doi: 10.14529/ctcr160220
8. Shang, J., Chen, M., Zhang, H., et al.: Increment-based recursive transformed component statistical analysis for monitoring blast furnace iron-making processes: An index-switching scheme. *Control Engineering Practice*. Vol. 77, pp. 190–200 (2018). doi: 10.1016/j.conengprac.2018.05.012
9. Zhou, B., Ye, H., Zhang, H., Li, M.: Process monitoring of iron-making process in a blast furnace with PCA-based methods. *Control Engineering Practice*. Vol. 47, pp. 1–14. (2016). doi: 10.1016/j.conengprac.2015.11.006
10. Barbasova, T., Zagorskina, E.: Blast-Furnace Melting Blast Control. 2019 IEEE Russian Workshop on Power Engineering and Automation of Metallurgy Industry: Research & Practice (PEAMI), Magnitogorsk, Russia, pp. 13-17. (2019). doi: 10.1109/peami.2019.8915284
11. Zhou, B., Ye, H., Zhang, H., Li, M.: A new qualitative trend analysis algorithm based on global polynomial fit. *AIChE Journal*, Vol. 63(8): pp. 3374-3383. (2017). doi: 10.1002/aic.15706
12. Vorotnikova, Z.J., Novoselzev, M.A.: Cast iron fusion quality control functional monitoring. Reporter of Priazovskiy State Technical University: Collection of scientific papers. Vol. 26: pp. 221-227 (2013). (In Russian)
13. Vorotnikova, Z.Ye.: Forming and use of an archived database in the system “Adviser of a blast furnace operator”. *Scientific works of Donetsk National Technical University*. Vol. 1 (26): pp.14-23 (2014) (In Russian).
14. Shnayder, D. A., Kazarinov, L. S., Barbasova, T. A., Lipatnikov, A. V.: Data mining and model-predictive approach for blast furnace thermal control. In 2017 Intelligent Systems Conference (IntelliSys), pp. 653-660 (2017). doi: 10.1109/IntelliSys.2017.8324364

15. Nose, K., Takemura, M., Morita, T.: Systems engineering approach to the optimization of hot blast stoves. IFAC Proceedings Volumes. Vol.: 16(15), pp. 47-54. (1983). doi: 10.1016/S1474-6670(17)64255-1/
16. Lee, Y.-H., Min, K. G., Han, C., Chang, K. S., Choi, T. H.: Process improvement methodology based on multivariate statistical analysis methods. Control Engineering Practice. Vol. 12(8), pp. 945–961 (2004). doi:10.1016/j.conengprac.2003.11.004
17. Jang, G.-S., Jeon, J.-H.: A Six Sigma Methodology Using Data Mining: A Case Study on Six Sigma Project for Heat Efficiency Improvement of a Hot Stove System in a Korean Steel Manufacturing Company. Cutting-Edge Research Topics on Multiple Criteria Decision Making. MCDM 2009. Communications in Computer and Information Science. Vol. 35: pp. 72-80 (2009). doi: 10.1007/978-3-642-02298-2_12
18. Koifman, A., Simkin, A.: Development and Software Implementation of the Hot Blast Stove Computer Model. CEUR Workshop Proceedings, Vol. 2353: pp. 440-454 (2019). Online <http://ceur-ws.org/Vol-2353/paper35.pdf>.

Situational management of urban engineering networks with intelligent support for dispatching decisions

Aleksandr Stenin^{1[0000-0001-5836-9300]}, Irina Drozdovych^{2[0000-0002-4216-2417]},
Mariya Soldatova^{3[0000-0003-1233-1272]}

^{1,3} Igor Sikorsky Kyiv Polytechnic Institute, Kyiv, Ukraine.
alexander.stenin@yandex.ua
benten1093@gmail.com

² Institute of telecommunications and global information space NAS of Ukraine, Kyiv, Ukraine.
irina.drozdowicz@gmail.com

Abstract. To improve the efficiency of urban engineering networks management, an intelligent system for supporting dispatching decisions (IDSS UEN) proposed, based on a situational approach using fuzzy logic. The developed algorithm for situational fuzzy control is the basis of the IDSS UEN and is based on the original method of mixed estimates of alternatives and fuzzy situational network UEN

Keywords: urban engineering networks, weighted graph of situations, intelligent decision support system, fuzzy situation management algorithm.

1 Introduction.

The work of the dispatcher of urban engineering networks (UEN) (power grids, gas pipelines, water pipes, sewers, etc.) is an extremely difficult task [1, 4, 5].

Decision-making by the UEN dispatcher takes place in a complex environment characterized by the following difficulties [2]:

- impossibility to get reliable, complete and accurate information about an emergency;
- the transience of changes in emerging situations in UEN;
- obsolescence of information used for decision making;
- the presence of a large number of factors;
- a multivariate approach to decision-making;
- compromise between economic benefits for the city and the quality of services for residents based on decisions made by the dispatcher;
- responsibility for the sole decision;

It is known [1-4], when a large amount of initial information used for decision-making, the quality of decisions significantly reduced. Research shows that decision-makers without additional analytical support tend to use simplified and contradictory decision-making rules. In this sense, effective decisions of the dispatcher in the work of UEN largely depend on the capabilities of technology and software tools that implement methods and methods of intellectual support for decisions. The use of the

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

intellectualization tools, which based on the experience of experts in this subject area, can significantly improve the efficiency of UEN operation.

The most appropriate method for implementing effective solutions is an intelligent decision support system in urban engineering networks (IDSS UEN), built on a situational approach using fuzzy logic [4, 6].

2 Problem statement

It knew [7] that the decision-making task (DMT) can represent as:

$$DMT = \langle T, A, Q, Y, F, G, D \rangle, \quad (1)$$

where: T – problem statement; A – the set of alternatives; Q – a set of selection criteria (evaluating the effectiveness of solution options); Y – multiple methods for measuring relationships between options; F – mapping a set of valid options to a set of ratings; G – the system preferences of experts; D – a decision rule that reflects the preference system.

Assume that the current situation that has developed in the process of UEN operation described as a fuzzy situation of the following type [9]:

$$S_{TEK} = \{M_{S_i}(x_i) / x_i\}, x_i \in X, \quad (2)$$

$M_{S_i}(x_i)$ – the function of belonging to a linguistic variable x_i , that characterizes the current situation S_{TEK} .

Also, assume that for each linguistic variable x_i corresponds to the j -th term of the set of terms in the knowledge base. Then the formula (2) can be written as:

$$S_i = \{M_{S_i}(x_i)(T_j^i) / T_j^i\}, j = \overline{1, M}; i = \overline{1, N}; x_i \in X, \quad (3)$$

T_j^i – j -th term i -th linguistic variable.

To determine the current state of urban engineering networks it is necessary to compare this fuzzy situation with each fuzzy situation from a certain set of existing situations in the knowledge base of this subject area $S = (S_1, \dots, S_K)$. The result of the comparison should be a dispatcher's decision on the strategy for further UEN management with the involvement of IDSS. If the current situation is "regular", the control implemented automatically according to the existing algorithm for this situation. If occurs in an "emergency" situation, the IDSS will help you choose a management strategy that is close to this situation, or, if one is not found, form a new management strategy.

3 Literature review

Decision-making in most cases consists of generating possible alternative solutions, evaluating them, and selecting the best option. In most cases, the choice of the optimal solution made under conditions of information uncertainty and conflicting factors. Uncertainties in a particular subject area can be divided into uncertainties associated with incomplete knowledge of the problem solved and uncertainties associated with the inability to fully account for the state of the environment [4, 5]. Inconsistency occurs due to ambiguity in the assessment of situations, errors in the choice of priorities, which greatly complicates decision-making. Research shows that decision-makers without additional analytical support tend to use simplified and sometimes contradictory decision-making rules. In this case, the most effective tool for making a potentially better decision is decision support systems (DSS). DSS provides assistance decisions make in the analysis of initial information (assessment of the current situation and restrictions imposed by the external environment). DSS provides identification and ranking of priorities, take into account the uncertainty in the estimates of the decision-maker and forming preferences, generation of possible solutions (forming a list of alternatives); evaluation of possible alternatives; analysis of the possible consequences of the decisions taken; selection of the potentially best solution [6,21].

Formalization of methods for analysis and generation of solutions, their evaluation and coordination is quite a difficult task. Its solution made possible due to the widespread use of computer technology and largely depends on the capabilities of technical software tools that implement methods and methods of intellectual support [8].

The decision-making process can proceed according to two main schemes: intuitive-empirical (based on comparing the problem situation with similar situations that have previously occurred) and formal-heuristic (based on building and researching a model of the problem situation for this particular subject area).

Under building a model of a problem situation, we examine the structure of the DSS, which determined by such elements as the state of the initial data of the problem, the model of the decision-making situation, restrictions, decision options, and their consequences, and external factors of an objective and subjective nature. The combination of these elements forms a specific environment (system). In other words, a DSS is a system that provides the decision-maker with the data, knowledge, conclusions, and/or recommendations necessary for making a decision [10].

Considering the existing conceptual models of the DSS, the authors identify approaches based on the use of the ideology of information systems, artificial intelligence, and the instrumental approach.

Within the framework of the ideology of the information approach, DSS classified as a class of automated information systems whose main purpose is to improve the activity of knowledge workers in organizations by using information technology.

Within the framework of the ideology of intelligent decision support systems (IDSS), knowledge-based systems differ significantly from similar systems, primarily expert systems (ES), in their target orientation [3, 19]. IDSS designed to help decision-makers solve problems, and ES designed to replace people in solving specific problems [3, 18].

The tool approach, depending on the specifics of the tasks to be solved and the technological tools used, we take into account applied DSS, which serve to support decision-making for individual applied tasks in specific situations. Application DSS are packages of software tools for searching and issuing data, modeling, etc., which used by developers to create specialized systems and objects.

Modern IDSS based on the use of specialized information storage (Data Warehouse) and OLAP (On-Line Analytical Processing) technologies - operational data analysis. The main purpose of OLAP technologies is dynamic multidimensional data analysis using an effective data-mining tool, modeling, and forecasting [2.20].

4 Building an IDSS based on a fuzzy situational network

The degree of proximity of a fuzzy situation S_{TEK} and situation $S_k (k = \overline{1, K})$ from the knowledge model determine as:

- degree of fuzzy enabling a fuzzy situation S_{TEK} in a fuzzy situation S_k ;
- the degree of fuzzy equality S_{TEK} and S_k ;
- degree of the fuzzy community S_{TEK} and S_k ;

Setting one of the chosen proximity measures, we can set some fuzzy relationships between situations, not only about the current one STEK but also between existing ones in the knowledge base of this subject area.

From this subject area, the most convenient measure of proximity can be considered the degree of fuzzy inclusion of the situation, which characterized by a certain threshold of inclusion determined by the expert, based on the conditions in which the work of urban engineering networks takes place. The inclusion threshold is defined, as well as the membership functions, in the normalized range $[0, 1]$ as follows:

$$t_{on} \in [\alpha_{min}, 1] \quad (4)$$

α_{min} - the lower limit of the range of the degree of inclusion, usually $\alpha_{min} = 0.6/0.7$. In this case, we can talk about how fuzzy signs of the current situation S_{TEK} are included in the fuzzy values of the corresponding signs of the situation $S_k (k = \overline{1, K})$ [4, 15].

Further, for the existing knowledge base of this subject area and the conditions of GIS operation in different operating modes, we form typical "regular" situations for which control actions on the UEN are developed in detail on the expert methods. In this case, the possible transition from one "regular" situation S_k to another S_i carried out using some R_{ki} solution. In this case, each possible solution determined by the degree of preference for this solution $\gamma_{kl} (S_k, R_{kl})$. In this way, you can build a fuzzy situational network (Fig. 1), which clearly shows possible transitions from one "regular" situation to another and the degree of preference for these transitions [10, 15].

It notes that both the knowledge base of this subject area and the situation network are evolutionary and are periodically updated both when a new "regular" situation appears (in fact, this is a variant of the "non-standard" situation developed by experts) and new requirements for the UEN operation.

As a result, based on the knowledge base of this subject area and conditions robots UEN fuzzy situational network is a fuzzy directed graph, whose vertices correspond to existing "regular" fuzzy situations, arc-weighted possible solutions required to transition situations, and degrees of preference of these decisions (fig.1).

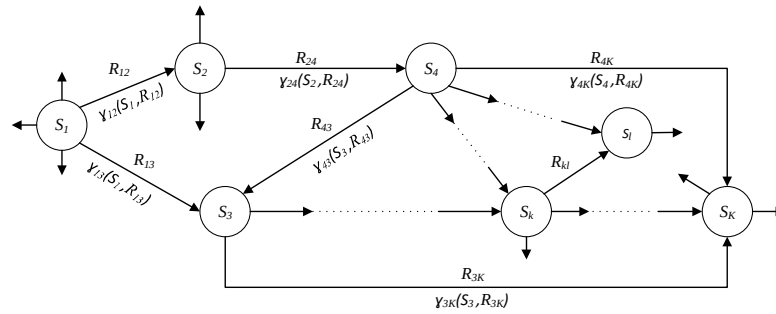


Fig. 1. Fuzzy situational network

The degrees of preference for solutions in each situation either unchanged and determined by an expert survey, or depended in some way on the situation and determined by the preferred solution in this situation. The control solution corresponding to the current state is a sequence of decisions necessary for the transition from the current situation to a given target situation along the "optimal route" in a fuzzy situation network. The criteria for minimizing time and/or energy costs can be used as a preference for solutions in the case of UEN. Then the "optimal route" corresponds to the minimization of the total time and/or energy costs when moving along this route. In this case, you can use the dynamic programming method.

Thus, the output of the solution is divided into setting a goal (target situation) and building a management strategy that corresponds to the optimal translation of the UEN to the target state. This conclusion of the decision is valid if the current situation can be included in one of the "regular" situations.

Otherwise, if the current situation has a degree of fuzzy inclusion less than α_{min} in the formula (4), then this "non-regular" situation presented to the experts to bring it to the rank of "regular" and form variants of transition situations. Thus, the existing fuzzy situational network corrected, which ensures its evolution. In fig.2 shown a block- scheme of the IDSS UEN containing the main blocks (modules) of the software: a knowledge base (KB), a block for generating control solutions, a block of hints, a knowledge acquisition block and an intelligent interface presented.

The intelligent interface combines linguistic, information, and software tools for interaction between the dispatcher, knowledge engineer (analyst), and UEN experts with the corresponding components of the Toolkit. The hint block is intended to

show, if necessary, in a form that is understandable to the dispatcher, the progress of the "reasoning" (working scenario) of the output mechanism to justify the management decision made by it.

Before the main work, IDSS UEN creates a knowledge base (KB). KB UEN includes two components: long-term knowledge about UEN, which can be represented as a set of rules of logical inference, hierarchical frame structures, semantic networks, or other information structures that combine the above; operational knowledge (operational data) that describes the current situation.

Note once again that the UEN database is evolutionary and is periodically updated either with new information in this subject area or with the emergence of new situations in the management of UEN.

The formation KB of the UEN is as follows. In the knowledge acquisition block, goals (sub-goals), criteria, requirements, and working conditions of the GIS formed. Next, the term sets of the domain formed based on the descriptors of this domain, i.e. UEN, which defined according to the work [11].

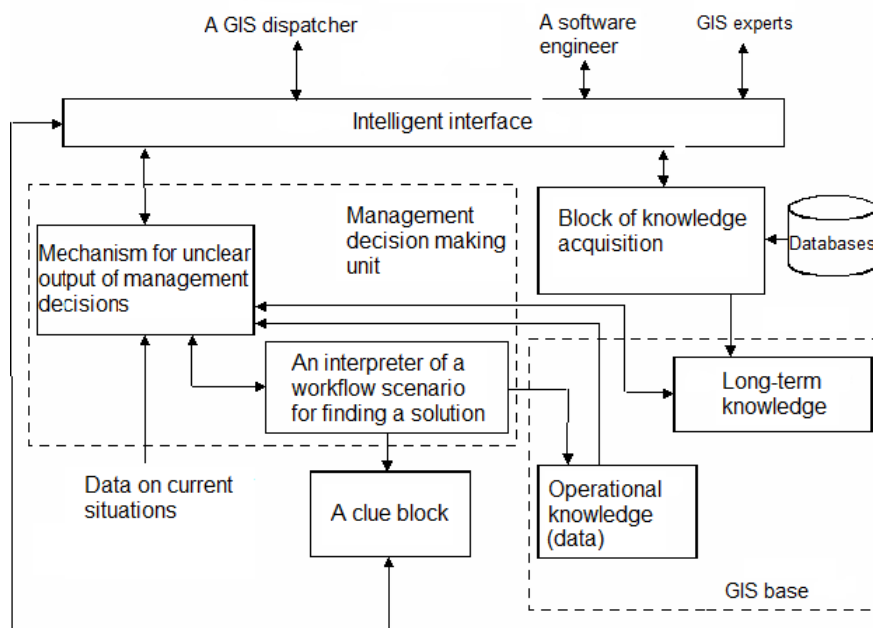


Fig. 2. Block-scheme IDSS UEN

From the selected set of descriptors, the most significant ones are selected, which are used in the software module according to the work [12] for the intelligent multi-agent subsystem (IMS) for searching information on the Internet, which one of the main components of the database (DB), for accumulating the UEN knowledge base. In this IMS synthesizes a generalized model of data collection and analysis based on a genetic approach and a multi-agent method for synthesizing decision trees and a neural

network [13] that uses the decision-making specificity index. The decrease in search time in IMS is due to an algorithm for selecting control solutions that use estimates of sets of types of situations and specifics of logics, rather than the sets themselves. Improving the quality of information achieved at each iteration by selecting behaviors with a high frequency of use and cutting off the area of superposition estimates by the logic specificity index and the situation index, which increases the level of ontological representation of information.

Information obtained from the Internet and other sources based on latent semantic analysis form the most significant content of the UEN KB [11] following the purpose and conditions of the UEN operation.

Next, a fuzzy situational network of this subject area, i.e. UEN, built in the interpreter of working scenarios for finding solutions in the block for generating control solutions. To do this, we extract a set of "regular" fuzzy situations from the UEN database, for which possible transitions between situations are determined based on information about the goals/sub-goals of GIS work, criteria for work, and requirements for UEN work. The result is a fuzzy situational network that is a fuzzy oriented weighted graph. The fuzzy situational network, as the KB UEN periodically updated with new information, clarifying objectives, criteria and the emergence of "nonstandard" situations missing BRS in GIS, i.e. in effect, the training of IDSS UEN.

The mechanism for fuzzy inference of management decisions based on operational data about the current situation, based on long-term knowledge, predicts the actions required by the current situation, planning a step-by-step working scenario for finding a solution. The block for forming management decisions implements the fuzzy situational management algorithm shown below (Fig.3) [15].

5 The algorithm of fuzzy situational control

In fig.3 shows a block-scheme of the UEN fuzzy situational management algorithm, which, if an "emergency" situation occurs, will help you choose a management strategy that is close to this situation, or, if one is not found, form a new management strategy.

In block 1, enter information about the current situation in the GIS, which characterized by many factors. In this case, some factors entered into the system automatically based on the readings of the corresponding measuring instruments, and some are targeted monitoring information. In General, information about the current situation will be characterized by both quantitative and qualitative values of factors. In block 2, the situation compared with a set of "regular" situations that are in the knowledge base (KB) and characterized by the same set of factors as the current one. Thus, this block allows you to limit the number of possible situations for which you need to calculate the degree of proximity to the current situation. This will significantly reduce the running time of the algorithm. In block 3, the degree of proximity $\nu = t_{on}$ of the current situation is calculated and the "regular" situations from block 2 that are close to it are determined. Determining the degree of proximity of fuzzy situations necessary to provide the required data for the algorithm for selecting control solutions

that operate on the control decision matrices stored in the database of "regular" fuzzy situations. In block 4 compares the calculated degree of proximity $\nu = t_{on}$ (the degree of belonging) to the specified normalized inclusion interval according to the formula (4), where $t_{on} = \xi$. This makes it possible to select the "regular" situations that are closest to the current one.

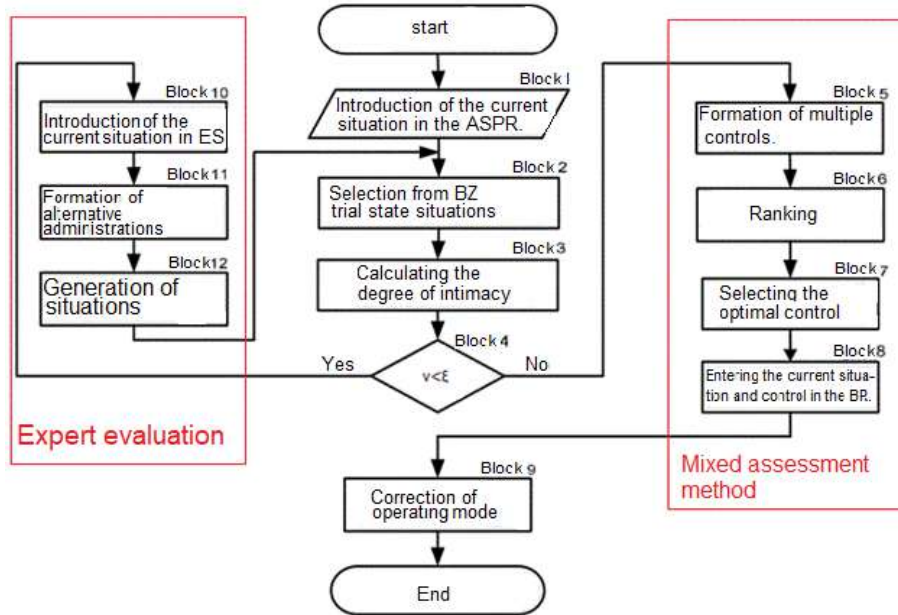


Fig. 3. Block-scheme of the fuzzy situational control algorithm

In block 5, management decisions generated that correspond to the "regular" situations selected in block 4. In blocks 6,7, the effectiveness of control actions ranked in descending order within a subset of selected "regular" situations and the optimal one is determined among them. In block 8 the optimal control and the corresponding situation are entered as "regular" in the knowledge base about the functioning of the UEN. In block 9 the operation mode is adjusted to take into account the current situation and the criteria for the effectiveness of the UEN.

In block, 10 the current situation is placed in the expert environment (ES) and the knowledge base of the subject area. In block 11, possible control actions are requested from experts in this subject area placed in the knowledge base. In block 12, possible situations and relevant management decisions generated and entered into the knowledge base. On the method of mixed assessments developed by the authors, new "regular" situations are formed and entered into the knowledge base about the functioning of UEN.

6 Mixed assessment method

The proposed method differs from the known methods PARK, ZAPROS, ORCLASS, and SHNUR [14] in that the entire set of alternatives evaluated for each criterion at once. The main stages of the method include:

Stage1. Construction of a set of hypotheses-bases $\{Z_i\}$ corresponding to a set of criteria $\{K_i\}$ and a set of hypotheses-consequences $\{z_{ij}\}$ corresponding to a set of criteria values.

Stage2. Formation of quantitative and linguistic scales of criteria for evaluating alternatives.

Stage3. Drawing up a block structure of the questionnaire and filling it out on the hypotheses-bases and hypotheses-consequences.

Stage4. Normalization of scales of criteria for evaluating alternatives.

Attributes that form alternatives to A_i contain both numeric (quantitative) and linguistic (qualitative) variables. The membership function $\mu: x \rightarrow [0,1]$ quantitatively grades the membership of elements of the set of alternatives A to the fuzzy set $\bar{A}$, with normalized variables, i.e. $\tilde{A} = \{(x, \mu_A(x)) | x \in X\}$.

A value of 0 means that the element not included in the fuzzy set, and 1 means that the element is fully described by the given set. Among the most well-known and used auxiliary functions, the most convenient and universal for the variables under consideration is a triangular function of the form [7,18]:

a) for the max-min preference scale of the A_i alternative

$$\mu(x_{ij}) = \begin{cases} 0, & \text{if } x \leq a_{ij}; \\ \frac{x_{ij} - a_{ij}}{b_{ij} - a_{ij}}, & \text{if } a_{ij} < x < b_{ij}; \\ 1, & \text{if } x \geq b_{ij}, \end{cases} \quad (5)$$

where $a_{ij} = x_{\min ij}$; $b_{ij} = x_{\max ij}$

б) for the min-max preference scale of the A_i alternative

$$\mu(x_{ij}) = \begin{cases} 1, & \text{if } x_{ij} \leq a_{ij}; \\ \frac{b_{ij} - x_{ij}}{a_{ij} - b_{ij}}, & \text{if } a_{ij} < x_{ij} < b_{ij}; \\ 0, & \text{if } x_{ij} \geq b_{ij}, \end{cases} \quad (6)$$

where $a_{ij} = x_{\min ij}; b_{ij} = x_{\max ij}$

Rationing of estimates of the compared alternatives carried out on the formulas (5) and (6). For this:

- for all quantitative estimates the maximum and minimum values of the variable under consideration determined;
- for all linguistic (qualitative) assessments, the maximum and minimum verbal value of the variable determined;
- the values of the criteria are determined following formulas (5) and (6).

Stage 5. Identifying a potentially better alternative

The values of ratings on the normalized scale according to the formulas (5) and (6) are formed as follows:

$$v_{ij}' = \mu(ij), \quad (7)$$

where i – alternative number, j – index of the value of a quantitative or linguistic scale.

Then the sum of s -th alternative can be calculated by adding the sum of scores for each j -th question, which is summed across all issues of the i -th block and all blocks of the questionnaire:

$$r_s = \sum_{S=1}^n \sum_{i=1}^m \sum_{j=1}^k v_{ij}' \quad (8)$$

As a result, can formulate for the initial set of alternatives As an ordered set of their ranks $R = \{r_1, r_2, \dots, r_n\}$. If it is necessary to take into account the importance of a particular criterion (question), it is necessary to enter weight coefficients in formula (8), which can be determined on the known methods of expert assessments [18].

7 Modeling optimal transitions in a fuzzy situational network

Let be a weighted oriented graph (Fig.4) is a fragment of a fuzzy situational network (Fig.1). Necessary to transfer the UEN from the "regular" situation S_1 to the "regular" situation S_{10} with the minimum total value of the preference functions $\gamma_{ij}(S_i, R_{ij})$, $i = 1, \dots, 10$; $j = 1, \dots, 10$, which reflect the time and material costs of the transition from S_i to S_j . Numerical values of preference functions indicated on the arcs of the graph in conventional units.

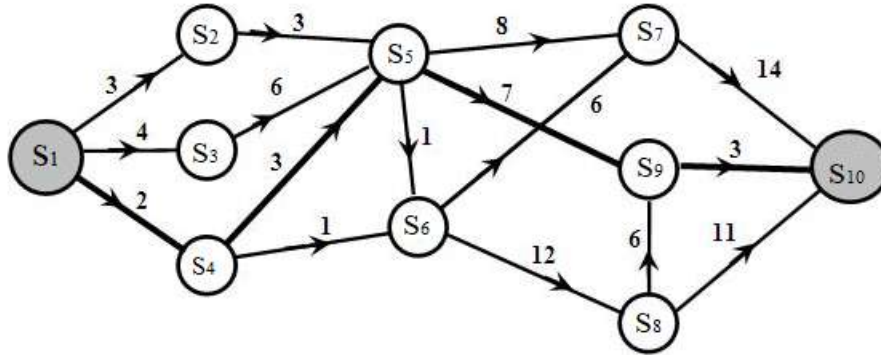


Fig.4. Fragment of a fuzzy situational network

To simplify and systematize the compilation of Bellman functions will number the vertices of S_i so that the arc leaves the vertex with a smaller number. In this case, we consistently find the f_i functions for each vertex of the oriented graph from the Bellman functional equation

$$f_1 = \min\{S_4 + f_1\}$$

$$f_1 = 0, f_2 = \min\{S_{21} + f_1\} = \min\{3 + 0\} = 3,$$

$$f_3 = \min\{S_{31} + f_1\} = \min\{4 + 0\} = 4$$

$$f_4 = \min\{S_{41} + f_1\} = \min\{2 + 0\} = 2,$$

$$f_1 = \min \begin{Bmatrix} S_{54} + f_4 \\ S_{53} + f_3 \\ S_{52} + f_2 \end{Bmatrix} = \min \begin{Bmatrix} 3 + 2 \\ 6 + 4 \\ 3 + 3 \end{Bmatrix} = 5$$

$$f_6 = \min \left\{ \begin{matrix} S_{64} + f_4 \\ S_{65} + f_5 \end{matrix} \right\} = \min \left\{ \frac{1+2}{1+5} \right\} = 3,$$

$$f_7 = \min \left\{ \begin{matrix} S_{76} + f_5 \\ S_{75} + f_5 \end{matrix} \right\} = \min \left\{ \frac{6+3}{8+5} \right\} = 9$$

$$f_8 = \min \{ S_{36} + f_6 \} = \min \{ 12 + 3 \} = 15,$$

$$f_9 = \min \left\{ \begin{matrix} S_{75} + f_5 \\ S_{98} + f_6 \end{matrix} \right\} = \min \left\{ \frac{7+5}{6+15} \right\} = 12$$

$$f_{10} = \min \left\{ \begin{matrix} S_{10,7} + f_7 \\ S_{10,9} + f_9 \\ S_{10,8} + f_8 \end{matrix} \right\} = \min \left\{ \begin{matrix} 14+9 \\ 3+12 \\ 11+5 \end{matrix} \right\} = 15$$

The sum of costs for the optimal trajectory is 15 conventional units. To determine it, the f_i functions viewed in reverse order. In this instance

$$f_{10} = \min \left\{ \begin{matrix} S_{10,7} + f_7 \\ S_{10,9} + f_9 \\ S_{10,8} + f_8 \end{matrix} \right\} = \min \left\{ \begin{matrix} 14+9 \\ 3+12 \\ 11+5 \end{matrix} \right\} = 15$$

The minimum selected amount of $3+12=15$ corresponds to the vertex S_9 .

$$f_9 = \min \left\{ \begin{matrix} S_{75} + f_5 \\ S_{98} + f_6 \end{matrix} \right\} = \min \left\{ \frac{7+5}{6+15} \right\} = 12$$

When calculating f_9 , the vertex S_5 selected. Continuing in the same way, we get the shortest path from a vertex S_1 to vertex $S_{10}(S_1, S_4, S_5, S_9, S_{10})$. In fig.6 arcs of the optimal trajectory shown in bold lines.

8 Conclusion

The principle of building IDSS, similar to the UEN IDSS developed in this article, is universal for other subject areas and can serve as a powerful incentive for creating intelligent information technologies and developing innovative systems for managing

urban development in the territory of a municipality [14-16]. The situational algorithm of fuzzy control given in the article is the basis of the ISPR and based on the automated method of mixed assessments and the expert method of generating and evaluating alternatives to management decisions [17]. Optimal trajectories of transitions from one situation to another, taking into account the preference functions, implemented on the dynamic programming method.

References

1. Soloviev, N.A., Nabatov, N.N.: System analysis of problems of emergency dispatching service of housing and communal services. Questions of technical, physical, and mathematical Sciences in the light of modern research. Novosibirsk, Sibak, (2018). pp. 5-12.
2. Kirakovski, V.V.: Geoinformation technologies in the diagnostics of engineering networks. GEOPROFI. Moscow: Geoprofi. (2005). No. 2. pp. 52-53.
3. Basalin, P.D., Bezruk, K.V., Radaeva, M.V.: Models and methods of intellectual support of decision-making processes: A textbook. Nizhny Novgorod: Nizhny Novgorod State University, (2011).
4. Demidova, L.A., Kirakovski, V.V., Pylkin, A.N.: Algorithms and fuzzy inference system in the diagnostics of city engineering communications. Moscow: Hotline-Telecom. (2005). 592p.
5. Kuzin, E.S.: Information complex problems and technology of their solution. Artificial intelligence news. (2003). No. 1. pp. 24-29.
6. Melikhov, A.N.: Situational advising systems with fuzzy logic. Moscow. Nauka, (1990). 272p.
7. Bodrov, V.I.: Mathematical methods of decision-making. Taganrog. TSTU, (2004). 124p.
8. Trahtengerts, E.A.: Computer-aided decision support. Moscow. Nauka, (1998). 376p.
9. Averkin, A.N., Batyrshin, I.Z., Blishun, A.F., Silov, V.B., Tarasov, V.B.: Fuzzy sets in control and artificial intelligence models. Moscow. Nauka, (1986). 312p.
10. Chekinov, G.P.: Application of situational management in information support of decision-making in the design of organizational and technical systems. Information technologies in design and production. (2003). 215p.
11. Stenin, A.A.: Latent semantic method for extracting information from Internet resources. Eastern European journal of advanced technologies. (2013). No. 4(9). pp. 19-22.
12. Stenin, A.A., Pasko, V.P., Lemeshko, V.A.: Neurosemantic approach to building automated information retrieval systems. Adaptive automatic control systems. KPI. (2019). №. 1(34). pp. 125-130.
13. Subbotin, S.A., Oleynik, A.A., Hoffman, E.A., Zaitsev, S.A., Oleynik, A.A.: Intelligent information technologies for designing automated systems for diagnostics and image recognition. Kharkiv. Smith Company, (2012). p. 318
14. Stenin, A. A.: Automation of the process of making innovative decisions in sociotechnical systems. Problems of information technologies. (2016). No. 19. pp. 51-57.
15. Stenin, A.A., Lisovichenko, O.I., Drozdovich, I.G., Stenin, S.A.: Analysis of passenger traffic and situational management of urban transport Bulgarian Journal for Engineering

Design issue. The Mechanical Engineering Faculty. Technical University-Sofia. No.40. (2019). pp. 7-12.

16. Vdovenko, A.V., Protasevich, E.V.: Information technologies in the management system of urban development of the territory of the municipality. Vestnik TOGU. Khabarovsk, (2009), no. 4(15). pp. 81-88.
17. Naikhanova, L.V., Dambaeva, S.V.: The Question-answer relationship in the method of knowledge retrieval a Survey. Theoretical and practical issues of modern information technologies. Ulan-Ude publishing house of VSGTU. (2003). pp. 37-41.
18. Orlov, A.I.: Organizational and economic modeling. Part II Expert assessments. Moscow. Publishing house of Moscow state University Bauman, (2011). 486p.
19. Ashikhmin, I.V., Furems, E.M.: UniComBos intellectual DSS for comparison and selection of multi-criteria objects. Proceedings of ISA RAS. (2005), vol. 12. pp. 16-25.
20. Barsegyan, A.A., Skupinou, M., Stepanenko, V.V., Holod, I.I.: Technology of data analysis. Data Mining, Visual Mining, Text Mining, OLAP. (2007). pp. 384
21. Trofymchuk, O.M., Bidyuk, P.I.: Decision support systems, modeling, forecasting, risk estimation. LAP LAMBERT Academic Publishing. (2019). 179p.

Application of Genetic Algorithm in Problems of Approximation of Complex Multidimensional Dependencies and Identification of Parameters of Theoretical Models

Dmitry Glukhov^{1[0000-0003-4983-2919]}, Tatsiana Hlukhava^{2[0000-0002-3154-3863]},
Aliaksandr Lukyanau^{3[0000-0001-6812-2184]}

Polotsk State University, Blokhin str., 29, Novopolotsk, 211440, Republic of Belarus
¹d.gluhov@psu.by, ²t.gluhova@psu.by, ³a.lukyanov@psu.by

Abstract. The article proposes a method for constructing an analytical approximation of n-dimensional data, based on the use of a genetic algorithm. A feature of the method is that the encoding of the search space is performed in the form of a parsing tree for an algebraic expression by the parser of the context-free grammar of the class LR (1). In addition, during the evolutionary process, in addition to the use of structure mutations (subject to their positive influence), the stage of mutation of the coefficients is performed, which allows avoiding the target function falling into local extremum. And also at each step of the evolutionary process, there is a stage for searching for an extremum in the space of coefficients and a stage for simplifying the analytical model.

Keywords: genetic algorithm, approximation, theoretical models

1 Introduction

By the complexity of a multidimensional dependency, we mean the uncertainty regarding the functional dependencies present in the data between input variables and the output value of the function. Such tasks include problems of identifying parameters of analytical models of approximation of multidimensional dependencies, identifying parameters of theoretical probability distributions, identifying parameters of trend models in forecasting.

Data approximation in the absence of a priori information about the real model requires the use of universal approximators, such as artificial neural networks, fuzzy logical approximators. However, the analytical representation of the model is the most compact and most accurate. Approximation methods based on the identification of parameters of some theoretical model solve the problem using an immutable model adopted from any assumptions of reasonableness.

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

We are making an attempt to use the genetic algorithm to find the most accurate analytical model of the approximator. The genetic algorithm performs not only the identification of parameters of a theoretical model, but also the search for the model itself in the space of possible modifications of the structure. The purposeful movement of the population of models to models providing a qualitative approximation is carried out due to the law of evolutionary selection according to the criterion of minimizing the standard deviation.

2 Implementation of genetic algorithm

A genetic algorithm (GA) is a heuristic search algorithm used to solve optimization and simulation problems by sequentially selecting, combining, and varying the desired parameters using mechanisms that resemble biological evolution.

Previously, we made attempts to solve this problem on multi-agent systems, which for analytical models in the space of 3 variables showed results superior to results of artificial neural networks, but on a limited variety of models [1, 2, 3]. The main disadvantage of such approximators is a limited set of structure transformations formulated in the form of local heuristics. Heuristics are obvious ways for a human expert to change the structure. Accordingly, modifications that are not obvious to the expert become impossible. In our opinion, the use of GA removes this restriction.

In addition, we have built approximators based on fuzzy logic (FLA). Fuzzy logic systems proposed by Lotfi Zadeh [8] were developed in works [9, 10]. If we consider the unknown parameter as continuous, then in this case we can draw a parallel between the conclusion about the value of the unknown parameter and the approximation of the function and talk about the property of a fuzzy system to act as a linguistic approximator. The advantage of such approximators is their speed, and the main disadvantage – limited opportunities to improve approximation accuracy.

The context-free grammar Gexp of the LR(1) class of algebraic expressions proposed by us is presented below in the Beckus-Naur form:

```
formula:  exp      {ANode* o = (ANode*)$1; created-
Nodes.push_back(o);}
|         formula ';' exp  {ANode* o = (ANode*)$3; created-
Nodes.push_back(o);}
;
exp : NUM      {ANodeNUM* o = new ANodeNUM($1); $$ =
(void*) o;}
| '(' exp ')'  {$$ = $2;}
| exp '+' exp  {ANode* o = (ANode*) new AN-
odeOp((ANode*)$1, (ANode*)$3, PLUS); $$ = (void*) o;}
| exp '-' exp  {ANode* o = (ANode*) new AN-
odeOp((ANode*)$1, (ANode*)$3, MINUS); $$ = (void*) o;}
| exp '/' exp  {ANode* o = (ANode*) new AN-
odeOp((ANode*)$1, (ANode*)$3, DIV); $$ = (void*) o;}
```

```

| exp '*' exp {ANode* o = (ANode*) new ANodeOp((ANode*)$1, (ANode*)$3, MUL); $$ = (void*) o;}
| 'p' 'o' 'w' '(' exp ',' exp ')' {ANode* o = (ANode*) new ANodeOp((ANode*)$5, (ANode*)$7, POW);
    $$ = (void*) o;}
| 's' 'i' 'n' '(' exp ')' {ANode* o = (ANode*) new ANodeFunc((ANode*)$5, SIN); $$ = (void*) o;}
| 'c' 'o' 's' '(' exp ')' {ANode* o = (ANode*) new ANodeFunc((ANode*)$5, COS); $$ = (void*) o;}
| 'l' 'o' 'g' '(' exp ')' {ANode* o = (ANode*) new ANodeFunc((ANode*)$5, LOG); $$ = (void*) o;}
| 'e' 'x' 'p' '(' exp ')' {ANode* o = (ANode*) new ANodeFunc((ANode*)$5, EXP); $$ = (void*) o;}
| 't' 'a' 'n' '(' exp ')' {ANode* o = (ANode*) new ANodeFunc((ANode*)$5, TAN); $$ = (void*) o;}
| 'c' 't' 'a' 'n' '(' exp ')' {ANode* o = (ANode*) new ANodeFunc((ANode*)$6, CTAN); $$ = (void*) o;}
| NAME {ANode* o = (ANode*) new ANodeNAME(string((char*)$1)); $$ = (void*) o; delete $1;}
;

```

3 Parsing process

In the process of syntactic parsing, as we can see, the right-side parsing tree is built in the form of an object-oriented structure. In order to be able to compile an algebraic expression into a tree of objects nested into each other, we have developed a class system presented in the class diagram (Figure 1).

As a result of syntactic parsing the initialization file, an initial population of analytical n-dimensional models is created. Initial variations of the models are formulated by an expert in the formal language of the Gexp grammar.

Objects of the base class ANode and derived classes ANodeNUM, ANodeNAME, ANodeOp, ANodeFunc, ANodeProxy attribute semantic stack nodes and form a tree of right-side parsing of an algebraic expression.

The Genetic class implements the main stages of the GA. Namely:

1. initialization() - initialization of the population with expert-defined algebraic expressions for the case of n-variables;
2. calcAllErrors() - calculating of the approximation error for each individual in its current state;
3. coefficientMutation() – the process of random cloning of individuals with mutation of coefficients to ensure the possibility of exit from local extremes;
4. structureMutation() – the process of cloning with a structure mutation, subject to the positive effect of the mutation;
5. fitCoefficients() – the process of searching for the extremum of the target function in the coefficient space. In the implementation of this process, we have chosen the

Simplification is triggered when the algorithm detects the convergence of the coefficient to 0 or 1.

The block diagram of the evolutionary search algorithm is shown in Figure 2.

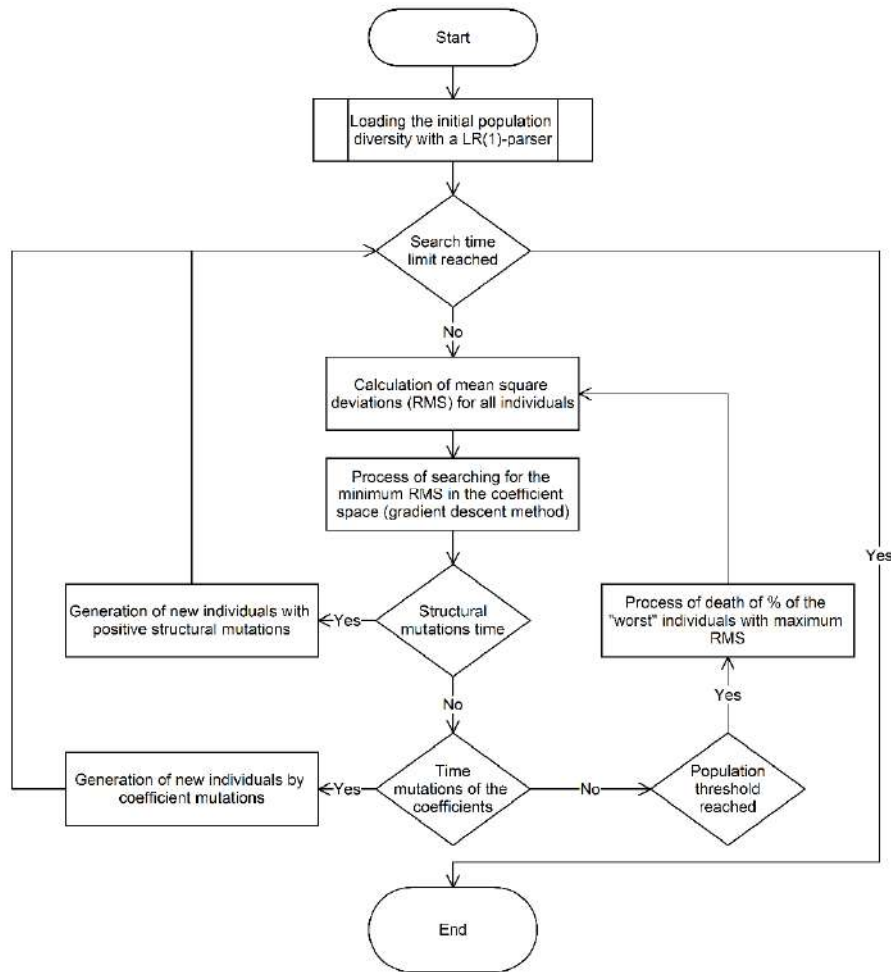


Fig. 2. Block diagram of the evolutionary search algorithm of the developed genetic algorithm

As the objective function we have chosen the standard deviation between the theoretical and calculated values of the function (SD). In the m-dimensional space of the coefficients of analytical models, the SD surface may have local extremes due to the limited data area and the complexity of approximated dependencies. However, the nature of the surface allows us to suggest the possibility of using the gradient descent algorithm to search for a local or global extremum. The nature of the SD surface in the space of two coefficients a , b of a function of one variable is illustrated in Figure 3.

In a number of works devoted to this topic, it is proposed to search for coefficients that minimize SD as a result of mutation of coefficients and to use the Nelder-Mead algorithm to search for a local extremum. [5, 6, 11]

Given that we are trying to minimize the standard error of the deviation of some analytical function from a discrete data set, we conducted a comparative analysis of the applicability of various numerical optimization methods and provide the user of the system with a choice of a specific method. Our approach is based on two points:

1. The mutation of coefficients is necessary for an abrupt transition to a new extremum search point, which allows the algorithm to run several extremum search strategies from different areas of the n-dimensional search space, and, therefore, to avoid getting into local extremes.
2. Application of the most qualitative methods for finding the extremum depending on the specifics of the problem (gradient descent method, simplex method, etc.) for identification of coefficients with a given accuracy.

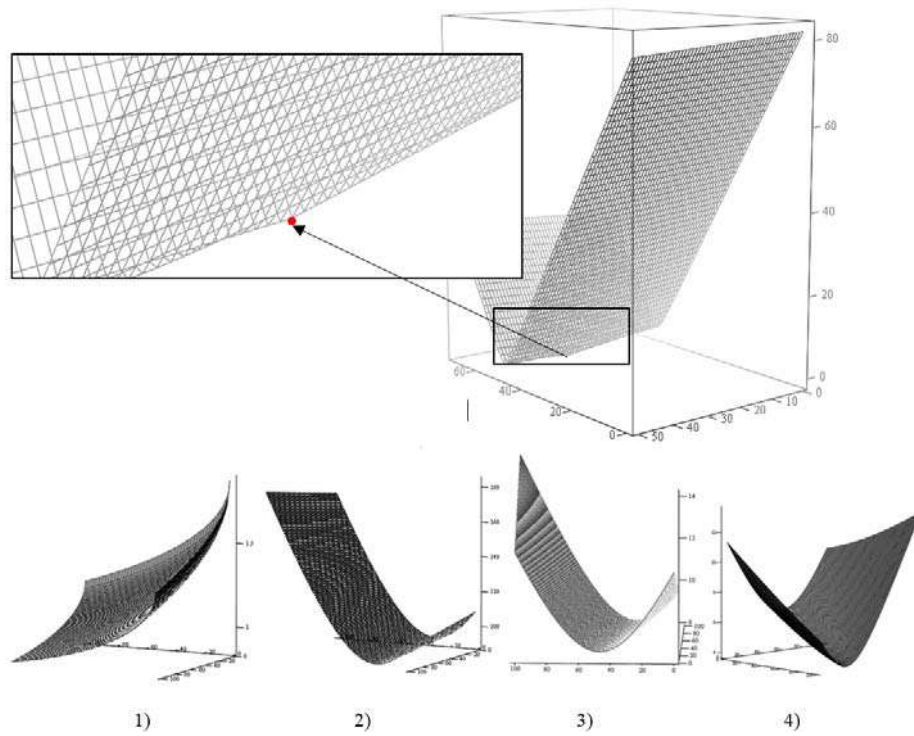


Fig. 3. The SD surface in the space of coefficients a, b of models:

1) $y(x) = \sqrt{\frac{a}{x}} + \sqrt{\frac{b}{x^2}}$, 2) $y(x) = ax^2 + bx$, 3) $y(x) = a \log(x) + bx$, 4) $y(x) = 5e^{1/x} + 3x$

Table 1. Example of the result of identifying model coefficients using the coordinate descent method

№	Formula	Standard deviation
Before identification of coefficients		
1	$5 * x_1 * x_1 + 3$	0.00000
2	$2 * x_1 + 3 * x_2 + 4$	56.92715
3	$2 * \text{pow}(x_2, 2) + 2$	154.52993
4	$15 * \text{pow}(x_1, 2) + 13$	166.70333
5	$2 * x_1 * x_1 + 3 * x_2 * x_2 + 2 * x_1 * x_2 + 3 * x_1 + 3 * x_2 + 3$	359.15568
6	$11 * \text{pow}(x_1, 3) + 3$	+INF
7	$\text{pow}(x_1, x_2) * 5$	+INF
After identification of coefficients		
1	$5 * x_1 * x_1 + 3$	0.00000
2	$27.9921 * x_1 + -2.02022 * x_2 + 3.12347$	17.37498
3	$1.8 * \text{pow}(x_2, 1.5) + 1.579$	57.33454
4	$5.00068 * \text{pow}(x_1, 1.99993) + 2.99844$	0.00100
5	$4.99973 * x_1 * x_1 - 0.48889 * x_2 * x_2 + 0.33344 * x_1 * x_2 - 3.33275 * x_1 + 4.86652 * x_2 + 3.2222$	0.00099
6	$5.00068 * \text{pow}(x_1, 1.99993) + 2.99844$	0.00100
7	$\text{pow}(x_1, x_2) * 3.3\text{e-}006$	50.52150

4 Algorithm operation example

In our numerical experiments, the best convergence rate was shown by the coordinate descent method, in which the step is performed along the coordinate giving the maximum gain. However, the search and implementation of more efficient optimization algorithms remains the subject of further research.

It is important to note that we have proposed a new method for screening out equivalent formulas. Since the same formula, due to the commutativity of operations (+, -, /, *) and the distributivity of the corresponding pairs of operations, as well as the possibility of arbitrary arrangement of brackets, can be written in many different ways, we proposed to form numerical hashes of formulas using special codes of variables, coefficients and functions and a mirror - modified algebra (+ instead of *, - instead of /, * instead of +, / instead of -).

An example of the operation of this hash function is shown in the Table 2:

Table 2. An example of the operation of this hash function

Formula	Hash
$2 * \log(x1) * x2$	16994
$\log(x1) * 2 * x2$	16994
$x2 * 2 * \log(x1)$	16994
$\log(x1) * x2 * 2$	16994
$2 * (\log(x1) + x2)$	2.89382e+007
$2 * \log(x1) + x2 * 2$	2.89722e+007

An example of the gradient descent method for 7 population individuals with an accuracy of 0.001 is shown in Table 1. The example also shows that we have developed a mechanism for processing results in the form of non-numbers and infinities, as large SD values. This approach allows you to continue searching without the occurrence of exceptional overflow situations and divisions by zero.

5 Algorithm application example

As an example of the application of the GA we constructed for solving an applied problem, we consider the problem of accounting for transients in stationary non-isothermal models of gas transportation.

Formulation of the problem: when the compressor station is switched on, gas heating begins and, despite the fact that the heated gas has not yet spread through the pipeline, the stationary model describes the state of the pipeline that will occur after the transition to steady state. Thus, significant sections of the pipeline in the model are found to have an overvalued gas temperature and, given the strong dependence of gas density on temperature, with an underestimated value of the gas reserve.

We calculated the correlation coefficient, which for the considered areas with temperature increase showed a high negative correlation, so for the 4 graphs given for example in Figure 4, the correlation coefficient was $\text{Correl}(X, Y) = \{-0,84369, -0,88839, -0,88509, -0,85517\}$.

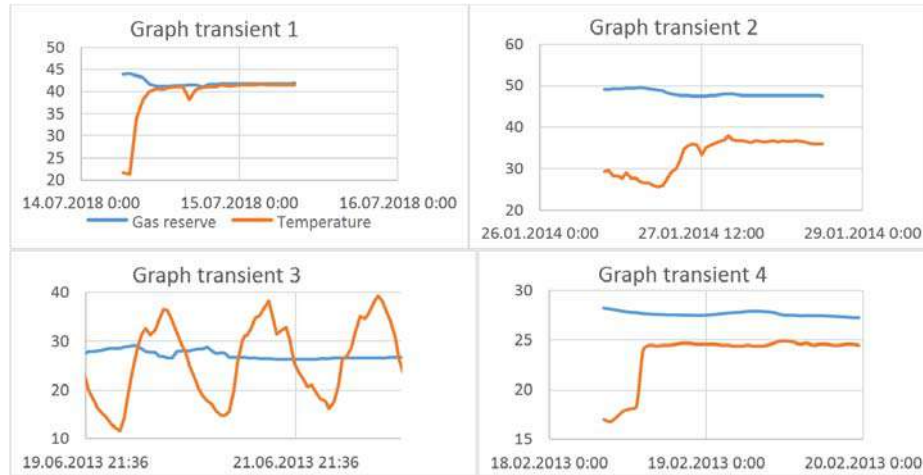


Fig. 4. Examples of transient graphs for the main gas pipeline model OJSC «Gazprom transgaz Belarus»

As the authors of the software package for calculating the gas reserve on the main gas pipeline of OJSC "Gazprom transgaz Belarus», we attempted to eliminate the influence of such transients by introducing the inertia of the temperature change at the compressor output during the time of system transition to the steady state.

We used GA to find an analytical relationship between the value of the transition time y and the current value of the gas flow Q , the ground temperature T_g , and the temperature difference at the output of the compressor station dT . The task of the approximator is to find such an analytical dependence that has a minimum SD for describing $y(Q, T_g, dT)$.

Comparative results of GA operation are given in Table 3.

Table 1. Comparative results of GA operation

№	Approximation method	Model	SD
	GA	$1.66597 * \text{pow}(1.7573, 1.1225 * \log(3.21177 * \text{pow}(dT * 4.23434, 1.1878))) / \text{pow}(\log(0.0858575 * Q), 0.0742341) - 4.15915 * dT$	1.078879
1	GA	$52 - 14.48 * \text{pow}(dT, 0.08961) / \text{pow}(Q, -0.1755)$	1.330954
2	GA	$102.607 * \text{pow}(dT, -0.369501) / \log(0.265488 * Q)$	1.200513
3	FLA	The model is described in our work [10]	1.080616

The solution found by GA exceeded in accuracy the fuzzy logical approximator with the selected optimal smoothing settings, while giving the most compact representation of the approximator.

The dynamics of improving the approximation quality can be demonstrated by the graph of positive mutations shown in Figure 5.

One of the parameters of the proposed GA is the limit size of the formula, which allows obtaining a wide variety of compact variants of approximators.

Developed GA has the possibility of flexible adjustment of limit thresholds of population size, percentage and death period of unpromising individuals, the probability of a period and type of mutations of the structure and coefficients, formulations of model simplification tools, algorithms of search of extremum. In particular, to solve the problem of accounting for transients in stationary non-isothermal models of gas transportation, we are interested not only in time, but also in the nature of temperature inertia, which requires collecting and processing of additional information.

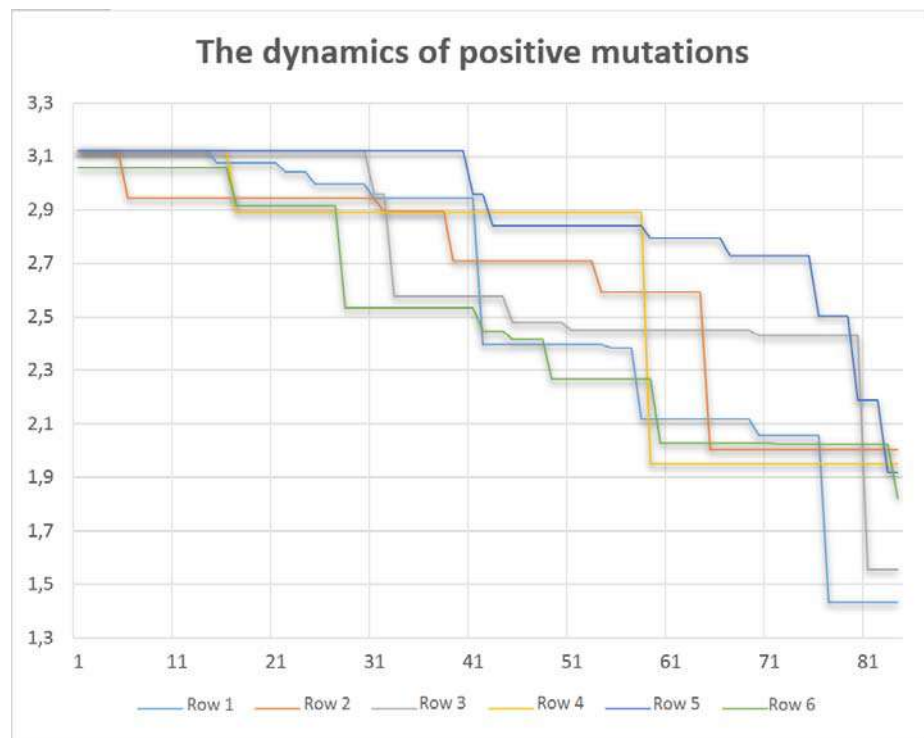


Fig. 5. SD change graph as a result of mutation steps

6 Conclusion

This article proposes a new method for constructing an analytical approximation of complex n-dimensional dependencies based on the use of the genetic algorithm.

The developed GA has the following features:

1. Search space encoding is performed as a parsing tree of an algebraic expression by the parser of the context-free grammar of the class LR(1);
2. In the course of the evolutionary process, in addition to the use of structure mutations (provided their positive effect), the stage of coefficients mutation is performed, which allows avoiding the target function falling into local extremes;
3. At each step of the evolutionary process there is a stage of searching for an extremum in the coefficient space and a stage of simplifying the analytical model.

The developed GA helped us solve the problem of accounting for transients in stationary non-isothermal gas transportation models. The resulting models were implemented as part of the software package for calculating the gas reserve on the main gas pipeline of OJSC "Gazprom transgaz Belarus», which reduced the impact of transients on the gas imbalance.

References

1. Glukhov A.O., Trofimov V.V.: Multi-agent structures for solving the traveling salesman problem. Problemy menedzhmenta: sb. nauchn. tr. Vypusk №3; pod obshch. red. prof. O.A. Strakhovoy, pp. 71–76. Izd-vo SPbGUEF, Saint-Petersburg (2000).
2. Trofimov V.V., Glukhov A.O.: Approximation on multi-agent structures. Ekonomicheskaya kibernetika: sistemnyy analiz v ekonomike i upravlenii: sb. nauchn. tr. Vypusk №1; pod red. D.V. Sokolova i N.N. Pogostinskoy, pp. 40–53. Izd-vo SPbGUEF, Saint-Petersburg (2000).
3. Glukhov A.O., Glukhov D.O., Trofimov V.V., Trofimova L.A.: Modified hybrid genetic algorithm of discreet optimization problems. 2017 XX IEEE International Conference on Soft Computing and Measurements (SCM 2017). Saint-Petersburg (2017).
4. Pozharskiy D. A., Zolotov N.B., Semenov I.E.: Genetic algorithm for finding the approximation coefficients of a function in contact problems for a cylinder. Molodoy uchenyy №24 (158), pp. 122–125 (2017).
5. Kildiushov M.S.: A program for reconstructing approximated algebraic functions from several variables from a set of discrete values of a function. Online magazine «NAUKOVEDENIE» Vol. 7, №5, (2015) <http://naukovedenie.ru/PDF/136TVN515.pdf>. doi: 10.15862/136TVN515
6. Kildiushov M.S.: The use of genetic algorithms for the restoration of approximated algebraic functions with a certain accuracy. Nauka i biznes: puti razvitiya. – Fond razvitiya nauki i kultury (Tambov) № 1 (55), pp. 25–31 (2016). ISSN: 2221-5182
7. Driankov D., Hellendoorn H., Reinfrank M.: An introduction to fuzzy control. Springer-Verlag (1993).
8. Lotfi A. Zadeh: Fuzzy Sets. Information & Control vol. 8, pp. 338–353 (1965).
9. Glukhov D.O.: Dynamic expert system by fuzzy inference rules to automations an examination of complex objects. Budownictwo i Inzynieria Srodowiska. – Zielonogorsk: Politechnika Zielonogorska, pp. 105–109 (1998).

10. Glukhov D.O., Glukhova T.M., Kundas S.P.: Soft calculations for organization of computer representation of nomograms on the example of calculating the limit creep coefficient. Vestnik Polotskogo gosudarstvennogo universiteta. – Series C: Basic Sciences, №3. pp. 2–6 (2010).
11. Zvonkov V.B., Popov A.M.: Comparative study of classical optimization methods and genetic algorithms. Vestnik Sibirskogo gosudarstvennogo aehrokosmicheskogo universiteta im. Akademika M.F. Reshetneva, pp. 23–27. Izdatelstvo: Federalnoe gosudarstvennoe byudzhethoe obrazovatelnoe uchrezhdenie vysshego obrazovaniya «Sibirskij gosudarstvennyj universitet nauki i tekhnologii imeni akademika M.F. Reshetneva», Krasnoyarsk (2013). ISSN: 1816-9724

Linear Quadratic Gaussian Control for Robotic Excavator

Alexander Gurko^[0000-0001-9905-8584], Igor Kyrychenko^[0000-0002-2128-3500],
Aleksandr Yaryzhko^[0000-0001-6398-8472], Oleksandr Kononykhin^[0000-0002-6396-6836],

Kharkiv National Automobile and Highway University,
Ya. Mudrogo str., 25, Kharkiv, 61002, Ukraine
gurko@khadi.kharkov.ua, igk160450@gmail.com, yaryzko@gmail.com,
makonon@i.ua

Abstract. When developing a robotic excavator, one of the main issues is tracking the given trajectories using its manipulator. This task is complicated by the fact that the system is subject to disturbances and measurement noise, which can most naturally be modelled as stochastic white noise processes. This paper describes an LQG regulator aimed at improving tracking performance at the levelling operation made by the robotic excavator. Simulation is performed on the attached backhoe equipment of the Boreks 2201 excavator. The results show that the proposed control algorithm is effective for improving the trajectory tracking accuracy.

Keywords: robotic excavator, trajectory tracking, LQG, Kalman filter

1 Introduction

Hydraulic excavators are widely used in many areas, such as construction, mining and agriculture, as well as in dangerous areas for rescue and recovery operations. Strict requirements for improving the quality of digging operations and reducing energy and time costs for the entire workflow lead to a change in the look of traditional hydraulic excavators. For instance, the number of degrees of freedom for working equipment has increased; articulated booms and sticks have appeared; multi-handed excavators have become available on the market, etc. However, growing complexity of the excavator requires great skill to control it, and now even an experienced operator cannot realize the full capabilities of the machine. Thus, automation of excavators is considered as a challenge in modern construction machinery. Moreover, as excavators are often used to accomplish dangerous tasks, an unmanned robotic excavator can greatly improve the safety.

Nevertheless, the lack of information about the environment and unpredictable mechanical parameters such as flexibility, friction, various nonlinearities in hydraulic actuators prevent the extensive use of unmanned excavators. To mitigate the above problems varieties of studies have been made recently concerning adaptive and robust control of an excavator arm.

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

In particular, Kim [1] has proposed a nonlinear proportional-derivative (PD) controller coupled with the μ -controller to compensate parametric and unstructured uncertainties. Wang [2] has introduced the nonlinear proportional-integral (PI) controller combined with a cross-coupled pre-compensation algorithm to improve tracking accuracy by an unmanned excavator bucket edge. PI-based adaptive velocity controller for each cylinder of a mini excavator has been presented by Wind in [3].

Some researchers have utilized Genetic Algorithms [4], artificial neural networks and fuzzy logic [5–9], ant colony optimization [10, 11] in order to adjust the parameters of the traditional proportional-integral-derivative (PID) and PD controllers. Sliding-mode and Time delay controls have been applied to compensate for the non-linearity of complex dynamics of hydraulic manipulators in [12, 13, 14]. Besides, in some studies, other intelligence algorithms are used, such as guaranteed cost control [15, 16], model predictive control [17], model reference adaptive control [18], a robust stochastic control, based on the method of analytical design of aggregated regulators (ADAR) [19], that were applied to control the trajectory of the excavator.

All the mentioned and other related works have contributed significantly to robotic excavator development. However, these controllers are quite complex; their use requires significant re-equipment of excavators, in particular, replacement of the hydraulic system. Therefore, no unmanned excavator is mass-produced today. Thus, automation of the excavators remains a topical issue.

Since in the process of digging the values of external disturbances and measurement noise are unknown, it is expedient to assume them as stochastic white noise processes. In such a case, a well-known in modern control theory Linear Quadratic Gaussian regulator (LQG), which is relatively easy to design and implement [20], can be used to control the excavator manipulator.

Thus, *this paper aims* at designing and applying the LQG in a hydraulic robotic excavator to control the manipulator trajectory.

The remainder of this paper is organized as follows. The formal statement of the problem is given in Section 2. Section 3 is dedicated to the excavator modelling; non-linear as well as linearized models of the excavator manipulator are described in this section. Section 4 presents the design procedure of the controller. The simulation results are provided in Section 5. Finally, Section 6 presents concluding remarks and future directions for this work.

2 Formal Problem Statement

It is assumed that the excavator manipulator model is linearized and the state equations of its motion look like:

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t) + w(t), \\ y(t) &= Cx(t) + v(t), \end{aligned} \quad (1)$$

where $y(t)$ is the $(r \times 1)$ observation vector of the true state $(n \times 1)$ vector $x(t)$ at time t ; $u(t)$ is the control input $(m \times 1)$ vector, $w(t)$ and $v(t)$ are stochastic processes associated with the working process and the manipulator state measurement, respectively.

The state matrix A ($n \times n$), control input gain matrix B ($n \times m$) and measured state matrix C ($r \times n$) are all linear time invariants, $n = 6$, $m = 3$, $r = 3$.

The process noise $w(t)$ and measurement noise $v(t)$ are white Gaussian random sequences with zero mean value, and $V(t)$ and $W(t)$ are the $v(t)$ and $w(t)$ covariance matrices.

It is necessary to obtain a control law in the form of static linear feedback:

$$u(t) = -Kx(t), \quad (2)$$

where the gain matrix K is meant to minimize a quadratic cost function formulated as follows:

$$J = \int_0^T (x(t)^T Q x(t) + u(t)^T R u(t)) dt, \quad (3)$$

where Q and R are symmetric positive definite matrices of compatible dimensions.

3 Excavator manipulator modelling

In this paper, the attached working equipment of the Boreks 2201 excavator is considered as an example. To study this excavator, its 3D model was originally built in Autodesk Inventor 2016, and afterwards it was imported into MATLAB Simscape Multibody [21]. The general view of the model is given in Fig. 1. The adequacy of the model was proved by the comparison with the results of a real excavator testing. The inputs of the model are the desired rods displacements of the boom, the stick and the bucket hydro cylinders, calculated on the base of the desired rotations of the corresponding joints [21].

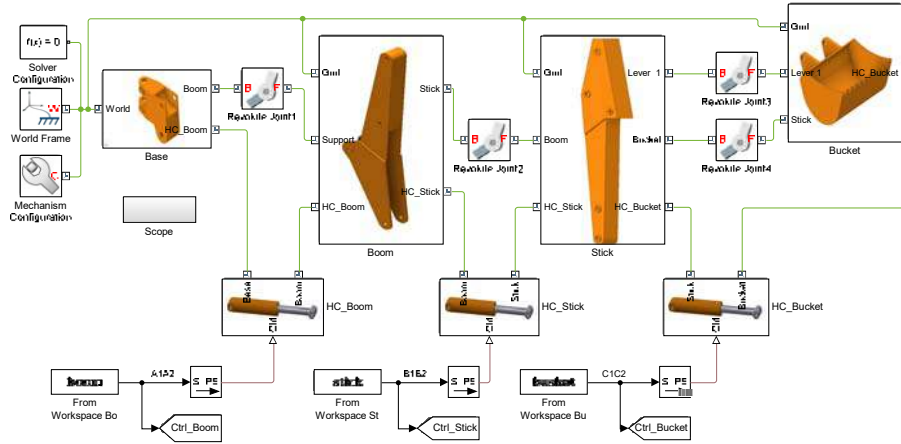


Fig. 1. General view of the excavator manipulator model

The model in Fig. 1 is a nonlinear one; however, the LQG theory is developed for linear systems described by (1). Therefore, in this section, the linear model of the excavator manipulator is under consideration. In robotics the feedback linearization [22] is widely known, though the obtained model is quite complex and requires knowledge of all the dynamic parameters of the manipulator. On the other part, a model of the manipulator link can be simplistically described by the second order system [14, 22]:

$$G_i(s) = \frac{1}{m_i s^2 + b_i s + c}, \quad (4)$$

where m_i , b_i , c_i are positive parameters to be determined, $G_i(s)$ is a transfer function from the desired θ_{di} to the actual θ_i rotation angle of the excavator manipulator links, and $i = 1, 2, 3$.

The transfer function (4) can be rewritten in the state-space form by the following relation:

$$\begin{aligned} \dot{x}_i(t) &= \begin{bmatrix} 0 & 1 \\ -c_i/m_i & -b_i/m_i \end{bmatrix} x_i(t) + \begin{bmatrix} 0 \\ 1/m_i \end{bmatrix} u_i(t), \\ y_i(t) &= [1 \quad 0] x_i(t), \end{aligned} \quad (5)$$

where $x_i(t) = \theta_i(t)$, $i = 1, 2, 3$.

In order to obtain the linear dynamic model of the excavator manipulator, step responses of the system in Fig. 1 were obtained for each joint (Fig. 2).

The values of m_i , b_i and c_i in (4) can be estimated using any of the numerical optimization methods that minimize the error between the outputs of the real system and its model based on the known information about the inputs and outputs of the real system. Here, to automate this procedure, the MATLAB System Identification Toolbox was used. The estimated values are given in Table 1, and Fig. 2(d) presents the errors between the actual outputs of each link and the outputs of the corresponding linear models.

Table 1. Estimated values of the linearized model for each joint

	m	b	c
Boom	$1.1285 \cdot 10^{-4}$	0.020057	1.000
Stick	$1.0065 \cdot 10^{-4}$	0.020011	1.000
Bucket	$1.0001 \cdot 10^{-4}$	0.020000	1.000

Fig. 2(d) shows that the results of the linear model coincide with the nonlinear one. The response of the linear model is somewhat slower but the agreement is satisfactory with respect to the high degree of simplification in the linear model. Therefore, it can be concluded, that the established linear model is accurate and credible.

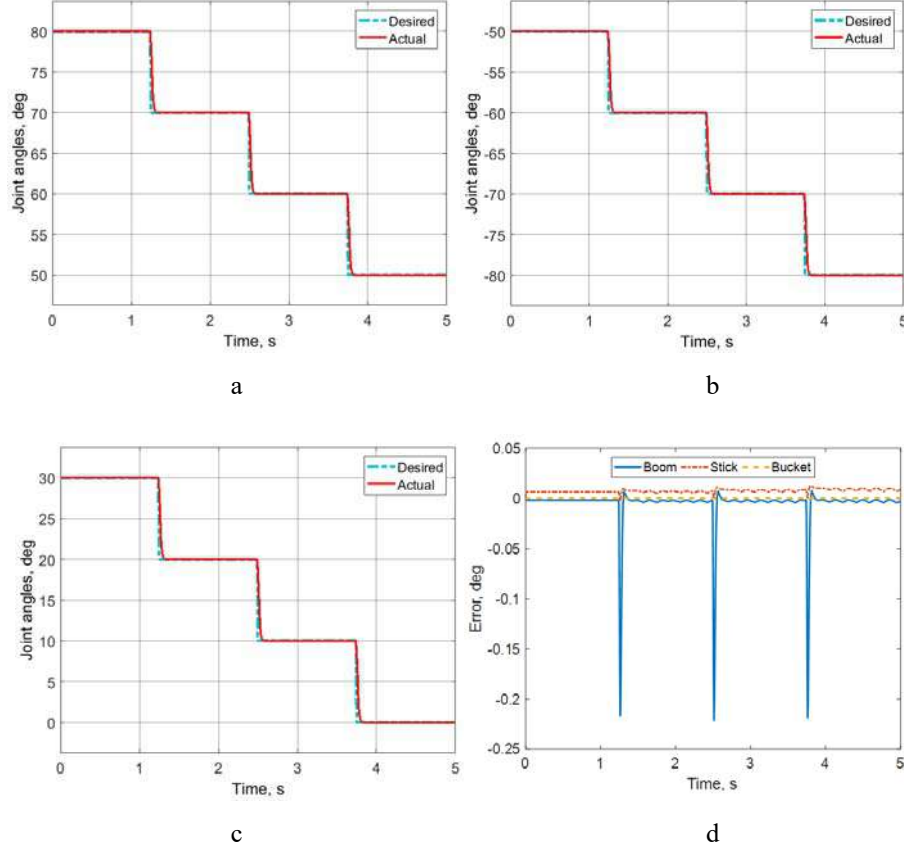


Fig. 2. Desired and actual joint angles of the boom (a), the stick (b) and the bucket (c), and the errors of the linearized model (d)

4 Controller design

4.1 LQG general form

LQG control is a modern state-space technique for designing optimal dynamic regulators [20]. It is rooted in optimal stochastic control theory and combines both the concepts of Linear Quadratic Regulators (LQR) for full state feedback and Kalman filters for state estimation. The plant is described by the state and output equations (1) and the controller design methodology enables a controller to be synthesized which is optimal with respect to the quadratic cost function (3), where Q and R are weighting matrices that define the trade off between regulation performance and control efforts. Taking into consideration the second term in (3) allows getting control signals of limited amplitude at its minimization, which is especially important when designing control systems of such unmanned vehicles as excavators.

The gain matrix K in the control law (2) has the following form

$$K = R^{-1} B^T P, \quad (6)$$

and is obtained by solving the algebraic Riccati equation (CARE):

$$A^T P + PA - PBR^{-1} B^T P + Q = 0, \quad (7)$$

where P is the unknown ($n \times n$) symmetric matrix.

Due to uncertainties, the Kalman filter is used to obtain the estimate of the state vector $\hat{x}$ (Fig. 3):

$$\dot{\hat{x}}(t) = A\hat{x}(t) + Bu(t) + L(y(t) - C\hat{x}(t)). \quad (8)$$

The Kalman filter gain L in (8) equals:

$$L = V^{-1} C^T S, \quad (9)$$

where S is the solution of the following matrix Riccati equation:

$$A^T S + SA - SCV^{-1} C^T S + W = 0. \quad (10)$$

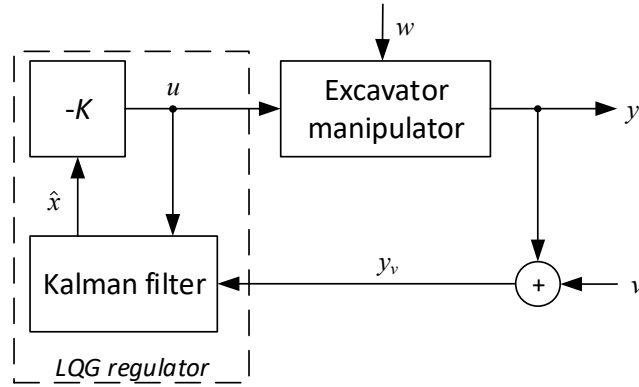


Fig. 3. Block diagram of the excavator control system with LQG regulator

4.2 Controller design

The controller design was accomplished in MATLAB. According to the common approach [20], the LQG design was split into two independent steps: first, the LQR was designed, using the combined state space model of the manipulator as (1):

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ -c_{bo}/m_{bo} & -b_{bo}/m_{bo} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & -c_{st}/m_{st} & -b_{st}/m_{st} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & -c_{bu}/m_{bu} & -b_{bu}/m_{bu} \end{bmatrix},$$

$$B = \begin{bmatrix} 0 & 0 & 0 \\ 1/m_{bo} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1/m_{st} & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1/m_{bu} \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix},$$

where x_1, x_3, x_5 are joint angles of the boom, the stick and the bucket, x_2, x_4, x_6 are their angular velocities, respectively.

The initial values of the weighting matrices in (3) were selected by using Bryson's rule, relating the reciprocal of the maximum squared values of the states with Q and the reciprocal of the maximum squared values of the control inputs with R :

$$Q = \frac{1}{\max(x^2)}, \quad (11)$$

$$R = \frac{1}{\max(u^2)}, \quad (12)$$

and then were refined experimentally. Satisfactory closed-loop responses were obtained using weighting matrices

$$Q = \begin{bmatrix} 1.5 & 0 & 0 & 0 & 0 & 0 \\ 0 & 5 \cdot 10^{-4} & 0 & 0 & 0 & 0 \\ 0 & 0 & 1.5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 5 \cdot 10^{-4} & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 \cdot 10^{-4} \end{bmatrix}, \quad R = \begin{bmatrix} 1 \cdot 10^{-5} & 0 & 0 \\ 0 & 1 \cdot 10^{-5} & 0 \\ 0 & 0 & 1 \cdot 10^{-5} \end{bmatrix}.$$

The obtained gain matrix K is:

$$K = \begin{bmatrix} 386.30 & 7.06 & 0 & 0 & 0 & 0 \\ 0 & 0 & 386.30 & 7.06 & 0 & 0 \\ 0 & 0 & 0 & 0 & 446.21 & 4.46 \end{bmatrix}.$$

4.3 Kalman filter design

At the second step, the Kalman filter was designed in MATLAB using noise covariances $W = \text{diag}\{0.1\}$, $V = \text{diag}\{0.02\}$. The Kalman gain matrix is:

$$L = \begin{bmatrix} 0.0651 & 0 & 0 \\ -2.4979 & 0 & 0 \\ 0 & 0.0626 & 0 \\ 0 & -2.4980 & 0 \\ 0 & 0 & 0.0625 \\ 0 & 0 & -2.4980 \end{bmatrix}.$$

5 Simulation results

The effectiveness of the proposed controller was verified by simulation, which was performed for levelling works using the nonlinear model in Fig.1. The desired bucket edge path and the initial and final configurations of the manipulator are given in Fig. 4.

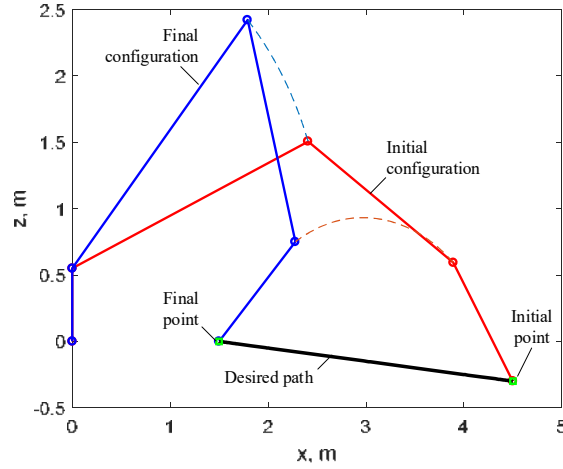


Fig. 4. The desired bucket edge path and extreme positions of the manipulator

During digging, resistance force F_r acts at the cutting edge of the bucket teeth, this force is a resultant reaction force of the tangential, F_t and the normal F_n forces. At the simulation, the tangential force was simplistically determined as

$$F_t = k_c b \sum h \Delta x, \quad (13)$$

where k_c is the specific cutting force, that takes into account soil resistance to cutting, frictional resistance of the bucket with the soil, resistance to the movement of the

prism of soil and all other forces; h and b are the thickness and width of the cut slice of soil; Δx is an increment of the bucket teeth path along the x-axis [15].

The normal component F_n was calculated as:

$$F_n = \psi F_t, \quad (14)$$

where ψ is a dimensionless factor depending on the digging angle, digging conditions and the cutting edge where $\psi = 0.1\text{--}0.45$. Higher values of ψ correspond to more blunting of the bucket teeth edge.

It was assumed that $k_c = 181.3 \text{ kN/m}^2$, $b = 0.75 \text{ m}$, $h = 0.1 \text{ m}$, $\psi = 0.1$. The forces F_t and F_n were unknown for the controller during simulation.

Figs. 5(a)–(c) show the desired, actual and measured trajectories of the boom, the stick and the bucket respectively, and the tracking errors are depicted in Fig. 5(d). Fig. 6 describes the levelling error.

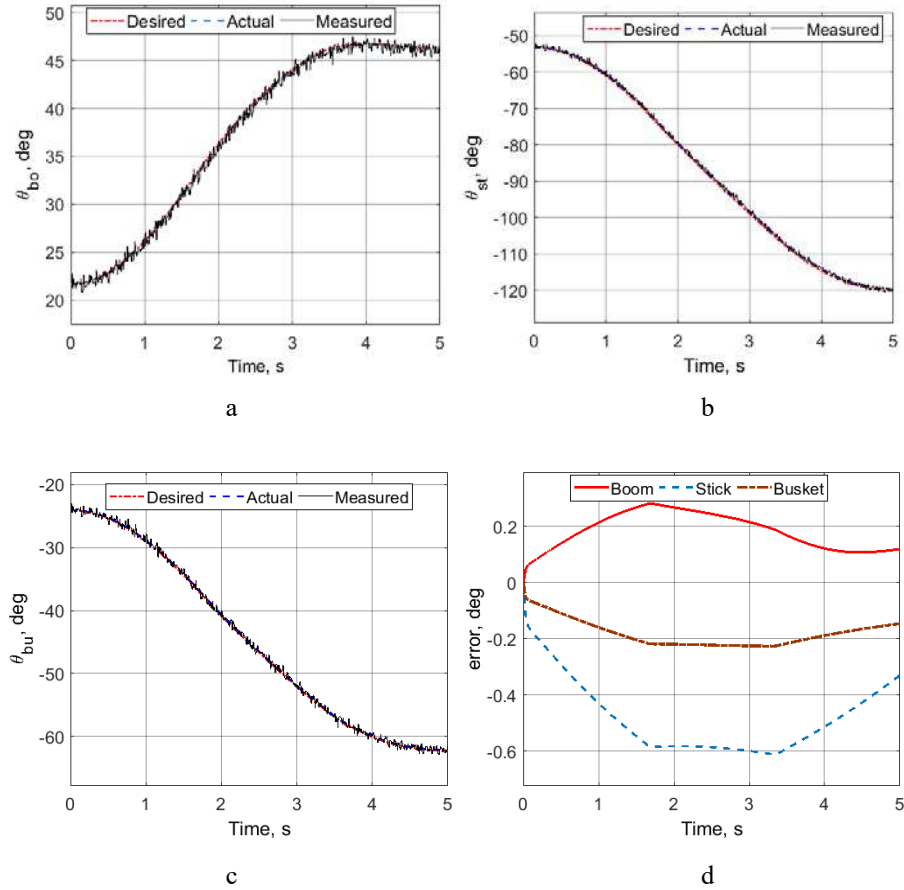


Fig. 5. Displacements of the manipulator links: a – for the boom, b – for the stick; c – for the bucket; d – links trajectories tracking errors

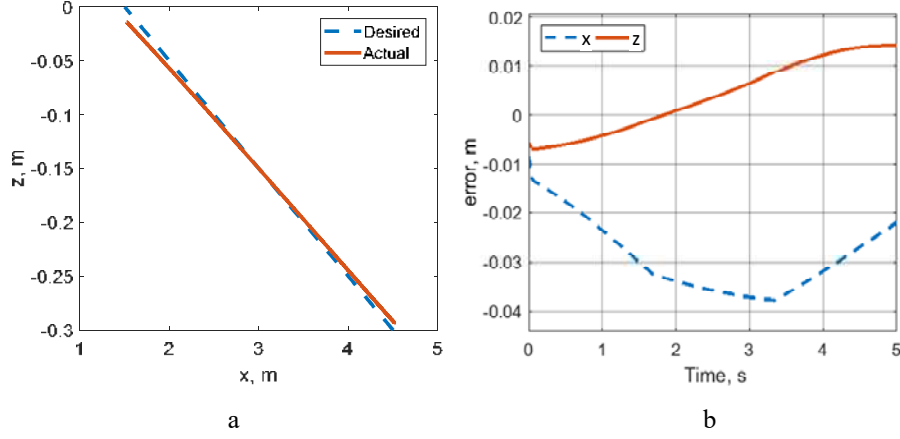


Fig. 6. Levelling error: a – in x - z plane; b – on each of the coordinates

There is a good agreement between the desired and measured displacements of the manipulator links (Fig. 5), and the fit to the desired trajectories is near 99%. As can be seen from Fig. 6, the error of the levelling does not exceed 4 cm along the x -axis and 1.5 cm – along the z -axis. These values are far higher than those obtained using the PID controller [21] (1 cm and 0.1 cm, respectively), however, in [21] the simulation was performed under ideal conditions, without taking into account the disturbances. Hence, the results prove that the LQG control ensures good system performance. It should be noted, though, that backlashes, which have not taken into the consideration in the model in Fig. 1, as well as variability of soil properties can increase the tracking error.

6 Conclusion and future work

In this paper, we have shown how an LQG regulator can be used to control robotic excavator trajectory. For this purpose, a linear model of the excavator manipulator was obtained, and its parameters were identified by the comparison with the nonlinear model, described in the previous work.

The control system is designed in the way that it provides a deviation of the bucket edge from a given path no more than 4 cm along the x -axis and 1.4 cm along the z -axis at the levelling operation in the presence of random measurement noise and unknown forces appearing when the bucket touches the soil. Tracking accuracy can be improved by adding the integrator to the controller. However, this can cause difficulties associated with time delays [20], which may not be taken into account in the model in Fig. 1. Moreover, the simulation was carried out for uniform soil, so more significant tracking deviations can be expected in real conditions. Therefore, our future work will be focused on reducing tracking errors under unknown external disturbances, including the variability of soil properties.

References

1. Kim, S., Park, J., Kang, S. et al. A Robust Control Approach for Hydraulic Excavators Using μ -synthesis. *Int. J. Control Autom. Syst.* **16**, 1615–1628 (2018). <https://doi.org/10.1007/s12555-017-0071-9>
2. Wang, D., Zheng, L., Yu, H. et al. Robotic excavator motion control using a nonlinear proportional-integral controller and cross-coupled pre-compensation. *Automation in Construction*, **64**, 1–6 (2016). <https://doi.org/10.1016/j.autcon.2015.12.024>.
3. Wind H., Jenisch L., Renner A. et al. Design of an Adaptive Velocity Controller for a Hydraulic Mini Excavator using a Moving-Horizon-Estimator. In: *IEEE Conference on Control Technology and Applications*, pp. 685–690 (2019). <https://doi.org/10.1109/CCTA.2019.8920496>
4. Feng, H., Yin, C.-B., Weng, W. et al. Robotic excavator trajectory control using an improved GA based PID controller. *Mechanical Systems and Signal Processing*. **105**, 153–168 (2018). <https://doi.org/10.1016/j.ymssp.2017.12.014>
5. Le Hanh, D., Ahn, K.K., Kha, N.B. et al. Trajectory control of electro-hydraulic excavator using fuzzy self tuning algorithm with neural network. *J Mech Sci Technol* **23**, 149–160 (2009). <https://doi.org/10.1007/s12206-008-0817-7>
6. Yan, J., Li, B., Guo, G. et al. Adaptive control for bucket position of excavator based on single neuron PID. *Journal of PLA University of Science and Technology (Natural Science Edition)* **3**, 316–319 (2012).
7. Park, J., Cho, D., Kim, S. et al. Utilizing online learning based on echo-state networks for the control of a hydraulic excavator. *Mechatronics* **24**(8), 986–1000 (2014). <https://doi.org/10.1016/j.mechatronics.2014.10.004>
8. Li, Y., Wang, Q. Adaptive neural finite-time trajectory tracking control of hydraulic excavators. *Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering* **232**(7), 909–925 (2018). <https://doi.org/10.1177/0959651818767770>
9. Bo, L., Jun, Y., Gang, G., et al. High performance control of hydraulic excavator based on fuzzy-pi soft-switch controller. In: *IEEE international conference on Computer science and automation engineering*, pp. 676–679 (2011). <https://doi.org/10.1109/CSAE.2011.5952558>
10. Hassan, M. Y., Sugban, A. F. Ant Colony Optimization Based Type-2 Fuzzy Force-Position Control for Backhoe Excavator Robot. *Al-Nahrain Journal for Engineering Sciences* **21**(1), 1-11, (2018). <https://doi.org/10.29194/NJES21010001>
11. Feng, H., Yin, C., Ma, W., et al. Parameters identification and trajectory control for a hydraulic system. *ISA Transactions* **92**, 228–240. (2019). <https://doi.org/10.1016/j.isatra.2019.02.022>.
12. Xu, J., Yoon, H. Sliding Mode Control of Hydraulic Excavator for Automated Grading Operation. *SAE International Journal of Commercial Vehicles*. **11**(2), 113–124 (2018). <https://doi.org/10.4271/02-11-02-0010>
13. Wang, J., Lee, M. C., Kallu, et al. Trajectory tracking control of a hydraulic system using TSMCSPPO based on sliding perturbation observer. *Applied Sciences*, **9**(7), 1455 (2019). <https://doi.org/10.3390/app9071455>
14. Kim, J., Jin, M., Choi, W., et al. Discrete time delay control for hydraulic excavator motion control with terminal sliding mode control. *Mechatronics* **60**, 15–25 (2019). <https://doi.org/10.1016/j.mechatronics.2019.04.008>.
15. Gurko, A. G., Sergiyenko, O. Yu., Hipólito, J. I. N., et al. Guaranteed Control of a Robotic Excavator During Digging Process. In: *12th International Conference on Informatics in Control, Automation and Robotics*, **2**, pp. 52–59. SciTePress (2015).

16. Gurko, A. G., Sergiyenko, O. Yu., Hipólito, J.I.N., et al. Trajectory Tracking Control of an Excavator Arm Using Guaranteed Cost Control. In: *Lecture Notes in Electrical Engineering*, **383**, 177–196 (2016) Springer, Cham https://doi.org/10.1007/978-3-319-31898-1_10
17. Tomatsu, T., Nonaka, K., Sekiguchi, K., et al. Model predictive trajectory tracking control for hydraulic excavator on digging operation. In: *IEEE Conference on Control Applications*, pp. 1136–1141 (2015). <https://doi.org/10.1109/CCA.2015.7320765>
18. Chiang, M., Huang, C. Experimental implementation of complex path tracking control for large robotic hydraulic excavators. *Int J Adv Manuf Technol* **23**, 126–132 (2004). <https://doi.org/10.1007/s00170-003-1636-8>.
19. Kolesnikova, S., Tsvetnitskaya S., Pobegailo, P. An Extension of the Method of Aggregated Regulators to a Discrete Stochastic Object. In: *International Siberian Conference on Control and Communications*, pp. 1–6 (2019). <https://doi.org/10.1109/SIBCON.2019.8729595>
20. Kwon W.H., Park P. LQG Optimal Controls. In: *Stabilizing and Optimizing Control for Time-Delay Systems. Communications and Control Engineering*. Springer, Cham (2019). https://doi.org/10.1007/978-3-319-92704-6_7
21. Gurko, A., Kyrychenko, I., Yaryzhko, A. Trajectories Planning and Simulation of a Backhoe Manipulator Movement. In: *CEUR Workshop Proceedings*, pp. 771–785 (2019). Online <http://ceur-ws.org/Vol-2353/paper61.pdf>.
22. Siciliano B., Khatib O. (Eds.): *Springer Handbook of Robotics*. Springer (2016).

Packing Equal Spheres by Means of the Block Coordinate Descent Method

G. Yaskov^[0000-0002-1476-1818], A. Chugay^[0000-0002-4079-5632]

A. Pidgorny Institute of Mechanical Engineering Problems of the National Academy of Sciences of Ukraine, 2/10 Pozharsky St., 61046 Kharkiv, Ukraine
yaskov2004@gmail.com, chugay.andrey80@gmail.com

Abstract. The paper deals with the problem of packing a large number of equal spheres into a container of spherocylindrical shape with a cylindrical prohibition area which arises in chemical industry. The problem is presented as a mathematical programming problem on the ground of the Stoyan's phi-functions method. Solving the problem is reduced to solving a sequence of non-linear programming problems making use of the block coordinate descent method and analysis of Lagrange multipliers to realize sequential addition of spheres. Numerical examples for up to two millions of spheres are given.

Keywords: packing, sphere, spherocylinder, catalyst, optimization

1 Introduction

Packing spheres consists to arrange equal or non-equal non-overlapping spheres within a containing space or into a larger three-dimensional domain (container). A typical objective is to fill as much of the space (container) as possible. The optimization sphere packing problem is studied in discrete and computational geometry.

C.F. Gauss proved in 1831 [1] that the hexagonal lattice sphere-packing configuration has the highest density amongst all possible lattice packings with the asymptotic value $\pi/(3\sqrt{2})$. Given a bounded container, the maximum packing density decreases depending on the ratio of the sphere diameter to the container sizes.

The sphere packing problems are of interest in studying structure of crystals [2], liquids [3], glassy materials [4], catalysts and fuel elements [5-8], etc. The packing density of equal spheres is a significant factor in the study of processes in thermal heat exchangers, chemical and nuclear reactors [9]. Well-known are applications of packing spheres in medicine and engineering.

The aim of the paper is to develop a method of dense random packing a large number of equal spheres into a composed container being a geometric model of a chemical reactor.

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

2 Problem Statement

Let there be spheres S_i , $i \in I_N = \{1, 2, \dots, N\}$ with radius r and a container C which is a specific composition of cylinders C_1 , C_2 and a spherical segment S_0 (Fig. 1):

$$C = (C_1 \cup S_0) \setminus \text{int } C_2$$

where $c(C_2)$ is the compliment of C_2 , $\text{int}(\cdot)$ is the interior of the set,

$$C_1 = \{(x, y, z) \in \mathbb{R}^3 \mid x^2 + y^2 \leq R^2, 0 \leq z \leq H\},$$

$$C_2 = \{(x, y, z) \in \mathbb{R}^3 \mid x^2 + y^2 \leq r_c^2, -R \leq z \leq -R + h\},$$

$$S_0 = \{(x, y, z) \in \mathbb{R}^3 \mid x^2 + y^2 + z^2 \leq R^2, -R \leq z \leq \min\{0, H\}\},$$

$R > 0$ is radius of S_0 and of base of C_1 , H is height of C_1 , $h > 0$ is height of C_2 and $r_c > 0$ is radius of its base, $-R + h < H$.

The bottom of C is of spherocylindrical shape and the top of C is bounded above by the plane $z = H$. In general, H can be negative. In this case $C_1 = \emptyset$.

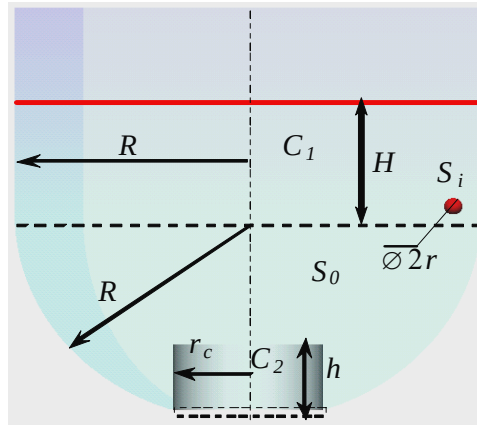


Fig. 1. Container and a sphere (catalyst) to be packed

The problem is formulated as follows: pack the maximal number n^* of spheres from the set S_i , $i \in I_N$ into the container C without overlappings and calculate their center coordinates $v_i^* = (x_i^*, y_i^*, z_i^*)$, $i = 1, 2, \dots, n^*$.

3 Literature review

A review of numerical simulation methods to solve sphere packing problems is presented in [8]. A major part of investigations make use of the Monte-Carlo method [9,10], Discrete Element Models [11] and sequential addition techniques [6,12,13]. The papers consider as a rule cylindrical or rectangular containers.

Mueller [13] applies a sequential addition technique to pack equal spheres into a right circular cylinder and develops the packing algorithm based on a dimensionless packing parameter. The algorithm yields a layerwise structure of the desired number of spheres. The spheres are packed in stable positions under gravity with the maximum dimensionless packing parameter.

Catalytic reactors (packed bed reactors) are used in chemical industry [14]. Reactors with a single adiabatic bed are traditionally used in either exothermic or endothermic reactions. Catalyst pellets have the shape of sphere or cylinder. Packed beds of catalysts should avoid unstable and inefficient arrangements. The packing density of catalysts influences on chemical reactions.

The concept of phi-functions and quasi phi-functions is known to be an efficient tool for mathematical modeling of 3D packing problems for geometrical objects [15-17]. Powerful tools to solve problems of packing geometric objects of various spatial shapes are proposed in [18-24].

Paper [19] proposes a method of solving the sphere packing problem based on homothetic transformations of spheres. The radii of the spheres are assumed to be variable and the auxiliary problems are solved, the sum of the sphere volumes being maximized. The optimization process continues until the sphere radii reach their original values. The method allows to jump from one local extremum to another.

In this paper we consider a method to obtain a dense random packing of a large number of equal spheres into a composed container being a geometric model of the catalytic reactor with a single adiabatic bed. The Stoyan's phi-function method [15,16] is applied to describe packing constraints. Solving the problem is reduced to solving a sequence of nonlinear programming problems making use of the block coordinate descent method [25] and analysis of Lagrange multipliers [26]. We realize sequential addition of spheres. Groups of variables are formed by center coordinates of spheres to be packed. A local minimum point is calculated for each group of variables.

4 Mathematical model

A mathematical model of the problem is presented as follows:

$$n^* = \max \sum_{i=1}^N \psi_i(v_i) \text{ s.t. } v \in G \quad (1)$$

where $v = (v_1, v_2, \dots, v_N) \in \mathbb{R}^{3N}$, $v_i = (x_i, y_i, z_i) \in \mathbb{R}^3$, $i \in I_N$;

$$\psi_i(v_i) = \begin{cases} 1 & \text{if } \Phi_i(v_i) \geq 0, \\ 0 & \text{otherwise;} \end{cases}$$

$$G = \{v \in \mathbb{R}^{3N} : \Phi_{ij}(v_i, v_j) \geq 0, i, j \in I_N, i < j\};$$

$$\Phi_{ij}(v_i, v_j) = (x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2 - 4r^2;$$

$\Phi_i(v_i)$ is a phi-function of S_i and $c(\text{int } C)$, $c(\circ)$ and $\text{int}(\circ)$ are the compliment and the interior of the set $(\circ)$ respectively; $\Phi_{ij}(v_i, v_j)$ is a phi-function of S_i and S_j , $i, j \in I_N, i < j$.

The inequality $\Phi_i(v_i) \geq 0$ provides the sphere S_i , $i \in I_N$, being within the container and the inequality $\Phi_{ij}(v_i, v_j) \geq 0$ specifies non-overlapping the spheres S_i and S_j , $i, j \in I_N, i < j$.

The phi-function $\Phi_i(v_i)$ can be constructed according to [15]:

$$\Phi_i(0, v_i) = \min\{\Phi_{1i}(v_i), \Phi_{2i}(v_i)\},$$

where

$$\Phi_{1i}(v_i) = \max\{\alpha_{1i}(v_i), \alpha_{2i}(v_i)\}, \quad \Phi_{2i}(v_i) = \max\{\beta_{1i}(v_i), \beta_{2i}(v_i)\},$$

$$\alpha_{1i}(v_i) = \max\{\chi_{1i}(v_i), \chi_{2i}(v_i), \varphi_i(v_i)\},$$

$$\chi_{1i}(v_i) = H - z_i, \quad \chi_{2i}(v_i) = z_i, \quad \varphi_i(v_i) = R - r - \sqrt{x_i^2 + y_i^2},$$

$$\beta_{1i}(v_i) = \sqrt{\left(\sqrt{x_i^2 + y_i^2} - r_c\right)^2 + (z_i + R - h)^2} - r,$$

$$\beta_{2i}(v_i) = \sqrt{\left(\sqrt{x_i^2 + y_i^2} - r_c\right)^2 + (z_i + R)^2} - r.$$

The objective function $\psi_i(v_i)$ is piecewise constant. The problem (1) is multiextremum, all local maxima being non-strict due to the axial symmetry. The number of non-linear inequalities is $n(n+1)/2$.

5 Solution Method

The number of spheres that can be packed into the container C is not defined in advance. It can be only evaluated making use of the maximum packing ratio $\pi/(3\sqrt{2}) \approx 0.74$ [1] as an upper bound. If the number of spheres is up to 100, one can adopt the method for packing spheres into a cylinder proposed in [19]. However, if the dimension of the problem is large, then the sequential addition scheme is most suitable to obtain a dense sphere packing.

According to the scheme solving the problem (1) is reduced to solving a sequence of problems. On the first step the sphere S_1 from the set S_i , $i \in I_N$ is packed into C being located at a position with the minimum value of z-coordinate. Then, the sphere S_2 , $i \in I_N$ is packed into C , the sphere S_1 being immovable, and so on. The number of the last sphere S_{n^*} from the set S_i , $i \in I_N$ which can be packed into C gives an evaluation of the maximum n^* of the problem (1). The sequential addition scheme can be realized by applying the block coordinate descent method [25].

Let the spheres $S_1, S_2, \dots, S_{k-1}$ be packed at the previous $k-1$ steps, the center coordinates being $v_k^* = (x_k^*, y_k^*, z_k^*)$, $i = 1, 2, \dots, k-1$ respectively. Coordinates of the sphere S_k form a group of variables at the step k :

$$v_k^* = \arg \min \kappa_k(v_k), \text{ s.t. } v^k \in G^k \subset \mathbb{R}^3, k = 1, 2, \dots, N, \quad (2)$$

where $\kappa_k(v_k) = z_k$;

$$G_k = \{v_k \in \mathbb{R}^3 : \Phi_k(v_k) \geq 0, \Phi_{jk}(v_j^*, v_k) \geq 0, j = 1, 2, \dots, k-1\}. \quad (3)$$

We search for a local minimum point of the problem (2)–(3). If a feasible point of the problem (2)–(3) for $k = n^* + 1$ cannot be found, then the solution of the previous problem ($k = n^*$) solved is a solution $N = n_0$ of the problem (1).

We indicate some peculiarities of the problem (2)–(3): the objective function $\kappa_k(v_k) = z_k$ is linear and minima are reached at extreme points of the feasible region G_k .

The frontier of the feasible region G_k ($fr(G_k)$) consists of points satisfying the equations $\Phi_{jk}(v_j^*, v_k) = 0$, $j = 1, 2, \dots, k-1$, or $\Phi_k(v_k) = 0$:

$$fr(G_k) = \{v = (x, y, z) \in \mathbb{R}^3 : v \in (T \cap G_k)\}, k = 1, 2, \dots, N,$$

where

$$T = T_0 \cup T_1 \cup T_2 \cup \dots \cup T_{k-1},$$

$$T_j = \left\{ v \in \mathbb{R}^3 : \Phi_{jk}(v_j^*, v) = 0 \right\}, \quad j = 1, 2, \dots, k-1,$$

$$T_0 = T_{01} \cup T_{02} \cup T_{03} \cup T_{04} \cup T_{05} \cup T_{06} \text{ (Fig. 2).}$$

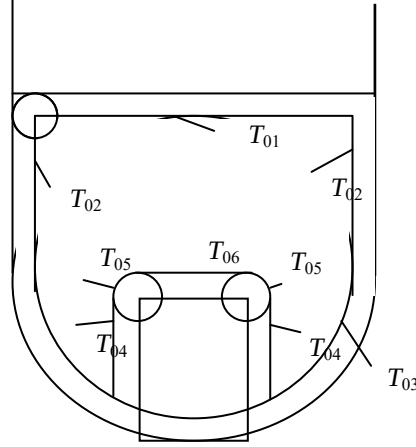


Fig. 2. Illustration of T_0

T_{02} and T_{04} lie on planes:

$$T_{01} = \begin{cases} \left\{ v \in \mathbb{R}^3 : z = H - r, x^2 + y^2 \leq (R - r)^2 \right\} & \text{if } H \geq 0, \\ \left\{ v \in \mathbb{R}^3 : z = H - r, x^2 + y^2 \leq (R - r)^2 - H^2 \right\} & \text{otherwise,} \end{cases}$$

$$T_{06} = \begin{cases} \emptyset & \text{if } H < -R + h + r, \\ \left\{ v \in \mathbb{R}^3 : z = -R + h + r, x^2 + y^2 \leq (r_c)^2 \right\} & \text{otherwise.} \end{cases}$$

T_{02} and T_{04} lie on cylindrical surfaces:

$$T_{02} = \begin{cases} \left\{ v \in \mathbb{R}^3 : x^2 + y^2 = (R - r)^2, 0 \leq z \leq H - r \right\} & \text{if } H > 0, \\ \emptyset & \text{otherwise,} \end{cases}$$

$$T_{03} = \left\{ v \in \mathbb{R}^3 : x^2 + y^2 = (r_c + r)^2, \sqrt{(R - r)^2 - (r_c + r)^2} \leq z \leq \min\{(-R + h), (H - r)\} \right\}.$$

T_{03} lies on spherical surface

$$T_{03} = \left\{ v \in \mathbb{R}^3 : x^2 + y^2 + z^2 = (R-r)^2, \right. \\ \left. \sqrt{(R-r)^2 - (r_c + r)^2} \leq z \leq \min\{0, (H-r)\} \right\}.$$

T_{05} lies on torus surface

$$T_{05} = \begin{cases} \emptyset & \text{if } H \leq -R+h, \\ \left\{ v \in \mathbb{R}^3 : \left(\sqrt{x^2 + y^2} - r_c \right)^2 + (z + R-h)^2 = r^2, \right. \\ \quad \left. x^2 + y^2 \geq r_c^2, -R+h \leq z \leq \min\{(-R+h+r), (H-r)\} \right\} & \text{otherwise.} \end{cases}$$

The set of extreme points can be presented as

$$E_k = T_{02} \cup T_{03} \cup \bigcup_{\substack{i,j,l=0 \\ i < j < l}}^{k-1} (T_i \cap T_j \cap T_l) \cap frG_k, \quad k = 1, 2, \dots, N.$$

The set of local minima L_k , $k = 1, 2, \dots, N$, consists of points of $(fr(G_k))$ whose coordinate values are solutions either of the system:

$$\begin{cases} x^2 + y^2 + z^2 = (R-r)^2, \\ x^2 + y^2 = (r_c + r)^2 \end{cases}$$

or the systems:

$$\begin{cases} f_i = 0, \\ f_l = 0, \quad i, l, m = 0, 1, \dots, k-1, \quad i < l < m, \quad (i, l, m) \in I. \\ f_m = 0, \end{cases}$$

where f_i, f_l, f_m are either $\Phi_k(v_k)$ or $\Phi_{jk}(v_j^*, v_k)$, $j = 1, 2, \dots, k-1$; the set I consists of triples of indices for which the condition of minimum [26] is fulfilled.

A sufficient condition of local minimum of the problem (2)–(3) is $\lambda_1 > 0$, $\lambda_2 > 0$, $\lambda_3 > 0$ [26] where the Lagrange multipliers $\lambda_1, \lambda_2, \lambda_3$ can be calculated by solving the following system of linear equations:

$$\begin{pmatrix} \frac{\partial f_i}{\partial x_k} & \frac{\partial f_l}{\partial x_k} & \frac{\partial f_m}{\partial x_k} \\ \frac{\partial f_i}{\partial y_k} & \frac{\partial f_l}{\partial y_k} & \frac{\partial f_m}{\partial y_k} \\ \frac{\partial f_i}{\partial z_k} & \frac{\partial f_l}{\partial z_k} & \frac{\partial f_m}{\partial z_k} \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

Thus, set

$$L_k = (T_{03} \cap T_{04}) \bigcup_{\substack{i,l,m=0 \\ i < l < m \\ (i,l,m) \in I}}^{k-1} (T_i \cap T_l \cap T_m) \cap frG_k.$$

The number of local minima not belonging to the set is less than $2n(n-1)(n-2)/6$.

The solution strategy of the problems (2)–(3) consists of the following stages.

1. Random choice of an initial point of the feasible region G_k .
2. Movement to the frontier of G_k decreasing the objective.
3. Movement on the frontier of G_k to an extreme point of G_k decreasing the objective.
4. Searching for an extreme point of G_k being a local minimum point of the problem (2)–(3).

6 Searching for a local minimum of subproblem

We consider a procedure to calculate a local minimum point of the problem (2)–(3).

1. A point $v_k^0 = (x_k^0, y_k^0, z_k^0)$ such that $(x_k^0)^2 + (y_k^0)^2 \leq R^2$ and $z_k^0 = H - r_k$ is constructed.
2. If $v_k^0 \notin G_k$, then $n^* = k - 1$ is an approximate solution of the problem (1).
3. Otherwise ($v_k^0 \in G_k$), a point $v_k^1 = (x_k^1, y_k^1, z_k^1) = (x_k^0, y_k^0, z_k^0 - \Delta z_q) \in frG_k$ at which S_k touches frC or S_p , $p \in \{1, 2, \dots, k-1\}$. The value Δz_q , $q = 1, 2, \dots$, is defined by bisection within the segment $[0, R + H]$.
4. If point v_k^1 is a local minimum point, then set $v_k^* = v_k^1$ and go to the next k .
5. Otherwise, let point $v_k^1 \in T_i$, $i \in \{0, 1, \dots, k-1\}$. A point $v_k^2 = (x_k^2, y_k^2, z_k^2) \in (T_i \cap T_j)$ is calculated where $i \neq j \in \{0, 1, \dots, k-1\}$ and $z_k^2 < z_k^1$ such that S_k touches frC

and S_j , $j \in \{1, 2, \dots, k-1\}$ or S_i and S_j . To this end we solve the following systems:

$$\begin{cases} f_i(x, y, z) = 0, \\ z = z_k^1 - \Delta z_q, \quad q = 1, 2, \dots \\ ax + by = c, \end{cases}$$

where the equation $f_i(x, y, z) = 0$ describes the frontier of T_i and $ax + by = c$ is an equation of a plane passing through the point $v^{k1} \in T_i$. The value Δz_q is defined by bisection within the segment $[0, R + H + z_k^1]$. If v^{k2} is a local minimum point, then set $v_k^* = v_k^2$ and go to the next k .

6. An extreme point $v_k^3 = (x_k^3, y_k^3, z_k^3) \in (T_i \cap T_j \cap T_l)$ is calculated where $i \neq j \neq l \in \{0, 1, \dots, k-1\}$ and $z_k^3 < z_k^2$ such that S_k touches frC and two spheres or three spheres. To this end we solve the following systems:

$$\begin{cases} f_i(x, y, z) = 0, \\ f_j(x, y, z) = 0, \\ z = z_k^2 - \Delta z_q, \quad q = 1, 2, \dots, \end{cases}$$

where the equations $f_i(x, y, z) = 0$ and $f_j(x, y, z) = 0$ describe the frontiers of T_i and T_j respectively. The value Δz_q is defined by bisection within the segment $[0, R + H + z_k^2]$. If v_k^3 is a local minimum point, then set $v_k^* = v_k^3$ and go to the step $k+1$.

7. Otherwise ($\lambda_1 < 0$ or $\lambda_2 < 0$ or $\lambda_3 < 0$) define $i_0 \in \{1, 2, 3\}$ such that $\lambda_{i_0} \leq \lambda_1$, $\lambda_{i_0} \leq \lambda_2$, $\lambda_{i_0} \leq \lambda_3$.
8. If $i = i_0$, take $i = j$, $j = l$ and $v_k^2 = v_k^3$ and go to the item 6.
9. If $j = i_0$, take $j = l$ and $v_k^2 = v_k^3$ and go to the item 6.
10. If $l = i_0$, take $v_k^2 = v_k^3$ and go to the item 6.

A special strategy of verifying the feasibility during optimization process is proposed. The strategy forms a list of inactive constraints analyzing distances from the sphere to be packed to the spheres already packed. This allows to essentially decrease the computational complexity of the algorithm.

To improve the value of the objective of the problem (2)–(3) we realize the multistart method on each step and choose the best local minimum.

7 Experimental results and discussion

In order to numerically test the packing method proposed a software is developed. Several examples with up to 2 million spheres are calculated. The Open Graphics Library for visualization of the optimization process.

Example 1. The metric characteristics of the container and the spheres to be packed are $R = 250$, $r_c = 80$, $H = 0$, $h = 250$ and $r = 15$. We use 30 starting points for each sphere. The number of spheres packed is $n^* = 1017$ (Fig. 3). The runtime is 1 sec.

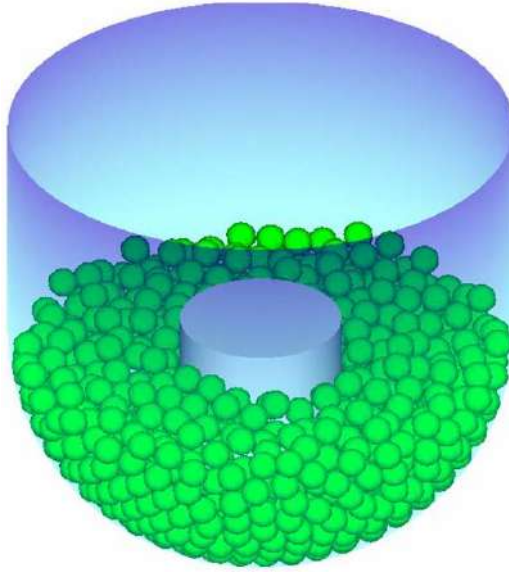


Fig. 3. Illustration of Example 1

Example 2. The metric characteristics are $R = 250$, $r_c = 80$, $H = -120$, $h = 80$ and $r = 5$. 30 starting points are chosen for each sphere. The number of spheres packed is $n^* = 9696$ (Fig. 4). The runtime is 20 sec.

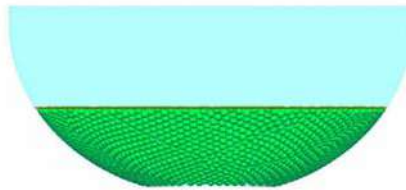


Fig. 4. Illustration of Example 2

Example 3. The metric characteristics are $R = 250$, $r_c = 50$, $H = 0$, $h = 270$ and $r = 2$. 30 starting points are chosen for each sphere. The number of spheres packed is $n^* = 539778$ (Fig. 5). The runtime is 1 h 20 min.

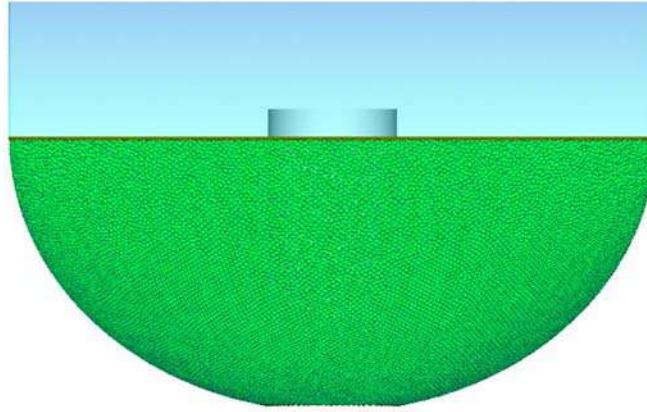


Fig. 5. Illustration of Example 3

Example 4. The metric characteristics are $R = 250$, $r_c = 80$, $H = -10$, $h = 80$ and $r = 1.25$. One local minimum is calculated for each sphere. The number of spheres packed is $n^* = 2\,063\,007$. The runtime is about 2 h.

The large-dimension problem reduces to solving a sequence of subproblems by means of the block coordinate descent method. The multistart method allows to choose a better local position for each sphere, thus increasing the total packing density.

As experimental results show (see Fig.4 and Fig.5) the spheres form regular structures along the packing domain frontier (the container wall).

It is obvious that the packings obtained in the examples do not, in general, correspond to local minimum points of the problem minimizing the height of the filled part of the container when all the spheres are moveable.

8 Conclusion

The problem of packing equal sphere into a bounded container is considered in the paper.

The method suggested allow to solve the problem of large dimension due to applying the block coordinate descent method which allow to reduce solving the problem to a sequence of non-linear programming problems.

The optimization method realizes search of a frontier point of the feasible region, an extreme point and movement to a local minimum point. Choice of descent direction is defined by analyzing Lagrange multipliers of active constraints.

Numerical examples show effectiveness of the method for up to two millions equal spheres.

The Stoyan's phi-function method allows to adopt the developed method to pack spheres into containers of arbitrary spatial shapes.

The method can be also extended to pack unequal spheres.

References

1. Gauss, C. F.: Untersuchungen über die Eigenschaften der positiven ternären quadratischen Formen von Ludwig August Seber, Göttingische gelehrte Anzeigen (1831)
2. Frenkel, D.: Computer simulation of hard-core models for liquid crystals. *Computer Physics Communications*, vol. 44(3), pp. 243–253 (1987) doi: 10.1016/0010-4655(87)90079-8
3. Bernal, J.A.: Geometrical Approach to the Structure Of Liquids. *Nature* 183, pp. 141–147 (1959) doi: 10.1038/183141a0
4. German, R.M.: Particle packing characteristics. Metal Powder Industries Federation, NJ (1989)
5. Gan, Yi., Kamlah, M., Reimann, J.: Computer Simulation of Packing Structure in Pebble Beds. *Fusion Eng. Des.* 85(10-12), pp. 1782–1787 (2010) doi: 10.1016/j.fusengdes.2010.05.042
6. Mueller, G.E.: Radial porosity in packed beds of spheres. *Powder Technol.* 203, pp. 626–633 (2010) doi: 10.1016/j.powtec.2010.07.007
7. Chigada, P.I.: Characterisation of flow and heat transfer patterns in low aspectratio packed beds by a 3D network-of-voids model. *Chemical Engineering Research and Design* 89, pp. 230–238 (2011) doi: 10.1016/j.cherd.2010.05.002
8. Burtseva, L., Valdez Salas, B., Werner, F., Petranovskii, V.: Packing of monosized spheres in a cylindrical container: models and approaches. *Rev. Mex. Fis.*, vol. 61(1) pp. 20–27 (2015)
9. Abreu, C. R. A., Macias-Salinas, R., Tavares, F. W., Castier, M.: A Monte Carlo simulation of the packing and segregation of spheres in cylinders. *Brazilian Journal of Chemical Engineering*, vol. 16(4), pp. 395–405 (1999) doi: 10.1590/S0104-66321999000400008
10. Roozbahani, M.M., Huat, B.B.K., Asadi, A.: Effect of rectangular container's sides on porosity for equal-sized sphere packing. *Powder technology* 224, pp. 46-50 (2012) doi: 10.1016/j.powtec.2012.02.018
11. Wensrich, C.M.: Boundary structure in dense random packing of monosize spherical particles. *Powder Technology*, vol. 219, pp. 118–127 (2012) doi: 10.1016/j.powtec.2011.12.026
12. Magnico, P.: Hydrodynamic and transport properties of packed beds in small tube-to-sphere diameter ratio: pore scale simulation using an Eulerian and a Lagrangian approach. *Chemical Engineering Science*, vol. 58(22), pp. 5005–5024 (2003) doi: 10.1016/S0009-2509(03)00282-3
13. Mueller, G.E.: Numerically packing spheres in cylinders. *Powder Technology* 159, pp. 105–110 (2005) doi: 10.1016/j.powtec.2005.06.002
14. Major, A.: Modeling of a Catalytic Packed Bed Reactor and Gas Chromatograph Using COMSOL Multiphysics. Corpus ID: 18618192 (2009) <https://www.semanticscholar.org/author/A.-Major/50519443>
15. Stoyan, Y., Scheithauer, G., Romanova, T.: Mathematical modeling of interactions of primary 3D geometric objects. *Cybern Syst Anal* 41, pp. 332-342 (2005) doi: 10.1007/s10559-005-0067-y

16. Stoyan, Y., Scheithauer, G., Gil, M., Romanova, T.: Φ -function for complex 2D objects. 4OR 2, pp.69-84 (2004) doi: 10.1007/s10288-003-0027-1
17. Stoyan, Y.G., Semkin, V.V., Chugay, A.M.: Modeling Close Packing of 3D Objects. Cybern Syst Anal 52, pp. 296-304 (2016) doi: 10.1007/s10559-016-9826-1
18. Stoyan, Y.G., Chugay, A.M.: Packing different cuboids with rotations and spheres into a cuboid. Advances in Decision Sciences, vol. 2014 (2014) doi: 10.1155/2014/571743
19. Stoyan, Y.G., Scheithauer, G., Yaskov, G.N.: Packing Unequal Spheres into Various Containers. Cybern Syst Anal 52, pp. 419–426 (2016) doi: 10.1007/s10559-016-9842-1
20. Yaskov G., Romanova T., Litvinchev I., Shekhovtsov S.: Optimal Packing Problems: From Knapsack Problem to Open Dimension Problem. In: Vasant P., Zelinka I., Weber GW. (eds) Intelligent Computing and Optimization. ICO 2019. Advances in Intelligent Systems and Computing, vol 1072. Springer, Cham, pp. 671–678 (2020) doi: 10.1007/978-3-030-33585-4_65
21. Yakovlev, S.V.: The Method of Artificial Space Dilation in Problems of Optimal Packing of Geometric Objects. Cybern Syst Anal 53, pp. 725–731 (2017) doi: 10.1007/s10559-017-9974-y
22. Yakovlev, S., Kartashov, O., Korobchynskiy, K., Skripka, B.: Numerical Results of Variable Radii Method in the Unequal Circles Packing Problem. 2019 IEEE 15th International Conference on the Experience of Designing and Application of CAD Systems (CADSM), Polyana, Ukraine, pp. 1–4 (2019) doi: 10.1109/CADSM.2019.8779288
23. Yakovlev, S.: Configuration Spaces of Geometric Objects with Their Applications in Packing, Layout and Covering Problems. In: Lytvynenko V., Babichev S., Wójcik W., Vynokurova O., Vyshemyrskaya S., Radetskaya S. (eds) Lecture Notes in Computational Intelligence and Decision Making. ISDMCI 2019. Advances in Intelligent Systems and Computing, vol 1020, Springer, Cham, pp. 122-132 (2020) doi: 10.1007/978-3-030-26474-1_9
24. Yaskov, G.: Methodology to solve multi-dimensional sphere packing problems. Journal of Mechanical engineering, vol. 22 (1), pp. 67-75 (2019) doi: 10.15407/pmach2019.01.067
25. Bertsekas, D.P.: Nonlinear Programming, 2nd ed., Athena Scientific, Belmont, MA (1999)
26. Gill, P.E., Murray, W., Wright, M.H.: Practical Optimization, Academic Press (1981)

Bi-Soft Open Sphere Topology Model of Configuration Space for Reactive Joint Motion Planning of Unmanned Vehicles

Volodymyr Sherstjuk^{2[0000-0002-9096-2582]}, Maryna Zharikova^{1[0000-0001-6144-480X]}, and
Ruslan Levkivskiy^{3[0000-0001-9280-8098]}

¹Kherson National Technical University, Berislav Road, 24, Kherson, 73008, Ukraine
¹vgsherstyuk@gmail.com, ²marina.jarikova@gmail.com,
³Kherson State Maritime Academy, Ushakova Str., 20, Kherson, 73000, Ukraine
levka.ru55555@gmail.com

Abstract. This work presents a model of configuration space for the reactive joint motion planning of unmanned vehicles' ensemble. This model is based on a dynamic bi-soft topology, which allows describing the safety conditions of the joint motion of a multitude of vehicles. The approximation space of the topology is constructed in the system of nested open spheres based on an angular coordinate system. The spheres are discretized onto sector-like cells. The proposed bi-soft topology is based on a two-level breakdown of the configuration space. On the first level, the configuration space is broken down onto "free to move" and obstacle subspaces taking into account uncertainties of observations. On the second level, the obtained soft subspaces are broken down onto safety zones with blurred boundaries. The offered topological model can be used in a hybrid motion planner combining potential field and random sampling methods. The use of the proposed bi-soft topology allows reducing the computational complexity of the motion planning task using partitioning of the configuration space and path search heuristics based on the calculation of the cell volume.

Keywords: soft set, soft topology, open sphere, unmanned vehicles, reactive path planning, collision avoidance, safety zones

1 Introduction

The great progress of the latest years in the development of unmanned vehicles (UV) has led to their use in large teams to address a sufficient number of important real-time tasks. UVs belonging to the large teams can be characterized by different sizes, functionalities, roles, motion parameters, and ever environments. For example, forest fire-fighting [1] has recently become an important domain to use the large teams of UVs where unmanned aerial vehicles of various types can provide fire detection, localization, and monitoring missions, while a variety of unmanned ground vehicles (e.g. bulldozers, excavators, tanks, etc.) can be simultaneously involved in fire extinguishing missions. Another example is a smart fishery task, where unmanned aerial

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

vehicles can be involved in searching for fishing flocks, unmanned underwater vehicles can identify fish in flocks, direct and accompany them, as well as unmanned boats can carry fishing gear.

The authors are concerned with the teams of UVs of various types providing different roles and functions while executing their missions jointly and simultaneously to achieve the common objective. Such a team is often called a heterogeneous ensemble due to differences in vehicles' features and their roles in the team [2]. Unlike the widely used concepts of swarms, which bring together a significant number of homogeneous UVs performing the same role, the concept of the ensemble has a crisp structure represented by its spatial and functional configurations taking into account different roles of involved UVs. Regarding the given role, each UV executes a certain scenario of the mission together with other members of the ensemble complementing each other.

However, the missions usually involve a certain number of UVs in the joint simultaneous movement within a confined space, which is open to any other moving objects (manned or unmanned), which may disturb not only the movement of the certain UVs but also the execution of the ensemble mission at all. For example, in the smart fishery mission, the invasion of third-party moving objects (MOs) into the area of mission can require changes in the trajectories of some UVs, which is not an easy task because of the limitations of safe motion, communication distance, fishing gear cover, etc. Furthermore, the change of the trajectories of some members of the fishery ensemble due to safety reasons entails the changes of the trajectories of other members of the ensemble due to requirements on keeping the spatial configuration.

The presented considerations pose the problem of dynamic control of the joint motion of the ensemble members. Thus, using the "sense and avoid" techniques [3], UVs must move to the given target avoiding obstacles and collisions but keeping the predetermined spatial configuration. The more complicated the structure and joint mission of the ensemble, the more difficult the motion control of UVs.

The problem addressed in this paper relates to the coordinated joint motion of heterogeneous UV ensembles. Further, we will consider any obstacles as static objects and any vehicles as moving objects (MO) irrespective of the environment.

2 Literature analysis

The problem of dynamic control of the joint motion of UVs is the subject of sufficient interest of many researchers investigated and reflected in numerous publications. In general, features of the joint ensemble motion differ from the other known formation structures such as swarms, flocks, leader-follower structures, and virtual structures [4] where the entire formation is always considered as a single virtual body [5].

Let us consider joint UV motion within a space C . Executing a given scenario, each UV A_i moves along a pre-planned path P_i represented as a sequence of waypoints (WP) or as a sequence of pairs "timepoint-waypoint" ($TP-WP$), which define desired spatial locations at the given time moments described by TP s. Thus, the path P_i can be almost always represented as a time-arranged sequence

$P_i = [(TP_{i1}, WP_{i1}), \dots (TP_{ij}, WP_{ij}), \dots (TP_{in}, WP_{in})]$, $TP_{i1} < TP_{ij} < TP_{in}$, which can be conveniently used to control the A_i motion.

In the literature, the most of the suitable approaches to build the path are inspired from global path planning methods, the systematic review of which is presented in [6]: the roadmap approach based on the visibility graph algorithm or Voronoi diagram algorithm, cell decomposition approach, potential field approach, and random sampling approach including probabilistic roadmap method (PRM) and Rapidly Exploring Random Trees (RRT). The last two methods are used most often.

Recent motion planning developments focused on environments that change over time or are not quite known are more tangible to the problem of the joint UV motion. When UV moves, it is exposed to a significant number of both dynamic (e.g., wind) and situational (MOs that disrupting its trajectory) perturbations. When such a perturbation is observed, UV must react and maneuver to avoid the collision. At the same time, it should be worried about keeping both the given spatial configuration (i.e., hold its relative position during the mission execution) and the safe distance from other MOs including the members of the ensemble and various obstacles. Therefore, the initial path P_i can be changed by the displacement of certain WPs forcing other ensemble members to adjust their motion trajectories.

Thus, we need to update the motion paths to adapt them to dynamic changes in the environments. It can be reduced to the dynamic path planning task [7], which can be solved by re-planning some fragments of the path or adjusting its parameters (in our case – locations of WPs). It allows us to obtain a new path ahead of time.

Taking into account the restricted computing capabilities of UVs' on-board computers as well as a lack of time to avoid collisions and obstacles, we should use quick enough methods to re-plan the path P_i . If UV needs to be capable of planning the path of its motion in time to react to dynamic environments as well as situational perturbations, we should provide reactive (real-time) planning immediately during the motion [7]. Thus, the real-time motion planning algorithm could not be iterative.

The potential field method considers obstacles as repulsive fields while targets as attractive fields, so the UV motion is guided to the attractive points avoiding the repulsive points. However, due to the use of iterative optimization techniques this method is so computationally intensive that it is not suitable for reactive collision avoidance in general [7].

Random sampling has emerged as a powerful tool for path planning in high-dimensional configuration spaces, its algorithms are both efficient and simple to implement. However, multi-query planning algorithms (PRM, DRM) have important issues such as a straightforward uniform sampling distribution and the implicit representation of space that need a pre-computation [8]. In contrast to multi-query planning, there is no pre-computation in the single-query algorithms like RRT, which can construct small roadmaps on the fly. However, this algorithm is prone to oversampling while the sampling distribution should be defined explicitly. Unfortunately, random sampling algorithms are not complete in the sense that they cannot consider such a situation when no existing paths are found [6]. Due to the use of lookup tables saving motion primitives, such algorithms are infeasible for resource-limited UVs [9].

To overcome the above-mentioned disadvantages, it is advisable to develop a hybrid method combining RRT and potential field complementing each other. To level the sampling issues, we should develop a model of the configuration space that can reduce the computational complexity by abandoning iterative calculations. A key aspect to combine the RRT and potential field methods is to build an approximate topological model endowed with metric properties to avoid intensive calculations. Well-known geometric approaches based on relative spatial estimates (so-called "point of collision") such as "cones", "safety domains", "closest points to approach", etc. [9-12] can be used to do this.

This work aims to develop the approximate model of the configuration space in the context of the control of the joint motion of unmanned vehicles within confined areas. This model is needed for subsequent implementation of a hybrid reactive path-planning method combined with both RRT and potential field methods. To overcome the computational complexity issue, we must develop a topological model using soft discretization. The developed model allows analyzing big data streams coming from remote sensors and representing them in a user-friendly style.

3 Formal problem statement

Let configuration be a set of k parameters that determine the location of the UV in the space of the joint motion (and, possibly, its motion parameters) uniquely. Each UV can be described by a moving point in a k -dimensional space C called the configuration space. A configuration q is a single point within C . A certain configuration q is free if the UV located at q does not collide with any obstacles or MOs.

A "free to move" subspace $\mathcal{F}$ is a subset of all free configurations in C , while an obstacle subspace $\mathcal{B}$ is a complement of $\mathcal{F}$ to C : $\mathcal{B} = C \setminus \mathcal{F}$ [6].

Planning is always performed within the configuration space C .

The motion planning problem concerning the UV A_i can be defined as finding a path P_i from a start configuration q_s to a target configuration q_t such that P_i lies entirely within the free subspace $\mathcal{F}$. A path P_i is defined by a continuous sequence of configurations [7]. A path planning algorithm is complete if it finds a path whenever one exists and reports none exists otherwise [6].

Reactive motion planning at a time t can be defined as changing the path P_i from a current configuration $q(t)$ to the target configuration q_t whenever a change of the configuration space C takes any fragments of the path P_i beyond the free subspace $\mathcal{F}$ (i.e., a fragment of P_i collides with the obstacle subspace $\mathcal{B}$).

An example is shown in Fig. 1, where UV A_i moving to the target configuration q_t along the path P_i with velocity v meets an obstacle U at current configuration $q(t)$. The collision conditions can be described by a collision cone constructed by dropping tangents r_1, r_2 from $q(t)$ to B representing the safety zone around U .

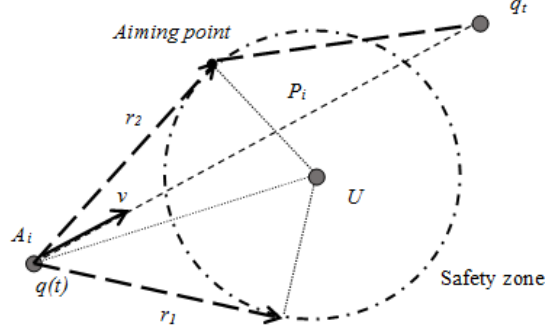


Fig. 1. Reactive motion planning

If the velocity vector v lies within the collision cone, A_i violates B . Decomposition of v in terms of the tangents r_1, r_2 gives us $v = ar_1 + br_2$. If $a > 0$ and $b > 0$, then the obstacle U is critical to A_i . Thus, P_i must be replanned by adding a new waypoint, which should be the nearest aiming point determined on B by tangents r_1, r_2 .

4 Building a basic spatial model

Consider a three-dimensional linear uniform space C . Let Y be a set of certain elements, and T be a set of time points t strictly ordered by $<_t$ having the initial count t_0 . Let us introduce a norm $\|y\|_c = \min_{t \in [0, T)} (y(t))$ within C , where $y \in Y$, $t \in T$, and a corresponding metric $\xi_C(y_1, y_2) = \|y_1 - y_2\|$ such as:

1. $\xi_C(y_1, y_2) = \|y_1 - y_2\| = 0 \Leftrightarrow y_1 = y_2$;
2. $\xi_C(y_1, y_2) = \|y_1 - y_2\| = \|y_2 - y_1\| = \xi_C(y_2, y_1)$;
3. $\xi_C(y_1, y_2) = \xi_C(y_1 + a, y_2 + a)$;
4. $\xi_C(\lambda y_1, \lambda y_2) = \lambda \cdot \xi_C(y_1, y_2)$.

Define the basis e_1, e_2, e_3 within C so that the metric ξ_C remains uniform. Therefore, the decomposition of a vector $v = \alpha_1 e_1 + \alpha_2 e_2 + \alpha_3 e_3$ gives the coordinates $v(\alpha_1, \alpha_2, \alpha_3)$ of a certain point of the space C describing the position of MO. Currently, we have built a basic continuous spatial model based on the Cartesian coordinate system that needs to be sampled to provide sufficient performance.

Using the metric ξ_C , we impose a metric grid of coordinate lines of size $\delta = \Delta\alpha_1 = \Delta\alpha_2 = \Delta\alpha_3$ onto space C with an origin $\alpha_1 = \alpha_2 = \alpha_3 = 0$, so that the coordinate lines form a set D of isometric cubic cells d of size $\delta \times \delta \times \delta$. Thus, we obtain a space

discretized by the grid $D = \{d_{xyz}\}$, in which the cells d_{xyz} are the smallest homogeneous spatial objects, whose coordinates x, y, z correspond to e_1, e_2, e_3 .

Although the basic model we obtained is consistent with the information received from the sensors observing the navigation situation, including the rectangular cell coordinate system, however, to represent the safe motion conditions adequately, this model must be adequately transformed.

5 Building an open sphere topology

Let us choose the position of the UV as an observer and endow it with the properties of origin. Let us construct a sphere $V \subseteq C$ with an infinite radius around this position and define the angular coordinate system, as it is shown in Fig. 2.

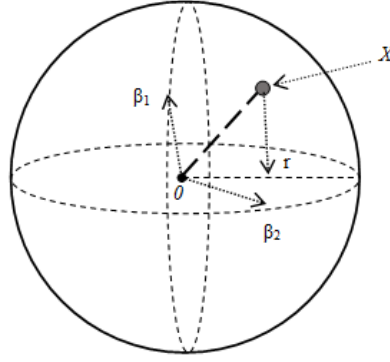


Fig. 2. Angular coordinates on a sphere

Accordingly, the coordinates of UV can be defined as $Crd(A) = (\beta_1, \beta_2, r)$, where β_1 is latitude, β_2 is longitude, and r is a distance from the origin to the UV position.

Let's create a metric ξ_B within the angular coordinate system with the properties similar to ξ_C using an isometric bijection $\chi: \xi_C \rightarrow \xi_B$. Obviously, $\chi: \nu(\alpha_1, \alpha_2, \alpha_3) \rightarrow Crd(\beta_1, \beta_2, r)$, so we can transform the coordinates of the observation space C to the angular coordinates in the constructed sphere V .

At the next stage, we construct a discrete sphere model using metric ξ_B and mapping ϕ , and place an angular grid of coordinate lines with equal angles and equal discrete of the radius $\phi = \Delta\beta_1, \Delta\beta_2, \Delta r$ on the sphere V with the origin point $Crd(0, 0, 0)$, so that the coordinate lines form a set W of cells w (Fig. 3).

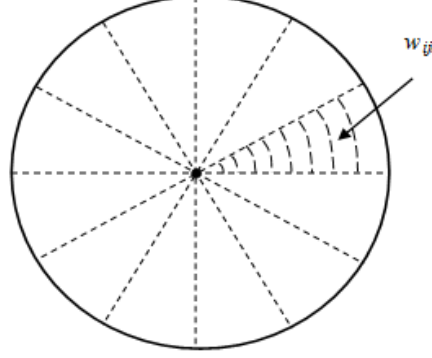


Fig. 3. Discretization of a sphere with infinite radius

Thus, we have got a sampled sphere $W = \{w_{ijk}\}$, whose cells w_{ijk} are the smallest sectors of the sphere V with the angular coordinates i, j, k . The cells w_{ijk} are homogeneous objects concerning their interior.

Obviously, the sampling of a sphere V by a set of cells W constitutes a topological space $\mathcal{T}_V^W = (V, \text{Def}(W))$ if the cells of each pair (w_l, w_m) are internally homogeneous and disjoint, that is $\forall l, m \quad w_l \cap w_m = \emptyset$, and their union completely covers the sphere V , so that $V = \cup_{w_l \in W} w_l$. The resulting topology $\text{Def}(W)$ is nonlinear since the further from the center of the sphere, the larger the volume of each subsequent cell.

6 Building the first level of soft open sphere topology

Let $Y = \{y_i\}_{i=0}^k$ be a set of $k+1$ possible states of the cell $w_{ijk} \in W$. For example, the state y_0 corresponds to an “empty” cell containing no objects, the state y_1 corresponds to a cell containing a certain obstacle, the state y_2 corresponds to a cell that is a target for the UV motion, the state y_3 corresponds to a cell containing UV from its ensemble, and the state y_4 corresponds to a cell containing MO considered as “intruder” because it does not belong to the ensemble.

Thus, a subset of states $\{y_1, y_2, y_3, y_4\} \in Y$ attributes the appropriate cells to the category of “occupied”, as well as $y_0 \in Y$ attributes the cell to the category of “free to move”. Correspondingly, the set of cells having the state y_0 constitutes the sampling of “free to move” subspace $\mathcal{F}$ of configuration space C , while the set of cells having the state $\{y_1, y_2, y_3, y_4\}$ constitutes the sampling of the obstacle subspace $\mathcal{B}$ of configuration space C .

Furthermore, the set of cells having the state y_2 corresponds to an attractive manifold in terms of potential fields while the set of cells having the state $\{y_1, y_3, y_4\}$ corre-

sponds to a repulsive manifold respectively. Since the state of a cell depends on time, each above-mentioned is dynamic, consequently, the partitioning of configuration space C onto $\mathcal{F}$ and $\mathcal{B}$ subspaces is also dynamic.

Assume that the set W is a universe and the set Y is a set of parameters. A pair (Y, Y) constitutes a soft set of cells if Y is a certain mapping Y into the set of all subsets of the set W , $Y: y_i \rightarrow 2^W$ [14]. In other words, the soft set is a parameterized family of subsets of the set of cells W . A certain set (Y, y_i) , $y_i \in Y$ from such family can be regarded as the set of y_i -approximated elements of the soft set [15], or, in other words, as an y_i -element of the soft set denoted by Y_i .

Thus, the universe W can be partitioned by the soft set (Y, Y) that is the union of all k of its y_i -elements, where $Y = \bigcup \{Y_i\}_{i=1}^k$, which constitute a multitude of pairs $Y_i = \{(Y, y_i): y_i \in Y, (Y, y_i) \in 2^W\}$. The soft set is often associated with the set of equivalence classes induced by a certain indiscernibility relation that depends on time [16]. We can define such relation of y_i -indiscernibility dynamically given on the set of cells W by $(\forall y_i \in Y) \mathfrak{R}_W^{y_i}(t) = \{(w_m, w_n) \in W \times W \mid y_i(w_m, t) = y_i(w_n, t)\}$.

Therefore, each y_i -element of the soft set Y_i provides the partitioning of the set of cells W into equivalence classes given by the y_i -indivisibility relation $\mathfrak{R}_W^{y_i}(t)$ at the moment t . In other words, the parameterized family of subsets of the set W , which constitutes the y_i -element of the set Y_i , can be considered as a factor-set $W / \mathfrak{R}_W^{y_i}(t)$ consisting of all equivalence classes of W induced by relation $\mathfrak{R}_W^{y_i}(t)$. Thus, the pair $apr_W = (W, \mathfrak{R}_W^{y_i}(t))$ defines the dynamic approximation space [17].

Let $\emptyset$ be an empty set, W be the universe, and elements $W / \mathfrak{R}_W^{y_i}(t)$ are elementary sets so that the finite union of one or more of elementary sets is a compound set. If the family of all compound sets is $Def(apr_W)$, then the dynamic approximation space uniquely defines the dynamic topological space $\mathcal{T}_W^{\mathfrak{R}_W^{y_i}}(t) = (W, Def(apr_W))$.

It is known that $Def(apr_W)$ is a topology on W if its subsets satisfy the following conditions for any $A, B \in W$ [18]:

1. $\emptyset \in Def(apr_W)$, $A \in Def(apr_W)$;
2. $A, B \in Def(apr_W) \Rightarrow A \cap B \in Def(apr_W)$;
3. $A, B \in Def(apr_W) \Rightarrow A \cup B \in Def(apr_W)$.

Consequently, $Def(apr_W)$ is the family of open sets and $\mathcal{T}_W^{\mathfrak{R}_W^{y_i}}(t) = (W, Def(apr_W))$ is the dynamic topological space. The mapping $Y(t)$ in the given interpretation is also dynamic and uniquely assigns each cell of the universe W to a certain y_i -element of

the soft set $(Y(t), Y)$ at the time moment t . If the results of the observation are uncertain, we can interpret them by “degrees of confidence” that the cells are “free to move” or contain obstacles at the moment of consideration. For this purpose, the function $Y(t)$ can be represented as a fuzzy set, so that $\tilde{Y}(t): y_i \rightarrow [0, 1]$, where the degree of confidence has a range of values in the interval $[0, 1]$. Accordingly, we need to soften the relation $\mathfrak{R}_W^y(t)$ using a tolerance relation $\tilde{\mathfrak{R}}_W^y(t)$ instead of the equivalence one. Thus, we have got a dynamic fuzzy soft set of cells $(\tilde{Y}(t), Y)$ instead of the soft set [19]. Specifying a certain threshold $\tau \in [0, 1]$, we will be able to cut off from the consideration all those cells $w \in W$, for which the degree of confidence at the moment t is lower than a given threshold τ , i.e. $\Upsilon_\tau(y_i, t) = \{w \in W : (\tilde{Y}, y_i)(w) \geq \tau\}$ [20]. Each y_i -element $(\Upsilon_\tau(t), y_i)$ of the corresponding fuzzy soft set consists of only those cells $w \in W$, for which the degree of confidence that their state is $y_i \in Y$, is greater than the threshold τ at the moment t . As a result, we obtain the approximation space $apr_W = (W, \tilde{\mathfrak{R}}_W^y(t))$ and the dynamic fuzzy soft topology $\mathcal{T}_s(t)$ that describe configuration space C at the moment t .

7 Safety zones assessment

Safety domains usually defined around obstacles and moving objects significantly affect the solution of the reactive motion planning problem.

Let $\rho = \{\rho_0, \dots, \rho_m\}$ be a set of distance limits such that $\xi_B(A, B) = \|Crd(A) - Crd(B)\| \rightarrow \rho_i$ for each pair (A, B) of moving objects.

Let $p = \{\tau_0, \tau_1, \tau_2\}$ be a set of time limits and ξ_T be a metric given on T as $\|t_i - t_j\|_T \rightarrow p$ with the following properties $\forall t_i, t_j, t_k \in T$:

1. $\xi_T(t_i, t_j) = 0 \Leftrightarrow t_i = t_j$;
2. $\xi_T(t_i, t_j) = \xi_T(t_j, t_i)$;
3. $\xi_T(t_i, t_k), \xi_T(t_i, t_j) + \xi_T(t_j, t_k)$.

In the case of the stationary obstacle, the assessment of the safety motion based on the collision cone is illustrated in Fig. 1 (Section 3). In the case of two moving objects, the assessment becomes more complex and usually is based on evaluation of the closest approach point (Fig. 4) concerning the time T_{CPA} and distance D_{CPA} [9]. It is possible to evaluate the potential collision based on time T_{CPA} and distance D_{CPA} to the closest approach point using both metrics ξ_B and ξ_T , for example, if $\xi_B(D_{CPA}, A_1) < \rho_m$ and $\xi_T(T_{CPA}, D_1 / V_1) \leq \tau_1$, there is a threat of collision.

If $\xi_B(D_{CPA}, A_i) < \rho_{mj}$, we assume that A_i interacts with A_j . It interacts with A_j dangerously with danger degree k , if $\xi_B(D_{CPA}, A_i) < \rho_{kj}$, $k > 0$, and $\xi_T(T_{CPA}, D_1 / V_1) \leq \tau_2$. Furthermore, A_i interacts with A_j critically, if $\xi_B(D_{CPA}, A_i) < \rho_{0j}$ and $\xi_T(T_{CPA}, D_1 / V_1) \leq \tau_0$. Finally, the motion of the A_i is forbidden if $\xi_T(T_{CPA}, D_1 / V_1) \leq \tau_0$, so A_i must turn off immediately.

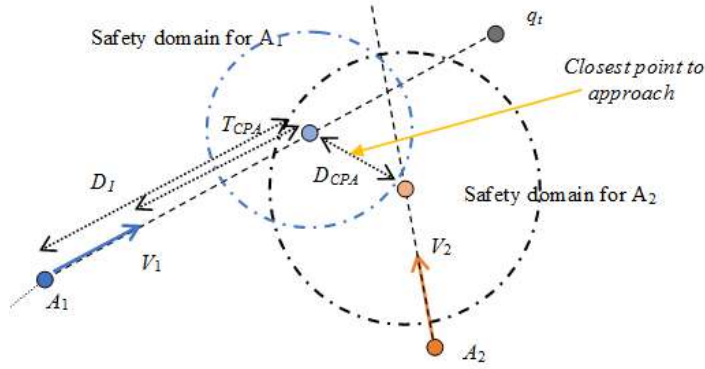


Fig. 4. Safety Assessment based on the Closest Point to Approach

It should be noted that such interactions are not always symmetric. In general, it depends on the vehicle safety domains, which usually are calculated based on its course and velocity, as well as several environmental factors such as weather [13]. Thus, in Fig. 4 vehicle A_1 interacts with the vehicle A_2 dangerously while the vehicle A_2 interacts with the vehicle A_1 not dangerously due to the difference in the size of their determined safety domains.

To take into account spatial zones with different danger levels related to the safety domain of the vehicle, we can define a set of safety assessments $\lambda = \{\lambda_0, \dots, \lambda_k, \dots, \lambda_n\}$ and assessment function $\gamma(A, B) \rightarrow \lambda_i$ that returns safety assessment λ_i for a vehicle A to the moving object B using both metrics ξ_B and ξ_T , as well as respective courses and velocities of A and B .

Based on the set of safety assessments λ , we can restrict a discretized sphere W by representing it as a system of $m+1$ nested spheres $W_n, \dots, W_0$ having a common center at the origin and corresponding radiuses from λ_m to λ_0 . Using such a system of nested spheres, we will be able to take into account certain spatial zones, which are determined by the safety conditions of the UV joint motion. For example, based on the UV motion parameters with respect to the conditions of collision avoidance, there can be defined forbidden h_A , dangerous h_B , restricted h_C , and free to move (unlimited) h_D zones bounded by corresponding spheres with boundary lines B_3, B_2, B_1 that meet the assessment levels $\lambda_3, \lambda_2, \lambda_1, \lambda_0$ respectively (Fig. 5) [21].

8 Building a second level of soft open sphere topology

At the first level, we have parameterized the y_i -subset family of the set of cells W , which constitutes the soft set of cells $(Y(t), Y)$ (or possibly fuzzy soft one $(\tilde{Y}(t), Y)$) related to the partitioning the cells into equivalence classes by their category.

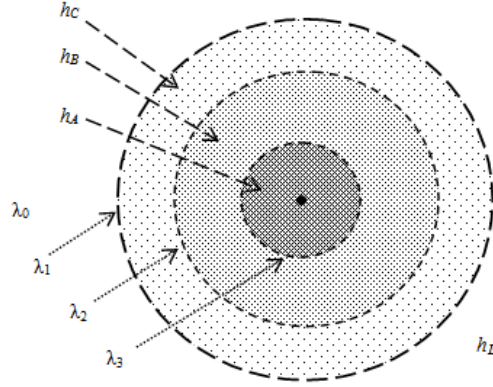


Fig. 5. Definition of safety zones

At the second level, it is necessary to break down each y_i -subset of the set W into safety zones according to the constructed system of spheres having radiuses $\lambda_0, \dots, \lambda_n$ to take into account the distribution of cells between the corresponding safety zones of the nested spheres $W_n, \dots, W_0$. Since the safety assessments for interacting vehicles can differ due to the differences in the size of their safety domains, there is no possibility to break down the entire set of cells W into safety zones. Thus, we need to partition each y_i -subset of W onto each $\lambda_j \in \lambda$ separately.

Let $(Y(t), Y)$ be a soft set on W such that $Y(t) = \cup \{Y_i(t)\}_{i=1}^k$, which consists of the union of all k of its y_i -elements $Y_i(t) = \{(Y(t), y_i) : y_i \in Y, (Y(t), y_i) \in 2^W\}$.

Let $\mathfrak{R}_{Y_i(t)}^{\lambda_j}$ be an λ_j -indiscernibility relation given on the y_i -element $Y_i(t)$ of the soft set $(Y(t), Y)$, such that $(\forall \lambda_j)_{j=0}^m \mathfrak{R}_{Y_i(t)}^{\lambda_j} = \{(w_m, w_n) \in W_j \times W_j \mid w_m \in Y_i(t), w_n \in Y_i(t)\}$. Thus, we can construct an λ_j -approximation of each y_i -element of the set $(Y(t), Y)$. Therefore, each λ_j -element of the soft set $(Y_i(t), Y)$ provides the partitioning of the y_i -element of the set $(Y(t), Y)$ into correspondent equivalence classes given by the λ_j -indivisibility relation $\mathfrak{R}_{Y_i(t)}^{\lambda_j}$ at the moment t . In other words, we parameterize first the family of subsets of the set W , which constitutes the y_i -elements of the set $Y_i(t)$, then we parameterize the family of subsets of each set $(Y_i(t), Y)$, which con-

stitutes the λ_j -elements of the set $Y_i(t)$. Just like the pair $apr_W = (W, \mathfrak{R}_W^{y_i}(t))$ defines the dynamic approximation space at the first level, the pair $apr_{Y_i(t)} = (Y_i(t), \mathfrak{R}_{Y_i(t)}^{\lambda_j})$ defines the dynamic approximation space at the second level.

Thus, we obtain a two-leveled soft set of cells $((Y(t), Y), \lambda)$ also named as a bi-soft set. The correspondent dynamic bi-soft topology $\hat{\mathcal{T}}(t) = (apr_W, Def(apr_{Y_i(t)}))$ is the partition of y_i -elements of the soft set of cells $(Y(t), Y)$ into approximated subsets of cells $Y_{ji}(t)$, which are λ_j -elements of the y_i -elements of the set $((Y(t), Y), \lambda)$. Clearly, $\hat{\mathcal{T}}(t) = ((W, \mathfrak{R}_W^{y_i}(t)), Def(Y_i(t), \mathfrak{R}_{Y_i(t)}^{\lambda_j}))$.

Each element Y_{ji} forms a soft topology $\mathcal{T}_{ji}(t)$. Thus, each λ_j -element is represented by the soft set of cells $(Y(t), Y)$ contained in the corresponding sphere V_j . The sphere is broken down in turn into y_i -elements of the set $((Y(t), \lambda), Y)$, each of which contains the cells belonging to a certain category $y_i \in Y$. The dynamic bi-soft topology that defines configuration space C can be represented as $\mathcal{T}_S = \bigcup_{j=0}^m (\bigcup_{i=0}^k \mathcal{T}_{jis})$. In the case of the fuzzy soft set of cells $(\tilde{Y}(t), Y)$ be a carrier set for the second level of topology, we have got the dynamic fuzzy bi-soft set of cells $((\tilde{Y}(t), Y), \lambda)$ and the corresponding topology.

9 Implementation of the model

The proposed model has been implemented in the reactive path planning module within UV onboard control system prototype based on embedded microcontroller STM32F429 (180 MHz Cortex M4, 2Mb Flash/256Kb RAM internal, QSPI Flash N25Q512). Information about the navigational situation comes from onboard sensors while information about team members and their targets is received from the radio communication channel.

Using the proposed model, a reactive planner transforms coordinates of all observed objects into the angular coordinate system within the configuration space. Then it creates an open sphere for each observed vehicle and discretizes it separately. Using the safety domains constructed for each observed vehicle, the planner creates the open sphere bi-soft topology. Since the environment can change, each observed event updates the partitioning of the configuration space to distinguish “free to move” subspace, in which the planner search for the next fragments of the path if a potential collision has been detected or the spatial configuration of the UV ensemble has violated.

The planner superposes open sphere topologies for each vehicle and eliminates all insignificant subsets of the cell. The set of target cells having the state equal to y_2 is considered as attractive manifold while the set of “occupied” cells having the state

equal to y_1 , y_3 , or y_4 is considered as repulsive manifold (Fig. 6).

Then the search for a path fragment to the next waypoint is provided using the potential fields-based algorithm, which builds a path choosing the cells of the constructed bi-soft topology that have the highest degree of safety motion.

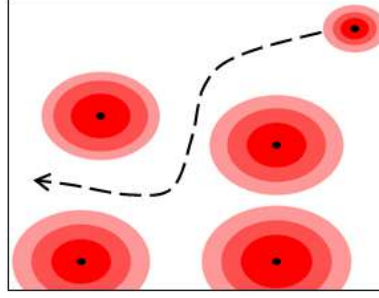


Fig. 6. Search for a path using potential fields

Finally, the path fragments will be transformed back to the cartesian coordinate system. The search algorithm is driven by the heuristic based on the fact that the greater the volume of the cells the farther the cell from the center of the sphere. It allows reducing the computational complexity essentially.

The efficiency of the proposed model has been examined compared to using the ordinary cartesian spatial model during the computer simulation based on the UV onboard control system and GIS-based model of the real terrain having the square 4 km². During the simulation, the number of vehicles has been varied from 10 to 100. The total time of the path search has been evaluated, the results are shown in Fig. 7.

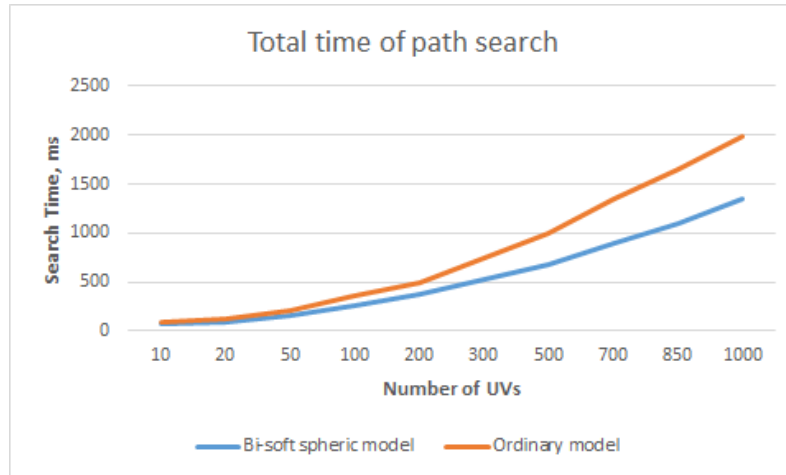


Fig. 7. The results of the simulation

The results of the experiment show that the proposed bi-soft topological model pro-

vides acceptable performance in terms of the path search time, its efficiency is about 45 percent higher than the efficiency of the ordinary model.

10 Conclusions

The problem of modeling of configuration space for the reactive joint motion planning of unmanned vehicles' ensemble is addressed in the paper.

The proposed model is based on a novel dynamic bi-soft topology, which allows describing the safety conditions of the joint motion of a multitude of vehicles. The two-leveled approximation space of this topology is constructed within the system of nested open spheres based on an angular coordinate system. The spheres are sampled into sector-like cells. The first level of the soft topology breaks down the configuration space onto "free to move" and obstacle subspaces represented by soft sets of cells taking into account uncertainties of observations. The second level of the soft topology breaks down the determined soft subspaces into safety zones also represented as soft sets. As a result, the bi-soft (fuzzy) set of cells is defined and corresponding dynamic (fuzzy) bi-soft topology is constructed.

The proposed topological model can be used in a hybrid motion planner combining potential field and RRT methods. The use of the proposed bi-soft topology allows reducing the computational complexity of the motion planning task using sophisticated partitioning of the configuration space as well as path search heuristics based on the calculation of the cell volume. The computer simulation has shown that the proposed model provides enough performance, its efficiency is about 45 percent higher than the efficiency of the ordinary cartesian model.

References

1. Sherstjuk, V., Zharikova, M., Sokol, I.: Forest Fire-Fighting Monitoring System Based on UAV team and Remote Sensing. In: Proc. of 2018 IEEE 38th International Conference on Electronics and Nanotechnology, pp. 663–668, Kyiv (2018). DOI: 10.1109/ELNANO.2018.8477527
2. Sherstjuk, V.: Scenario-Case Coordinated Control of Heterogeneous Ensembles of Unmanned Aerial Vehicles. In: Proc. of 2015 IEEE 3rd Int. Conf. on Actual Problems of Unmanned Aerial Vehicles Developments (APUAVD 2015), pp.275–279, Kyiv (2015). DOI: 10.1109/APUAVD.2015.7346620
3. Skowron, M., Chmielowiec, W., Glowacka, K., Krupa, M., Srebro, A.: Sense and avoid for small unmanned aircraft systems: Research on methods and best practices. Proceedings of the Institution of Mechanical Engineers. Part G: Journal of Aerospace Engineering 233(16):6044–6062 (2019). DOI: 10.1177/0954410019867802
4. Abbasi, Y., Moosavian, S., Novinzadeh, A.: Formation control of aerial robots using virtual structure and new fuzzy-based self-tuning synchronization. Transactions of the Institute of Measurement and Control 39(12):1–14 (2017). DOI: 10.1177/0142331216649021
5. Kang, S., Choi, H., Kim, Y.: Formation flight and collision avoidance for multiple UAVs Using concept of elastic weighting factor. International Journal of Aeronautical and Space Sciences 14:75–84 (2013). DOI: 10.5139/IJASS.2013.14.1.75

6. Gonzalez-Banos, H., Hsu, D., Latombe, J.C.: Motion planning: Recent developments. In: Autonomous Mobile Robots: Sensing, Control, Decision-Making and Applications. CRC Press (2006). DOI: 10.1201/9781315221229-13
7. Short, A., Pan, Z., Larkin, N., van Duin, S.: Recent progress on sampling based dynamic motion planning algorithms. In: Proc. of 2016 IEEE International Conference on Advanced Intelligent Mechatronics (AIM), pp. 1305–1311, USA (2016). DOI: 10.1109/AIM.2016.7576950
8. González, D., Pérez, J., Milanés, V., Nashashibi, F.: A Review of Motion Planning Techniques for Automated Vehicles. IEEE Transactions on Intelligent Transportation Systems 17(4):1135–1145 (2016). DOI: 10.1109/TITS.2015.2498841
9. Mujumdar, A., Padhi, R.: Reactive Collision Avoidance Using Nonlinear Geometric and Differential Geometric Guidance. Journal of Guidance, Control, and Dynamics 34(1):303–310 (2011). DOI: 10.2514/1.50923
10. Chakravarthy, A., Ghose, D.: Obstacle avoidance in a dynamic environment: a collision cone approach. IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans 28(5):562–574 (1998). DOI: 10.1109/3468.709600
11. Pietrzykowski, Z., Uriasz, J.: The Ship Domain – A Criterion of Navigational Safety Assessment in an Open Sea Area. Journal of Navigation 62:93–108 (2009). DOI: 10.1017/S0373463308005018
12. Song, L., Chen, Z., Dong, Z., Xiang, Z., Mao, Y., Su, Y., Hu, K. Collision avoidance planning for unmanned surface vehicle based on eccentric expansion. International Journal of Advanced Robotic Systems. 16(3):1–9 (2019). DOI: 10.1177/1729881419851945
13. Zhang, M.: Formation flight and collision avoidance for multiple UAVs based on modified tentacle algorithm in unstructured environments. PLoS ONE 12(8): e0182006 (2017). DOI: 10.1371/journal.pone.0182006
14. Molodtsov, D.: Soft Set Theory – first results. Computers and Mathematics with Applications 37:19–31 (1999). DOI: 10.1016/S0898-1221(99)00056-5
15. Maji, P. K., Roy, A. R., Iswas, R. B.: An application of soft sets in a decision-making problem. Computers and Mathematics with Applications 44(8-9):1077–1083 (2002). DOI: 10.1016/S0898-1221(02)00216-X
16. Allam, A.A., Bakeir, M.Y., Abo-Tabl, E.A.: Some Methods for Generating Topologies by Relations. Bull. Malays. Math. Sci. Soc. Series (2) 311(1), 35–45 (2008).
17. Feng, F., Li, Y., Leoreanu-Fotea, V.: Application of level soft sets in decision making based on interval-valued fuzzy soft sets. Computers and Mathematics with Applications 60:1756–1767 (2010). DOI: 10.1016/j.camwa.2010.07.006
18. Varol, B.P., Aygun, H.: Fuzzy soft topology. Hacettepe Journal of Mathematics and Statistics 41(3):407–419 (2012).
19. Wang, L., Qin, K.: Incomplete Fuzzy Soft Sets and Their Application to Decision-Making. Symmetry 11:535 (2019). DOI: 10.3390/sym11040535
20. Mahanta, J., Das, P.K.: Fuzzy soft topological spaces. Journal of Intelligent & Fuzzy Systems 32(1):443–450 (2017). DOI: 10.3233/JIFS-152165
21. Zharikova, M., Sherstjuk, V.: Case-based Approach to Intelligent Safety Domains Assessment for Joint Motion of Vehicles Ensembles. In: Proc. of the 4th International Conference on Methods and Systems of Navigation and Motion Control, p. 245–250, Kyiv (2016). DOI: 10.1109/MSNMC.2016.7783153

Genetic Algorithms as an Optimization Approach for Managing Electric Vehicles Charging in the Smart Grid

Sergiy Korotunov^{1[0000-0001-5184-4455]}, Galyna Tabunshchik^{2[000-0003-1429-5180]},
Vyacheslav Okhmak^{3[0000-0003-1969-6063]}

¹²³ National University “Zaporizhzhia Polytechnic”, Zhukovsky str., 64, Zaporizhzhia, 69063,
Ukraine
skorotunov@yahoo.com, galina.tabunshchik@gmail.com,
slavas490@gmail.com

Abstract. Usage of the genetic algorithms for solving electric vehicles optimization problem in the scope of smart grid is an extremely actual problem nowadays. Electric vehicles are modern and promising alternative to conventional vehicles. They are characterized by lower operation cost and environment-friendly ability to use renewable sources of energy. Smart grids can be used in order to avoid undesirable impact of electric vehicles. Such grids require optimization and correct scheduling to handle growing number of electricity consumers. This can be achieved with implementation of specifically designed genetic algorithms. The goal of the paper is to select optimal method and propose it for using for optimization of the digital twin of the electric vehicles charging infrastructure. As a result of paper such method is proposed. Moreover, as a scientific novelty, genetic algorithm functions are compared and analyzed applying to the problem in consideration.

Keywords. Genetic algorithm, optimization, electric vehicle, smart grid, crossover, selection, mutation.

1 Introduction

Electric vehicles (EVs) can be viewed as an eco-friendly and cost-effective alternative to conventional vehicles with internal combustion engines (ICE). EVs are produced and designed by number of different manufacturers and their production amount is expected to grow rapidly in the coming years [1]. EVs have lower operating costs with respect to ICE vehicles and can be charged with locally produced renewable energy sources (RESs) [2]. However, number of challenges to wide spreading of EVs exist. Although their operating costs are less, EVs are still more expensive to buy than ICE vehicles. In addition, access to charging stations is limited, and large capital investment is required for developing a public charging infrastructure [2]. More than that, EVs consume comparatively high power from the grid during charging. Therefore, uncoordinated charging of a large number of EVs can have an adverse impact on the grid operation (power outages, unacceptable voltage fluctuations) [3]. Power gen-

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

eration can be increased in order to handle the peak demand of EVs. However, this will lead to significant infrastructure cost. As an alternative cost-effective solution, smart grid allows EVs to coordinate their charging operations, which can improve frequency regulation [4], smooth out intermittent power generation from RESs, and make the electric power usage efficient [5].

2 Formal problem statement

Smart grid [6, 7] is a concept aimed at providing the next generation electricity network that stands out by high configurability, reactivity and self-control capability. This is a complex infrastructure that characterizes by following attributes [8]:

- multidisciplinary character,
- spatial distribution,
- network systems heterogeneousness,
- implementation of evolutionary development,
- functioning and controlled separation of its elements.

Smart grid is expected to be a key part of the global system of interacting actors, which in its turn will lead to improved management of available resources and increased energy efficiency. Precise monitoring and management are needed to achieve this goal [9]. Smart grid is created on the basis of advanced information and two-way digital communication technologies. Therefore, it cannot just intelligently deliver electricity but also manage power facilities taking advantage of real-time information exchange and interaction between providers and consumers [10].

Meanwhile, EVs are new and important targets for the smart grid to manage. EVs are equipped with batteries having high energy-storage capacity, so EV charging imposes large electric load on the power system [11]. EVs are forced to be equipped with an electronic interface for grid connection to allow controlled energy exchanges. Despite high number of the researches in area, EVs still need to be charged more often and it takes at least tens of minutes [12]. That makes it necessary to build a large charging infrastructure which embraces fast charging stations, battery swapping stations, and individual charging points for slower charging [11].

Different types of interactions are possible between the power grid and an EV:

- grid to vehicle (G2V),
- vehicle to grid (V2G).

In G2V, an EV's battery can be charged from the grid using stored electricity from external power sources. That means that the power flow is always unidirectional. In V2G, the power flow is bidirectional, i.e., from the grid to an EV while charging and from an EV to the grid while discharging. V2G-enabled EVs may earn incentives and sell power while discharging to the grid and make payments while charging batteries from the grid. Therefore, V2G-enabled EVs can facilitate the supply/demand balance by discharging during peak hours (peak clipping) and charging during off-peak hours (valley filling) as can be seen at Fig. 1.

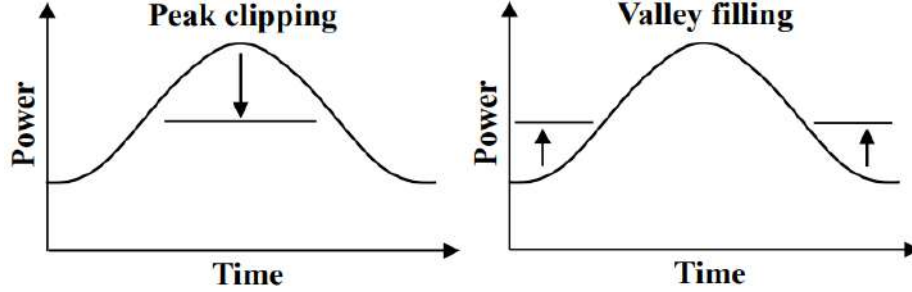


Fig. 1. Peak clipping and valley filling (Adapted from [13])

The impact of EVs on the power grid has been studied [14]. One solution to decrease the impact of EVs on the grid is to schedule their charging/discharging profiles. This can be done by aggregating different sets of EVs for charging or discharging with different start times and durations such that grid constraints are maintained. However, the aggregation of EVs differs from the aggregation of more traditional power resources [15]. In particular, the temporal availability of EVs along with their location information is an important parameter to consider while aggregating EVs for possible grid overload planning and management. Thus, determining and optimizing the appropriate charge and discharge times of EVs that do not violate grid constraints while maintaining acceptable degrees of user satisfaction is a challenging problem.

3 Literature review

The EV charge scheduling problem formulates as following. It takes a set of EVs, grid, user, and aggregator-side parameters as input and outputs a charging schedule. By charging schedule, it is meant the starting and ending times of charging of each EV in the set. It is an optimization problem that optimizes some grid, user, or aggregator-side parameters (or a mix of them) subject to different number of constraints. No single mathematical formulation exists for the problem in general, forming different optimization problems.

The optimization intends to provide the best local or global solutions. Generally, the mathematical formulation of an optimization problem is to maximize or minimize an objective function while satisfying all considered constraints related to the integrated components in the model [16].

Depending on the complexity and the difficulty, optimization can be addressed by means of exact or approximate methods.

The exact mathematical methods can generate an optimal solution when they are specified in a feasible region. There are two basic categories: linear and non-linear model. They are based on all implemented constraints and the objective functions. The linear models are divided in three types: linear programming, integer program-

ming and mixed integer linear programming, according to the variables if they are real, integer, or both variable types, correspondingly.

The approximate methods have an advantage that can simply manage the nonlinear constraints and objective functions. In the same time, they cannot always guarantee the quality of the obtained results because they generally employ random search methods [17]. More than that, the possibility to find the global solution decreases as soon as the size of the considered problem increases [18].

In this paper we consider the genetic algorithms (GAs) as a part of approximate methods [19]. We study those for the modeling of the EV scheduling problem because it generally permits to employ the characteristics of the integrated distributed energy resources with employing integer and binary variables to make a decision on the operation status of the production systems, battery storage system, EVs and smart appliances of the microgrid.

3.1 Analyze GA approaches for EV optimization problem

GA are becoming one of the most popular search techniques for the problems having large search space [20]. One of such problems can be EV charging scheduling GA can create a reasonable quality solution within a controllable time bound by means of a population-based directed search, which is inspired by biological evolution principles.

In the description of GA, the workflow is a sequential change of populations consisting of a fixed number of individuals corresponding to the trial points of the solution space. Individuals who respond to high-quality (more appropriate) solutions receive the advantage of producing offspring in the next population [21].

From the population consisting of feasible solutions called chromosomes, the next-generation population is created through genetic operations such as selection, crossover, and mutation. This evolution process continues for the given number of iterations or until a termination condition is met [22].

Authors in [23] integrated vehicular networks with smart grid and designed a charge scheduler for EVs based on heuristic-based approaches and GAs, which minimizes the load at a charging station. Each request consists of:

- vehicle type,
- estimated arrival time,
- desired service completion time (deadline),
- current battery charge.

On receiving the request, the power consumption profile of the vehicle is retrieved from the repository of vehicular information. Then, the charging station verifies whether it can satisfy the new request along with the other requests already submitted to the scheduler. The result is communicated back to the vehicle. On receiving the result, the driver may accept the schedule, initiate a renegotiation session, or choose another charging station.

From the vehicle-side viewpoint, entering the station, the vehicle is assigned and plugged to a charger. The controller connects or disconnects power to each vehicle

according to the schedule generated by either a scheduler within a charging station or a remote charging server running in the Internet.

In [24] same authors evaluated the performance of a charging task scheduler for EV, aiming at reducing the peak load and improving the service ratio in charging stations. They try to achieve better results by making the initial population include both heuristic-generated schedules for fast convergence and randomly generated schedules for diversity loss compensation. The performance measurement result obtained from that work reveals that scheme in consideration can reduce the peak load for the given charging task sets by up to 4.9%, compared with conventional schemes.

Authors in [25] established a stochastic procedure for modeling and analyzing an EV fleet to generate an accurate charging and discharging profiles. They focused on the uncertainties that may affect EV charging/discharging profiles:

- EV type – battery electric vehicle (BEV) which is supplied by an electrical source to feed its energy storage unit and plug-in hybrid electric vehicle (PHEV) which has ability to utilize ICE along with electric one,
- EV battery capacity – authors assumed that the battery pack for a BEV has a 20-30 kWh pack and for a PHEV, 5-15 kWh,
- time duration availability and scheduling time – authors hypothesized that 50% of EV of the fleet were plugged-in to be charged at the workplace in the parking lots and the other EV were connected to the grid to be charged at home in the evening.

Comparing described previously works it can be said that in first two works authors presented a chromosome as a single feasible schedule. It was represented by a fixed-length string of an integer-valued vector. The charging task could be started, suspended, and resumed at a slot boundary. They used the roulette wheel as selection method, random operation as crossover method. Due to the fact that each element has a different permissible range, the mutation was prohibited.

In contrast, authors of third work used number of selection methods. Namely, stochastic universal selection, roulette-wheel selection, tournament selection and ranking selection [26]. Like the first group of authors they used random operator as crossover method. But, unlike them they used mutation operator that would change randomly the genes of the chromosomes which could also change their characteristics.

Another important distinction is that the proposed by last group of authors GA algorithm allows to make the optimal trade-off between V2G and G2V operations cost to highly increase benefits from EV batteries by scheduling the charging mode in the low power price periods and discharging mode in the high-power price periods. This contradicts to first two works where authors only utilize and optimize charging phase of the EV, ignoring discharging ability.

Therefore, it can be concluded that after analyzing and comparing GA approaches one can be selected as optimal. It is the third work which has critical advantages above other ones.

- it allows and supports usage of the mutation operator which should help in creating better GA and prevent falling into local minimums,

- it provides support both for V2G and G2V, which can be critical in modern EVs and smart grids.

4 Model selection

After analyzing, following method can be used for the future optimization of the EV charging infrastructure along the horizon T with t time steps. The aim of the EV charging/discharging scheduling is to minimize the smart grid total cost and to minimize the overall cost of the power in the G2V operations of EVs. Which is given by following function:

$$\min f_{\text{cost}} = \sum_{t=1}^T \left\{ \sum_{n=1}^{EV_{\text{Char}}} \eta_{\text{Char}} \times E_n \times C(t) - \sum_{n=1}^{EV_{\text{Disch}}} \eta_{\text{Disch}} \times E_n \times C(t) \right\} \quad (1)$$

Here:

- EV_{Char} and EV_{Disch} are number of charging and discharging EVs respectively,
- η_{Char} and η_{Disch} are charging and discharging efficiency respectively,
- E_n is the power received or delivered from vehicle n ,
- $C(t)$ is the hourly price of electricity.

From that model, EV can be in one of the following states:

- in the charging state,
- in the discharging state,
- not charging or discharging (battery is in idle mode).

Number of constraints need to be defined in order to achieve complete model of the charging process.

$$SoC_n^t = SoC_n^{t-1} + \eta_{\text{Char}} \times E_n \quad (2)$$

$$SoC_n^t = SoC_n^{t-1} - \eta_{\text{Disch}} \times E_n \quad (3)$$

$$\eta_{\text{Char}} \times E_n \leq SoC_n^{\text{max}} \quad (4)$$

$$-\eta_{\text{Disch}} \times E_n \leq -SoC_n^{\text{min}} \quad (5)$$

$$\sum_{t=1}^T \eta_{\text{Char}} \times E_n + \sum_{t=1}^T \eta_{\text{Disch}} \times E_n = SoC_n^{\text{need}} \quad (6)$$

The (2) constraint describes the cost of the power delivered to the EV battery, while the (3) describes the cost of the power supplied from the EV battery to the grid. Constraints (4) and (5) ensure that the state-of-charge (SoC) is scheduled within a predefined range between $SoC_n^{\min}$ and $SoC_n^{\max}$ (generally the customer can pre-set this range). Constraint (6) guarantees that the system meets each single EV's energy need at the unplugging time (SoC_n^{need}) at any recharging cycle.

5 EV charging infrastructure in consideration

Considered optimization method can be applied to the developed EV charging station. From the perspective of the EV, the global smart grid can be separated into smaller EV systems. The EV system has three levels: device level, communication level and application level as shown in Fig. 2 [27].

Micro-grid was developed in order to consider EV-charging system [28] as a part of ISRT (Interactive platform for Embedded Software Development Study, developed in Zaporizhzhia Polytechnic National University [29]) infrastructure. The ISRT-server is a platform for remote laboratories, in order to train students in IoT-tasks. It is ideal for the scalability of the charging station case. A ready-made black-box charging station module is available in the ISRT so that students can work on the communication and client layer.

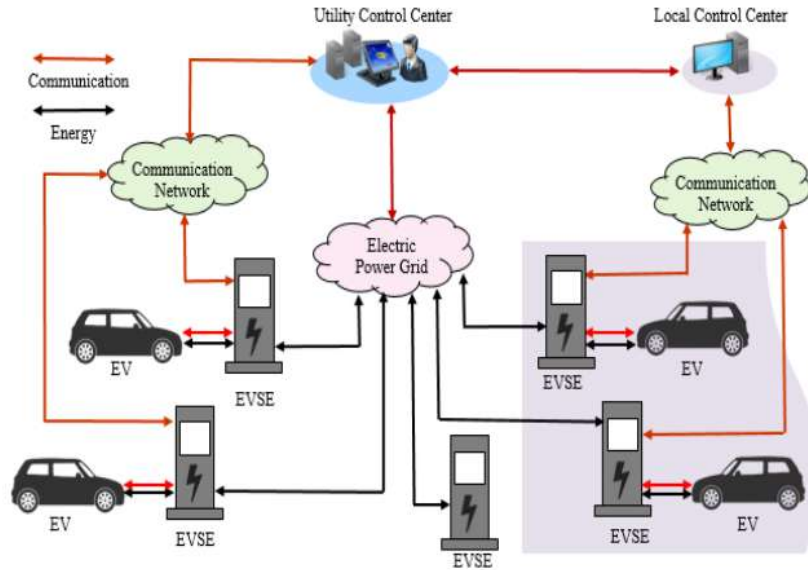


Fig. 2. Cyber-physical nature of EV system [27]

The hardware part is emulated with a simple electronic circuit to charge 1.25V Li-ion-batteries. Energy input side represented with following elements:

- net-powered,
- wind-powered,
- solar energy.

The charging system itself is both a consumer on the net and an energy provider in a V2G setup. For simulation purposes of the micro grid system users can variate input variables such as:

- the battery charge level,
- turn on and off different sources of energy;
- decide to start selling energy to the smart grid.

This system is displayed at Fig. 3. Where power sources noted as the letter A, the load as a letter B, the sale of surplus shown in the form of LED as letter C, battery and charge controller as letter D. With digits 1 – 8 different relays are depicted.

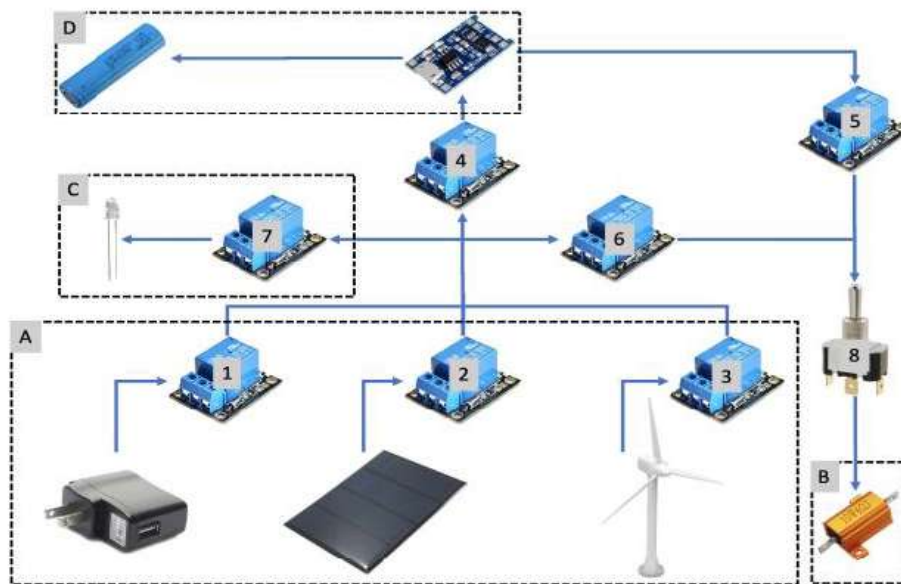


Fig. 3. Different power supplies for the charging station in a smart grid

The developed model of a charging station allows to:

- choose a power source (any separately or 2-3 sources at a time),
- control the process of battery charging (protection against undercharge or overcharge),
- “sell” residual energy and switch load.

The power source is indicated by letter A, load – B, surplus sales are indicated by LED – C, battery and charge controller – D.

As can be seen, developed digital twin of the EV charging infrastructure can be used by students to design optimal smart grid.

6 Experiments and results

Considered optimization method is simulated in the scope of the discussed charging station microgrid. To check efficiency of the specific GA some test data are used. Task of the optimization is slightly simplified because of the lack of SoC needed data for the set of EVs.

Therefore, goal of optimization can be formulated as to minimize cost of charging and discharging profiles of the set of EVs during given number of time slots.

Price of the electricity is also given for each time period. As well as charging and discharging efficiency. Each EV is characterized by state-time vector which represents if specified EV is charging, discharging or idle.

Input data for GA is matrix that contains EVs power for each time slot in given boundaries.

Number of available selection, scaling, mutation and crossover functions are compared in order to find optimal set for specified optimization problem. There are 200 iterations for each experiment.

First set of experiments has the goal to determine optimal scaling function. Fitness scaling converts the raw fitness scores that are returned by the fitness function to values in a range that is suitable for the selection function. Following functions are considered:

- rank function,
- proportional function,
- top function with quantity 0.4,
- top function with quantity 0.6.

Results can be seen at Fig. 4.

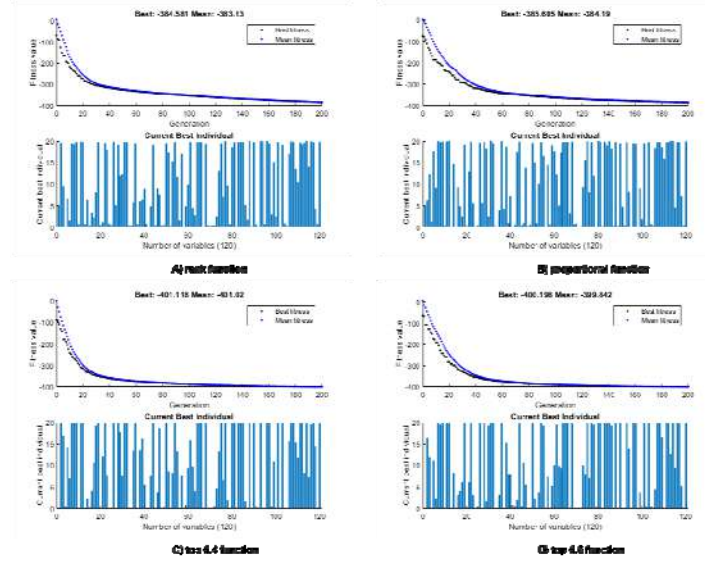


Fig. 4. Comparison of the different scaling functions:
A) rank, B) proportional, C) top 0.4, D) top 0.6

As can be seen, best optimization result is achieved using top scaling function with quantity of 0.4. Top scaling assigns 40 percent of the fittest individuals to the same scaled value and assigns the rest of the individuals to value 0. That means that only 40 percent of the fittest individuals can be selected as parents.

Next step is to determine selection GA function. Selection function specifies how the GA chooses parents for the next generation. Following functions are considered:

- stochastic uniform function,
- roulette function,
- tournament function,
- uniform function.

Results can be seen at Fig. 5.

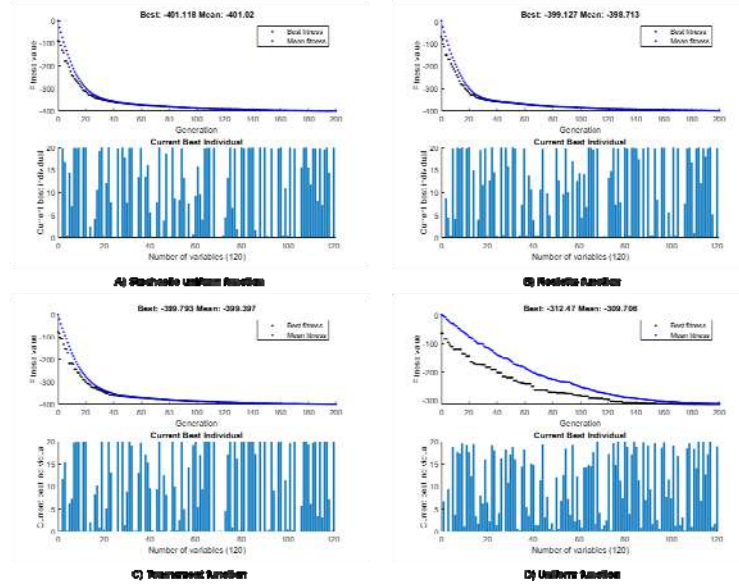


Fig. 5. Comparison of the different selection functions:
A) stochastic uniform, B) roulette, C) tournament, D) uniform

From results, best optimization result is achieved using stochastic uniform selection function. Stochastic uniform function creates a line in which each parent corresponds to a section of the line of length proportional to its scaled value. The algorithm moves along the line in steps of equal size. At each step, the algorithm allocates a parent from the section it lands on. The first step is a uniform random number less than the step size.

Following set of experiments is related to selection of the best mutation function for specified problem. Mutations specify how the GA makes small random changes in the individuals in the population to create mutation children. Mutation provides genetic diversity and enables the GA to search a broader space and does not allow it to fall into local minimum. Following functions are considered:

- adaptive feasible function,
- uniform function.

Comparison of those can be seen at Fig. 6.

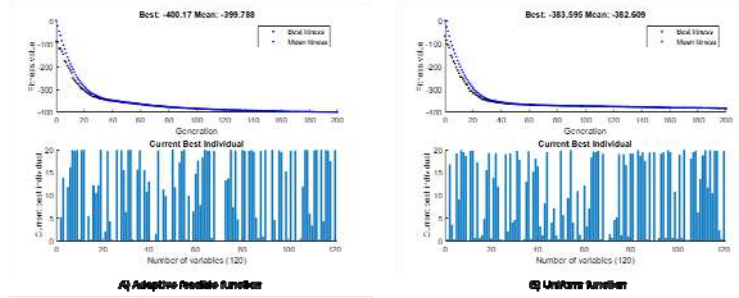


Fig. 6. Comparison of the different mutation functions:
A) adaptive feasible and B) uniform

It can be seen that best optimization result is achieved using adaptive feasible mutation function. Adaptive feasible randomly generates directions that are adaptive with respect to the last successful or unsuccessful generation. The mutation chooses a direction and step length that satisfies bounds which are defined.

Last set of experiments has as its goal to determine the best crossover function for specified problem. Crossover functions specify how the GA combines two individuals, or parents, to form a crossover child for the next generation. Following crossover functions are considered:

- two-point function,
- constraint dependent function,
- heuristic function.

They can be seen at Fig. 7.

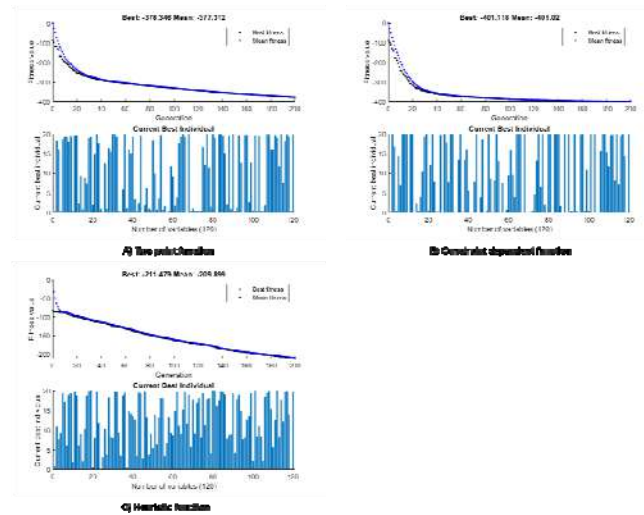


Fig. 7. Comparison of the different crossover functions:
A) two-point, B) constraint dependent, C) heuristic

It can be seen that best optimization result is achieved using constraint dependent function. Constraint dependent chooses scattered function when there are no linear constraints, and chooses intermediate function when there are linear constraints. These choices ensure that feasible parents give rise to feasible children, where feasibility is with respect to bounds and linear constraints. Scattered function is used externally because there are no linear constraint functions. Scattered crossover function creates a random binary vector. It then selects the genes where the vector is a 1 from the first parent, and the genes where the vector is a 0 from the second parent, and combines the genes to form the child.

After executing specified experiments, best GA functions for specified EV charging problem can be selected. Thus, best optimization results achieved using top scaling function with quantity 0.4, stochastic uniform selection function, adaptive feasible mutation function and scattered crossover function.

7 Conclusion

In the paper the authors considered the possibility of utilization of the GAs for developed in the previous researches EV charging infrastructure. Review of the main characteristics of the EVs, charging, smart grid and GAs was carried out by the authors. Current researches related to implementing GAs for managing EVs charging are considered and analyzed. After investigation and executing simulation experiments it is possible to make an assumption that utilization of the GAs for EVs charging problem is promising optimization approach. As the future work, it is planned to develop own GA which will handle optimization of the developed EV charging infrastructure.

Scientific novelty of the work is that number of GA scheduling techniques were considered and analysed. One optimization method on the basis of GA was selected as optimal after comparison. Simulation experiments were executed for that method using different GA options. Those were compared and selected optimal options for specified problem. This method is proposed to use for later researches to provide optimization of the developed EV charging infrastructure.

8 References

1. Global EV Outlook 2019 – Analysis - IEA, <https://www.iea.org/reports/global-ev-outlook-2019>
2. Sortomme, E., El-Sharkawi, M.: Optimal Charging Strategies for Unidirectional Vehicle-to-Grid. *IEEE Transactions on Smart Grid.* 2, 131-138 (2011). doi: 10.1109/tsg.2010.2090910
3. Clement-Nyons, K., Haesen, E., Driesen, J.: The Impact of Charging Plug-In Hybrid Electric Vehicles on a Residential Distribution Grid. *IEEE Transactions on Power Systems.* 25, 371-380 (2010). doi: 10.1109/tpwrs.2009.2036481
4. Sekyung Han, Soohee Han, Sezaki, K.: Development of an Optimal Vehicle-to-Grid Aggregator for Frequency Regulation. *IEEE Transactions on Smart Grid.* 1, 65-72 (2010). doi: 10.1109/tsg.2010.2045163

5. Sundstrom, O., Binding, C.: Flexible Charging Optimization for Electric Vehicles Considering Distribution Grid Constraints. *IEEE Transactions on Smart Grid*. 3, 26-37 (2012). doi: 10.1109/tsg.2011.2168431
6. Internet of Energy: ICT for energy markets of the future, https://www.iese.fraunhofer.de/content/dam/iese/en/mediacenter/documents/BDI_initiative_IoE_us-IdE-Broschuere_tcm27-45653.pdf
7. Smartgrids: Strategic deployment document for europe's electricity networks of the future, https://ec.europa.eu/research/energy/pdf/smartgrids_en.pdf
8. Maier, M.: Architecting principles for systems-of-systems. *Systems Engineering*. 1, 267-284 (1998). doi: 10.1002/(sici)1520-6858(1998)1:4<267::aid-sys3>3.0.co;2-d
9. Karnouskos, S.: Cyber-Physical Systems in the SmartGrid. 2011 9th IEEE International Conference on Industrial Informatics. (2011). doi: 10.1109/indin.2011.6034829
10. Gellings, C.: The smart grid. The Fairmont Press, Lilburn (2009)
11. Lopes, J., Soares, F., Almeida, P.: Integration of Electric Vehicles in the Electric Power System. *Proceedings of the IEEE*. 99, 168-183 (2011). doi: 10.1109/jproc.2010.2066250
12. Tribioli, L., Barbieri, M., Capata, R., Sciubba, E., Jannelli, E., Bella, G.: A Real Time Energy Management Strategy for Plug-in Hybrid Electric Vehicles based on Optimal Control Theory. *Energy Procedia*. 45, 949-958 (2014). doi: 10.1016/j.egypro.2014.01.100
13. Lampropoulos, I., Kling, W., Ribeiro, P., van den Berg, J.: History of demand side management and classification of demand response control schemes. 2013 IEEE Power & Energy Society General Meeting. (2013). doi: 10.1109/PESMG.2013.6672715
14. Gong, Q., Midlam-Mohler, S., Marano, V., Rizzoni, G.: PEV charging impact on residential distribution transformer life. *IEEE 2011 EnergyTech*. (2011). doi: 10.1109/energytech.2011.5948498
15. Jansen, B., Binding, C., Sundstrom, O., Gantenbein, D.: Architecture and Communication of an Electric Vehicle Virtual Power Plant. 2010 First IEEE International Conference on Smart Grid Communications. (2010). doi: 10.1109/smartgrid.2010.5622033
16. Iqbal, M., Azam, M., Naeem, M., Khwaja, A., Anpalagan, A.: Optimization classification, algorithms and tools for renewable energy: A review. *Renewable and Sustainable Energy Reviews*. 39, 640-654 (2014). doi: 10.1016/j.rser.2014.07.120
17. Chen, Y., Wu, Y., Song, C., Chen, Y.: Design and Implementation of Energy Management System With Fuzzy Control for DC Microgrid Systems. *IEEE Transactions on Power Electronics*. 28, 1563-1570 (2013). doi: 10.1109/tpel.2012.2210446
18. Lin, M., Tsai, J., Yu, C.: A Review of Deterministic Optimization Methods in Engineering and Management. *Mathematical Problems in Engineering*. 2012, 1-15 (2012). doi: 10.1155/2012/756023
19. Leoshchenko, S., Oliynyk, A., Subbotin, S., Shylo, S., Shkaruplyo, V.: Method of Artificial Neural Network Synthesis for Using in Integrated CAD. 2019 IEEE 15th International Conference on the Experience of Designing and Application of CAD Systems (CADSM). (2019). doi: 10.1109/cadsm.2019.8779248
20. Sivanandam S, Deepa S.: *Introduction to Genetic Algorithms*. Springer-Verlag Berlin Heidelberg, Berlin, Heidelberg (2008)
21. Desprez, C., Chu, F., Chu, C.: Minimising the weighted number of tardy jobs in a hybrid flow shop with genetic algorithm. *International Journal of Computer Integrated Manufacturing*. 22, 745-757 (2009). doi: 10.1080/09511920902810938
22. Subbotin, S.: THE NEURO-FUZZY NETWORK SYNTHESIS WITH THE RANKING AND SPECIFIC ENCODING OF FEATURES FOR THE DIAGNOSIS AND AUTOMATIC CLASSIFICATION ON PRECEDENTS. *Radio Electronics, Computer Science, Control*. 0, (2016). doi: 10.15588/1607-3274-2016-1-6

23. Lee, J., Kim, H., Park, G., Jeon, H.: Genetic Algorithm-Based Charging Task Scheduler for Electric Vehicles in Smart Transportation. *Intelligent Information and Database Systems*. 208-217 (2012). doi: 10.1007/978-3-642-28487-8_21
24. Lee, J., Park, G.: Genetic algorithm-based demand response scheme for electric vehicle charging. *International Journal of Intelligent Information and Database Systems*. 7, 535 (2013). doi: 10.1504/ijiids.2013.057420
25. Elmehdi, M., Abdelilah, M.: Genetic Algorithm for Optimal Charge Scheduling of Electric Vehicle Fleet. *Proceedings of the 2nd International Conference on Networking, Information Systems & Security - NISS19*. (2019). doi: 10.1145/3320326.3320329
26. Goldberg, D., Deb, K.: A Comparative Analysis of Selection Schemes Used in Genetic Algorithms. *Foundations of Genetic Algorithms*. 69-93 (1991). doi: 10.1016/b978-0-08-050684-5.50008-2
27. Ahmed, M., Kim, Y.: Performance Analysis of Communication Networks for EV Charging Stations in Residential Grid. *Proceedings of the 6th ACM Symposium on Development and Analysis of Intelligent Vehicular Networks and Applications - DIVANet '17*. (2017). doi: 10.1145/3132340.3132352
28. Arras, P., Tabunshchyk, G., Okhmak, V., Korotunov, S.: Modeling and Simulation of the Services for Vehicle Charging Infrastructure Interaction. *2019 10th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS)*. (2019). doi: 10.1109/idaacs.2019.8924449
29. Arras, P., Tabunshchyk, G., Korotunov, S., Okhmak, V.: Cost Optimization Simulation for Electric Vehicle Charging Infrastructure. *2020 IEEE European Technology & Engineering Management Summit*. (2020)

Implementation of Web System Optimization Method

Galina Kirichek¹[0000-0002-0405-7122], Stepan Skrupsky¹[0000-0002-9437-9095], Mariia Tiahunova¹[0000-0002-9166-5897] and Artur Timenko¹[0000-0002-7871-4543]

¹National University "Zaporizhzhia Polytechnic", 64 Zhukovsky str., Zaporizhzhia, 69063, Ukraine

kirgal08@gmail.com, sskrupsky@gmail.com, mary.tyagunova@gmail.com, timenko.artur@gmail.com

Abstract. The method of optimization of the network web system client interface has been considered in the paper. The system model demonstrates the specific sequences of subject's actions and their interactions. The proposed work uses modern technologies and techniques, including React library and its features to create a "smart" client. It can partially perform the functions of server logic, thereby providing less server load and improving the interactive user interface. The research methods are based on interaction schemes modeling of the system and its modules, the methods of virtualization of long lists and the shouldComponentUpdate React lifecycle method. Optimization with Front-End technologies does not affect the functionality and integrity of the system.

Keywords: optimization, Front-End, React, Laravel, client, interface

1 Introduction

Nowadays, with ever-increasing competition in information technology and the increasing popularity of mobile devices for accessing to Internet, the issue of user-friendliness is relevant [1, 2]. Therefore, the development of interactivity and simplification of web systems, when using them, is the reason of the creation of new optimization methods and development of user-friendly and informative design of network systems. By providing intuitive interfaces, it allows you to use systems more efficiently and reduce the threshold for the onset of user interaction with the system [3]. Equally important, to create a user-friendly interface, is the latency of the data that the user interacts with. The shorter the waiting time, the more efficient and user-acceptable is the system [4, 5].

Every year more developers implement complicated network systems, which available for consumers and exist in web environment. When network web systems are in process of implementation, the development is divided to Front-End (development of user interface) and Back End (development of server side). For ensuring a cooperation among them an interface is used which gives an opportunity to transfer data from the client side to the server one and vice versa, independently of programming languages.

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

The most popular examples of these “bridges” are technologies JSON and XML [6, 7]. The leader in development of client’s applications is a programming language JavaScript. After transition to the new (ES6) version, JavaScript has an opportunity to create libraries and frameworks when client side parts are implemented for network web systems [1]. Although new libraries and frameworks are being built to develop user interfaces, the most popular technologies are React and VueJS libraries and Angular framework.

A framework Angular, which uses Type Script language, is an analog and the main competitor of these libraries in implementation the client side of the system. React uses a few smart methods to minimize numbers of DOM operations and makes an interface faster without using additional optimization [8]. During component performance analysis, in development mode, you can see how components are mounted, updated, and unmounted by using browser performance tools [9, 10].

2 Problem statement

In solving the main task of the research it is necessary to provide a reduction of the waiting time when performing typical tasks that do not require a page reload; optimize the first page load of a user's browser; ensure the distribution of client (Front-End) and server (Back End) parts; optimize the client part using the DOM tree; increase the usability of the system usage by the client taking into account the latest hardware, software and social needs, as well as the need to use advanced Front-End technologies.

During optimization of the implemented system we pay much attention on the models of the system, which demonstrate its structure and behavior. In this case we have the system that has been implemented with using the framework Laravel and have the client side, which consists of Blade templates. UML models help to structure the modules of the system while performing its optimization. Here is use-case diagram for the system (Fig.1) and information model of its organization (Fig.2).

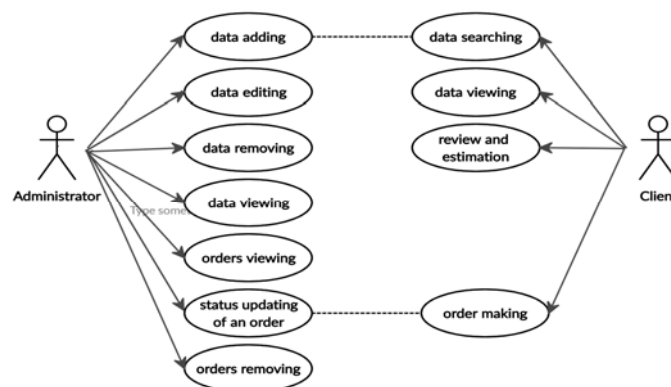


Fig. 1. Use-case diagram

A technology of maintaining a process relatively interacting with organizational structures of the network system consists of next stages: information publishing about items of the system by administrator; searching the information by a client; creating an order by client; processing the order by administrator; client's feedback and assessment. This system uses the library jQuery, which works with a DOM tree directly and increases the waiting time of working system.

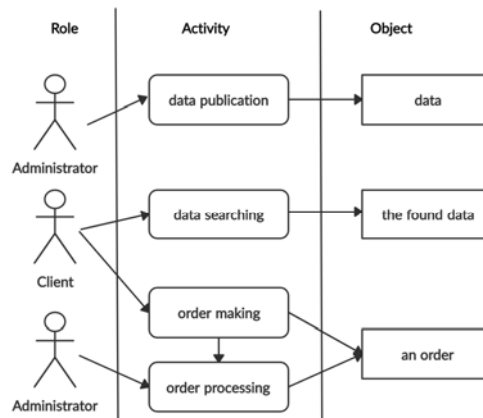


Fig. 2. Informational model of the system organization

The goal of this work is an optimization of the network web system, which has been developed by using Laravel. An object of research is client interface optimization process. Subject is models, methods and optimization tools of the system client interface.

The methods of research are based on modeling of the system and modules interaction schemes, methods of long lists virtualization and shouldComponentUpdate React lifecycle method.

3 Modeling of system

Features which were implemented by using the jQuery library will have been written by using JavaScript ES6 and will have been minimized for faster downloading. If the application makes rendering of long rows, we use a method of "Window's accessing". This method makes rendering only a small amount of a subset of rows during one timeframe and allows considerably reducing time which is needed for repeating rendering of components and numbers of creating DOM knots [10].

React components of application have the structure in a shape of tree and the root element, which includes children components of hierarchical structure. If the main component is updated, children components forcibly are updated too. It loads the system, as not all components have to be updated. To avoid this, we use the Should-

ComponentUpdate component lifecycle method to configure the comparison of previous and existing Props.

If this comparing return true, then the current component does not have to be updated. Also, for optimization of components it is possible to apply the automation of the comparing props by the method of prohibition or permissions of updating the component with using of the pureComponents, which are imported from the React library. For understanding of interaction client-server and a database we will build deployment model (Fig.3).

Optimizing the Laravel-developed system with Front-End technologies does not affect the functionality and integrity of the system. System must fulfill its tasks, use design patterns and have the same routes as before optimization. Optimization will provide a user-friendly interface and reduce server load.

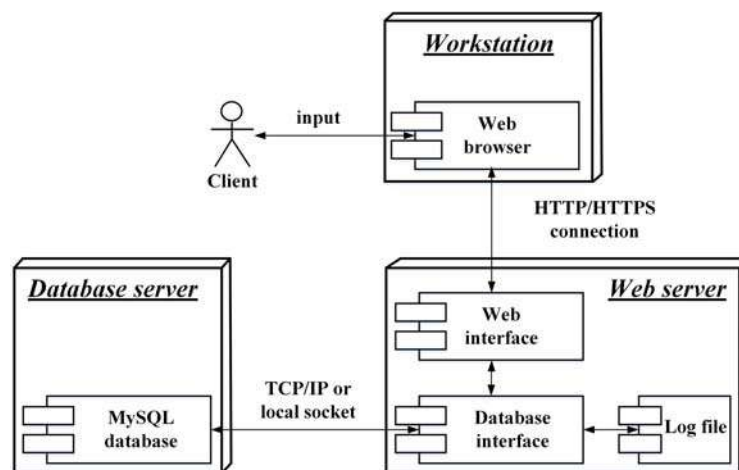


Fig. 3. Deployment model

Let's pass to inclusion of React. Install NodeJs, create the file "package.json", which responsible for modules which are needed for future development of the application and set up the module Yarn as a package manager. Next step is installing of the React library, modules react-dom and babel-preset-react. Make corrections in the file web-pack.mix.js: replace a row of code mix.js ('resources/assets/js/app.js', 'public/js') to mix.react('resources/assets/js/ app.js','public/js').

There is a point of entering to JavaScript files, compiled files are placed in the directory public/js. For installing dependencies execute the command "npm install" in the command line.

After settings were installed, create React components in the directory resources/assets/js/component/. Every time a component is implemented, it is imported to the entry point to the JavaScript file app.js. Import of components to the entry point is given below.

```
require('./bootstrap'); /* Import the Main component */
```

```
import Main from './components/Main';
```

As the system uses the design pattern MVC, it is necessary to ensure its usage [11, 12].

Let's pass to the views, which are placed in the directory resources/views of project. Import an entry point of js files, in which all application components are defined, to the file "welcome.blade.php". The fragment of code is given below.

```
<!doctype html>
<html lang="{{ app()->getLocale() }}">
  <head>
    <meta charset="utf-8">
    <meta http-equiv="X-UA-Compatible" content="IE=
edge">
    <meta name="viewport" content="width=device-width,
initial-scale=1">
    <title>Laravel React application</title>
    <link href="{{mix('css/app.css')}}" rel="stylesheet"
type="text/css">
  </head>
  <body>
    <h2 style="text-align: center"> Laravel and React
application </h2>
    <div id="root"></div>
    <script src="{{mix('js/app.js')}}" ></script>
  </body>
</html>
```

The network system uses routes of the Laravel framework, so after the React and modules that are necessary for future development are connected, the system reloads pages every time when the user uses other routes [13, 14].

4 Optimization of system

After examining the precedents of roles and sequence diagrams, we modeled the principle of operation of the system, taking into account the Laravel framework and its architectural style (Fig.4). Zone "A" is the client, "B" is the server, "C" is the database.

By default, the Laravel framework is implemented using the Blade Template. This template helps simplify system partitioning into client and server parts. In this case, a conditional distribution is implied, since the view as well as the controller and model are located on the server and can theoretically fulfill the logic of the server.

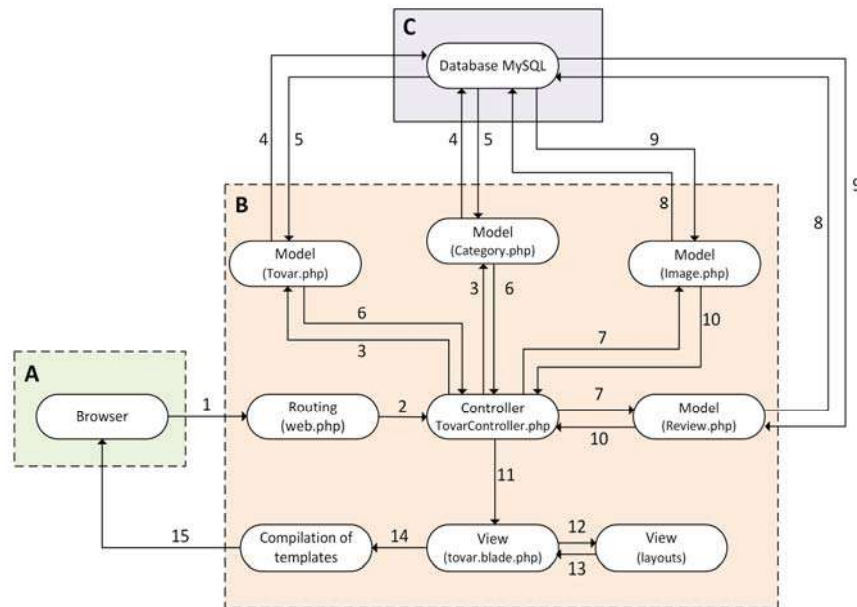


Fig. 4. In-app request processing algorithm

Fig. 4 shows a model of a system based on Laravel architectural style that involves the use of an MVC pattern (model, view and controller). Consider a modified model that will use the same pattern, but with the inclusion of Front-End technology to client. Now it is possible to speak not only about the formal division of the system into client and server (as in Fig. 4), but also about the physical one, since now the view (template) is fully browser-based and has its own logic.

Consider changes in the algorithm of the application. Fig.5 shows the changes in operation of system:

- User requests are managed by the React library (A).
- User request is tracked by JavaScript code and transmitted to controller (1).
- Controller is connected to model (2).
- Model creates a request using JSON technology (3).
- Server has a special intermediary – the API that meets requests from the client, decodes the string in JSON format, parses the essence of the request, checks the headers for identification of the client who sent them.
- API creates the required request and sends it to Laravel (4).
- Controller Laravel TovarController framework transfers the data to the API (14).
- API encodes the data received from the controller and returns them to the client part (15).
- Client receives the data and adapts it to the user's display (16, 17).

The user browser only takes steps to display the content of the page on the screen. However, modern browsers contain enough resources to perform more complex tasks.

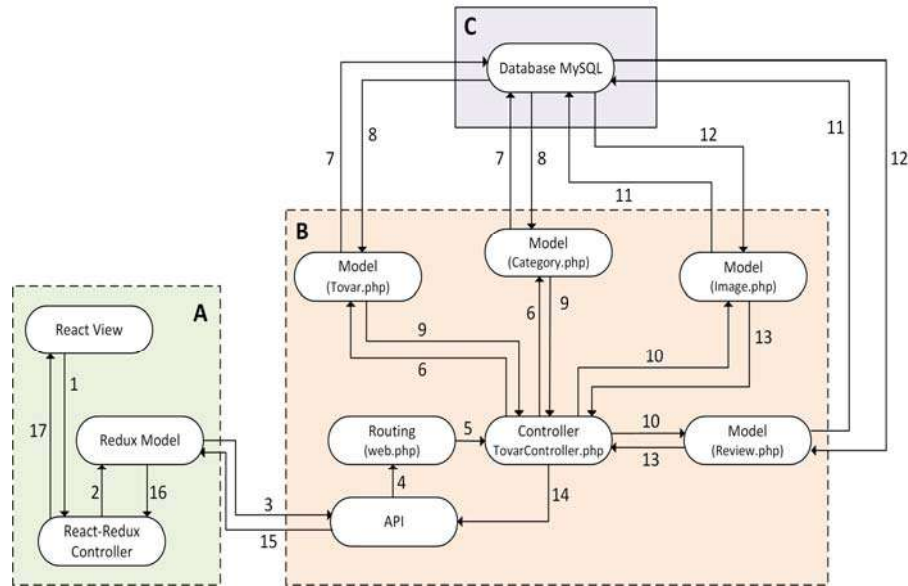


Fig. 5. Request processing on the system with React library

Having a system model using the React library and algorithms to change client-server interactions, let's move on to optimization of the system. General stages are the following:

- Measurement of page load time, external resources in a non-optimized system.
- Measurement of time of script execution.
- Implementation of the react library in the system.
- Configuring load balancing between server clusters using JavaScript code.
- Replacing the jQuery library with native JavaScript code.
- Setting the sequence of loading of third-party resources and basic JavaScript of the system.
- Setting up parallel load of resources.
- Providing interactive download.
- Minimizing js, css, html files. Image optimization.
- Setting of data caching. Setting up Gzip data archiving.
- Measurement of page load time, external resources after system optimization.
- Comparing optimized and non-optimized systems with slower internet connections.

The optimized system with Front-End technologies involves a large amount of JavaScript code, React libraries, which in turn may also have third-party modules, can primarily affect the first time the client is loaded and, moreover, the user's expectations, so let's provide interactive download.

Based on the classification above, the first thing that should appear in the client's browser is the information that the page loads. The more interactive this information, the more likely that the client will wait for the full page loading.

Another way to optimize web system is to minimize code, both HTML, CSS and JS. To minimize the code, there are some services that allow you to remove all the spaces, paragraphs and formatting of the code text.

However, this kind of optimization is too time consuming and requires every time the system is upgraded and changed. The solution to this problem is to use a special WebPack tool with specific settings.

Compressing files will make them smaller in size and speed up loading. Combining some files will affect the number of files, which will also speed up the download, since fewer new requests and answers are required to download the next files.

To optimize the reload of each page with similar content change the application style to Single Page Application. We have two views - user and admin. View user – welcome.blade.php, admin – admin.blade.php. Depending on authorization, certain components will run on a particular page. We make changes to controller code and now they return welcome.blade.php or admin.blade.php depending on user rights. Next step is changing of links on those that are controlled by JavaScript. A program code below is showing usage of the fetch function during creating a React request and receiving a response from it:

```
this.state = {
/* Initialization State, which will include all products,
by default - empty array */
  products: [],
}
componentDidMount() {          /* fetch API in action */
  fetch(`/category/${categoryName}`)
    .then(response => {
      return response.json();
    })
    .then(products => {
      //Fetched product is stored in the state
      this.setState({ products });
    });
}
```

When you create a query, response is provided, component changes the structure of DOM tree, and fetch function fills the pages with database data. Similarly, we are optimizing other pages and currently have a working system [15, 16].

The next optimization step is to create a balancing act between the cluster of servers and a client-side using the following algorithm:

- create a list of servers to connect;
- calculate the total number of possible servers;
- develop a method for random selection of the server;
- connect to the server;

- process the response from the server;
- in case of server failure write it in the list of non-working servers;
- make the same connection to another server;
- clear information about unavailable servers after a successful connection.

To do this, we implement a separate js file that contains an array listing the servers that are imported into the server selection program file with the following variables: brokenServers and randNum; the first is an empty array that contains non-working servers, the second is a random value. Next, we create a function that will determine a random number equal to the number of servers minus one (numbering in the array from zero). Next, check for a randomly selected number in the brokenServers array.

If the values are the same, the random number function is called [17]. It invokes the fetch function with the server address parameter that you want to connect to. The next step is to process the response from the server. If the server answered, the system continues to work, if not, the server index in the server array is written in the brokenServers array and a new server is selected for connection. Server selection code for the connection:

```
let brokenServers = [];
let randNum = 0;
function getRandomNum() {
    randNum = Math.floor(Math.random()
        * servers.length);
}
getRandomNum ();
if(brokenServers.length > 0) {
    for (let i=0; i < brokenServers.length; i++) {
        if(brokenServers[i] === randNum) {
            getRandomNum();
        }
        else {
            getFetch(this);
            break;
        }
    }
}
else {
    getFetch(this);
}
function getFetch(state) {
    fetch(servers[randNum] + 'api/')
        .then(response => {
            brokenServers = [];
            return response.json();
        }, function(reason) {
            brokenServers.push(randNum);
```

```

        return brokenServers;
    })
    .then(jsonResponse => {
        state.setState({categories: jsonResponse.categories});
        state.setState({products: jsonResponse.tovar});
    });
}

```

Test the work of balancing the server cluster on the client side. To do this, we specify two server addresses [14].

```

const servers = [
    "http://test-react-mag.loc/",
    "http://s1.test-react-mag.loc/"
];

```

The first address is the existing address where the system is hosted, the second is the missing non-response address. When connecting to the test-react-mag.loc server, the necessary data was received, otherwise, when connecting to the server located at s1.test-react-mag.loc, the server does not return an answer and its serial number is added to the non-working array servers. After successfully testing the load sharing between the servers on the client side, to complete this idea, you must deploy the system to multiple servers and specify their addresses instead of test-react-mag.loc and s1.test-react-mag.loc. This kind of balancing is characteristic only of the client side of the system. The averaged results of an experimental verification of the proposed system optimization method are presented in the table 1.

Table 1. The averaged results of an experimental verification of the proposed method

Parameters	Before optimization	After optimization
Load Time	3.606s	2.555s
First byte	0.539s	0.205s
Start render	1.900s	1.000s
First contentful paint	1.939s	0.997s
Speed index	3.030s	1.594s
Last painted hero	3.600s	1.866s
First CPU idle	> 1.939s	> 0.996s
Document complete		
Time	3.606s	2.555s
Requests	37	23
Bytes in	1.313 KB	1.017 KB
Fully loaded		
Time	3.853s	2.655s
Requests	38	24
Bytes in	1.313 KB	1.017 KB

As a result of implementation of the proposed web system optimization method we were able to obtain an improvement of the web system work of all considered criteria.

5 Conclusion

This paper describes methods for optimization of network web system which includes reduced load on server, developed sophisticated features that help improve user interface, save server traffic and funds. The scientific novelty of the paper lies in developed web system multicriteria optimization method which uses modern Front-End technologies. The practical value is optimized web system which uses a "smart" client, which can partially perform server logic, thereby providing fewer loads on the server and improve interactive properties of user interface. Optimization of system does not affect its features and consistency. The existing application architecture has been preserved, the conditions for the further improvement of system and its development have been created, and set tasks have been fully fulfilled.

References

1. Henderson, P.: Functional Programming Application and Implementation. Prentice Hall, 1980.
2. Tajima, M., Goto, K., Toyama, M.: Non-procedural generation of web pages with nested infinite-scrolls in superSQL. In: Proceedings of the 19th International Conference on Information Integration and Web-based Applications & Services, pp. 289-295 (2017). doi:10.1145/3151759.3151806
3. Larsen, H.L., Blanco, J.M., Pastor, R.P., Yager, R.R.: Using Open Data to Detect Organized Crime Threats: Factors Driving Future Crime. Springer. (2017)
4. Gorodnyaya, L.: On the Presentation of the Results of the Analysis of Programming Languages and Systems. In: CEUR Workshop Proceedings 2260, pp. 152-166 (2018)
5. Barki, H., Rivard, S., Talbot, J.: Toward an assessment of software development risk. *Journal of management information systems* 10(2), 203-225 (1993)
6. Zaninotto, F.: React is Slow, React is Fast: Optimizing React Apps in Practice. <https://medium.com/dailyjs/react-is-slow-react-is-fast-optimizing-react-apps-in-practice-394176a11fba> (2017) Accessed 07 Jan 2020
7. Ahmed, T.M., Bezemer, C.P., Chen, T.H., Hassan, A.E., Shang, W.: Studying the effectiveness of application performance management (APM) tools for detecting performance regressions for web applications: an experience report. In: IEEE/ACM 13th International Conference on Mining Software Repositories (MSR), pp. 1-12. IEEE (2016)
8. Elbaum, S., Karre, S., Rothermel, G.: Improving web application testing with user session data. In: 25th International Conference on Software Engineering IEEE Computer Society, pp. 49-59. IEEE (2003) doi:10.1109/ICSE.2003.1201187
9. Arora, A., Sinha, M.: Web application testing: A review on techniques, tools and state of art. *International Journal of Scientific & Engineering Research* 3(2), 1 (2012)
10. Atterer, R., Schmidt, A.: Tracking the interaction of users with AJAX applications for usability testing. In: Proceedings of the SIGCHI conference on Human factors in computing systems, 1347-1350 (2007)

11. Kirichek, G., Kurai, V.: Implementation quadtree method for comparison of images. In: 14th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering (TCSET), pp. 129–132. IEEE (2018). doi:10.1109/TCSET.2018.8336171
12. Zhongsheng, Q.: Test case generation and optimization for user session-based web application testing. *Journal of Computers* 5(11), 1655–1662 (2010)
13. Kirichek, G., Tymoshenko, V., Rudkovskyi, O., Hrushko, S.: Decentralized System for Run Services. In: Luengo D., Subbotin S., Arras P., Bodyanskiy Ye., Henke K., Izonin I., Levashenko V., Lytvynenko V., Parkhomenko A., Pester A., Shakhovska N., Sharpanskykh A., Tabunshchik G., Wolff C., Wuttke H.-D., Zaitseva E. (eds.) *Proceedings of the Second International Workshop on Computer Modeling and Intelligent Systems (CMIS-2019)*, Zaporizhzhia, Ukraine, April 15–19, 2019, CEUR-WS.org 2353, pp. 860–872 (2019)
14. Kirichek, G., Harkusha, V., Timenko, A., Kulykovska, N.: System for detecting network anomalies using a hybrid of an uncontrolled and controlled neural network. In: Kiv, A.E., Semerikov, S.O., Soloviev, V.N., Striuk, A.M. (eds.) *Proceedings of the 2nd Student Workshop on Computer Science & Software Engineering (CS&SE@SW 2019)*, Kryvyi Rih, Ukraine, November 29, 2019, CEUR-WS.org, pp. 138–148 (2019)
15. Qian, Z., Miao, H., Zeng, H.: A practical web testing model for web application testing. In: *Third International IEEE Conference on Signal-Image Technologies and Internet-Based System*, pp. 434–441. IEEE (2007). doi:10.1109/SITIS.2007.16
16. Polpong, J., Kansomkeat, S.: Syntax-based test case generation for web application. In: *2015 International Conference on Computer Communications and Control Technology (I4CT)*, pp. 389–393. IEEE (2015). doi: 10.1109/I4CT.2015.7219604
17. Kulykovska N.A, Timenko A.V. A Structure of Semantic Service in a Distributed Knowledge Based System. In: Luengo D., Subbotin S., Arras P., Bodyanskiy Ye., Henke K., Izonin I., Levashenko V., Lytvynenko V., Parkhomenko A., Pester A., Shakhovska N., Sharpanskykh A., Tabunshchik G., Wolff C., Wuttke H.-D., Zaitseva E. (eds.) *Proceedings of the Second International Workshop on Computer Modeling and Intelligent Systems (CMIS-2019)*, Zaporizhzhia, Ukraine, April 15–19, 2019, CEUR-WS.org 2353, pp. 533–544 (2019)

Decentralized Electronic Voting System Based on Blockchain Technology Developing Principals

Kateryna Isirova^[0000-0002-0250-7636], Anastasiia Kiian^[0000-0003-2110-010X],
Mariia Rodinko^[0000-0003-4692-9811] and Alexandr Kuznetsov^[0000-0003-2331-6326]

V. N. Karazin Kharkiv National University, 4 Svobody Sq., Kharkiv, 61022, Ukraine
KaterinaIsirova@gmail.com, nastyak931@gmail.com,
m.rodinko@gmail.com, kuznetsov@karazin.ua

Abstract. Electronic trust services are becoming an integral part of the information space. With the reliable implementation of basic services as an electronic signature and electronic authentication, it is possible to build more complex systems that rely on them, particularly the electronic voting system. In the paper, the new concept for developing a decentralized electronic voting system using blockchain technology is proposed. The two-level architecture provides a secure voting process without redundancy of existing (not based on blockchain) systems. The presented blockchain-based voting protocol has six steps that ensure all requirements that are put forward to such types of protocols including voting transparency and anonymity.

Keywords: Decentralized Electronic Voting System, Decentralized identification, Electronic Voting Protocol, Blockchain Technology.

1 Introduction and Formal Problem Statement

Electronic trust services are becoming an integral part of the information space. Their use is governed by Regulation (EU) No 910/2014 of the European Parliament and of the Council of 23 July 2014 on electronic identification and trust services for electronic transactions in the internal market and repealing Directive 1999/93/EC [1] which establishes terms and conditions. With the reliable implementation of such basic services as an electronic signature and electronic authentication, it is possible to build more complex systems that rely on them, for example, an electronic voting system.

Remote (electronic) voting has many advantages. It is assumed that they are more convenient for end-users because people can vote without leaving home; this increases the activity of voters. Maintenance of electronic voting is cheaper: instead of permanently printing ballots, it's enough to develop a system once. In addition, the assumption that no one can interfere with the program on the voting device implies that electronic voting is less susceptible to corruption, administrative pressure, and human factors. However, this raises a number of specific problems that hinder the integrity of elections. Remotely, it is much more difficult to authorize a voter or make

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

sure that no one has influenced the voting process. On the other hand, the Internet provides more opportunities for checking by ordinary voters whether the voice is correctly taken into account. Currently, electronic voting is fully legal or partially applicable in many countries of the world [2]. Since more and more people are involved in them, the need for safer and more efficient methods for their implementation is increasing, which is what special cryptographic protocols are designed for [3-8].

It should be noted that today the developing process of any system has to take into account the evolution of quantum computers and as a result the growth of computational speed. In the conditional of current cyber threats secure of the system should not base only on key parameters cryptographical secure [9-11]. Important point is to ensure the resilience of the system. From this point of view blockchain technology might be useful.

The main purpose of this paper is to formulate the development principles for a decentralized e-voting system that would prevail over existing e-voting systems without a decentralized structure.

2 Electronic Voting System Principals Structure

The system of electronic voting is a set of interrelated rules, methods, processes, tools, and technologies, as well as legal norms that together provide and regulate the remote legitimate voting of authorized users (voters).

Components (subsystems / levels) of the electronic voting system (see Fig. 1):

- legal level (laws and other regulatory documents);
- organizational level (e-voting system architecture);
- process level (processes for participants);
- technological level (methods, tools, protocols, technologies).

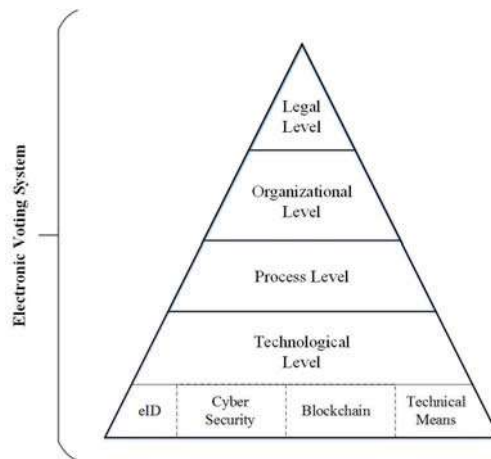


Fig. 1. Levels of electronic voting system Requirements for secret e-voting systems [12].

Required:

- no one except the voter should know his choice;
- only legitimate members can vote, and moreover only once;
- the decision of the voter cannot be secretly or explicitly changed by anyone (except, perhaps, by himself). In addition, the desired requirements are set out [12]:

Desirable:

- each legitimate participant can check whether his voice is correctly counted;
- each legitimate participant may change his mind and change his choice within a certain period of time;
- the system should be protected from the sale of votes by voters;
- in case the vote is incorrectly counted, each legitimate participant can report this to the system without revealing his anonymity;
- it is impossible to track where the voter remotely voted from;
- operator authentication;
- you can find out who participated in the vote, and who - no;
- maintaining the system should not require a lot of resources;
- the system must be fault tolerant in case of technical malfunctions (loss of power supply), unintended (loss of the key by the voter) and malicious (intentionally disguising itself as another voter, DoS / DDoS) attacks.

The major threats to systems of this type are:

- legitimate voter cannot vote;
- loss of voter anonymity;
- registration of non-existent voters;
- the use of blank ballots that registered but did not participate in the election.

3 The architecture of the decentralized voting system

The architecture of the decentralized e-voting system is two-level and consists of two intersecting blockchain networks, the lower network is a decentralized electronic identification infrastructure (DI eID), and the upper network is a decentralized infrastructure for voting itself and counting the results (DI voting) (see Fig. 2).

3.1 Decentralized Electronic Identification Infrastructure (DI eID)

This infrastructure should provide a procedure for the reliable identification of users and a list of legitimate voters' establishment. It consists of providers of identification services (hereinafter - IdP providers). It is necessary to ensure the implementation of the identification using:

- BankID;
- MobileID;

- e-passport of the citizen;
- electronic signature (including both software and hardware implementation (token))

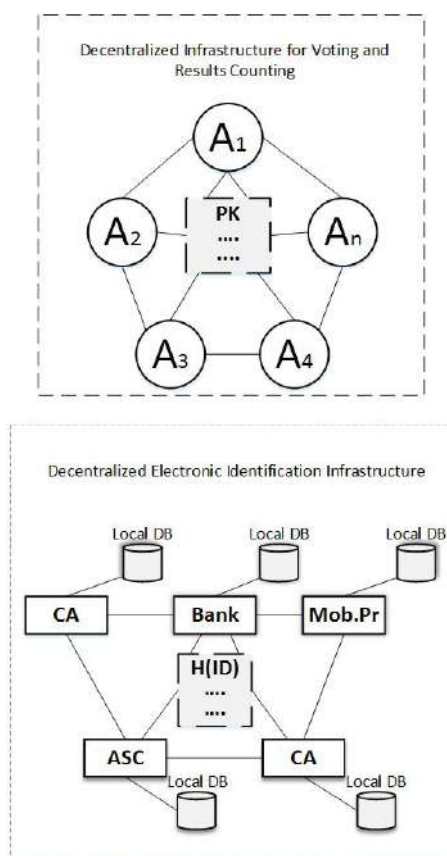


Fig. 2. Decentralized Electronic Voting System Architecture

According to the requirements, the following entries may act as identity providers:

- Bank institutions;
- Mobile operators;
- Migration Service Centers (Administrative Service Centers - ASC);
- Certification authorities of the national digital signature system.

Provisions for the operation of providers are established by the Law of Ukraine “On Electronic Trust Services”, implemented by the EU Regulation and other international and national normative documents.

Each identity provider has a pre-generated local database of its users, which contains their identities and possibly local IDs. The responsibility for the secure storage and correct use of local databases rests with identity providers.

In order to organize the identification infrastructure within a decentralized electronic voting system, the identity providers are combined into a separate private permissioned blockchain network. In this network, each of the identity provider acts as a validator node. It should be noted that complex and energy-intensive consensus protocols are not required for such a network because the network connects trusted ("honest") nodes.

3.2 Decentralized Infrastructure for Voting and Results Counting

The infrastructure should provide for the process of the remote voting of registered (authorized) legitimate voters and the process of results counting. In addition, generation processes for voters' wallets and candidates' wallets should be organized in this infrastructure. For the organization voting infrastructure under the decentralized electronic voting system, the representative offices responsible for conducting the election process ($A_1, A_2, \dots, A_n$), like identification providers, are combined into a separate private blockchain network, in which each of A_i acts as a validator node - collectively they represent a decentralized Agency (A). Similar to the upper blockchain network, the lower one also does not need to use complex and energy-intensive consensus protocols because the network connects trusted ("honest") nodes. Validator nodes form purses for legitimate voters and carry out voter authentication. They are also responsible for the process of wallet generation for alternatives (candidates).

3.3 Voting Protocol in a Decentralized Electronic Voting System

The voting protocol in a decentralized electronic voting system consists of the following steps:

1. Formation of the of legitimate voters' list in a decentralized electronic identification infrastructure;
2. Generation of legitimate voters' wallets in a decentralized infrastructure for remote voting and results counting;
3. Candidates registration in a decentralized infrastructure for remote voting and results counting;
4. Voters' authentication in decentralized infrastructure for remote voting and results counting;
5. Voting in a decentralized infrastructure for voting and results counting;
6. Counting of votes in a decentralized infrastructure for voting and results counting.

The implementation of this protocol using blockchain technology allows depending on the needs of the target system to change the order of some stages (basically the fourth and fifth) without loss of reliability. The direct sequence (fourth to fifth) implies that only authenticated users (legitimate voters) are allowed to vote. The reverse

sequence (fifth to fourth) allows participation in the voting process of potential violators (illegitimate voters), but due to the peculiarities of the implementation of the transaction consensus mechanism, and accordingly, the votes of illegitimate users will not be taken into account. This is based on the assertion that in any blockchain network, a transaction is considered validated only if both conditions are fulfilled:

1. the format and signatures of the transaction are verified;
2. validator nodes have reached consensus on including this transaction in the block chain.

The principles of building a decentralized infrastructure for remote voting and counting results do not allow validation nodes to include a transaction from an illegitimate voter in the blockchain since the first condition will not be fulfilled (transaction signature will not be valid).

Stage One (Formation of the of legitimate voters' list in a decentralized electronic identification infrastructure).

Forming lists of legitimate voters occurs in the lower blockchain network. Each potential voter independently generates a key pair (SK; PK). Then he/she sends a request to be included in the voters' list to one of his/her available identity provider, in which he/she provides his/her with his identification information and public key.

The format of the request depends on the available communication channels between the voter and the identity provider. It can be made remotely via the Internet provided there is a reliable communication channel (see Fig. 3b), or such an identification request can be made personally by a potential voter within the identity provider controlled zone. If the request is made remotely, the responsibility for complying with the key pair generation rules rests with the user. If the request is made personally within the controlled zone, the identity provider is responsible for complying with the conditions of the generation of the key pair of the user.

If a potential voter already has generated key pair as required by one of the identification providers, he or she may use it. In this case, the public key certificate must be included in the request to the provider (see Fig. 3a).

If a positional voter does not have a local ID in any of the identity provider database, he or she must undergo a primary identification procedure with one of the identity providers and only then be included in the list of legitimate voters (see Fig. 3c). The initial identification procedure should be conducted in accordance with the rules of a certain identity provider.

Thus, when the time allotted for forming legitimate voters' list has run out, an anonymous (depersonalized) list of potential legitimate voters is created in the lower blockchain, and the Agency receives a list of all registered legitimate voters, but voters remain anonymous. The identification processes for different types of users are shown on the figures 4a-4c.

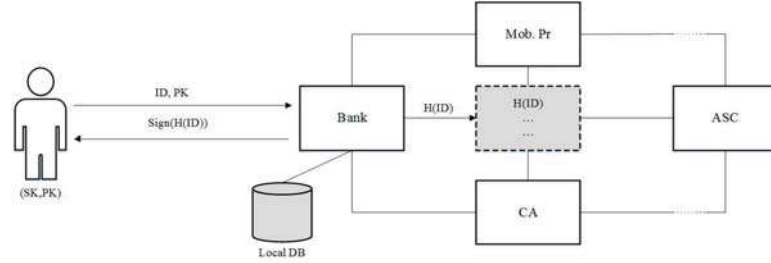


Fig. 3a. Identification procedure based on user's local ID

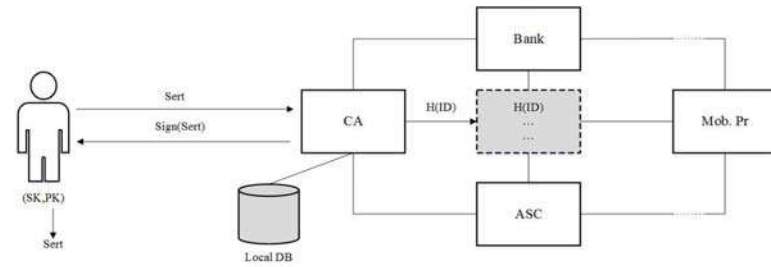


Fig. 3b. Identification procedure based on user's Certificate

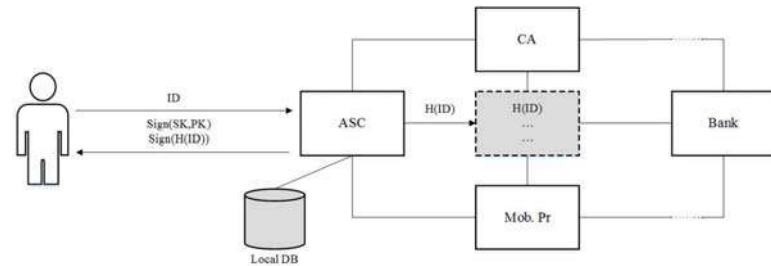


Fig. 3c. Identification procedure based on user's ID

Depersonalized Identity Format should be established in the following form:

$$H(ID),$$

where ID is the user ID in the following sequence: (series and passport number of the citizen); and H is a cryptographic hash function

The ID format must be the same for all identity providers. This condition makes it impossible for voters to re-register with different identity provider.

Stage Two (Generation of legitimate voters' wallets in a decentralized infrastructure for remote voting and results counting).

The process of generating a voter wallet (see Fig. 5) is initiated by the validator node of the upper blockchain network when it receives the public key from any of the identity providers in the form of a transaction. The voter wallet's initial balance is 0.

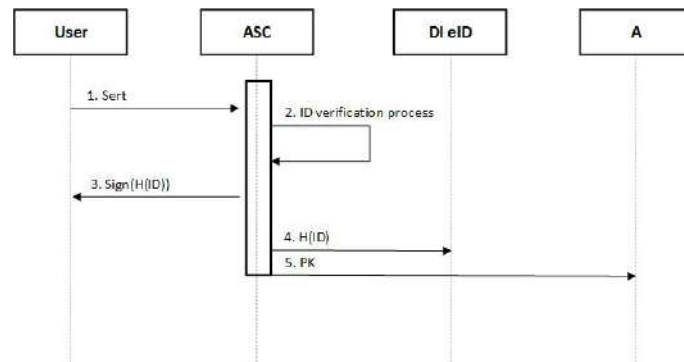


Fig. 4a. Identification process based on user's Certificate

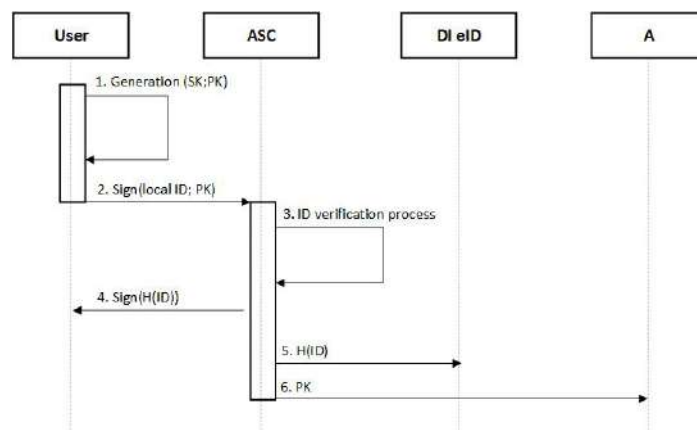


Fig. 4b. Identification process based on user's local ID

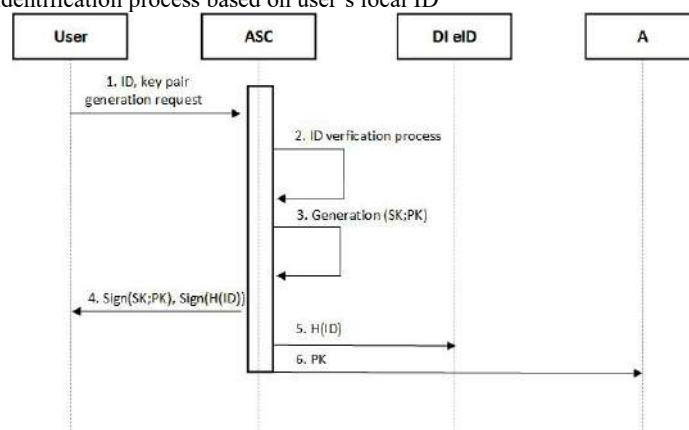


Fig. 4c. Identification process based on user's ID

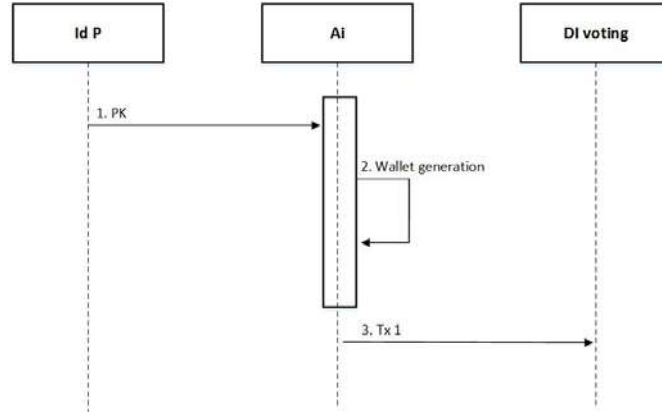


Fig. 3. User's wallet generation process

Stage Three (Candidates registration in a decentralized infrastructure for remote voting and results counting).

Candidates are registered in the top blockchain network. Hereinafter, the Agency will be understood to mean the totality of territorial polling stations combined into a separate private permissioned blockchain.

Responsibility for the candidates' registration rests on the Agency (upper blockchain validators) (see Fig. 6).

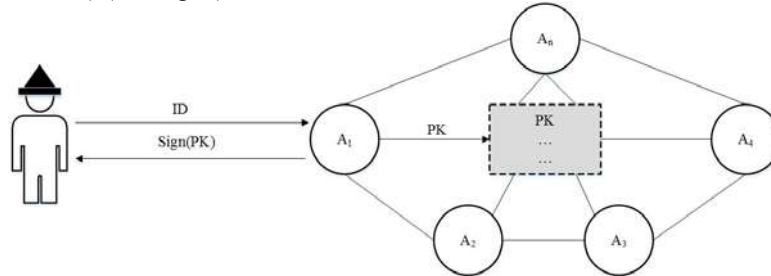


Fig. 4. Candidate registration procedure

Representatives of the Agency are responsible for conducting the initial identification of the candidates and initiate a transaction for the inclusion of the candidate which means the generation of the candidate wallet with zero starting balance (see Fig. 7).

Stage Four (Voters' authentication in decentralized infrastructure for remote voting and results counting).

A voter who has received confirmation from the identity provider for admission to the voting process and wishes to participate in the voting will contact one of the

Agency's nodes for the authentication procedure (see Fig. 8). The user can only authenticate with a private key (SK), provided that there is a corresponding public key (PK) in the Agency's blockchain network.

If the authentication process (see Fig. 9) was successful, the balance of the voter's wallet is increased by 1 token.

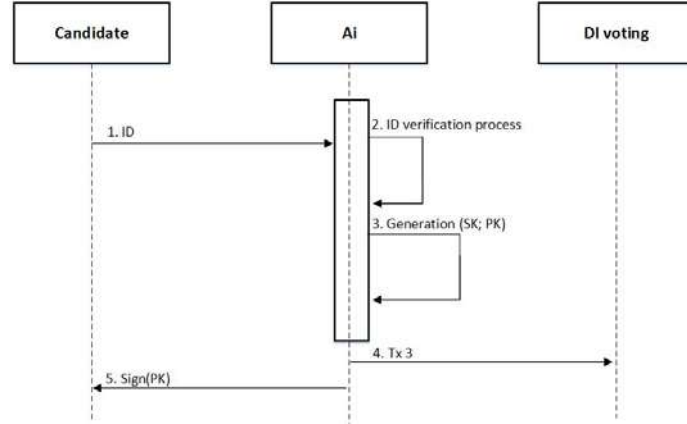


Fig. 5. Candidate wallet generation

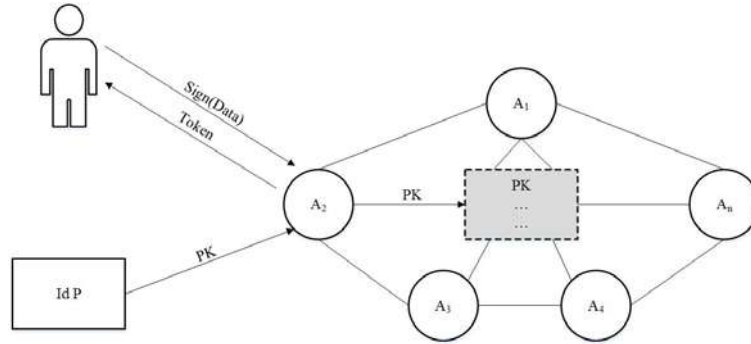


Fig. 6. User's authentication procedure

Stage Five (Voting in a decentralized infrastructure for voting and results counting).

Voters who have undergone an authentication procedure will make a choice by forwarding a token to one of the wallets which corresponds to the registered candidate by forming a corresponding transaction, which they sign with their own private key (see Fig. 10).

Stage Six (Counting of votes in a decentralized infrastructure for voting and results counting).

Vote counting is done automatically. The results will become available to everyone after the voting time has elapsed.

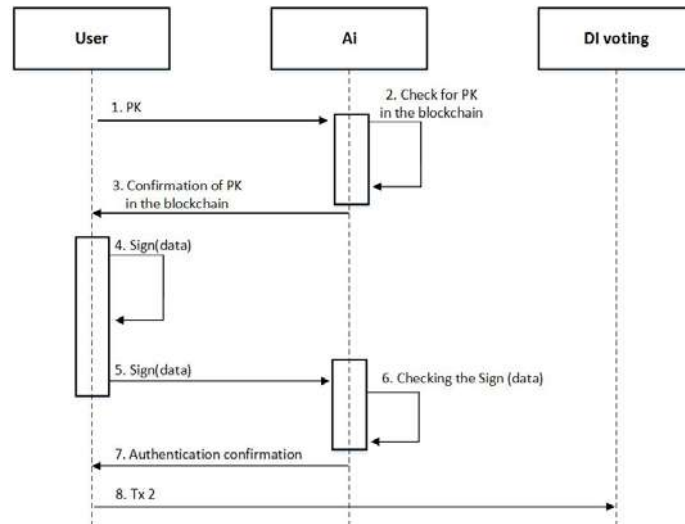


Fig. 7. User's authentication process

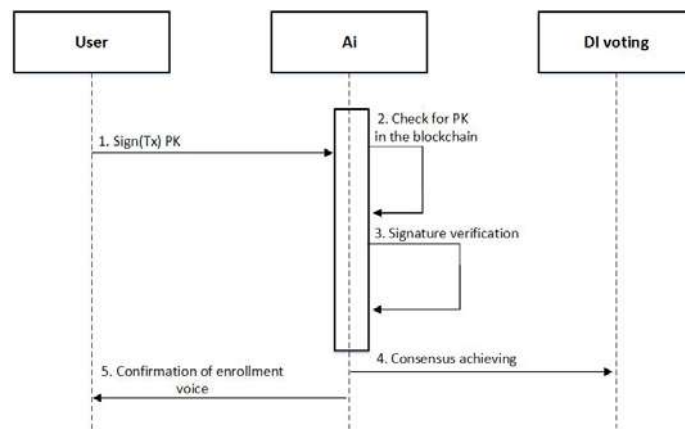


Fig. 8. Voting process

4 Conclusions

In the face of cyber threats, the deployment of reliable electronic services systems becomes an important task. The analysis showed that blockchain technology can be useful for this purpose. Particularly, for developing electronic voting systems. Classical systems do not meet all desired requirements for voting systems (for example, a

voter cannot check whether his voice is correctly taken into account and, if necessary, inform the authorized bodies about this).

In the paper, the new concept for developing a decentralized electronic voting system using blockchain technology is proposed. The two-level architecture provides a secure voting process without redundancy of existing (not based on blockchain) systems. The presented blockchain-based voting protocol has six steps that ensure all requirements that are put forward to such types of protocols including voting transparency and anonymity. Proposed blockchain-based approach has several advantages: central trust point absence, as a result there is no directed attack aim; reducing material costs for each stage of voting (since there is no need to print ballots, deliver them to polling stations).

Moreover, it should be noted that blockchain technology is more convenient for switching to post-quantum crypto primitives. Such an opportunity is also an important advantage in conditions of rapid evolution of quantum computers. It can also be used in other important applications [13-17].

References

1. Regulation (EU) No 910/2014 of the European Parliament and of the Council of 23 July 2014. <https://www.eid.as/Regulation>
2. E-voting world map. <https://www.e-voting.cc/en/it-elections/world-map/>
3. Menezes, A.J., van Oorschot, P.C., Vanstone, S.A.: Handbook of Applied Cryptography (2018). doi: 10.1201/9780429466335
4. Kuznetsov, O., Potii, O., Perepelitsyn, A., Ivanenko, D., Poluyanenko, N.: Lightweight Stream Ciphers for Green IT Engineering. In: Green IT Engineering: Social, Business and Industrial Applications. pp. 113–137. Springer International Publishing (2018). doi: 10.1007/978-3-030-00253-4_6
5. Gorbenko, Yu.I., Isirova, K.V.: Improved mechanism of one-time keys for post-quantum period based on the hashing functions. Telecommunications and Radio Engineering. Volume 77, Issue 14, 1277–1296 (2018)
6. Andrushkevych, A., Gorbenko, Y., Kuznetsov, O., Oliynykov, R., Rodinko, M.: A Prospective Lightweight Block Cipher for Green IT Engineering. In: Green IT Engineering: Social, Business and Industrial Applications. pp. 95–112. Springer International Publishing (2018). doi: 10.1007/978-3-030-00253-4_5
7. Gorbenko, I., Kuznetsov, A., Gorbenko, Y., Vdovenko, S., Tymchenko, V., Lutsenko, M.: Studies on Statistical Analysis and Performance Evaluation for Some Stream Ciphers. International Journal of Computing. 18(1), 82–88 (2019)
8. Bernstein, D.J., Buchmann, J., Dahmen, E. eds: Post-Quantum Cryptography. Springer Berlin Heidelberg (2009)
9. Pass, R., Seeman, L., Shelat, A.: Analysis of the blockchain protocol in asynchronous networks. Annual International Conference on the Theory and Applications of Cryptographic Techniques. 643-673 (2017)
10. Isirova, K., Potii, O.: Decentralized public key infrastructure development principles. In: 2018 IEEE 9th International Conference on Dependable Systems, Services and Technologies (DESSERT). IEEE (2018). doi:10.1109/dessert.2018.8409149
11. Kovalchuk, L., Kaidalov, D., Nastenko, A., Rodinko, M., Shevtsov, O., Oliynykov, R.: Decreasing Security Threshold Against Double Spend Attack in Networks with Slow Syn-

- chronization. In: IEEE INFOCOM 2019 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS). IEEE (2019). doi: 10.1109/INFOCOMW.2019.8845301
12. Nurmi, H., Salomaa, A.: Conducting secret ballot elections in computer networks: Problems and solutions. *Annals of Operations Research*. 51, 185–194 (1994). doi:10.1007/bf02032763
 13. Bondarenko, S., Liliya, B., Oksana, K., & Inna, G.: Modelling instruments in risk management. *International Journal of Civil Engineering and Technology*. 10(1), 1561-1568 (2019)
 14. Runovski, K., Schmeisser, H.: On the convergence of fourier means and interpolation means. *Journal of Computational Analysis and Applications*. 6(3), 211-227 (2004)
 15. Tkach, B.P., Urmancheva, L.B.: Numerical-analytic method for finding solutions of systems with distributed parameters and integral condition. *Nonlinear Oscillations*. 12, 113–122 (2009). doi:10.1007/s11072-009-0064-6
 16. Chornei, R.K., Daduna V.M., H., Knopov, P.S.: Controlled Markov Fields with Finite State Space on Graphs. *Stochastic Models*. 21, 847–874 (2005). doi:10.1080/15326340500294520
 17. Kuznetsov, A., Shekhanin, K., Kolhatin, A., Kovalchuk, D., Babenko, V., Perevozova, I.: Performance of Hash Algorithms on GPUs for Use in Blockchain. In: 2019 IEEE International Conference on Advanced Trends in Information Theory (ATIT). IEEE (2019). doi: 10.1109/ATIT49449.2019.9030442

Linguistic Analysis Method of Ukrainian Commercial Textual Content for Data Mining

Oleg Bisikalo¹[0000-0002-7607-1943], and Victoria Vysotska²[0000-0001-6417-3689]

¹Vinnytsia National Technical University, Vinnytsia, Ukraine
obisikalo@gmail.com

²Lviv Polytechnic National University, Lviv, Ukraine
Victoria.A.Vysotska@lpnu.ua

Abstract. This article deals with the scientific and practical task of automatically detecting significant keywords and rubricating Ukrainian content in Internet systems based on the method of linguistic analysis of text information. The article presents theoretical and experimental substantiation of the method of linguistic analysis of Ukrainian content using Porter's stemming. The method is aiming to automatically detect significant keywords of Ukrainian content on the basis of the proposed formalization of components of analysis - grammatical (grapheme), morphological, syntactic, semantic, referential, and structural.

Keywords: Text, Ukrainian, Algorithm, Content Monitoring, Keywords, Content analysis, Porter's Stemmer, Linguistic Analysis, Syntactic Analysis.

1 Introduction

In practical terms, the analysis of the symbolic level of the organization of naturalistic text is limited to the separation of syntactic punctuation from the word itself, the allocation of abbreviations, abbreviations, etc. Analysis of existing texts shows that at the level of sign organization of the text one is using the descriptive capabilities of the semiotic system to encode knowledge about fragments of real reality. For instance, the use of quotation marks (for example, the movie theater "Star") indicates that the token in quotation marks cannot be considered in the meaning given in the dictionary. The proper names given in the text may coincide with the spelling of common words, but have different meanings (for example, the group Black September, Black Friday, student Sophia Vovk (wolf), assistant Andriy Krolik (rabbit), teacher Nadiya Kohut (cock), singer Katya Chile, singer Vinnytska Alona, Actor David Duchovny as Fox William Mulder, Liberty Avenue, May 1 Street, Actress Sarah Gabriel, Actress Nastya Zadorozhna, etc.). In addition, a number of tokens in the text are not subject to the grammatical rules of the language, but act as semantic units of sign level (for example, the number 30, the percentage value of 15%, the reduction of millions, thousand or kg, etc.). These features of naturalistic text make it necessary to develop a significant level of text organization as the initial stage of building a model for under-

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

standing text. A linguistic method of processing textual information to automatically detect meaningful keywords consists of six steps.

1. Grammatical (graphemic) analysis of textual content that is, parsing text with regard to the features of graphs of different languages.
2. Morphological analysis of textual content.
3. Syntax analysis of textual content.
4. Semantic analysis of textual content.
5. Reference analysis for the formation of interphase unities.
6. Structural analysis of textual content.

The input of the graphemic (two-graphemic) analysis or parsing step is the current text file and the a priori reference models (lines and characters). Separation of such units of text as a name, designation, title, etc. allows us to identify at this stage some functional elements of the structure of concepts. Therefore, it is advisable to begin with an analysis of character-level text to solve the urgent problem of forming effective domain-specific knowledge recognition procedures. The electronic component of this is electronic dictionaries of abbreviations, geographical names, names. This approach is caused by the diversity of sign (grapheme) representation of lexical units in the text, which defines their different semantic functions in one context or another. For the automated processing of naturalistic information, it is also essential to define the structure of the text - to separate service information, highlight paragraphs, headings, and more. The text is considered as a sort of organized sequence of lines and graphemes.

2 Relate the Highlighted Issue to Important Scientific and Practical Work

The article deals with the scientific and practical task of automatically detecting significant keywords and rubricating Ukrainian-language content in Internet systems based on the method of linguistic analysis of textual information. The work was performed within the framework of joint scientific researches of the Department of Information Systems and Networks of the Lviv Polytechnic National University on the topic «Research, development and implementation of intelligent distributed information technologies and systems based on database resources, data warehouses, data spaces and knowledge in order to accelerate the formation processes of modern information society», as well as the department of automation and information-measuring technique of Vinnytsia National Technical University within spine Research Center of Applied and Computational Linguistics. The results of the research were carried out within the framework of the state budget research works on the topics "Development of methods, algorithms and software for modeling, designing and optimization of intellectual information systems based on Web technologies «WEB» and "Intelligent information technology of image analysis of text and synthesis of integrated knowledge base language content". Scientific research was also carried out

within the framework of the initiative topics of the ISM Department of Lviv Polytechnic National University on the development of intelligent distributed systems based on an ontological approach to integrate information resources.

3 Analysis of Recent Research and Publications

Text content (article, commentary, book, etc.) contains a considerable amount of data in natural language, some of which is abstract [1-7]. The text is presented as a unified sequence of character units, the main properties of which are information, structural and communicative connectivity / integrity, which reflects the content / structure of the text [8-22]. The method of text processing is linguistic content analysis (e.g., comments, forums, articles, etc.) [23-30]. The process of text processing divides the content into tokens using finite state machines (Fig. 1).

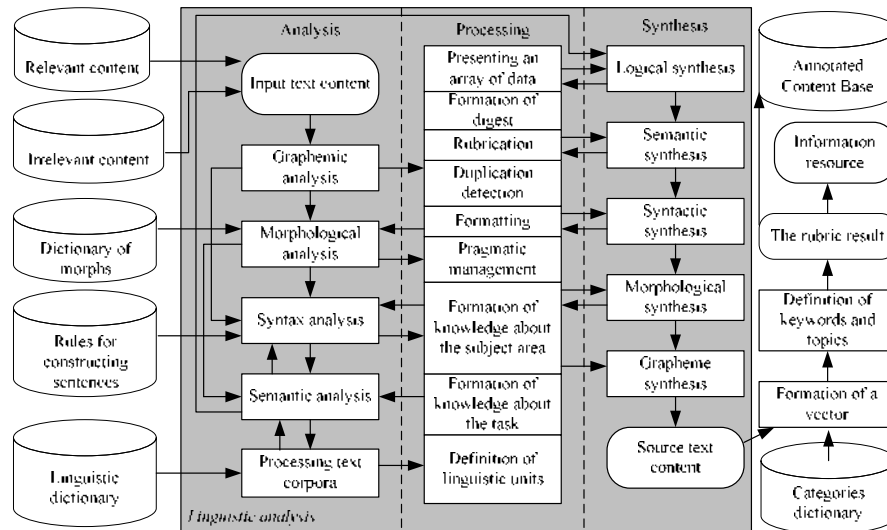


Fig. 1. Structural diagram of linguistic analysis of textual content

4 Analysis of Scientific Results

As a functional-semantic-structural unity, the text conforms to the rules of construction, reveals the regularities of content and formal connection of constituent units. Cohesiveness is manifested through external structural indicators and formal dependence of the text components, and integrity through thematic, conceptual and modal dependence. Integrity leads to a meaningful and communicative organization of text, and coherence to a form, a structural organization. Therefore, it is proposed to analyze the multilevel content structure in the analysis: linear sequence of characters; linear

sequence of morphological structures; linear sequence of sentences; a network of interconnected unities (Alg. 1).

Algorithm 1. Linguistic analysis of textual content.

Stage 1. Grammatical (graphemic) analysis of textual content C_1 .

Step 1. Divide textual commercial content $C_1 \rightarrow C_2$ into sentences and paragraphs.

Step 2. Divide the content character chain C_2 into words.

Step 3. Allocate numbers, numbers, dates, unchanged turns, and content cuts C_2 .

Step 4. Publish non-text content C_2 characters.

Step 5. Formation and analysis of linear content enhancement technology for content C_2 .

Stage 2. Morphological analysis of textual content C_2 .

Step 1. Obtained the basic (word form with cut offs).

Step 2. A grammatical category is formed for different words (collection of grammatical meanings: rarity, deviation, deviation).

Step 3. Formation of linear ability of morphological structure.

Stage 3. Syntax analysis $\alpha_4 : (C_2, U_K, T) \rightarrow C_3$ of textual content C_2 .

Stage 4. Semantic analysis of textual content C_3 .

Step 1. The word matches the semantic classes in the dictionary.

Step 2. Selection of morphosemantic alternatives required for this review.

Step 3. Cut the words into a single structure.

Step 4. Generate an orderly number of superposition entries with basic lexical functions and semantic classes. The accuracy of results is the most commonly used / corrective dictionary.

Stage 5. Reference analysis for interphase unities.

Step 1. Contextual analysis of text commercial content C_3 . With it, the resolution of local references (the one that is, his) is realized and the expression of the expression is the kernel of unity.

Step 2. Thematic analysis. Separation of statements on a theme and a rheum allocates thematic structures which are used, for example, at formation of a digest.

Step 3. Determine the regular repetition, synonymization and re-nomination of keywords; the identity of the reference, that is, the ratio of words to the subject of the image; presence of implication based on situational connections.

Stage 6. Structural analysis of textual content C_3 . The prerequisites for use are a high degree of coincidence of terms of unity, a discursive unit, a sentence in a semantic language, utterance, and an elementary discursive unit.

Step 1. Identify the basic set of rhetorical connections between content unities.

Step 2. Building a nonlinear unity network. The openness of a link set involves its extension and adaptation to analyze the structure of the text C_3 .

Let us consider in detail each of the stages of the proposed algorithm.

Step 1. Grammatical (graphemic) analysis of textual content. The grapheme is called the minimum content unit of written text. The objective of this level of recognition is to build a formalized representation of the grapheme structure of the text and to

Щ, Ъ, Ю, Я, Ё, І, Ї, Ы, Э, Ь, а, б, в, г, д, е, ж, з, и, й, к, л, м, о, н, п, р, с, т, у, ф, х, ц, ч, ш, щ, ъ, ю, я, ё, і, ї, ы, э, ь, '>' and a list of tuples for those defined in Table 2 elements of the set of reference models of graphemes:

Table 1. Rules for recognizing lines in text in the grapheme stage

Line name	First position	Last position	All positions
Empty string (<i>EmpStr</i>)	–	–	Gap
Full string (<i>FulStr</i>)	Symbol	Symbol	–
Incomplete right (<i>IncRgt</i>)	Symbol	Gap	–
Incomplete left (<i>IncLgt</i>)	Gap	Symbol	–
Symmetric incomplete (<i>SmtInc</i>)	Gap	Gap	–

Table 2. Denomination of the elements of the set of reference models of graphemes

N	Name	Abbr.	N	Name	Abbr.
1	Grammar	G	35	Ukrainian small letter	<i>Usm</i>
2	Alphabet	<i>V</i>	36	Russian capital letter	<i>Rcp</i>
3	Term	<i>T</i>	37	Russian small letter	<i>Rsm</i>
4	Initial character	<i>S</i>	38	Consonant letter	<i>Cnl</i>
5	Production rules	<i>P</i>	39	Vowel letter	<i>Vwl</i>
6	Symbol	<i>Sb</i>	40	Latin capital consonant letter	<i>Lcc</i>
7	Space	<i>Sp</i>	41	Latin small consonant letter	<i>Lsc</i>
8	Digit	<i>Dgt</i>	42	Latin capital vowel letter	<i>Lcv</i>
9	Special symbol	<i>Ssb</i>	43	Latin small vowel letter	<i>Lsv</i>
10	Syntactic sign	<i>Ssg</i>	44	Cyrillic capital consonant letter	<i>Ccc</i>
11	Letter	<i>Ltr</i>	45	Cyrillic small consonant letter	<i>Csc</i>
12	Latin letter	<i>Lat</i>	46	Cyrillic capital vowel letter	<i>Ccv</i>
13	Cyrillic letter	<i>Cyr</i>	47	Cyrillic small vowel letter	<i>Csv</i>
14	English alphabet	<i>Eng</i>	48	English capital consonant letter	<i>Ecc</i>
15	German alphabet	<i>Ger</i>	49	English small consonant letter	<i>Esc</i>
16	Polish alphabet	<i>Pol</i>	50	English capital vowel letter	<i>Ecv</i>
17	Ukrainian alphabet	<i>Ukr</i>	51	English small vowel letter	<i>Esv</i>
18	Russian alphabet	<i>Rus</i>	52	German capital consonant letter	<i>Gcc</i>
19	Official symbol	<i>Os</i>	53	German small consonant letter	<i>Gsc</i>
20	Brackets	<i>Bsb</i>	54	German capital vowel letter	<i>Gcv</i>
21	Mathematical symbol	<i>Msb</i>	55	German small vowel letter	<i>Gsv</i>
22	Capital letter	<i>Cpl</i>	56	Polish capital consonant letter	<i>Pcc</i>
23	Small letter	<i>Sml</i>	57	Polish small consonant letter	<i>Psc</i>
24	Latin capital letter	<i>Lcp</i>	58	Polish capital vowel letter	<i>Pcv</i>

21. $Vwl := \langle Ecv \cup Esv \cup Gcv \cup Gsv \cup Pcv \cup Psv \cup Ucv \cup Usv \cup Rcv \cup Rsv \rangle$ is a set of uppercase and lowercase letters of the respective languages;
22. $Lcp := \langle Lcc \cup Lcv \cup Q \cup V \cup X \rangle$ is Latin capital letters;
23. $Lsm := \langle Lsc \cup Lsv \cup q \cup v \cup x \rangle$ is Latin lowercase letters;
24. $Ccp := \langle Ccc \cup Ccv \cup Ъ \cup Ы \rangle$ is Cyrillic capital letters;
25. $Csm := \langle Csc \cup Csv \cup ь \cup ы \rangle$ is Cyrillic lowercase letters;
26. $Ecp := \langle Lcc \cup Lcv \cup Q \cup V \cup X \rangle$ is English capital letters;
27. $Esm := \langle Lsc \cup Lsv \cup q \cup v \cup x \rangle$ is English lowercase letters;
28. $Gcp := \langle Lcc \cup Lcv \cup \ddot{A} \cup \ddot{O} \cup \ddot{U} \cup Q \cup V \cup X \rangle$ is German capital letters;
29. $Gsm := \langle Lsc \cup Lsv \cup \ddot{a} \cup \ddot{o} \cup \ddot{u} \cup \beta \cup q \cup v \cup x \rangle$ - German lowercase letters;
30. $Pcp := \langle Lcc \cup Lcv \cup \text{Ą} \cup \text{Ć} \cup \text{Ę} \cup \text{Ł} \cup \text{Ń} \cup \text{Ó} \cup \text{Ś} \cup \text{Ź} \cup \text{Ż} \rangle$ is Polish capital letters;
31. $Psm := \langle Lsc \cup Lsv \cup \text{ą} \cup \text{ć} \cup \text{ę} \cup \text{ł} \cup \text{ń} \cup \text{ó} \cup \text{ś} \cup \text{ź} \cup \text{ż} \rangle$ is Polish lowercase letters;
32. $Ucp := \langle Ccc \cup Ccv \cup \text{Є} \cup \text{І} \cup \text{Ї} \cup \text{Ї} \rangle$ is Ukrainian capital letters;
33. $Usm := \langle Csc \cup Csv \cup \text{є} \cup \text{і} \cup \text{ї} \cup \text{ї} \rangle$ is Ukrainian lowercase letters;
34. $Rcp := \langle Ccc \cup Ccv \cup \text{И} \cup \text{Э} \cup \text{Ђ} \rangle$ is Russian capital letters;
35. $Rsm := \langle Csc \cup Csv \cup \text{ы} \cup \text{э} \cup \text{ѣ} \rangle$ is Russian lowercase letters;
36. $Lcc := \langle B \cup C \cup D \cup F \cup G \cup H \cup J \cup K \cup L \cup M \cup N \cup P \cup R \cup S \cup T \cup W \cup Z \rangle$ is terminal Latin capital consonant letters;
37. $Lcv := \langle A \cup E \cup I \cup O \cup U \cup Y \rangle$ is terminal Latin capital loud letters;
38. $Lsc := \langle b \cup c \cup d \cup f \cup g \cup h \cup j \cup k \cup l \cup m \cup n \cup p \cup r \cup s \cup t \cup w \cup x \cup z \rangle$ is terminal Latin lowercase consonant letters;
39. $Lsv := \langle a \cup e \cup i \cup o \cup u \cup v \cup y \rangle$ is terminal Latin lowercase loud letters;
40. $Ccc := \langle Б \cup В \cup Г \cup Д \cup Ж \cup З \cup К \cup Л \cup М \cup Н \cup П \cup Р \cup С \cup Т \cup Ф \cup X \cup Ц \cup Ч \cup Ш \cup Щ \rangle$ is terminal Cyrillic capital consonant letters;
41. $Csv := \langle А \cup Е \cup И \cup О \cup У \cup Ю \cup Я \rangle$ is terminal Cyrillic capital loud letters;
42. $Csc := \langle б \cup в \cup г \cup д \cup ж \cup з \cup к \cup л \cup м \cup н \cup п \cup р \cup с \cup т \cup ф \cup х \cup ц \cup ч \cup ш \cup щ \rangle$ is terminal Cyrillic lowercase consonant letters;
43. $Csv := \langle а \cup е \cup и \cup о \cup у \cup ю \cup я \rangle$ is terminal Cyrillic lowercase loud letters.

The following production rules are offered to recognize the language of the text:

$P := \langle S \rightarrow S \text{ Gr}, S \rightarrow \Lambda, \text{Gr} \rightarrow \text{Gr Sb}, \text{Gr} \rightarrow \text{Gr Sp}, \text{Gr} \rightarrow \Lambda, \text{Sp} \rightarrow \_, \text{Sb} \rightarrow \text{Ltr}, \text{Sb} \rightarrow \text{Dgt}, \text{Sb} \rightarrow \text{Ssb}, \text{Sb} \rightarrow \text{Ssg}, \text{Ltr} \rightarrow \text{Lat}, \text{Ltr} \rightarrow \text{Cyr}, \text{Ltr} \rightarrow \text{Eng}, \text{Ltr} \rightarrow \text{Ger}, \text{Ltr} \rightarrow \text{Pol}, \text{Ltr} \rightarrow \text{Ukr}, \text{Ltr} \rightarrow \text{Rus}, \text{Ltr} \rightarrow \text{Cpl}, \text{Ltr} \rightarrow \text{Sml}, \text{Ltr} \rightarrow \text{Cnl}, \text{Ltr} \rightarrow \text{Vwl}, \text{Ltr} \rightarrow \cdot, \text{Ssb} \rightarrow \text{Os b}, \text{Ssb} \rightarrow \text{Bsb}, \text{Ssb} \rightarrow \text{Msb}, \text{Cpl} \rightarrow \text{Lcp}, \text{Cpl} \rightarrow \text{Ccp}, \text{Cpl} \rightarrow \text{Ecp}, \text{Cpl} \rightarrow \text{Gcp}, \text{Cpl} \rightarrow \text{Pcp}, \text{Cpl} \rightarrow \text{Ucp}, \text{Cpl} \rightarrow \text{Rcp}, \text{Sml} \rightarrow \text{Lsm}, \text{Sml} \rightarrow \text{Csm}, \text{Sml} \rightarrow \text{Esm}, \text{Sml} \rightarrow \text{Gsm}, \text{Sml} \rightarrow \text{Psm}, \text{Sml} \rightarrow \text{Usm}, \text{Sml} \rightarrow \text{Rsm}, \text{Lat} \rightarrow \text{Lcp}, \text{Lat} \rightarrow \text{Lsm}, \text{Cyr} \rightarrow \text{Ccp}, \text{Cyr} \rightarrow \text{Csm}, \text{Eng} \rightarrow \text{Ecp}, \text{Eng} \rightarrow \text{Esm}, \text{Ger} \rightarrow \text{Gcp}, \text{Ger} \rightarrow \text{Gsm}, \text{Pol} \rightarrow \text{Pcp}, \text{Pol} \rightarrow \text{Psm}, \text{Ukr} \rightarrow \text{Ucp}, \text{Ukr} \rightarrow \text{Usm}, \text{Rus} \rightarrow \text{Rcp}, \text{Rus} \rightarrow \text{Rsm}, \text{Lcp} \rightarrow \text{Lcc}, \text{Lcp} \rightarrow \text{Lcv}, \text{Lcp} \rightarrow \text{Q}, \text{Lcp} \rightarrow \text{V}, \text{Lcp} \rightarrow \text{X}, \text{Lsm} \rightarrow \text{Lsc}, \text{Lsm} \rightarrow \text{Lsv}, \text{Lsm} \rightarrow \text{q}, \text{Lsm} \rightarrow \text{v}, \text{Lsm} \rightarrow \text{x}, \text{Ccp} \rightarrow \text{Ccc}, \text{Ccp} \rightarrow \text{Csv}, \text{Ccp} \rightarrow \text{b}, \text{Ccp} \rightarrow \text{Ў}, \text{Csm} \rightarrow \text{Csc}, \text{Csm} \rightarrow \text{Csv}, \text{Csm} \rightarrow \text{ь}, \text{Csm} \rightarrow \text{ѣ}, \text{Ecp} \rightarrow \text{Lcc}, \text{Ecp} \rightarrow \text{Lcv}, \text{Ecp} \rightarrow \text{Q}, \text{Ecp} \rightarrow \text{V}, \text{Ecp} \rightarrow \text{X}, \text{Esm} \rightarrow \text{Lsc}, \text{Esm} \rightarrow \text{Lsv}, \text{Esm} \rightarrow \text{q}, \text{Esm} \rightarrow \text{v}, \text{Esm} \rightarrow \text{x}, \text{Gcp} \rightarrow \text{Lcc}, \text{Gcp} \rightarrow \text{Lcv}, \text{Gcp} \rightarrow \ddot{A}, \text{Gcp} \rightarrow \ddot{O}, \text{Gcp} \rightarrow \ddot{U}, \text{Gcp} \rightarrow \text{Q}, \text{Gcp} \rightarrow \text{V}, \text{Gcp} \rightarrow \text{X}, \text{Gsm} \rightarrow \text{Lsc}, \text{Gsm} \rightarrow \text{Lsv}, \text{Gsm} \rightarrow \ddot{a}, \text{Gsm} \rightarrow \ddot{o}, \text{Gsm} \rightarrow \ddot{u}, \text{Gsm} \rightarrow \beta, \text{Gsm} \rightarrow \text{q}, \text{Gsm} \rightarrow \text{v}, \text{Gsm} \rightarrow \text{x}, \text{Pcp} \rightarrow \text{Lcc}, \text{Pcp} \rightarrow \text{Lcv}, \text{Pcp} \rightarrow \text{Ą}, \text{Pcp} \rightarrow \text{Ć}, \text{Pcp} \rightarrow \text{Ę}, \text{Pcp} \rightarrow \text{Ł}, \text{Pcp} \rightarrow \text{Ń}, \text{Pcp} \rightarrow \text{Ó}, \text{Pcp} \rightarrow \text{Ś}, \text{Pcp} \rightarrow \text{Ź}, \text{Pcp} \rightarrow \text{Ż} \rangle$

number of identical bases. There is a well-known algorithm for English-language texts - Porter's stemmer [9], but for Ukrainian texts it does not work perfectly.

Porter's Stemmer is a Stemming algorithm published by Martin Porter in 1980. The original version of Stemmer was designed for English and was written in BCPL. Subsequently, Martin created the Snowball project and, using the basic idea of the algorithm, wrote Stemmer for common Indo-European languages, including Russian [10-17]. The algorithm does not use the bases of words, but only, following a series of rules, cuts off endings and suffixes, based on the features of the language, and therefore works quickly, but not always error-free. The algorithm was very popular and duplicated, often changed by different developers, and not always successfully. Around 2000, Porter decided to "freeze" the project and continue to distribute a single implementation of the algorithm (in several popular programming languages) from his site [10-17]. For example, this algorithm takes into account in Ukrainian-language texts only the presence of an ending, and the suffixes - not then, the words search, search identifies, and search - no. The form of the flexion determines the type of word, for example,

```
var $ADJECTIVE =
'/(ими|ій|ий|а|е|ова|ове|ів|е|ій|ее|ее|я|ім|ем|им|ім|их|іх|ою|йми|іми|у|
ю|ого|ому|ої)$/' ; //http://uk.wikipedia.org/wiki/Прикметник +
http://wapedia.mobi/uk/Прикметник
var $PARTICIPLE = '/(ий|ого|ому|им|ім|а|ій|у|ою|ій|і|их|йми|их)$/' ;
//http://uk.wikipedia.org/wiki/Дієприкметник
var $VERB =
'/(сь|ся|ив|ать|ять|у|ю|ав|али|учи|ячи|вши|ши|е|ме|ати|яти|е)$/' ;
//http://uk.wikipedia.org/wiki/Дієслово
var $NOUN =
'/(а|ев|ов|е|ями|ами|еи|и|ей|ой|ий|й|иям|ям|ием|ем|ам|ом|о|у|ах|иях|ях|ы|
ь|ию|ью|ю|ия|ья|я|і|ові|ї|ею|єю|ою|є|еві|ем|єм|ів|їв|\''ю)$/' ;
//http://uk.wikipedia.org/wiki/Іменник
```

Features of the algorithm. The algorithm works with individual words, so the context in which the word is used is unknown. Linguistics categories such as word structure (root, suffix, etc.) and parts of language (noun, adjective, etc.) are also not available. We currently have the following techniques for analyzing words:

- The term is cut off from the word, for example, ending the *увати* turns the word *критикувати* into a *критик*.
- The word has a constant ending. Words with this ending remain unchanged. Example – *ск* and constant words *блиск*, *тиск*, *обеліск* and more.
- The word changes the ending. This rule applies to words in which certain letters fall out during cancellation (*ядро* and *ядер* – ending *ер* changes by *р*) or change (*чоловік* and *чоловіче* – *к* changes by *ч*).
- The word corresponds to a regular expression. This is an attempt to combine several rules into one difficult one. Perhaps this technique will not live up to the final version of the algorithm. But now, the code contains expressions similar to: *(ов)*, у(ва|вши|вшись|ла|ло|ли|ння|нні|нням|нню|ти|вся|всь|лись|лися|тись|тися)*

- The word does not change when it is being staged, but it is an exception to the rules. This is an undesirable case for the algorithm. It forces the vocabulary of exception words to hold. Examples of the *віче, наче*.
- The word changes during stemming, but is also an exception. This is the worst case for the algorithm because it forces two words to be stored in the dictionary at once: the original and the stemmed. For example, the word *відер* should be changed to *відр*, although other words ending as *ер* are not so categorized (*авіадиспетчер, вітер, правер* etc.).
- Short words remain unchanged. Service parts of a language (prepositions, conjunctions, parts) are usually very short words that are ignored by the algorithm (words up to 2 letters inclusive).

All of these techniques are applied by groups that form the rules of stemming. But this significantly complicates the algorithm for finding keychains. Therefore, it is first suggested to consider common endings - not traditional endings, as part of a word, but the sequence of letters that end a word (Table 3-4). In the Table 3-4 endings of words 1 to 4 letters long are given. Five or more letters are not taken into account, since there are not enough such words (for the maximum of 5 *їтєсь* (6837), for 6 - *ванням* (4656), etc.). This has created a kind of map for the project of stemming. The purpose of the project is to build a static termination tree and to capture the algorithm of all branches of the tree. Generally, a more detailed tree can be built [18-22], but for commercial content we choose a weighted level of detail - from 500 words with common ending. Consider in more detail the idea of a Porter's stemmer, namely finding the basis of a word for a given source word [23-30]. The algorithm does not use the bases of words, but works consistently using a number of rules for truncating endings and suffixes (Fig. 3).

Table 3. Statistic table of common word-endings

Word-endings (number)					
я (164062)	ння (9001)	мось (20536)	али (10666)	ному (19112)	ові (17191)
ся (148160)	в (32681)	лось (10231)	ними (19089)	о (90454)	сті (8731)
лися (10338)	ня (9765)	тись (10366)	м (119779)	мо (33568)	ості (7636)
ося (30769)	а (68134)	лись (10337)	т (2980)	го (31445)	ю (80877)
ься (25211)	ь (151355)	тєсь (19105)	ім (31343)	ло (17238)	ою (39616)
ися (21940)	сь (111459)	лась (10229)	им (31166)	ймо (11229)	ню (10075)
єся (19105)	ть (33055)	ість (7606)	ам (20154)	ємо (11136)	ною (20280)
шся (11775)	ось (30788)	и (123402)	ом (17018)	ого (31389)	кою (7497)
ася (10235)	ись (22656)	ми (62080)	ям (15717)	ало (10465)	нню (9054)
вся (10076)	єсь (19114)	ти (20025)	нім (19333)	ного (19090)	стю (7648)
юся (8044)	ась (10239)	ли (17711)	ним (19093)	і (90275)	у (94504)
лася (10230)	всь (10016)	ими (31121)	ням (9434)	ні (31679)	му (35023)
мося (20532)	сть (7688)	ами (20106)	нням (8975)	ві (22543)	ну (23125)

Word-endings (number)					
лося (10233)	юсь (8047)	ями (9844)	ку (11624)	ті (12596)	ній (19549)
ться (25036)	ють (11222)	ати (10819)	ому (31585)	нні (9909)	ний (19042)
их (31127)	ах (20023)	ях (9855)	них (19092)	ї (34702)	ої (31421)
тися (10379)	ів (15898)	ав (10547)	ш (19163)	єш (11138)	є (11466)

Table 4. Statistic word-endings tree whose total weight is less than 1%

Word-endings (number)					
р (2709)	ч (959)	г (636)	п (341)	щ (110)	
н (2531)	с (914)	з (581)	б (281)	ц (34)	
л (1038)	л (754)	ж (353)	ф (214)	г (4)	

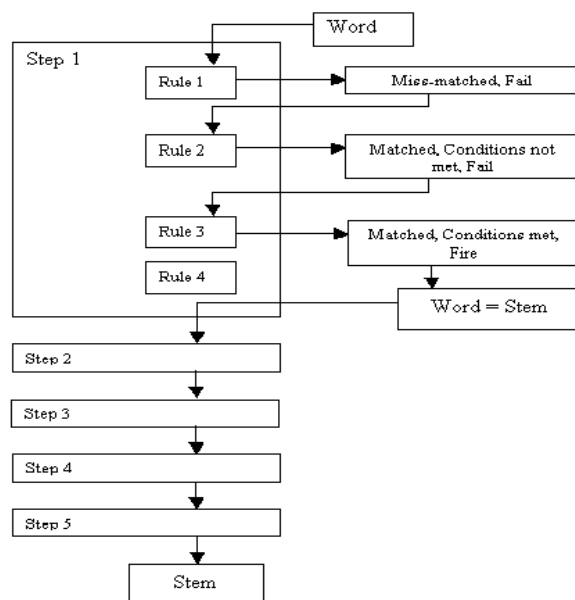


Fig. 3. Structural diagram of Porter's stemmer algorithm

First, let's introduce some definitions:

- **Vowels letters** are *a, e, i, ĭ, o, y, u, e, ю, я*.
- **RV** is part of the word after the first vowel. It is empty if there are no vowels in the word.
- **R1** is part of the word after the first combination is vowel-consonant.
- **R2** is part of **R1** after the first combination is vowel-consonant.

For example, in the word *інформаційний*: RV = *нформаційний*, R1 = *формаційний*, R2 = *маційний*. Now let's define several classes of word endings, leaving their original names in the original description of the algorithm.

Class 1. PERFECTIVE GERUND

- **Group 1:** *в, виши, вишися*. The ending should be preceded by the letter а or я.
- **Group 2:** *ив, ивиши, ивишися*.

Class 2. ADJECTIVE

а, е, і, и, ими, їми, їй, ий, їм, ім, им, ього, ого, ьому, ому, їх, их, ую, юю, ая, яя, ою, єю.

Class 3. PARTICIPLE

- **Group 1:** *ви, юва, ува, уч, юч, л*. The ending should be preceded by letter а or я.
- **Group 2:** *нн, н, ячи, ачи, ова, ову, єм*.

Class 4. REFLEXIVE are *ся, сь*.

Class 5. VERB

- **Group 1:** *ла, є, єте, йте, ли, люю, й, в, єм, ємо, ний, ло, ть, но, ють, ні, ть, єш*. The ending should be preceded by the letter а or я.
- **Group 2:** *ила, ела, ена, йте, ите, єте, юй, уй, їй, ай, ало, ив, или, имо, ений, ило, їло, ено, ють, ать, ені, ять, їть, ить, иш, ую, ю*.

Class 6. NOUN are *а, ев, ов, і, тя, е, ами, іями, ями, єї, єю, ями, ям, її, и, ою, їй, ой, ий, й, им, им, ім, ам, ом, о, у, ах, ях, ую, ю, ія, я*.

Class 7. SUPERLATIVE (*найдовший, милиший, більший*) are *ш, іш*.

Class 8. DERIVATIONAL (*милість, щедрість, малість, крайність*) is *ість*.

Class 9. ADJECTIVAL defined as ADJECTIVE or PARTICIPLE + ADJECTIVE. For example: *падюча* = *пада* + *юч* + *а*.

Rules. When looking for an ending, the longest one is chosen. For example, in the word *інформація* *ія* needs to be chosen, not *я*. All inspections are conducted on a part **RV**. So, when checking for PERFECTIVE GERUND the previous letters а and я should also be inside the **RV**. Letters before **RV** don't take part in inspections at all.

Step 1. To find an ending of PERFECTIVE GERUND. If it exists, then delete it and complete step. In other words, delete the ending of REFLEXIVE (if exists). Then in the following order check and if there is an ending delete: ADJECTIVAL, VERB, NOUN. Once one of them is found, then the step is completed.

Step 2. If the word ends with *і* – delete *і*.

Step 3. If in the **Step 2** there will be an end DERIVATIONAL, then delete it.

Step 4. One of three options is possible:

1. If the word ends with *н*, delete the last letter.
2. If the word ends with SUPERLATIVE, delete it and delete the last letter again if the word ends with *н*.
3. If the word ends with *ь*, delete it.

Stage 3. Syntax analysis of textual content. Syntax is known to be a set of rules that make it possible to construct formulas and recognize the correct formulas in the sequence of characters. It is important for the symbolic computation system that all but one of the expression logic operations are binary. This will be based on a parser. We will consider the process of revising the input sequence of characters in order to parse the grammatical structure according to the given formal grammar. A parser is a program or part of a program that performs parsing [10]. Generally (not just in the computer industry), the term syntactic parsing means the breakdown of text into parts of a language with the identification of their forms, purpose and syntactic relationship with other parts. This is largely determined by the stage of learning the differences and positioning parts of a particular language that can be quite difficult to formalize in inflected languages [22]. It is not at all easy to parse the sentences of such languages. For example, there are significant ambiguities in the structure of human language, that is, words and expressions that can themselves convey meaning in a vast number of variants, but only one of the meanings is relevant in a particular case. The success of choosing the right value in the vast majority of cases depends on many factors of contextual content, and it is almost impossible to predict all combinations of meaning. It is difficult to prepare formal rules for describing informal behavior, although, of course, there are strict rules, many of which form the basis of the grammar that forms the basis of the parser. During parsing, the text is framed into a data structure, usually a tree that matches the syntax structure of the input sequence and is well suited for further processing. As a rule, parser work in two stages: the first identifies meaningful tokens (lexical analysis is performed), the second creates a parse tree. For example (Fig. 4), for arithmetic expression $1 + 2 * 3$:

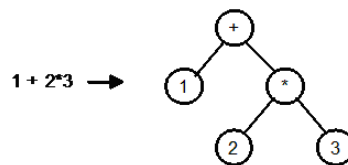


Fig. 4. An example of parsing an expression into a tree

A token is a sequence of one or more characters that stand out as an atomic object. The process of forming tokens is called tokenization or lexical analysis. Tokens are distinguished on the basis of the basic rules of the lexical analyzer (or lexer), which often differ depending on the scope [22]. Tokens are often classified by the position (location) of characters in the character sequence or context in the data stream. This is not just about highlighting a group of characters that are delimited by punctuation on either side (spaces or punctuation). Tokens are defined by the token rules and include grammatical elements of the language used in the data stream. In natural languages, these are usually categories of nouns, verbs, adjectives, or punctuation. The categories are used in the further processing of tokens with a parser or other functions in the program. The tasks of lexical analysis are as follows [20]:

- Convert a character set to a token sequence.

- Highlight each token as a logical part of the text (keyword, variable name, punctuation mark, etc.).
- Matching token and token - specific token text ("for", "variable", ":", etc.).
- Selection of additional token attributes (e.g. variable value).
- Formation of an output token sequence that will be used by the parser as input.

The lexical analyzer usually does nothing with the combination of tokens it has allocated. For example, a typical lexical analyzer recognizes parentheses as characters, but does not check for each open parenthesis "(" closed parenthesis ")". This task remains for the parser or parser.

Step 4. Semantic analysis of textual content. Latent-semantic analysis is a method of processing information in natural language that allows you to analyze the relationship between a collection of documents (messages, articles, i.e. textual content) and terms (keywords) that occur in them. Compares some factors (topics) to all documents and terms. Initially, words correlate with semantic vocabulary classes. Then, the morphosemantic alternatives needed for this sentence are selected. What follows is the linking of words into a single structure and the formation of an ordered set of superposition entries of basic lexical functions and semantic classes. The accuracy of the result is determined by the completeness / correctness of the dictionary.

Step 5. Reference analysis for the formation of interphase unities. A contextual analysis of textual content is carried out. With it, the resolution of local references (the one that is, his) is realized and the expression of the expression is the kernel of unity. The following is a thematic analysis. Separation of statements on a theme and a rheum allocates thematic structures which are used, for example, at formation of a digest. Define regular repeatability, synonymization and re-nomination of keywords; the identity of the reference, that is, the relation of the words to the object of reflection; presence of implication based on situational connections.

Step 6. Structural analysis of textual content. The prerequisites for use are a high degree of coincidence of terms of unity, a discursive unit, a sentence in a semantic language, utterance, and an elementary discursive unit. A basic set of rhetorical relationships between content unities is identified and a nonlinear unity network is constructed [1-7]. The openness of a link set involves its extension and adaptation to analyze the structure of the text. There are several ways to use semantic analysis to define keywords as a phrase, that is, to define terms $Noun \in U_{K1}$ is nouns, noun phrases, or noun adjective among a plurality of text content words. For example, by the rules:

1. If the keyword is an adjective (flexion of the word $u\ddot{u}$ - masculine noun). Then all the words used to the right of this adjective in any case are found in the text (the search follows the basis of this adjective) and a frequency dictionary is built for them. Those phrases that use more than a certain limit (but can be used less than the adjective itself) and are new keywords. The limit is determined by moderator.
2. If the keyword is a noun (flexion of the word is not $u\ddot{u}$), then all words to the right and left of it are analyzed.

3. First, all words to the left of him are checked for flexion *uŭ*. A frequency dictionary is also being built. It is determined by the set of words that occur most often with a certain moderator defined limit - these are the new keywords.
4. Then all the words on the right are analyzed - they should all be without flexion *uŭ*. Similarly, a frequency vocabulary is used to define many keywords.

Experimental studies. 100 scientific publications of the Bulletin of the Lviv Polytechnic National University of the series "Information Systems and Networks" (<http://science.lp.edu.ua/sisn>) from 783 and 805 (<http://science.lp.edu.ua/SISN/SISN-2014>, <http://science.lp.edu.ua/sisn/vol-cur-805-2014-2>) were selected as the linguistic base for the experimental study of the proposed method. The analysis of statistics of functioning of the system of detection of a set of keywords from 100 scientific articles was carried out in two stages, in particular:

1. Analyze all articles by checking common block words and thematic vocabulary.
2. Analyze all articles by checking refined blocked words and refined thematic vocabulary (with more startup, a set of unknown words (missing both in the thematic dictionary and in many blocked ones) is formed.

In addition, at each stage, the review was performed in two steps for each article: analysis of the entire article (<http://victana.lviv.ua/index.php/kliuchovi-slova>) and analysis of the article without beginning (title, authors, editors, annotations in two languages, authors keywords in two languages, authors' place of work), and no literature list to determine errors in the accuracy of multiple keyword formation (Fig.5).

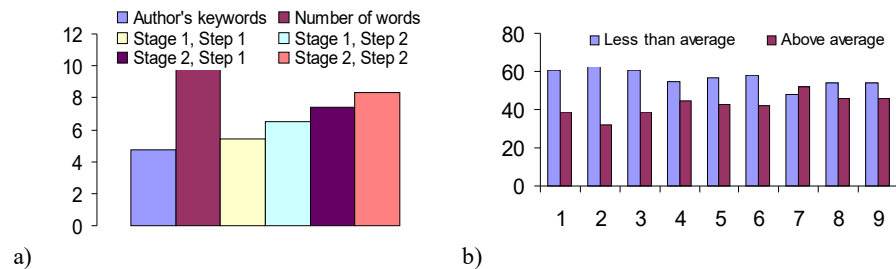


Fig. 5. The results of the review of 100 articles

In Fig. 5a) the diagram of analysis of statistics of formation by set of sets of all potential keywords in comparison with the set defined by authors of articles is presented. The first column is the average number of keywords identified by the author (4.77), and the second is the average number of words that make up those author keywords (9.82). The third column is the arithmetic mean of the potential keywords defined systematically in Step 1, Step 1 (5.46); the fourth is in step 1, step 2 (6.51); the fifth - in step 1, step 1 (7.43); sixth - in step 2, step 2 (8.35). Label these columns accordingly $A_1 \div A_6$. The value A_3 is different from the value A_1 by 0.69 (in quantity but not

in content); accordingly A_4 is different from A_1 by 1.74; A_5 from A_1 by 2.66; A_6 from A_1 by 3.58. The value A_2 is different from the value A_3 by 4.36; accordingly A_2 from A_4 by 3.31; A_2 from A_5 by 2.39; A_2 from A_6 by 1.47. Therefore, on average, the author of an article defines fewer keywords than is actually present in this work. Adjusting the system parameters increases the number of defined keywords by almost 2 times (with a similar comparison the value A_1 with the value A_3 more at 1.144654; A_4 – at 1.36478; A_5 – at 1.557652; A_6 – at 1.750524). The total increment of the value obtained by the system depending on the moderation of the dictionaries is accordingly $A_3 - 14.46541$; $A_4 - 36.47799$; $A_5 - 55.7652$; $A_6 - 75.05241$. If consistently compare A_2 with $A_3 \div A_6$ (in how many times the value A_2 is greater), then we get accordingly the range 1.7985; 1.5084; 1.3217; 1.176.

In Figure 5, b) shows a chart that contains statistics on more detailed textures in the analyzed articles, where 1 is an analysis of the page of the article page (husband's manual and average reference), 2 is paragraphs in the article, 3 is lines of text, 4 is words. 5 is characters, 6 is characters and spaces, 7 is words on the page, 8 - characters on the page, 9 - characters and prints on the page.

In Figure 6 shows the distribution diagram of the set of sets of all potential keywords for each article compared to the set defined by the authors of the articles.

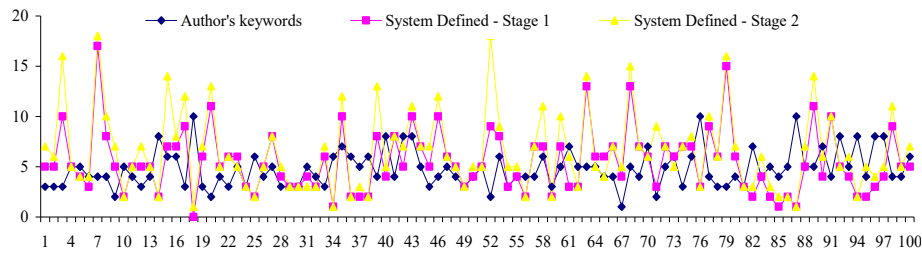


Fig. 6. The results of the inspection of 100 articles

Keyword accuracy is enhanced during the dictionary moderation process. The difference between the number of keywords identified by the author and the system defined in step 1, step 1 is 44.39919% (the percentage difference). The accuracy is improved in step 1, step 2 is 33.70672%, it is significantly improved in step 2, step 1 is 24.33809%, and in step 2, step 2 is already 14.96945%.

In the Table 5 the results of an analysis of the statistics of the set of sets of all potential keywords for each article compared to the set defined by the authors of the articles are presented, where A is for the author's keywords, B is for the system-defined keywords in stage 1 (step 1), C is for the key words system-defined keywords in stage 1 (step 2), D is for system-defined keywords in stage 2 (step 1), E is for system-defined keywords in stage 2 (step 2). In the Table 6-7 the statistics of the analysis of the text of the articles in the formation of sets of keywords for the construction of appropriate histograms for groups A-E are presented.

Table 5. Descriptive statistics for keyword formation for the texts studied

	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>
Average	4.8080808	5.5151515	6.5656566	7.5050505	8.4343434
Standard error	0.1808594	0.3103929	0.3903499	0.3012972	0.3246112
Median	4	5	6	7	8
Moda	4	5	5	7	8
Standard deviation	1.7995281	3.0883707	3.8839324	2.9978691	3.2298405
Sample dispersion	3.2383014	9.5380334	15.084931	8.9872191	10.43187
Kurtosis	0.6528151	1.7052728	0.7486433	-0.456455	-0.504378
Asymmetry	0.9479385	1.1253053	1.0657159	0.5375984	0.5170473
Interval	8	16	17	12	13
Minimum	2	1	1	2	3
Maximum	10	17	18	14	16
Sum	476	546	650	743	835
Account	99	99	99	99	99
The biggest(1)	10	17	18	14	16
The smallest(1)	2	1	1	2	3
Reliability level (95.0%)	0.3589095	0.6159647	0.7746366	0.5979144	0.6441803

Table 6. Statistics for histograms for group *A* and group *B*

<i>A</i>	Frequency	Integral %	<i>A</i>	Frequency	Integral %	<i>B</i>	Frequency	Integral %	<i>B</i>	Frequency	Integral %
1	0	0.00%	4	27	27.27%	1	2	2.02%	5	20	20.20%
2	4	4.04%	5	21	48.48%	2	10	12.12%	7	16	36.36%
3	20	24.24%	3	20	68.69%	3	12	24.24%	3	12	48.48%
4	27	51.52%	6	11	79.80%	4	4	28.28%	2	10	58.59%
5	21	72.73%	8	8	87.88%	5	20	48.48%	6	9	67.68%
6	11	83.84%	7	5	92.93%	6	9	57.58%	4	4	71.72%
7	5	88.89%	2	4	96.97%	7	16	73.74%	8	4	75.76%
8	8	96.97%	10	3	100.00%	8	4	77.78%	10	4	79.80%
9	0	96.97%	1	0	100.00%	9	2	79.80%	11	3	82.83%
10	3	100.00%	9	0	100.00%	10	4	83.84%	12	3	85.86%
11	0	100.00%	11	0	100.00%	11	3	86.87%	14	3	88.89%
12	0	100.00%	12	0	100.00%	12	3	89.90%	1	2	90.91%
13	0	100.00%	13	0	100.00%	13	2	91.92%	9	2	92.93%
14	0	100.00%	14	0	100.00%	14	3	94.95%	13	2	94.95%
15	0	100.00%	15	0	100.00%	15	1	95.96%	16	2	96.97%
16	0	100.00%	16	0	100.00%	16	2	97.98%	18	2	98.99%
17	0	100.00%	17	0	100.00%	17	0	97.98%	15	1	100.00%
18	0	100.00%	18	0	100.00%	18	2	100.00%	17	0	100.00%
Moi	0	100.00%	Moi	0	100.00%	Moi	0	100.00%	Mo	0	100.00%

The author of a scientific article usually chooses at his discretion the number of keywords in the range of 2 to 8 words (most often 3-5 keywords). The system also defines a different number of words, depending on the writing style of the particular author (there are articles in which the system does not find any keywords by Zipf law). For group B, most often the system determined the number of keywords 5, 7 and 3 (more than 10), although the distribution of found keywords ranged from 1 to 18 words (except 17). For Group B, the system most commonly identified keywords as 5, 7, and 3, although the distribution of keywords found ranged from 1 to 18 words (except 17), but the number of keywords found increased and the highest reliability was achieved. For Group C, the system most often determined the number of keywords 7, 6, 5, 10 and 8, although the distribution of found keywords ranged from 2 to

14 words (significantly narrowed the range). For Group D, the number of keywords 8, 5, 7, and 10 was most often determined by the system, although the distribution of the keywords found ranged from 3 to 16 words (improved accuracy).

Table 7. Statistics for histograms for group *C*, group *D* and group *E*

<i>C</i>	Frequency	Integral %	<i>C</i>	Frequency	Integral %	<i>D</i>	Frequency	Integral %	<i>D</i>	Frequency	Integral %
1	2	2,02%	5	20	20,20%	1	0	0,00%	7	15	15,15%
2	10	12,12%	7	16	36,36%	2	1	1,01%	6	14	29,29%
3	12	24,24%	3	12	48,48%	3	5	6,06%	5	13	42,42%
4	4	28,28%	2	10	58,59%	4	9	15,15%	10	12	54,55%
5	20	48,48%	6	9	67,68%	5	13	28,28%	8	11	65,66%
6	9	57,58%	4	4	71,72%	6	14	42,42%	4	9	74,75%
7	16	73,74%	8	4	75,76%	7	15	57,58%	12	6	80,81%
8	4	77,78%	10	4	79,80%	8	11	68,69%	3	5	85,86%
9	2	79,80%	11	3	82,83%	9	4	72,73%	14	5	90,91%
10	4	83,84%	12	3	85,86%	10	12	84,85%	9	4	94,95%
11	3	86,87%	14	3	88,89%	11	1	85,86%	13	3	97,98%
12	3	89,90%	1	2	90,91%	12	6	91,92%	2	1	98,99%
13	2	91,92%	9	2	92,93%	13	3	94,95%	11	1	100,00%
14	3	94,95%	13	2	94,95%	14	5	100,00%	1	0	100,00%
15	1	95,96%	16	2	96,97%	15	0	100,00%	15	0	100,00%
16	2	97,98%	18	2	98,99%	16	0	100,00%	16	0	100,00%
17	0	97,98%	15	1	100,00%	17	0	100,00%	17	0	100,00%
18	2	100,00%	17	0	100,00%	18	0	100,00%	18	0	100,00%
Mor	0	100,00%	Mor	0	100,00%	Mor	0	100,00%	Mor	0	100,00%

<i>E</i>	Frequency	Integral %	<i>E</i>	Frequency	Integral %
1	0	0,00%	8	14	14,14%
2	0	0,00%	5	12	26,26%
3	1	1,01%	7	11	37,37%
4	9	10,10%	10	11	48,48%
5	12	22,22%	4	9	57,58%
6	9	31,31%	6	9	66,67%
7	11	42,42%	9	9	75,76%
8	14	56,57%	11	5	80,81%
9	9	65,66%	14	5	85,86%
10	11	76,77%	12	4	89,90%
11	5	81,82%	13	4	93,94%
12	4	85,86%	15	3	96,97%
13	4	89,90%	16	2	98,99%
14	5	94,95%	3	1	100,00%
15	3	97,98%	1	0	100,00%
16	2	100,00%	2	0	100,00%
17	0	100,00%	17	0	100,00%
18	0	100,00%	18	0	100,00%
Mor	0	100,00%	Mor	0	100,00%

5 Conclusions

The article presents theoretical and experimental substantiation of the method of linguistic analysis of Ukrainian-language commercial content using Porter's stemming. The method is aimed at automatic detection of significant keywords of Ukrainian-language content on the basis of the proposed formalization of components of analysis is grammatical (graphemes), morphological, syntactic, semantic, referential and structural. To implement grammatical analysis, the rules of line recognition in the text are proposed, the set of standard graph models for 5 languages according to the nor-

mal Backus-Naur form is determined, and the corresponding grammar $G = \langle V, T, S, P \rangle$ for identifying meaningful units of textual commercial content analysis. The morphological analysis was implemented by adapting Stemming M. Porter's algorithm to the Ukrainian language, in particular, a static end-tree was constructed and a weighted level of detail was selected - from 500 words with common endings, rules for truncation of endings and suffixes were substantiated. The basic requirements and procedures of syntactic, semantic, referential and structural analysis of Ukrainian-language commercial content are defined. An experimental study of the method of linguistic analysis was conducted on the materials of 100 scientific publications from two issues (783 and 805) of the Bulletin of the Lviv Polytechnic National University of the series "Information Systems and Networks" (<http://science.lp.edu.ua/sisn>). Based on the proposed method, the keyword search system demonstrated the ability to improve itself by forming and refining a number of common blocked words and a thematic dictionary with the participation of moderators. It is found that, for the technical scientific texts of the experimental base, the authors of articles usually define fewer keywords on average than are actually present in this work. Numerous statistics show that debugging your system's keywords nearly doubles the number of defined keywords, without compromising accuracy or reliability. Further experimental research will require testing the proposed method to identify keywords from other categories of texts is scientific humanities, artistic, nonfiction, etc.

References

1. Bobicev, V., Kanishcheva, O., Cherednichenko, O.: Sentiment Analysis in the Ukrainian and Russian News. In: First Ukraine Conference on Electrical and Computer Engineering, 1050-1055 (2017)
2. Sharonova, N., Doroshenko, A., Cherednichenko, O.: Issues of fact-based information analysis. In: CEUR Workshop Proceedings, 2136, 11-19 (2018)
3. Cherednichenko, O., Babkova, N., Kanishcheva, O.: Complex Term Identification for Ukrainian Medical Texts. In: CEUR Workshop Proceedings, 146-154 (2018)
4. Khomytska, I., Teslyuk, V., Holovatyy, A., Morushko, O.: Development of methods, models, and means for the author attribution of a text. In: Eastern-European Journal of Enterprise Technologies, 3(2-93), 41-46 (2018)
5. Khomytska, I., Teslyuk, V.: Authorship and Style Attribution by Statistical Methods of Style Differentiation on the Phonological Level. In: Advances in Intelligent Systems and Computing III. AISC 871, Springer, 105-118 (2019)
6. Babichev, S.: An Evaluation of the Information Technology of Gene Expression Profiles Processing Stability for Different Levels of Noise Components. In: Data, 3(4), 48 (2018)
7. Babichev, S., Durnyak, B., Pikh, I., Senkivskyy, V.: An Evaluation of the Objective Clustering Inductive Technology Effectiveness Implemented Using Density-Based and Agglomerative Hierarchical Clustering Algorithms. In: Advances in Intelligent Systems and Computing, 1020, 532-553 (2020)
8. Lytvyn, V., Vysotska, V., Peleshchak, I., Basyuk, T., Kovalchuk, V., Kubinska, S., Chyrun, L., Rusyn, B., Pohreliuk, L., Salo, T.: Identifying Textual Content Based on Thematic Analysis of Similar Texts in Big Data. In: International Scientific and Technical Conference on Computer Science and Information Nechnologies (CSIT), 84-91 (2019)

9. Moseichuk, V.: Porter stemming algorithm for Ukrainian languages. http://www.marazm.org.ua/document/stemer_ua/
10. Vysotska, V., Lytvyn, V., Kovalchuk, V., Kubinska, S., Dilai, M., Rusyn, B., Pohreliuk, L., Chyrun, L., Chyrun, S., Brodyak, O.: Method of Similar Textual Content Selection Based on Thematic Information Retrieval. In: International Scientific and Technical Conference on Computer Science and Information Nechnologies (CSIT), 1-6 (2019)
11. Russian stemming algorithm. <http://snowball.tartarus.org>
12. Porter stemmer. <https://github.com/allaud/porter-stemmer>
13. The Porter Stemming Algorithm. <http://tartarus.org/~martin/PorterStemmer/>
14. Porter Stemming Algorithm. <http://snowball.tartarus.org/algorithms/porter/stemmer.html>
15. English stemming algorithm. <http://snowball.tartarus.org/algorithms/english/stemmer.html>
16. Porter, M. F.: An algorithm for suffix stripping. http://telemat.det.unifi.it/book/2001/wchange/download/stem_porter.html
17. Willett, P.: The Porter stemming algorithm: then and now. <http://eprints.whiterose.ac.uk/1434/>
18. Senyk, M. The Porter Stemming Algorithm for Ukrainian. <http://www.senyk.poltava.ua>
19. Vysotska, V., Fernandes, V.B., Lytvyn, V., Emmerich, M., Hrendus, M.: Method for Determining Linguometric Coefficient Dynamics of Ukrainian Text Content Authorship. In: Advances in Intelligent Systems and Computing, 871, 132-151 (2019)
20. Lytvyn, V., Vysotska, V., Pukach, P., Nytrebych, Z., Demkiv, I., Senyk, A., Malanchuk, O., Sachenko, S., Kovalchuk, R., Huzyk, N.: Analysis of the developed quantitative method for automatic attribution of scientific and technical text content written in Ukrainian. In: Eastern-European Journal of Enterprise Technologies, 6(2-96), 19-31 (2018)
21. Vysotska, V., Lytvyn, V., Hrendus, M., Kubinska, S., Brodyak, O.: Method of textual information authorship analysis based on stylometry. In: 13th International Scientific and Technical Conference on Computer Sciences and Information Technologies, 9-16 (2018)
22. Kulchytskyi, I.: Statistical Analysis of the Short Stories by Roman Ivanychuk. In: CEUR Workshop Proceedings, Vol-2362, 312-321 (2019)
23. Shandruk, U.: Quantitative Characteristics of Key Words in Texts of Scientific Genre (on the Material of the Ukrainian Scientific Journal). In: CEUR Workshop Proceedings, Vol-2362, 163-172 (2019)
24. Lovins, J.B.: Development of a stemming algorithm. In: Mechanical Translation and Computational Linguistics 11:22-31 (1968)
25. Jongejan, B., Dalianis, H.: Automatic training of lemmatization rules that handle morphological changes in pre-, in- and suffixes alike. <http://www.aclweb.org/anthology/P/P09/P09-1017.pdf>
26. Vysotska, V., Kanishcheva, O., Hlavcheva, Y.: Authorship Identification of the Scientific Text in Ukrainian with Using the Lingvometry Methods. In: Computer Sciences and Information Technologies, CSIT, 34-38 (2018)
27. Vysotska, V., Burov, Y., Lytvyn, V., Demchuk, A.: Defining Author's Style for Plagiarism Detection in Academic Environment. In: Proceedings of the 2018 IEEE 2nd International Conference on Data Stream Mining and Processing, DSMP, 128-133 (2018)
28. Lytvyn, V., Vysotska, V., Burov, Y., Bobyk, I., Ohirko, O.: The linguometric approach for co-authoring author's style definition. In: Intelligent Data Acquisition and Advanced Computing Systems, IDAACS-SWS, 29-34 (2018)
29. Hardcoded stemmer for Ukrainian. <https://github.com/vgrichina/ukrainian-stemmer>
30. Perestoronin, P.: The Porter Stemming Algorithm for Russian. <http://blog.eigene.in/post/49598738049/snowball>

Application of the Universal Meta-Formalism to Substantiate and Research Intelligence Properties

Alexander V. Sosnitsky¹[0000-0001-6468-9060], Anatoly I. Shevchenko²[0000-0002-0095-538X]

¹Berdiansk State Pedagogical University, Schmidta St., 4, 71110, Berdiansk, Ukraine, sosnitsky.ukr@yandex.ua

²Institute of AI Problems of the MES and NAS of Ukraine, Mala Zhytomyrska St., 11/54, 01001, Kyiv, Ukraine
ipai.kiev@gmail.com

Abstract. A formal scientific definition of Intelligence is a topical problem not only of the synthesis of AI but of the world science as a whole as a product of exclusively intellectual activity. It is due to the fundamental shortcomings of the modern general scientific paradigm that is limited to particular dogmatic concepts. This problem is solved by the transition to the universal general scientific paradigm by increasing the level of knowledge abstraction. Achieving a uniform initial Universal Axiom, in principle, allows you to deduce a complete system of Universe's concepts and get a Universal Definition and Formalism of Intelligence from it, which are supposedly applicable to all Universe's phenomena. The paper presents the main provisions of the new approach and deduces the initial components of the Universal Formalism of Intelligence, which follows directly from the Universal Axiom and is represented by exclusively the highest Universe's meta-concepts that can solve the said problem. Such Formalism for the first time qualitatively allows us to formally derive and objectively evaluate any of the innumerable properties of Intelligence, the initial of which are given in this paper. The obtained results are pioneering in this area and have important general scientific significance and the initial application in many technical and humanitarian research and development.

Keywords: The Universe, Universal Theory, Intelligence, meta-definition, meta-formalization, universalization

1 Introduction

Despite the enormous efforts, modern science cannot give a formal definition of Intelligence, which delegitimizes all research in this area and science in general as a product of exclusively intellectual activity [1].

This state has not been changed since the pioneering Dartmouth workshop in 1956, when, for the first time, J. McCarthy admitted: "The problem is that we cannot yet characterize in general what kinds of computational procedures we want to call intelligent. We understand some of the mechanisms of intelligence and not others" [2].

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

As a result, in the 1980s, J. Searle's compromise philosophical hypothesis about the essential unity of any Intelligence was supported, which allowed us to divide it into two different concepts: 1) weak (private, machine) Intelligence, which performs separate intellectual functions that can be formalized and implemented now, and 2) strong (general, natural) Intelligence, similar to human, the formalization (understanding) of which is postponed until future [3].

However, such a division did not eliminate, but only transferred the said uncertainty to the problem of distinguishing intellectual functions in the diversity of all actions of living beings. The McCarthy-Searle paradigm has advanced, but has not legalized the above-mentioned works, and a person remained an informal general scientific standard of Intelligence, according to the well-known Turing test [4].

Obviously, the problem of a scientific definition and formalization of Intelligence is a consequence of the fundamental shortcomings of modern science and it can only be solved by substantiated development of a general scientific paradigm [5-6].

For the first time this was done by the authors by increasing the level of knowledge abstraction and moving from the modern system of particular Universe's axioms (in more general terms – dogmas), producing limited systems of concepts in the corresponding fields, to the uniform initial Universe's Axiom (Dogma, UA), producing the entire pyramid-like system of concepts/categories of the Universe (UAP) as the uniform Universal (Meta-) Formalism (UF, UMF). If particular conceptual areas can describe only simple phenomena that fit entirely into one area, then the UMF must obligatory describe all phenomena, up to the most complicated ones, which the Universe and its Intelligence are (Fig. 1) [6-10].

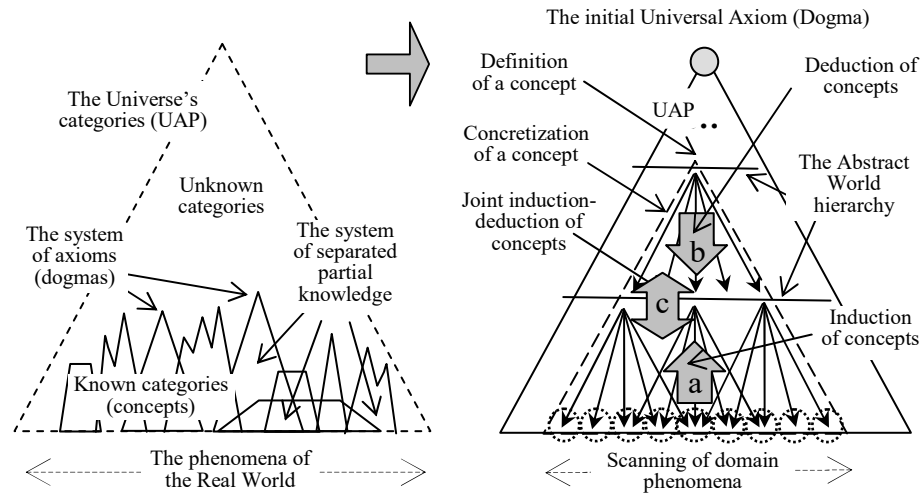


Fig. 1. The development of the world science scheme from dogmatization to universalization of knowledge. Directions of action: a - induction, b – deduction, c - joint induction-deduction

The UMF is a new scientific concept and its construction is problematic, however, it is simplified by the supposed homotropy and constancy of the categories of the Universe, allowing them to accumulate copies to form a uniform stable externally

(with the Universe) and internally (among themselves) consistent system of concepts that is then tested in practice. Naturally, such a UMF must be only one because our Universe has such a character, and the alternative attempts to create it must lead to a single result.

Thus, for the first time, it was possible to build the UMF and deduce from it the Universal (Meta-) Formalism of Intelligence (UMFI) that consistently coincides and extends the observed properties of Intelligence as its new cognition. The UMFI becomes a new standard of Intelligence instead of human, allowing for a deeper formal study of the origin, properties and classifications of the latter. Substantiation of the deduction of Intelligence concept is the goal of this work.

2 Universal Meta-Formalism of Intelligence

2.1 Initial Ontology of Intelligence

Intelligence comes from the concept of relation, which, by all the signs of an axiom, is accepted as the UA (Fig. 2). Thus, Universe's entities presumably have a structural nature and are adequately represented by exclusively multiple set-theoretic and ER-formalisms.

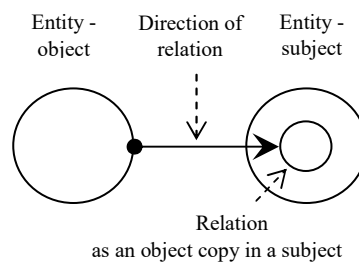


Fig. 2. The scheme of a meta-concept relation

Definition. A relation is a copy of a certain entity (object) in another entity (subject).

Definition. An entity is part of the Universe, singled out by a certain relation as a unitary whole.

Definition. The Universe is a complete set of directly or indirectly related entities.

Consequence. The Universe and its entities have a structural nature that is extremely infinite into breadth and depth.

Consequence. The Universe is its entity.

Consequence. The Universe is a self-determined entity.

Definition. Knowledge is a copy of an entity.

Definition. Cognition is the establishment of relations with entities.

Consequence. Cognition of the Universe is the establishment of relations by its entities with it as a unitary whole.

Consequence. The Universe is a mutual cognition of its entities.

An elementary combination of relations generates the initial ontology of cognition.

The advancement of an object copy generates 4 stages of cyclical cognition of a subject: 1) on the border of a subject (information), 2) inside a subject (knowledge), 3) coordination of the copy with other knowledge of a subject (understanding) and 4) additional cognition of an object (research) (Fig. 3).

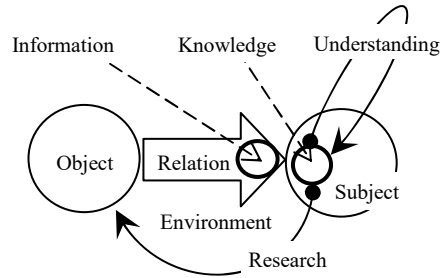


Fig. 3. The scheme of meta-concept cognition

Cognition of a multitude of objects gives rise to the scheme in Fig. 4.

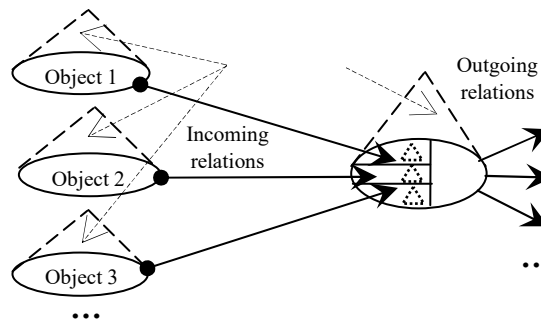


Fig. 4. The scheme of mutual meta-concept cognition

Transferring a copy of an object through an intermediate subject is teaching (Fig. 5).

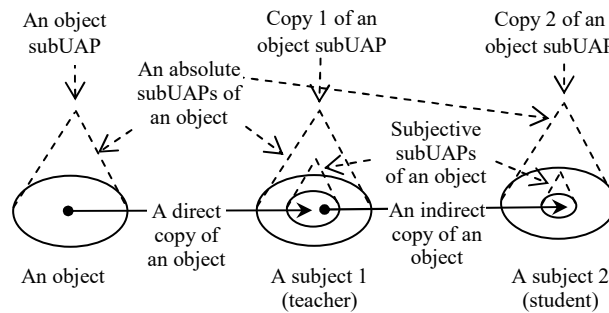


Fig. 5. The scheme of meta-concept teaching

Virtualization of a copy of an object within a subject (virtualizer) is an excess of the categories of an object on their copies (Fig. 6).

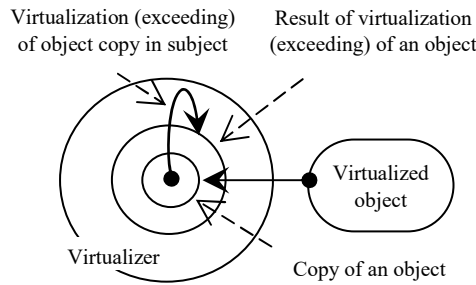


Fig. 6. The scheme of meta-concept virtualization

2.2 Abstract / Real Dichotomy of the Universe

According to Plato and Aristotle, the Universe is divided into the World of categories (Abstract World (AW)) and the World of phenomena (Real World (RW)), which, according to the modern understanding, have the scheme in Fig. 1. According to the universal ideas, the AW begins in the UA and develops further into a pyramid-like system of categories/abstracts/concepts, built into the RW in the form of Space-Time-Matter Complex (STM-Complex), consisting of interrelated Complexes of Space, Time and Matter, the initial properties of which are generally understood (Fig. 7).

Definition. The World is a relatively integral part of the Universe.

Definition. A category (abstract) is an entity of the AW.

Definition. A phenomenon is an entity of the RW.

Definition. A Complex is a phenomenon that is present in all the RW phenomena.

Definition. Space is a regular and invariable Complex of the RW.

Definition. Time is a regular and variable Complex of the RW.

Definition. Matter is an irregular Complex of the RW.

Definition. A thing is an invariable part of the Matter Complex.

Definition. A process is a variable part of the Matter Complex.

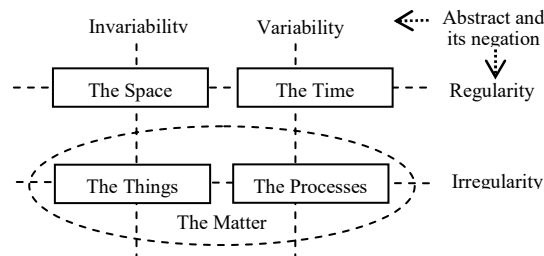


Fig. 7. The scheme of the Complexes ontology

The RW (STM-Complex) consists of interrelated phenomena with built-in categories that determine their properties. The existence of phenomena is determined solely by its categories. Presumably the categories of phenomena do not have internal contradictions. The nature of the category carriers is unknown, but it does not prevent formalization.

Consequence. Concepts are cognition of the abstract part (categories) of phenomena.

Consequence. Facts are cognition of the real part of phenomena.

Consequence. Due to indirect actions (through other categories), the categories are poorly recognized by phenomena.

Consequence. Due to the direct action, phenomena are strongly cognized by phenomena.

Hypothesis. The system of categories is the same throughout the (infinite) Universe.

Consequence. Facts determine the activated categories of phenomena.

Consequence. Knowledge is the concepts of facts and categories of phenomena.

Definition. A formula (formalism) is the system of categories / concepts of an entity.

Definition. An axiom (definition) is initial formula of an entity.

Definition. SubUAP is an abstract part of an entity.

2.3 Universal Harmonic Classification of Phenomena

Phenomena 1) exist in an STM-Complex and 2) are divided into 3 non-empty groups of relations: internal, external and intermediate among them (Fig. 8). Additionally, these relations are divided into the Past, Present and Future (Fig. 9).

Special states of internal relations development give rise to the universal harmonic classification of phenomena, the simplest (initial) of which is shown in Table 1.

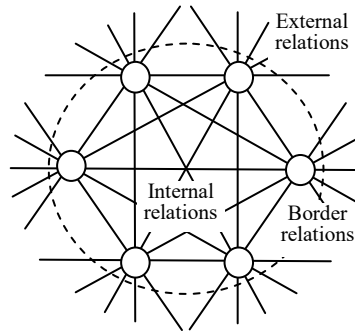


Fig. 8. The scheme of a phenomena harmonic structure

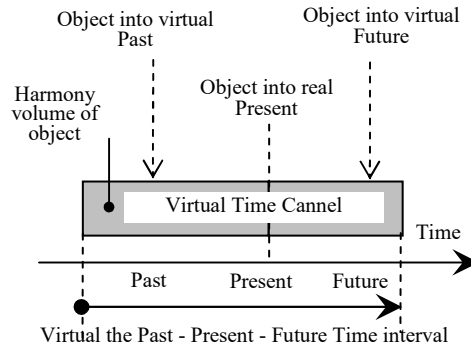


Fig. 9. The scheme of a phenomenon dividing into the Past, Present and Future and its Virtual Time Channel

Definition. A virtual relation is a copy of a real relation.

Class 1 (Thermodynamics) has relatively weak real intermediate relations and exists exclusively at the point of the Present, continuously moving from the Past to the Future.

Class 2 (Mechanics / Natural selection) has real intermediate relations comparable to internal and external ones at the point of the Present and is supplemented by virtual relations with the Past that make up an internal structure of a phenomenon (a copy of the Past in the Present of the phenomenon). Mechanics inherits and develops Thermodynamics.

Class 3 (Life / Intelligence) inherits Thermodynamics and Mechanics and is supplemented by virtual relations with the Future (the forecast of a phenomenon) by predicting the development of a phenomena over a certain time interval.

Class 4 (the Highest Reason) inherits Thermodynamics, Mechanics and Life and is supplemented by real relations along the entire axis of Time.

Thus, Classes 2-4 are directly related to Intelligence, have specific formalisms, definitions, properties, and further are concretized into the following subclasses. Class 2 conditionally got into this group due to the importance of natural selection for living phenomena.

Table 1. Universal harmonic classification of phenomena

Class	Name	Internal structure	Harmonious Resource	Harmony Type	Characteristic quantity
1	Quasi-chaos (Thermodynamics)	No	Real relations in the Present	Starting	Entropy
2	Natural selection (Mechanics)	Present	+ virtual relations during the interval in the Past	Passive	Energy
3	Life (Intelligence)	Copy of STM-Complex	+ virtual relations during the interval in the Future	Active	Harmony
4	The Highest Reason	Real STM-Complex	+ real present on all Time axis	Highest	Harmony
5	The Harmon	Absolute	Absolute connectedness	Absolute	Harmony
6	Absolute Chaos	No	There are no relations	No	No

2.4 Virtual Time Channel, Basic Dependencies and Characteristic Values of Classes

Intelligence sequentially forms and develops a Virtual Channel in Time (VTC), which is fundamentally different from the Shannon channel in Space [11] and provides additional internal (subjective) Time commutation of phenomena, creating a harmonious superiority of the upper Classes over the lower Classes (Fig. 9). The VTC is a key indication of Intelligence, in contrast to purely logical phenomena.

The VTC radically changes the interaction of phenomena (Fig. 10). Thermodynamic phenomena almost do not interact with external relations and pass them with small changes. Mechanical and living phenomena strongly influence external relations with an increase in the VTC and from the passive turn to active and destabilizing ones. The Universe is completely self-determined and depends only on internal relations.

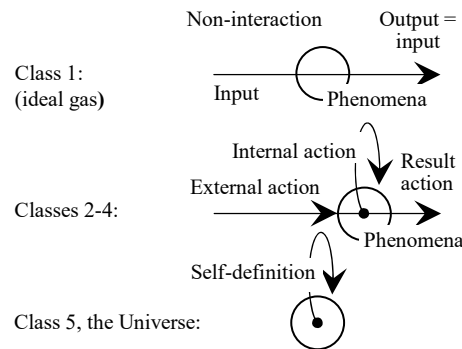


Fig. 10. Classification of the phenomena Classes interaction

The VTC radically changes the characteristic quantities of phenomena from passive entropy (that tends from complex to the simplest states) and conservative energy (that preserves phenomena) to active harmony that gathers itself back from the simplest to the highest states and thereby creates complex phenomena (Fig. 11).

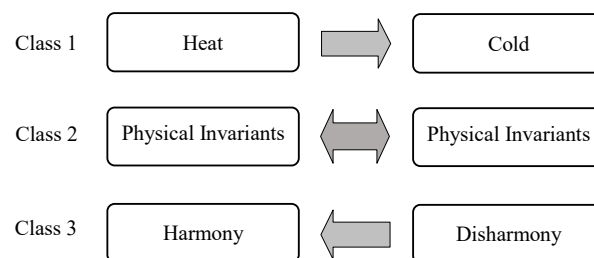


Fig. 11. The scheme of action of characteristic quantities on

phenomena Classes

Accordingly, Class 1 destroys, Class 2 saves, and Class 3 develops phenomena (Table 2).

The VTC and additional Time harmony correspond to the desired negative entropy of E. Schrödinger [12].

Table 2. Characteristic Quantities of Harmonic Phenomena Classes

Class	Class name	Characteristic quantity	Degree of phenomena harmony
1	Quasi-chaos (Thermodynamics)	Entropy	Degradation
2	Natural selection (Mechanics)	Energy	Conservation
3	Life (Intelligence)	Harmony	Development

2.5 The Harmony, Meta-Law and the Harmon

Cognition is embedded in the definition of the Universe and its UA, and therefore the highest Universe's characteristic quantity (that is based on connectedness) is naturally assumed. Let us call it 'harmony', the exact formula of which has not yet been substantiated (deduced), but it is qualitatively confirmed everywhere. The initial (Meta-) Law of the Universe assumes its successive increase (harmonization): all entities tend to increase their harmony.

The Meta-Law is the initial functionally complete method of harmonization that provides all the opportunities for ubiquitous increase in the connectivity of Universe's entities, the particular case of which is cognition.

The limit of an entity harmonization is the state of complete connectivity (local harmon) – a complete graph built on the components of an entity (harmony), the added arcs of which are the chaos of an entity (Fig. 12).

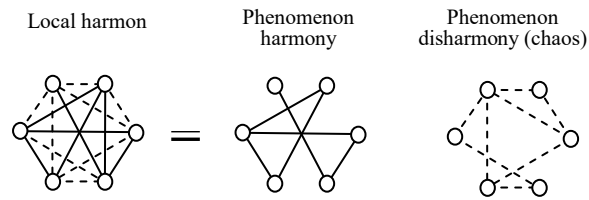


Fig. 12. The scheme of a dichotomy of a phenomenon into harmony (solid lines) and chaos (dotted lines) for $n = 6$

Definition. Harmony is connectedness of entities.

Definition. Chaos is the lack of connectedness of entities.

Consequence. Chaos is a resource of harmony of entities.

The limit of harmonization of the Universe is the state of the Harmon as the Complete Infinite Oriented Graph $\text{CIOG}(\text{CIOG}) = \lim \text{COG}_n(\text{COG}_n)$, number of vertices $n \rightarrow \infty$, whose vertices are the same graphs (Fig. 13). Harmon is a central concept of the

Universal Theory as an initial and final Universe's entity, the study of which is beyond the scope of this topic [9].

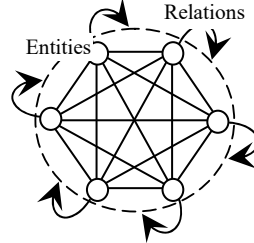


Fig. 13. The scheme of the Harmon (for $n = 6$), $n \rightarrow \infty$

2.6 The Universal Cognition Methodology

Harmonization begins with the cognition of phenomena, which in itself is an act of harmonization through the access of a subject to an object and copying its real and abstract parts. The real part activates the categories of an object, and the abstract part selects and transfers them into the concepts of a subject. This is implemented by various combinations of the available means of a subject at all stages of harmonization: access, selection, copying, formation of the VTC-representation, search and implementation of the required harmonic state of an object.

Due to the direct invisibility of the AW in the RW, the biggest problem is the abstraction of an object, which performs a subject with logical induction (a) and deduction (b) operations (id-operations (c)) that are understood to mean generalization and combination of selected concepts respectively (Fig. 1).

As a result, two opposing flows of inductive and deductive concepts arise in a subject, which coincide in nature and must coincide in a subject. The coinciding concepts form a certain subjective system of concepts, and the mismatched concepts are sent for additional cognition as initiators of new knowledge (Fig. 14).

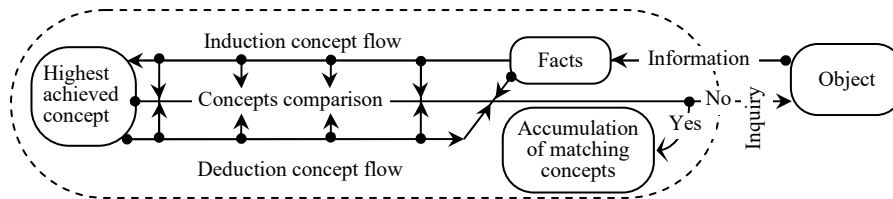


Fig. 14. The scheme of a conditioned reflex

This scheme corresponds to a conditioned reflex [13-14] and determines a typical dichotomy of the brain into two hemispheres [15], specializing in induction and deduction, which fundamentally ensures complete cognition, starting with the zero cognitive ability of any subject, which any Universe's phenomenon is (Fig. 15).

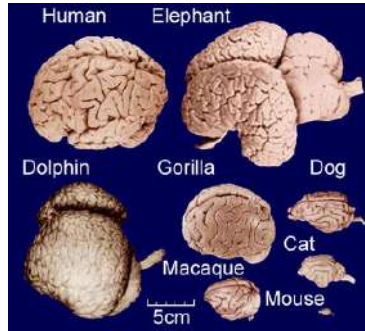


Fig. 15. Typical brain dichotomy
(<https://psychology.wikia.org/wiki/Brain>)

Consequence. Knowledge is subjective.

Consequence. Knowledge is hypothetical.

Definition. A hypothesis is a statement, the falsity of which has not been proved.

Consequence. The elimination of external (with the Universe) and internal (with themselves) contradictions of knowledge is a necessary method of cognition because our Universe has such a character.

Definition. The truth is the Universe.

Consequence. The criterion of the truth is the value of externally and internally consistent knowledge.

Consequence. A conditioned reflex provides unlimited knowledge of the Universe, starting with the zero ability of cognition.

A conditioned reflex and the elimination of contradictions substantiate the derivative method of sequential concretization of concepts by introducing additional hypothetical concepts with the supposedly overcome distance between the conflicting concepts to achieve the target formalization, which are then verified by the subsequent factual analysis as any consciousness does. The volume of such a system of concepts is the main measure of the truth of the required formalism (Fig. 16).

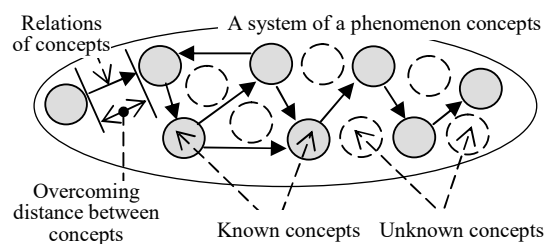


Fig. 16. The scheme of a method of sequential concretization of concepts

2.7 The Universal Meta-Definition of Intelligence

The Universe has 3 main divisions: 1) Abstract (AW, RW), 2) Time (Past, Present and Future) and 3) Categorical (categories as limiting harmonization), which are overcome only together under the influence of the Meta-Law through virtualization using the above-mentioned harmonization tools that make up the initial formalism (definition, relation) in the UMFI as an important tool of the Universe, further developed by the Highest Reason (Fig. 17).

Definition. Intelligence is a universal active joint harmonizer of the Abstract, Time and Categorical divisions of the Universe.

UMFI ontologically includes all of the above properties and mechanisms of action, suggesting further concretization and materialization to a variety of RW conditions.

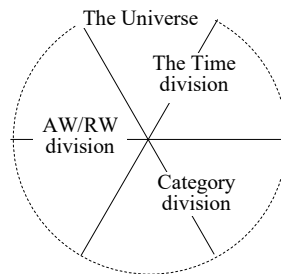


Fig. 17. Illustration of the Universe Abstract, Time and Categorical divisions

2.8 Concretization and Materialization of Intelligence

The UA is directly included in the definition of Intelligence, which is further specified, presumably, in all places of the Universe through an endless hierarchy of ecological niches of existence (ENE) with different categorical systems up to materialization at the lowest real level of abstract hierarchy.

Accordingly, the UMFI receives numerous additional concretizations up to and including materialization, which are consistently complicated various low-level formalisms up to the achievement of the status of phenomena that makes it possible to cognize the entire Universe (Fig. 18).

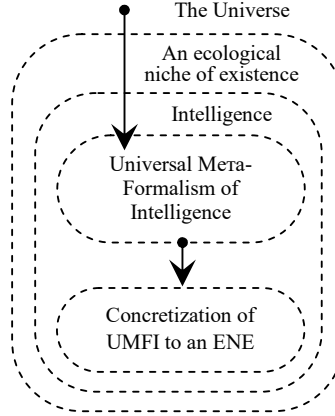


Fig. 18. Intelligence concretization scheme

3 Properties of Intelligence Formalisms

3.1 Ecological niches of Intelligence existence

Each category divides the Universe and creates a new ecological niche for the existence of Intelligence, the concretization of which effectively harmonizes this division under conditions of this niche with a harmonic advantage over the lower Classes of phenomena in it.

Definition. The division of the Universe is any difference from the state of the Harmon.

Thus, 1) the hierarchy of categories generates 2) the hierarchy of divisions, generating corresponding 3) hierarchies of ecological niches and further 4) the hierarchies of formalism 5) regarding its concretization and 6) materialization of Intelligence. As a result, a system of these hierarchies is formed, which is subsequently concretized and complicated as they become more detailed from the UA up to the RW (Fig. 19).

Intelligence ontologically originates from the UA, but, as far as categories are concerned, it arises at the level of the STM-Complex during the joint virtualization of the triad of divisions of 1) the AW/RW, 2) the UAP categories and 3) Time, and is further concretized up to the RW. Through meta-cognition, it is also able to virtualize the highest divisions up to the UA and thus achieves and harmonizes all Universe's divisions, which is fundamentally important for the Universal Cosmology.

All the said hierarchies sequentially inherit and develop their properties as the components are concretized. Intelligence formalisms develop from the UMFI to the RW with a loose mutual intersection of formalisms in conceptual and real niches with the obligatory inclusion of the UMFI, otherwise harmonization becomes unintellectual due to a fundamental reduction in harmonization resources (Fig. 20).

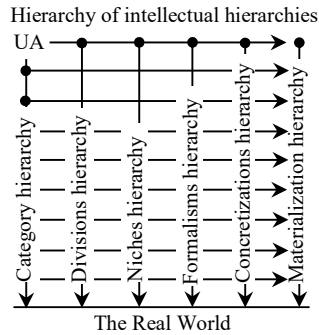


Fig. 19. The scheme of hierarchy of Universe's intellectual hierarchies

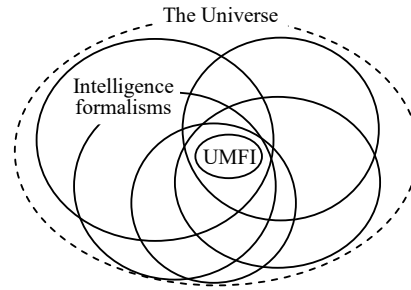


Fig. 20. The scheme of concretization of IMFI

3.2 Intelligence naturalization

Each niche is 1) a (maximal) simplification of the Universe, 2) including a niche harmonized separation and 3) a means of its harmonization 4) that also follow from the Universe. Obviously, intelligent formalisms inherit the Universe (macrocosm) and are its reduced copies (microcosms), which confirms the well-known philosophical hypothesis [16].

All the properties of Intelligence come exclusively from the Universe as a result of the naturalization of concepts, which explains and expands the base of mutual research of the Universe's phenomena and is part of universalization.

Naturalization obliges us to conceptually deduce any concept used from the sufficiently substantiated properties of the Universe, taking into account the fact that any knowledge is hypothetical and its truth is determined solely by the volume of the externally and internally consistent system of concepts and is relative to this system. Since any knowledge is incomplete and possibly contradictory, it is always ready to develop and eliminate contradictions.

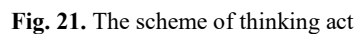
The analogy of the Universe, brain, computer and program is shown in Table 3 to confirm naturalization.

Table 3. Mutual correspondence of the components of the Universe, brain, computer and a program text

The Universe	Brain	Computer	Program text
Space	Neurons	RAM	Data area
Time	Biorhythms	Clock & command sequence	Algorithm
Matter	Neural network	Data	Data
Entity	Connected group of neurons	Variables	Variables
Relation	Neuron connections	Data structures	Data structures
Movement	Change of neurons/connections	Data change / command flows	Data change / functions call
Categories	Properties and rules	Machine commands	Language operators
Goal	Desires	Desired result	Desired result

The UMFI projects the Meta-Law on three reduced components of an object that are hardly visible in the RW: 1) the AW (UAP), 2) the RW (STM-Complex) and 3) the UAP categories.

The UMFI cyclically harmonizes an object by the following stages: 1) observing the object over a certain time interval, 2) building a model of an object's process, 3) predicting the development of the model over the next time interval, 4) predicting the required harmonic state of the object taking into account the target of a subject, 5) calculating the control action on the object, 6) returning from the subjective into real time, 7) implementing the control action on the object, 8) comparing the predicted and actual state of the object and 9) correction of the process model (Fig. 21).



As a result of the UMFI operation, the following classification of the discrepancy (disharmonization) of the predicted and actual state of an object occurs (Table 4).

Table 4. Classification of the degree of internal disharmonization of an object

Name	The degree of internal disharmonization of an object
Coincidence	No difference between the prediction and the fact of object harmonization
Difference	Little difference between fact and prediction for an object
Evolution	The loss of goal harmonization, which can be removed by the resources of an object-subject pair
Problem	The loss of harmonization of the goal, which is hard to remove by the resources of an object-subject pair
Contradiction	Nonremovable by the resources of an object-subject pair loss of harmonization of the goal, which can be removed by the resources of environmental means
Crisis	Balance of harmonizing and disharmonizing forces of an object
Revolution	Object downgrade
Disaster	Destruction of the definition of an object and the cessation of the existence of the object

3.5 Intelligence as an active converter of chaos into harmony

Class 1 reduces (clears of weak harmony), Class 2 preserves (accumulates), and Class 3 increases the harmony of the Universe under the influence of the Meta-Law and the built-in VCT (Fig. 10-12).

Thus, Class 1 is a passive transducer of harmony into chaos, Class 2 is a preserver, and Class 3 is an active transducer of chaos into harmony. Since these Classes inherit the previous Classes, they can establish mutual relations.

Class 3 harmonizing ability depends on both 1) the VCT transmission capacity and 2) the state of environmental harmony, which together perform 3) various non-linear positive / negative / oscillating harmonization of objects with a general prevalence of the former up to self-reinforcing explosion-like transition into a limiting state of the Harmon (anti-Big Bang), characteristic of Class 4 and important for the Universal Cosmology (Fig. 22) [9].

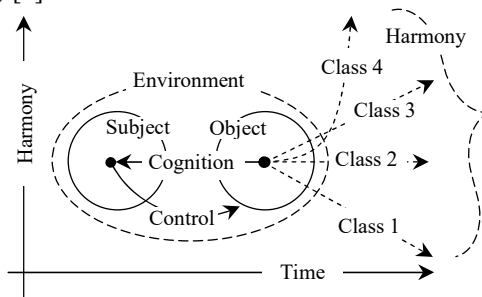


Fig. 22. The directions of phenomena harmony change of under the Classes 1-4 control

3.6 The prevalence of Intelligence in the Universe

The harmonious advantage of Class 3 over the other well-known Classes of phenomena should increase the presence of the former until the complete filling of the Uni-

verse, which is not really observed today. One of the reasons may be natural invisibility or a conscious hidden character of this Class under conditions of the infinity of the Universe in space and time, according to the hypothetical Universal Cosmology [9,17].

3.7 Intelligence as a subjective world

Living subjects move Harmony towards their subjective goal that must coincide exactly with the most harmonious state of an object. However, under conditions of always incomplete cognition, a goal is usually chosen causally approximately in this direction, or even vice versa if harmonization is included in the goal.

Under conditions of dogmatism, the goals are chosen inside subjects (the scheme of egoism), and under conditions of universalism – inside the society of subjects (the scheme of collectivism), between which there arise multiple combinations with the corresponding properties, the initial of which are given in Table 5. Thus, Intelligence creates multiple subjective worlds that ensure its special existence in real ecological niches.

Table 5. Classification of goals of subjects

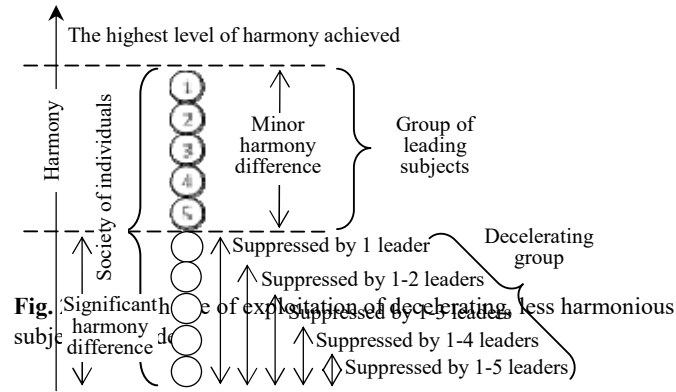
Target class	Target location	Object of harmonization
Egoism	Inside a subject	Subject
Altruism	Inside an object	Object
Collectivism	Inside a society	Society
Centralism	+ reasonable condensation of goals by special subjects	Society without duplication of control

3.8 Intelligence Aggressiveness

Living Classes (3-4) both produce and consume harmony from all the Classes with lower harmonizing ability. The social nature of Class 3 supports the interaction of living subjects with the transition of harmony to more harmonious subjects. Then they redirect harmony according to their goals.

Subjects contain multiple incoming/outgoing active relations and the corresponding flows of harmony/disharmony, the total amount/difference of which has a positive/zero/negative resulting value. Harmony is extracted from the environment/partner and is reciprocal for the latter, which determines the aggressiveness of subjects, which grows with intellectualization.

Society is usually very heterogeneous and ranked according to degree of harmony of its subjects with singling out of leading and decelerating subjects (Fig. 23). With a significant difference in their harmony, the former exploit the latter and increase this difference. Purely dogmatic societies are self-destructing and universalization consolidates society into a unitary organism.



Homogeneity and heterotrophy of Intelligence

Due to homogeneity of the AW and heterogeneity of the RW Intelligence has a supposedly single UMFI in different places (heterotrophy) of the Universe under any conditions, but in the order of successive filling of ecological niches (Fig. 24).

Homogeneity and heterotrophy generate Intelligence dichotomy on individuals and societies. Harmonization involves a variety of extremely important consistent actions and an individual must contain all of them (local Intelligence) in order to avoid their self-destruction. Society is a set of interacting individuals that can exist when part of individuals is lost.

Definition. An individual is a unitary whole Intelligence.

Definition. A society is a set of interacting individuals.

A society expands conceptual systems of individuals through the unification and distribution of individual concepts of all the members of a society and is, therefore, more harmonious (Fig. 25).

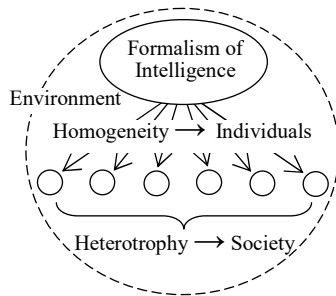


Fig. 24. The scheme of homogeneity/heterotrophy of Intelligence

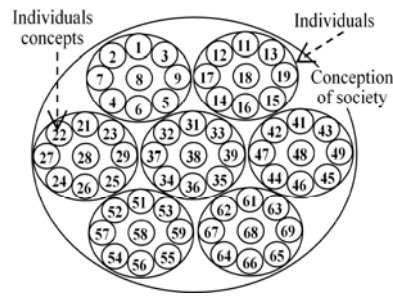


Fig. 25. Scheme of expanding a system of concepts of society regarding individuals

3.9 Passionarity of dogmatic Intelligence

Dogmatic Intelligence irregularly cognizes and causally identifies the higher concepts, the deduction from which, unexpectedly for a subject, produces new areas of knowledge and real possibilities for harmonization that are usually poorly understood but act strongly and change radically the activity of subjects like hit (impetus) (Fig. 26). This property is called passionarity that exercised a decisive influence on the formation of Eurasian ethnic groups and the development of world civilization [18-20].

Passionarity gives rise to unstable revolutionary oscillatory development of subjects and is the cause of the self-emergence/disintegration of peoples, states and empires for no apparent material reasons (Fig. 27).

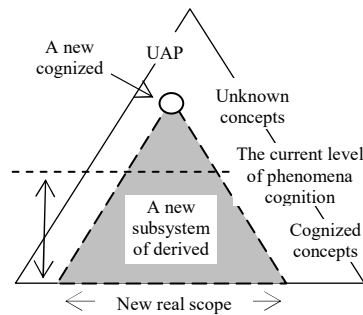


Fig. 26. The scheme of a passionary impetus

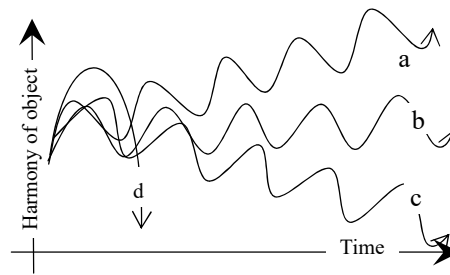


Fig. 27. The scheme of ascending (a), stagnating (b) and descending (c) oscillation or annihilation (d) of passionary processes

3.10 Instability/stability of dogmatic/universal Intelligence

Cognition begins with partial low-level dogmatic knowledge (Fig. 1) that naturally gives rise to 1) an unstable causal additional cognition with 2) a chaotic change in the direction of development/degradation, 3) an absent position of sustainable existence and 4) insoluble internal contradictions, 5) the only possible the method of resolving of which is the mutual suppression of all the opponents, except one.

On the contrary, universalization stabilizes cognition and the existence of subjects at the expense of the use of higher unifying concepts. On the one hand, it develops dogmatization, and, on the other – it radically changes knowledge up to the opposite one, so dogmatic and universal Intelligence can be considered different subclasses of Intelligence (Fig. 28).

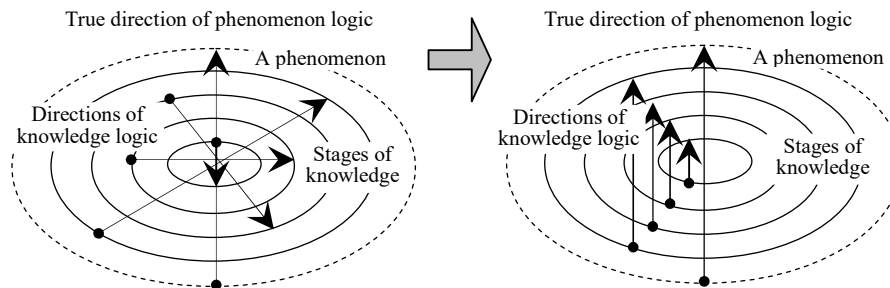


Fig. 28. The scheme of an instability of dogmatic phenomena cognition and its stabilization by universalization

4 Conclusion

Universalization uses additional resources of cognition and, for the first time, substantiates the Universal Meta-Formalism of Intelligence that allows to formally study and explain its basic properties and classifications. This is achieved exclusively by the transition from the particular dogmatic to the universal paradigm of cognition, revealing the highest universal meta-concepts and methods of cognition that change many traditional dogmatic conceptions. The meta-conceptual part of this work deduces the UMFI components, which are further concretized for different niches of materialization with the identification of important specific meta-properties that have their own fundamental practical and theoretical significance. Due to the special structure of the Universe, great originality and unusual nature of the topic and its content, this paper uses structural ER-formalisms that unambiguously allow both a natural verbal interpretation and set-theoretical representation up to the materialization of the obtained formalisms. This study has great scientific prospects and its further development leads to a full-fledged strong formalization of the living world in different environments with the execution on modern logical machines.

This study was carried out according to the plan of AI Problems Institute of the MES and NAS of Ukraine.

References

1. Legg, S., Hütter, M.: Universal Intelligence: A Definition of Machine Intelligence. *Minds and Machines*, 17(4): 391-444 (2007). doi:10.1007/s11023-007-9079-x
2. McCarthy, J.: What is artificial intelligence? Computer Science Department Stanford University, Stanford (2007)
3. Searle, J. R.: Minds, brains, and programs, *Behavioral and Brain Sciences*, 3 (3), pp. 417-457 (1980). doi:10.1017/S0140525X00005756
4. Turing, A. M.: Computing Machinery and Intelligence. *Mind*, 59, pp. 433-460 (1950)

5. Godfrey-Smith, P.: Theory and reality: an introduction to the philosophy of science. The University of Chicago Press, Chicago (2003). doi: 10.7208/chicago/9780226300610.001.0001
6. Sosnitsky, A.: Artificial Intelligence and Unresolved Scientific Problems. Information Theories and Applications, 18(1), pp. 82 – 92 (2011)
7. Sosnitsky, A.: Harmonious Foundations of Intelligence. Communication of SIWN, 7, pp. 66 – 72 (2009)
8. Sosnitsky, A.: Beginnings if the Universe Model and Deduction of Initial System of Information Concepts. Information Theories and Applications, 19(1), pp. 56–85 (2012)
9. Sosnitsky, A., Shevchenko, A.: The Chaotic Universe's Cosmogony: the Universe's Perpetuum Mobile, Multivariance of the Universe and the "Boiling" Universe Hypothesis. Journal of Applied Mathematics and Computation, 2(3), pp. 84–92 (2018). doi:10.26855/jamc.2018.03.002
10. Sosnitsky, A.: The Universal Theory as a New general Scientific Paradigm. Journal of Applied Mathematics and Statistical Applications, 2(1), pp. 11-18 (2018)
11. Shannon, C. E.: A Mathematical Theory of Communication. Bell System Technical Journal, 27 (3), pp. 379–423 (1948). doi: 10.1002/j.1538-7305.1948.tb01338.x
12. Schrödinger, E.: What is Life? The Physical Aspect of the Living Cell. Cambridge University Press, Cambridge (1944)
13. Pavlov, I. P.: Conditioned reflexes. An investigation of the physiological activity of the cerebral cortex, Oxford University Press, London (1927)
14. Boakes R.: From Darwin to behaviourism. Cambridge: University Press, 1984.
15. Saladin, K.: Human anatomy (3rd ed.), McGraw-Hill, New York (2011)
16. Hamilton E., Cairns H.: Collected Dialogues of Plato. Princeton University Press, Princeton (1962)
17. Shevchenko, A.: The God, Nauka i Osvita, Kiev (2017)
18. Sosnitsky, A.: The Universal Theory of Passionarity. Web Intelligence, 17(1), pp. 75-84 (2019). doi:10.3233/WEB-190402
19. Gumilev, L. N.: Les Fluctuations du niveau de la mer Caspienne. Cahiers du Monde Russe et Sovietique, Paris, 6(3), pp. 331–366 (1965)
20. Gumilev, L. N.: New Data on the History of the Khazars. Acta Archeologica Academiae Scientiarum Hungaricae, 19, pp. 61–103 (1967)

Designing Complex Intelligent Systems on the Basis of Ontological Models

Olha Tkachenko^{1[0000-0003-1800-618X]}, Oleksandr Tkachenko^{2[0000-0001-6911-2770]},
Kostiantyn Tkachenko^{1[0000-0003-0549-3396]}

¹ State University of Infrastructure and Technologies, I. Ogienko str., 19, Kyiv,
02000, Ukraine

oitkachen@gmail.com
tkachenko.kostyantyn@gmail.com

² National Aviation University, Liubomyra Huzara ave. 1, Kyiv, 03058, Ukraine
aatokg@gmail.com

Abstract: The article discusses the problems of automation of complex systems' design process, which include intelligent systems, using the example of designing a complex technical object – motor vehicle. The concept of designing the intelligent systems on the basis of modeling is considered. An intellectual system model is a composition of following models: tasks, subject area, user, presentation, dialogue script. The ontology is considered in the article as a model of the architecture of an intelligent system.

Keywords: intelligent system, subject area, system models, ontology, ontology model, knowledge base, architecture of intelligent system.

1 The Intelligent System and Ontologies

Information in the modern world has become one of the most important resources, and intelligent systems (IS) have become a necessary tool in various fields of human activity. By now, a huge amount of information has been accumulated, which is not fully utilized. In addition, such information can be represented in many different formats, which makes it difficult to use. All this led to the emergence of the problem of heterogeneous information unification.

To solve these problems, a description of the data, its processing tools and configurations that are the objects of modeling when creating a universal semantic model for representing and processing knowledge is used [1, 2, 5-7, 8]. All components of IS and the relationship between them can be represented in the form of semantic model.

Currently, ways to solve the problems of using semantic modeling are being studied to coordinate descriptions of interacting systems, methods and technical implementation of such semantic models [3, 4, 9, 12-15]. However, problems associated with practical implementation are still relevant, in particular: the lack of

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

agreed methods for the formation of ontological models, their integration and transformation, the lack of methods for creating and coordinating ontological models in the development (design) of IS.

The ontology is a detailed specification of the structure of a specific subject area (SA). The main function of ontologies is to integrate information. Ontologies suggest the implementation of the following aspects:

- they give a definition of the formal semantics of information, automating the processing of this information;
- they determine semantic relations with real world objects, allowing to connect the information that is presented in the form necessary for computer processing with information that presented in a convenient form for human perception, on the basis of general terminology [3].

In addition, in any area of human activity, ontologies allow:

- *refine the structure of knowledge*. In an ontological analysis, the concepts of the subject domain and the relationships between them are determined so that the result is a clear specification of the concepts and terms used in relation to the body of knowledge that must be built [4].
- *reduce conceptual and terminological ambiguity*. Ontological analysis provides the basis for the synthesis of the objects under consideration with different needs (or viewing angles) depending on the specific context [9].
- *provide knowledge sharing*. Thanks to ontological analysis, many conceptualizations of a certain subject area and a set of terms supporting it are achieved. Having adequate syntax, these conceptualizations and the relationships between them are expressed and encoded in ontology [8].

Intelligent systems belong to the class of complex systems. A characteristic feature of modern IS are the transition to the architecture of heterogeneous, including distributed objects, services and infrastructures. For example, autonomous external intelligent systems can be considered as objects. In this situation, one of the most important problems is the problem of creating a common IS architecture and maintaining integrity at all stages of its life cycle.

By architecture we mean the organization of a system that takes into account the composition of its components (elements, subsystems, etc.), their relationship with each other and the external environment, the principles of creation, management and development. When developing the architecture of IS, it is necessary to coordinate all the requirements for the system with the methods for their implementation, determined by the characteristics of the interacting subsystems. The complexity of creating IS is determined, in particular, by such reasons:

- there is no universally accepted conceptual base for the description of subsystems, the conceptual base for the description of the architecture of each subsystem reflects its subject area;
- a big influence on the architecture of IS is provided by the system that provides the process of IS developing; therefore, it is necessary to ensure the interaction at the level of the target and supporting systems when coordination of functional, technological and structural descriptions is required.

Consider the approach to creating a unified ontology of IS architecture, based on knowledge management technology, its usage to harmonize the functional and structural descriptions of the target and supporting systems. The knowledge base that

was formed on the basis of the proposed ontology allows one to accumulate the experience of building architectures and system relationships of IS and use the experience to develop new IS. The IS model is a composition of:

- model of tasks;
- model of domain;
- model of users;
- model of presentation;
- model of dialogue scenario.

Intelligent systems describe the behavior of managed objects in a complex dynamic environment. IS design is carried out on the basis of a conceptual model of the main properties of the studied subject area, the construction of the knowledge base. This approach provides the opportunity for a formal definition of concepts and laws of classification [8].

2 Conceptual Ontology Model

A conceptual model of knowledge about design models and methods is implemented on the basis of an integrated ontology with a reasonable structure and content. The system that describes the knowledge base when real-time information systems are functioning contains elements that provide ontology modification based on the extension of standard integrated ontology and metaontology procedures.

By a formal ontology model we mean a set consisting of three sets $O = \langle A, R, F \rangle$, where A is a finite set of concepts of SA; R is the finite set of relations between the concepts of SA; F is a finite set of interpretation functions defined on concepts [12].

A formal model of an integrated ontology assumes the presence of:

- FORMALIZATION OF COMPLEX ONTOLOGY, including:
 - modernization of the classic model;
 - modernization based on the concept of the system.
- FORMALIZATION OF METAONTOLOGY, including:
 - model based on the concept of a system;
 - model based on formal theory.
- FORMALIZATION OF THEORETICAL MODELS OF KNOWLEDGE ENGINEERING, including:
 - model “Field of knowledge”;
 - model “Pyramid of Knowledge”.
- FORMALIZATION OF THE AXIOM OF ONTOLOGY, including:
 - axioms of identification;
 - axioms of planning;
 - axioms of computing.

The ontology is based on the construction and analysis of a semantic model, the objects of which are systematized according to the functional attribute of their properties, which are the part of the class hierarchy. Relations determine the structure of the system, and elements determine the function of nodes in this structure. The root

quality structure can be expanded to more specific classes by dividing properties at each level in accordance with the division of the class.

The considered classification can be refined (renamed, expanded, supplemented by arbitrary classes) depending on the problem of processing information and changing knowledge about the designed dynamic object [4, 5-7, 9-12].

For example, the class **INFORMATION STRUCTURE** in the classification of properties can be refined by highlighting the properties of the STRUCTURE BY MANAGEMENT and the STRUCTURE BY DATA. This will entail highlighting in the classification of components within the **INFORMATION COMMUNICATION** class of the COMMAND TRANSFER and DATA TRANSFER subclasses, to which the corresponding notation can be assigned.

Thus, the systematic approach [13-15] allows you to use your own set of modeling tools: elements and relationships in solving each specific task of information processing. At the same time, it is possible to diversify instances of classes of information links and elements, preserving the corresponding components in the form of abstract classes.

The systematic approach to the interpretation of design results and the knowledge base which is based on a complex ontology allows us to represent the state tree of a complex system, which is also an IS.

IS integrates the achievements of artificial intelligence, methods of classical mathematics, fuzzy logic, the theory of neural systems and a genetic algorithm to solve difficult formalized decision-making problems in a complex dynamic environment.

Within the framework of such an interpretation, the principle is maintained that the properties of elements and relationships are determined by the hierarchy of classes that take these properties into account. This provides the possibility of system decomposition of the task of creating IS in the interpretation of complex dynamic situations. When formalizing the task of constructing a knowledge base, ontologies improve the interpreted characteristics of IS and simplify its usage for analysis and modeling of design decisions, especially in emergency and extreme situations.

To ensure the functionality of the IS, the law of system decomposition is used [13-15]: elements on the i -tier of the system are in relation to support the functional ability of the $(i + 1)$ -tier of the system (the system should support the subsystem, subsystems – the system, etc.). Compliance with this law is ensured by fulfilling the following system decomposition rules:

- the correct attachment of elements to each other in accordance with the qualitative and quantitative characteristics of the bonds (*rule of attachment*);
- ensuring a qualitative and quantitative balance of incoming and outgoing functional relationships (*rule of balance*);
- closure of the supporting bonds (*rule of closure*).

The systematic approach formulated in this way can be applied at the stage of substantiating the principles of construction and functioning of a system for modeling and analysis of complex situations. The used ontology model describes a computer interpretation (building models for formal systems) of the subject area related to the formalization of the task of monitoring the functioning of IS and modeling emergency and extreme situations that arise during operation.

3 Domain Ontology Model / Ontology Model of Subject Area

The development and formalization of a complex ontology during the creation of IS consists in a formalized representation of the SA model, which assumes a description of many objects and concepts, knowledge about them and the relationships between them.

Definition 1. *Subject area (SA)* – a part of reality displayed using intelligent technologies to obtain new information (knowledge). SA, when formalizing knowledge in IS, is interpreted as part of the real world that has a certain *semantic localization* – spatial, temporal, functional, etc. When considering the semantic space of the studied software, it is necessary to first carry out its semantic localization.

Definition 2. *Semantic localization* in IS is associated with the determination of the interface between SA in the semantic space of the subject areas. At the stage of formalizing knowledge, the components included in software can be considered as sets of their semantic properties $(M_{SA-1}, \dots, M_{SA-N})$ [16]:

$$M_{SA-1} = \{SS_{11}, SS_{12}, \dots, SS_{1m}\}, \dots, M_{SA-N} = \{SS_{N1}, SS_{N2}, \dots, SS_{Nm}\}.$$

In this case, the intersection of the sets of semantic properties of various SA

$$M_{SA-1} \cap \dots \cap M_{SA-N} \neq \{0\},$$

which allows us to write down the criteria for the localization of SA in the semantic space $M_{SA-1} \cap \dots \cap M_{SA-N} = \{0\}$.

A necessary criterion for the existence of SA is the distinguishability of its properties in the presented semantic localization. For the set of *SS* properties of this model of the studied SA, its unique identification follows $\forall s, s \in SS \Rightarrow M_{SA}$.

At the same time, it is believed that the properties remain identical to themselves for a time sufficient to build the SA model and use it in a formalized IS knowledge system, which is especially promising when implementing aspect-oriented modeling and programming technology.

A great importance in SA modeling at the development stage is functional completeness and logical integrity. The SA model is necessary for solving a certain class of design problems and their geometric and analytical interpretation. Therefore, it should include only the necessary and sufficient properties for this. The functional completeness of the SA model implies only fixing those properties of objects that are necessary and sufficient to solve the design problems.

The criterion for the functional completeness of the SA model depends on the class of tasks being solved and requires the determination of the criterion for the depth of detail of the SA. A formalized formulation of the tasks to be solved in a given SA allows you to highlight the features of the current situation, the modeling of which is necessary and sufficient for a rational choice of design solutions.

4 Formalization of the Domain Ontology

The conceptual foundations of the formalization of different SA knowledge ontology are considered in [1-3, 10, 11, 17-23]. Within the framework of the approach

proposed in [18], classes of terms, relations, and transformations corresponding to physical and abstract entities for solving SA problems are distinguished. The SA representation serves as a signature for creating a model of subject knowledge. The ontology of SA is denoted $Ont(SA)_S$ and defined in the following form [22]:

$$Ont(SA)_S = \langle T(S), R(S), Ax(S) \rangle,$$

where S (Subject) is a finite set of SA objects; $T(S)$ (Terms) – a finite set of terms (concepts) of software that have qualitative features that make up their distinctive feature in the ontology S ; $R(S)$ (Relations) – a finite set of relations between classes of terms; $Ax(S)$ (Axioms) is a finite set of axioms (interpretation functions) defined on the classes and relations of the ontology $Ont(SA)_S$.

The ontology formation approach proposed in [18] uses the principles of object-oriented analysis and consists in a phased refinement of constructions of the type object - attributes and interactions between objects.

Strong name groups (for example, in texts they can be nouns) help to describe a lot of objects that are combined into classes of terms, forming a factor set. To support the IS developer, a dictionary of word classes close to each other is used.

Linguistic problems are eliminated by the fact that each object in a given situation gets some direct name, which distinguishes it from various indirect ones – classifiers and indirect names. Ontology can be considered as a language-dependent conceptual model [22].

For the convenience of formalization of the information about the SA, the matrix K_{ONT} (Knowledge-Ontology) is used, in which the columns are the groups “object” and “facts”, and the rows are the corresponding records with the names of objects and the facts related to them.

General characteristics of groups and sets used in the matrix K_{ONT} :

- group “Object” (Object), where $i = 1, 2, \dots, n$; n is the number of identified objects;
- attributes group, where $j = 1, 2, \dots, m$; m is the number of attributes of the corresponding object;
- behavior group that is added for each object, where $l = 1, 2, \dots, p$; p is the number of behaviors of the corresponding object.
- group of interaction that defines the subject and object of the proposal.

5 Justification of the Approach to the Construction of IS Ontology

IS knowledge bases, developed on the basis of ontological models, represent the design process in the form of an ordered structure with clearly defined relationships between the elements (components, steps, tasks) of the design process.

Consider the ontology of SA – “Vehicle Design”, which contains the concepts of this SA, the interpretation of knowledge and relationships within this area.

The ontological approach in this area is described in the form of the following generalized algorithm of actions:

- compilation of a dictionary of SA (Thesaurus) based on the design components of a complex technical system, which is a vehicle;
- obtaining ontology according to the Thesaurus of SA “Vehicle Design” reflecting the “natural” connections between concepts;
- verification by experts in this field of the resulting ontology of the SA “Vehicle Design”, support and filling of the ontology.

Thesaurus SA “Vehicle Design” can be used as a tool for standardization and formalization of knowledge, as well as to provide access for users who solve the problems of optimal design of a vehicle.

Thesaurus of the SA “Vehicle Design” is designed to solve the following problems:

- classification and unification of the concepts of SA;
- classification of methods and tasks of designing vehicles;
- the construction of descriptions of the methods and tasks of vehicle (aircraft) design in the knowledge base to support the optimal design of a vehicle;
- classification and search for background information on this topic.

Ontology for SA “Vehicle Design” uses the Thesaurus and is necessary for:

- developing a common understanding of the field of knowledge in question;
- presentation of knowledge in a form that is convenient for their processing by appropriate IS;
- opportunities for obtaining and accumulating new knowledge;
- providing the possibility of reuse of knowledge.

Ontology describes the basic relationships and relationships between elements of the design process.

IS design involves the usage of resources of the supporting system (S_{SS}). The description of the SA of IS, including the conceptual model and the task model, is the input information, and the IS architecture is the result of the function $S = S_{SS}(S_D)$, where S is the developed (designed) IS providing a description of its components (elements, subsystems, etc.); S_D – system for describing the SA; S_{SS} – providing the system for the development and operation of IS S .

S_{SS} is a complex system in which each type of activity has its own separate, independent support system (subsystem). When developing an IS, it is necessary to solve the problems of identifying and coordinating problems solved within the framework of IS S , developing appropriate data structures, their processing tools, a user interface, and a model of transitions between individual elements (components) of the system. Then

$$S = \langle S_D, S_T, S_I, S_C, S_G, S_N \rangle,$$

where S_T is a system for describing the processes and tasks of various SA and their management; S_I – information model management system that allows you to describe information objects and data structures; S_C – component development

system; S_G – user interface development system; S_N – a system for developing a navigation model – a model of transitions between individual elements (components).

The development of IS (its architecture) involves modeling processes, including a description of ontological models, taking into account the interests (concerns) of interested parties (stakeholders), including models for describing the relationships between the functions of the system and its components.

For each of the interests, separate groups of system descriptions (views) are created. Each group of descriptions reveals a separate aspect of the system, and a set of groups forms its complete description.

The agreements by which a group of descriptions is created, displayed and analyzed, are established by the description method (viewpoint). Architecture (A_S) of the system S : $A_S = \{A_{DSC}\}$, where $A_{DSC} = \langle M_i, A_i, R_i, Z_i \rangle$; A_{DSC} – group of descriptions; Z_i is a description method; $i(i = 1, \dots, n)$ is an aspect characterizing the A_{DSC} corresponding to Z_i , connecting the set of elements M_i defined on the set of their attributes A_i through the relations R_i . Using the semantic description method, with each S_{SS} , we associate its description group and the ontology of the SA:

- O_S – ontology of the target system – the result of the development and description of the technical implementation of IS;
- O_D – ontology of SA;
- O_T – ontology of tasks;
- O_N – navigation ontology;
- O_I – ontology of information elements;
- O_C – ontology of components;
- O_G – ontology of the user interface.

Each of the groups of descriptions is characterized by its own set of models, tasks, tools, executors, structure, functioning, etc. For each group of descriptions, methods for considering and describing systems and rules for their application are determined. When reused for each description group, individual ontological models are defined that are elements of the description group ontology.

The joint usage of ontologies implies the need for an ontological system O^S :

$$O^S = \langle O_{meta}, \{O_{D\&T}\}, M_{MO} \rangle,$$

where O_{meta} is metaontology (top-level ontology), that is, the ontology of the architecture of the system S ; $\{O_{D\&T}\}$ – many ontologies of terms (concepts) and ontologies of SA problems; M_{MO} is a model of an output machine associated with the ontological system O^S .

O_{meta} describes classes that allow you to:

- represent the concepts of SA (terms, definitions) and their relationship;
- integrate knowledge of subsystems and use this knowledge in solving problems.

The ontology of IS (its architecture) involves a set of necessary ontological classes. Examples of definitions presented in the ontology of IS architecture, for example, may be:

- model of ontology = {ontology of concepts, ontology of operations, ontology of notations};
- system = {target system (S), providing system (S_{ss})};
- description group = {domainview, taskview, infoview, navview, guiview, compview}, where the following description groups are used:
 - domainview – conceptual model of software;
 - navview – the navigation structure of the application;
 - infoview – information elements and their data schemes;
 - guiview – user interface;
 - taskview – processes and tasks;
 - compview – components of IS.

The ontology consists of classes of SA and classes of classifiers. The set of classifier objects defines a partition of the set of objects of the class of SA into groups. Many classifiers, together with their relationships, define the class of the relationship template, since the relationship structure of their objects defines the relationship structure between objects of the SA classes. An inheritance system is established between classifiers, and a system of relations between specializations (specifications) of templates between templates. Example of relationship:

- the type of *model corresponds to* – the type of *primitive modeling*;
- the type of *the model of the user interface subclass* – the type of *the primitive modeling of the user interface*;
- the type of *the user interface model corresponds to* – the type of *the primitive modeling of the user interface*;
- *the model consists of* – *a primitive simulation*;
- *user interface model* – *subclass model*;
- *user interface model consists of* – *a primitive user interface modeling*.

The connections between S and S_{ss} (including between their components), as well as set M_{om} ontological models are determined by templates defined by the

PROJECT class and its classifiers. Elements of the set M_{om} when creating and using the knowledge base can be selected into separate subsets of models (by introducing the necessary classes and relationships) and reused in various projects.

The integration of description group ontologies is provided by the annotation mechanism. An *annotation* is a pair $\langle u, e \rangle$, where e is a concept from O_{meta} , or a composition of concepts from O_{meta} , u is an instance of concept e . Thus, the classes and ontologies of description groups are instances of concepts that are subtypes of metaontology concepts O_{meta} .

We will consider:

- O_{meta} metaontology, describing a set of designs used in wide class of information models;
- ontologies of description groups V and U ;
- sets of annotations A_V and A_U , respectively, in terms (concepts, definitions) of the O_{meta} metaontology associated with the classes of description group ontologies.

For the annotation operation, a special relationship is created between the subgraph from V or U and the concept from O_{meta} . For this a special class COMMUNICATION is introduced.

Ensuring the interoperability of models is supported by mapping the $M_{U(V)}$ model V to model U . These models are not connected in any way, so it is necessary to find semantically close constructions of models. We consider semantically close pairs of constructions $\langle u, v \rangle, u \in U, v \in V$ for which there exist annotations $\langle v, e \rangle \in A_V$ and $\langle u, d \rangle \in A_U$ such that e is a subring of d .

We define $task_m$, $info_j$, gui_i , $navlink_i$, f_k as elements from description groups: taskview, infoview, guiview, navview, compview, respectively. In the ontology of IS (its architecture), for each of the ontology description groups, we define classes of modeling primitives (in the general case, the class inheritance tree). Determining the relationships between the classes of primitives and applying the annotation mechanism will allow us to present models from different groups of descriptions and determine the architecture of IS.

6 Design Features of Ontology-Driven IS

The creation of ontologies is carried out not only in the development of environments aimed at sharing information between several users, but also in the design of knowledge bases, the creation of expert systems and decision support systems, and the development of various search engines.

When designing IS managed by ontologies, it is worth considering the features of the life cycle of the ontologies themselves. The life cycle of ontologies is intertwined with the life cycles of projects for the development of specific software products that are connected to the ontology of the same applied sphere. Both life cycles are supported by the respective roles: software engineer and ontology engineer.

M. Fernandez emphasized a certain analogy between the two processes and investigated well-known models of the software development life cycle, presenting them as potential paradigms for the development of ontologies. Among the considered paradigms are [5-7]:

- cascading model;
- phased model;
- evolutionary model.

The development of ontology does not imply the implementation of a plan sequence with intermediate results (only individual ontology components can be developed in this way). Therefore, cascading models are inappropriate.

A phased model is more appropriate. Starting from the root (an element of the ontology of the highest level), formalization of individual elements of the SA is carried out. However, a strong connection with the previous stages of development makes the final ontology redundant.

The most suitable for the development of ontology is the evolutionary model, which begins with a prototype ontology containing some basic definitions. These definitions can be presented in the form of a reference book (Thesaurus). Further, the implementation of each SA product project initiates a new ontology evolution cycle.

Ontologies are a promising tool for transferring knowledge from project to project; from one development cycle to the next project; from a project in one SA to a project in another SA.

7 Final Results and Conclusions

Researches show that the complex ontology used in the design of IS can be based on various formalizations.

The developed ontology allows us to formalize the description of the architecture of IS and to integrate models created in the framework of different ontological models.

The presented ontology can be used in the design of IS.

As a result of the ontology usage, the need for introducing special models of ontologies that are of a general nature has been identified. One such ontology is the ontology of the IS life cycle.

In the long term, ontologies can become an attractive paradigm of software engineering.

The described methodology will provide an opportunity to use it when designing compatible models and to reuse the IS elements, as well as to reduce the cost of designing and developing of software products.

References

1. Gavrilova, T.A., Choroshevskij, V.F.: Bases knowledge of intellectual systems. Piter, St. Petersburg (2000)
2. Kudryavtsev, D., Menshikova, A., Gavrilova, T.: Representing Strategic Organizational Knowledge via Diagrams, Matrices and Ontologies. International Journal "Information Theories & Applications", vol. 23, no 1, pp. 48-66 (2016)
3. Kostenko, K.I.: The rules for abstract knowledge spaces inference operator. Software engineering, vol. 7, no 6, pp. 258-267 (2016)
4. Kostenko, K.I.: Simulation of inference operator for hierarchical knowledge representation formalisms. Software engineering, vol. 7, no 9, pp. 424 – 431 (2016)
5. Fernandez, M., Gomez-Perez, A., Juristo, N.: Methontology: From ontological art towards ontological engineering. In: Symposium on Ontological Engineering of AAAI. Stanford California (1997)

6. Fernandez, M., Overbeeke, C., Sabou, M., Motta E.: What makes a good ontology? A case-study in fine-grained knowledge reuse. In: ASWC 2009 - Proceedings of the 4th Asian Semantic Web Conference 2009. Shanghai, China, pp.61-75 (2009)
7. Suarez-Figueroa, M.C., Gomez-Perez, A., Fernandez-Lopez, M.: The NeOn Methodology for Ontology Engineering. *Ontology Engineering in a Networked World*. pp. 9-34 (2012)
8. Denisenko, V.N., Krasina, E.A.: General theory of systems and linguistic systemology of professor G.P. Melnikov: Methodology and Method. *Bulletin of RUDN University. Series Theory of Language. Semiotics. Semantics*, no. 1, pp. 15-21 (2014)
9. Gelovani, V.A., Bashlykov, A.A., Britkov, V.B., Vyazilov, E.D.: Intelligent decision support systems in emergency situations using information on the state of the environment. Editorial URSS, Moscow (2001)
10. Bondarenko, M.F., Matorin, S.I., Solovieva, E.A.: Features of the theory and practice of solving complex problems based on ontology. *Artificial Intelligence*, no. 3, pp. 25-33 (2000)
11. Chaplinsky, Y.P., Subbotina, O.V.: Ontology and context in applied decision-making. *Artificial Intelligence*, no. 2, pp. 147-155 (2016)
12. Gehlert A., Pfeiffer D., Becker J. The BWV-model as method engineering theory. In: *AMCIS 13 th - Proceedings of the American Conference on Information Systems*, vol. 5, pp. 3389-3398 (2013)
13. Mesarovich, M., Takahara, Ya.: *General theory of systems: mathematical foundations*. Mir, Moscow (1978)
14. Ginis, L.A., Gordienko L.V.: *Modeling complex systems: a cognitive set-theoretic approach*. Publishing House of the Southern Federal University, Rostov-on-Don – Taganrog (2016)
15. Kononyuk A.E.: *Systemology. General theory of systems*. Osvita Ukraine, Kiev (2014)
16. Vostrov, G.N., Mezhuiev, V.I.: Problems of building information systems over subject areas. *Artificial Intelligence*, no. 4, pp. 736-746 (2008)
17. Palagin OV, Petrenko MG: On some features of construction of ontological models of subject areas. *Control Systems and Computers*, no 3, pp. 23-37 (2019). doi.org/10.15407/csc.2019.03.023
18. Yakimov, V.I., Dyakonov, G.N., Mashkov, A.V.: The formation of domain ontologies based on the analysis of the NFL-continuum. *Information Technologies*, no 3, pp. 36-39 (2006)
19. Tkachenko, O., Tkachenko, A., Ovcharuk, I., Tkachenko, K., Radionov, B., Chyhyr, I.: Modeling of process management in online lending systems. *Norwegian Journal of development of the International Science*, vol. 1, no 35, pp. 44-49 (2019)
20. Fridman, N., Hafner, C.: Ontology design: A survey and comparative review. *Artificial Intelligence Magazine*, no 18(3), pp.53–74 (1997)
21. Zhou, L.: Ontology Learning: State of the Art and Open Issues. *Information Technology and Management*, no 8(3), pp. 241-252 (2007)
22. Greger, S.E., Porshnev, S.V.: Building an ontology of information system architecture. *Fundamental Research*, no 10, pp. 2405-2409 (2013)
23. Uschold, M., Gruninger, M.: *Ontologies: Principles, methods and applications*. Knowledge Engineering Review, vol. 11, no 2 (1996).

Structuring of a Transaction Database Using the Rough Set Theory

Igor Kovalenko¹[0000-0003-2655-6667], Alyona Shved³[0000-0003-4372-7472],
Kateryna Antipova⁴[0000-0002-9012-5290] and Yevhen Davydenko²[0000-0002-0547-3689]

Petro Mohyla Black Sea National University, Mykolaiv, Ukraine

{ihor.kovalenko, avshved, antipova.katerina,
davydenko}@chmnu.edu.ua

Abstract. This paper describes an approach aimed at structuring of the generated transaction database in the process of forming a market basket using the rough set theory. The analysis of the recent publications and achievements showed that the structuring, categorization and classification at the stage of preliminary transaction analysis, before rules discovery, remain underdeveloped. The paper describes cases for a target set of the elements of the universe, that are classified into categories based on the equivalence relation. Approximations are proposed for the formal representation of such a set based on the rough set theory. A detailed example of the analysis and classification of transactions with various sets of objects is given.

Keywords: rough set theory, association rules, universe, transaction, market basket, approximation.

1 Introduction

The modern affinity analysis is one of the common techniques of data mining that discovers relationships between co-occurring activities.

Affinity analysis is mainly used for market basket analysis, the purpose of which is to detect associations between different events for the quantitative description of the connection between two or more events, which are called association rules. Such rules have two basic concepts: a transaction – a certain set of events co-occurring together (for example, a customer purchasing goods at a supermarket) and an itemset – a non-empty set of goods that have been bought in a single transaction.

Nowadays supermarkets collect information about purchases and store it in a database to be used later for association rules discovery.

However, the data gathered this way is not structured, since the transactions are written to the database one after another and their attributes, such as the frequency of transactions, the size of the itemset, its cost indicators etc., are not taken into account.

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

2 Formal problem statement

Thus, the database is loaded with the unstructured data where each storage unit can be represented by a finite number of attributes. This leads to a task of categorization and classification of unstructured data as well as the reducing of the amount of data by removing superfluous transactions.

The aim of this paper is to develop a methodology for structuring a transaction database at the preliminary stage of their analysis using the rough set theory.

3 Literature review

A series of recent publications [1, 2, 5-11, etc.] discusses the development of methods, techniques and algorithms for the analysis of transactions and the binary association rules discovery in them. So, in the work [1], a rather in-depth review of the existing approaches to the search for numerical, generalized, temporal and fuzzy associations was performed.

In [2], the problem of constructing bases of numerical associative rules is solved: a method for synthesizing bases of associative rules is developed, in which the transaction database is fuzzified, threshold support values are calculated, criteria are used to evaluate indirect associations, which reduces the degree of user or expert participation [3, 4] in the process of searching for associative rules, and also allows you to retrieve not only frequently found sets, but rarely arising interesting associative rules.

The work [5] is devoted to a comparative analysis of the tools for association rules discovery (Apriori, DHP, Partition, DIC, etc.). In the work [6], the combined use of associative analysis and the decision tree method for solving economic problems is considered.

It should be noted that the existing publications generally propose the tools for the direct search for binary association rules in transactions. At the same time, the task of structuring and classification at the stage of preliminary transaction analysis, before rules discovery, remains underdeveloped. The theory of rough sets of Z. Pawlak [12] can be used to perform this task. This theory was further developed in [13-21]. It operates with the arrays of disordered (rough) data and, through their categorization, gives them a structured form.

In [13-17], authors outline some selected past and present research directions of rough sets. In particular, has emphasize the importance of searching strategies for relevant approximation spaces as the basic tools in achieving computational building blocks (granules or patterns) required for approximation of complex vague concepts. In [18], proposes a hybrid multi-granulation rough sets based on variable precision tolerance relations. Basic properties of hybrid multi-granularity rough set of variable precision are discussed, which provides a new approach to deal with the incomplete information system. A new generalization of coarse fuzzy sets in the generalized approximation space is proposed in [19-21]. The equivalence relations are viewed as a special type of a binary relations of a universe. Then, the rough fuzzy sets in generalized approximation space is defined.

4 Basic framework of the rough set theory

This theory defines a knowledge base as $K=(U, R)$, where U – is a finite set of objects (the universe), R – is an equivalence relation on U . For any R there is an associated equivalence relation $IND(R)$. The relation $IND(R)$ is called R -indiscernibility relation. Each partition includes elements that have the same values of classification features (attributes). Within each partition, elements are considered indiscernible.

Let $X \in U$ be a target set while the objects of U are categorized based on the attribute R , then the following situations are to be considered:

1. A set X is called crisp (exact) with respect to R if and only if the boundary region of X is empty. The boundary region consists of those objects that can neither be ruled in nor ruled out as members of the target set X .
2. A set X is called rough (inexact) with respect to R if and only if the boundary region of X is nonempty.

In order to characterize the set X with respect to R , additional notation and basic concepts of rough set theory are presented below:

1. R -lower approximation of a rough set X is a subset of all objects which can be with certainty classified as members of X with respect to $IND(R)$:

$$\underline{R}X = \bigcup \{Y \in IND(R) : Y \subseteq X\} \quad (1)$$

2. R -upper approximation of a rough set X is a set of all objects which can be only classified as possible members of X with respect to $IND(R)$:

$$\overline{R}X = \bigcup \{Y \in IND(R) : Y \cap X \neq \emptyset\} \quad (2)$$

The R -lower approximation of X is called R -positive region of X :

$$POS_R(X) = \underline{R}X \quad (3)$$

The R -negative region of X is a subset of the objects of the universe that can be definitely ruled out as members of a target set X :

$$NEG_R(X) = U - \overline{R}X \quad (4)$$

The boundary region of a set X is a subset of all the objects that belong to the R -upper approximation of X :

$$BN_R(X) = \overline{R}X - \underline{R}X \quad (5)$$

5 Example

Let us consider a knowledge base $K=(U, R)$, where $U=\{x_1, x_2, \dots, x_{10}\}$ – is the universe, R – is an equivalence relation [13]. The selected equivalence classes are:

$$U/IND(R) = \{\{x_1, x_2\}, \{x_3, x_7, x_{10}\}, \{x_4\}, \{x_5\}, \{x_6\}, \{x_8\}, \{x_9\}\}.$$

The two target sets are: $X_1 = \{x_1, x_2, x_4, x_5\}$, $X_2 = \{x_1, x_2, x_3, x_4\}$.

The approximations, negative and boundary regions of the sets were obtained as follows:

$$\begin{aligned} \underline{R}X_1 &= \{x_1, x_2, x_4, x_5\}, \\ \overline{R}X_1 &= \emptyset; \\ NEG_R(X_1) &= \{x_3, x_6, x_7, x_8, x_9, x_{10}\}, \\ BN_R(X_1) &= \emptyset. \end{aligned}$$

$$\begin{aligned} \underline{R}X_2 &= \{x_1, x_2, x_4\}, \\ \overline{R}X_2 &= \{x_1, x_2, x_3, x_4, x_7, x_{10}\}; \\ NEG_R(X_2) &= \{x_5, x_6, x_8, x_9\}, \\ BN_R(X_2) &= \{x_3, x_7, x_{10}\}. \end{aligned}$$

In order to evaluate the accuracy of the rough set representation of the set X , the following estimates were introduced:

$$1. \alpha_R(X) = \frac{card \underline{R}X}{card \overline{R}X}, X \neq \emptyset.$$

The accuracy of the rough set representation of X displays the degree of completeness of existing knowledge and is in the range $\alpha_R(X) \in [0,1]$. If the boundary region of X is empty, i.e. $\overline{R}X = \underline{R}X$, then $\alpha_R(X) = 1$ and X is crisp (precise) with respect to R . And otherwise, if $card \overline{R}X > card \underline{R}X$, then $\alpha_R(X) < 1$ and X is rough (vague) with respect to R .

2. The value of roughness of X $\rho_R(X) = 1 - \alpha_R(X)$ was introduced as an alternative for the accuracy $\alpha_R(X)$. The value characterizes the degree of incompleteness of existing knowledge.

In general, the procedure of transaction classification using the rough set theory can be carried out in the following way:

- if the new transaction belongs to the lower approximation of a certain class, then it belongs to this class;

- if a new transaction belongs to the negative region of a particular class of transactions, then it can be with certainty identified as one that does not belong to this class;
- if a new transaction belongs to the boundary region of a particular class, then it is undecided whether it belongs to this class.

6 Practical application of the proposed idea

Let us consider a detailed example of the analysis and classification of transactions ($n=15$) with various itemsets (X_i). The universe of transactions, their itemsets, and attributes are presented in table 1.

Table 1. The universe of transactions, itemsets and attributes

Transactions (Tr_i)	Itemsets	Attributes								
		a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9
Tr_1	X_1	1	1	1	1	1	1	1	2	2
Tr_2	X_2	1	0	0	0	1	0	1	0	0
Tr_3	X_3	0	1	1	0	0	1	1	0	1
Tr_4	X_4	1	0	0	0	1	0	1	0	0
Tr_5	X_5	0	0	1	1	1	0	1	1	1
Tr_6	X_6	1	1	1	0	0	0	0	1	1
Tr_7	X_7	0	1	0	1	0	1	0	0	1
Tr_8	X_8	1	0	0	0	1	0	1	0	0
Tr_9	X_9	0	0	1	1	1	0	1	1	1
Tr_{10}	X_{10}	0	1	1	0	0	1	1	0	1
Tr_{11}	X_{11}	1	1	1	0	0	0	0	1	1
Tr_{12}	X_{12}	0	1	0	1	0	1	0	1	0
Tr_{13}	X_{13}	0	0	0	0	1	1	0	0	0
Tr_{14}	X_{14}	1	0	0	0	0	0	1	0	0
Tr_{15}	X_{15}	0	0	0	0	1	1	0	0	0

The following main product groups of the market basket were selected as attributes $a_i, i = \overline{1,9}$: a_1 – meat products; a_2 – dairy products; a_3 – vegetables; a_4 – fruits; a_5 – baked goods; a_6 – confectionery products; a_7 – drinks. The listed attributes are evaluated on a verbal-numeric scale: “the attribute is present in the transaction” – “1”, “the attribute is not present in the transaction” – “0”. Additionally, a_8 is included in the list of attributes – the size of the itemset with gradations: “small” – “0”; “average” – “1”; “large” – “2”, as well as a_9 – the cost of the itemset with gradations: “low” – “0”; “medium” – “1” and “high” – “2”.

Using the equivalence relation, let us divide the universe U of the table 1 into partitions:

$$\begin{aligned}
U / IND(R_1) &= \{X_1\}; \\
U / IND(R_2) &= \{X_2, X_4, X_8\}; \\
U / IND(R_3) &= \{X_3, X_{10}\}; \\
U / IND(R_4) &= \{X_5, X_9\}; \\
U / IND(R_5) &= \{X_6, X_{11}\}; \\
U / IND(R_6) &= \{X_7\}; \\
U / IND(R_7) &= \{X_{12}\}; \\
U / IND(R_8) &= \{X_{13}, X_{15}\}; \\
U / IND(R_9) &= \{X_{14}\}.
\end{aligned}$$

The base of transactions for the universe is defined by:

$$BTr = (U, E_1, E_2, E_3, E_4, E_5, E_6, E_7, E_8, E_9),$$

where $E_1 = \{X_1\}$; $E_2 = \{X_2, X_4, X_8\}$; $E_3 = \{X_3, X_{10}\}$; $E_4 = \{X_5, X_9\}$; $E_5 = \{X_6, X_{11}\}$; $E_6 = \{X_7\}$; $E_7 = \{X_{12}\}$; $E_8 = \{X_{13}, X_{15}\}$; $E_9 = \{X_{14}\}$ – are the families of equivalence classes of U ; the elements that belong to each of such classes are indiscernible.

The new transactions with the following itemsets were formed:

$$\begin{aligned}
X^{(1)} &= \{X_1, X_3, X_5, X_7, X_9, X_{10}, X_{12}, X_{14}\}, X^{(1)} \subseteq U; \\
X^{(2)} &= \{X_2, X_3, X_5, X_6, X_7, X_{11}, X_{13}, X_{15}\}, X^{(2)} \subseteq U; \\
X^{(3)} &= \{X_1, X_2, X_4, X_5, X_6, X_7, X_{12}, X_{14}\}, X^{(3)} \subseteq U; \\
X^{(4)} &= \{X_2, X_4, X_5, X_7, X_8, X_9, X_{12}, X_{14}\}, X^{(4)} \subseteq U; \\
X^{(5)} &= \{X_3, X_5, X_6, X_7, X_{11}, X_{12}, X_{14}, X_{15}\}, X^{(5)} \subseteq U.
\end{aligned}$$

Let us estimate the representation accuracy of the sets $X^{(1)}, X^{(2)}, X^{(3)}, X^{(4)}, X^{(5)}$.

The set $X^{(1)}$ can be defined with certainty as a union of transaction classes, that is:

$$\begin{aligned}
X^{(1)} &= E_1 \cup E_3 \cup E_4 \cup E_6 \cup E_7 \cup E_9 = \\
&= \{X_1\} \cup \{X_3, X_{10}\} \cup \{X_5, X_9\} \cup \{X_7\} \cup \{X_{12}\} \cup \{X_{14}\} = \\
&= \{X_1, X_3, X_5, X_7, X_9, X_{10}, X_{12}, X_{14}\}.
\end{aligned}$$

The set $X^{(2)}$ includes elements from the classes E_5, E_6, E_8 and one element from each of the classes E_2, E_3, E_4 . Therefore, this set is rough.

The set $X^{(3)}$ includes elements from the classes E_1, E_6, E_7, E_9 , two elements $\{X_2, X_4\}$ from the class E_2 , an element X_5 from the class E_4 , and an element X_6 from the class E_5 . Therefore, this set is also rough.

The set $X^{(4)}$ includes elements from the classes $E_2 = \{X_2, X_4, X_8\}$, $E_4 = \{X_5, X_9\}$, $E_6 = \{X_7\}$, $E_7 = \{X_{12}\}$, $E_9 = \{X_{14}\}$, therefore, it can be unambiguously represented by a union of the listed transaction classes, that is:

$$\begin{aligned} X^{(4)} &= E_2 \cup E_4 \cup E_6 \cup E_7 \cup E_9 = \\ &= \{X_2, X_4, X_8\} \cup \{X_5, X_9\} \cup \{X_7\} \cup \{X_{12}\} \cup \{X_{14}\} = \\ &= \{X_2, X_4, X_5, X_7, X_8, X_9, X_{12}, X_{14}\}. \end{aligned}$$

The set $X^{(5)}$ includes elements from the classes $E_5 = \{X_6, X_{11}\}$, $E_6 = \{X_7\}$, $E_7 = \{X_{12}\}$, $E_9 = \{X_{14}\}$ and one element from each of the classes $E_3 = \{X_3, X_{10}\}$, $E_4 = \{X_5, X_9\}$, $E_8 = \{X_{13}, X_{15}\}$, i.e. this set is also rough.

Using (1)-(5), let's calculate the approximations of the sets $X^{(1)}, X^{(2)}, X^{(3)}, X^{(4)}, X^{(5)}$ in the following way:

$$\begin{aligned} \underline{RX}^{(1)} &= \overline{RX}^{(1)} = \{X_1, X_3, X_5, X_7, X_9, X_{10}, X_{12}, X_{14}\}; \\ POS_R(X^{(1)}) &= \underline{RX}^{(1)} = \{X_1, X_3, X_5, X_7, X_9, X_{10}, X_{12}, X_{14}\}; \\ NEG_R(X^{(1)}) &= U - \overline{RX}^{(1)} = \{X_2, X_4, X_6, X_8, X_{11}, X_{13}, X_{15}\}; \\ BN_R(X^{(1)}) &= \overline{RX}^{(1)} - \underline{RX}^{(1)} = \emptyset; \\ \alpha_R(X^{(1)}) &= \frac{card \underline{RX}^{(1)}}{card \overline{RX}^{(1)}} = \frac{8}{8} = 1; \\ \rho_R(X^{(1)}) &= 1 - \alpha_R(X^{(1)}) = 0. \end{aligned}$$

$$\begin{aligned} \underline{RX}^{(2)} &= \{X_6, X_7, X_{11}, X_{13}, X_{15}\}; \\ \overline{RX}^{(2)} &= \{X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_9, X_{10}, X_{11}, X_{13}\}; \\ POS_R(X^{(2)}) &= \underline{RX}^{(2)} = \{X_6, X_7, X_{11}, X_{13}, X_{15}\}; \\ NEG_R(X^{(2)}) &= U - \overline{RX}^{(2)} = \{X_1, X_{12}, X_{14}\}; \\ BN_R(X^{(2)}) &= \overline{RX}^{(2)} - \underline{RX}^{(2)} = \{X_2, X_3, X_4, X_5, X_8, X_9, X_{10}\}; \\ \alpha_R(X^{(2)}) &= \frac{card \underline{RX}^{(2)}}{card \overline{RX}^{(2)}} = \frac{5}{12} = 0.42; \\ \rho_R(X^{(2)}) &= 1 - \alpha_R(X^{(2)}) = 0.58. \end{aligned}$$

$$\begin{aligned} \underline{RX}^{(3)} &= \{X_1, X_7, X_{12}, X_{14}\}; \\ \overline{RX}^{(3)} &= \{X_1, X_2, X_4, X_5, X_6, X_7, X_8, X_9, X_{11}, X_{12}, X_{14}\}; \\ POS_R(X^{(3)}) &= \underline{RX}^{(3)} = \{X_1, X_7, X_{12}, X_{14}\}; \\ NEG_R(X^{(3)}) &= U - \overline{RX}^{(3)} = \{X_3, X_{10}, X_{13}, X_{15}\}; \end{aligned}$$

$$\begin{aligned}
BN_R(X^{(3)}) &= \overline{RX}^{(3)} - \underline{RX}^{(3)} = \{X_2, X_4, X_5, X_6, X_8, X_9, X_{11}\}; \\
\alpha_R(X^{(3)}) &= \frac{\text{card } \underline{RX}^{(3)}}{\text{card } \overline{RX}^{(3)}} = \frac{4}{11} = 0.36; \\
\rho_R(X^{(3)}) &= 1 - \alpha_R(X^{(3)}) = 0.64.
\end{aligned}$$

$$\begin{aligned}
\underline{RX}^{(4)} &= \overline{RX}^{(4)} = \{X_2, X_4, X_5, X_7, X_8, X_9, X_{12}, X_{14}\}; \\
POS_R(X^{(4)}) &= \underline{RX}^{(4)} = \{X_2, X_4, X_5, X_7, X_8, X_9, X_{12}, X_{14}\}; \\
NEG_R(X^{(4)}) &= U - \overline{RX}^{(4)} = \{X_1, X_3, X_6, X_{10}, X_{11}, X_{13}, X_{15}\}; \\
BN_R(X^{(4)}) &= \overline{RX}^{(4)} - \underline{RX}^{(4)} = \emptyset; \\
\alpha_R(X^{(4)}) &= \frac{\text{card } \underline{RX}^{(4)}}{\text{card } \overline{RX}^{(4)}} = \frac{8}{8} = 1; \\
\rho_R(X^{(4)}) &= 1 - \alpha_R(X^{(4)}) = 0.
\end{aligned}$$

$$\begin{aligned}
\underline{RX}^{(5)} &= \{X_6, X_7, X_{11}, X_{12}, X_{14}\}; \\
\overline{RX}^{(5)} &= \{X_3, X_5, X_6, X_7, X_9, X_{10}, X_{11}, X_{12}, X_{13}, X_{14}, X_{15}\}; \\
POS_R(X^{(5)}) &= \underline{RX}^{(5)} = \{X_6, X_7, X_{11}, X_{12}, X_{14}\}; \\
NEG_R(X^{(5)}) &= U - \overline{RX}^{(5)} = \{X_1, X_2, X_4, X_8\}; \\
BN_R(X^{(5)}) &= \overline{RX}^{(5)} - \underline{RX}^{(5)} = \{X_3, X_5, X_9, X_{10}, X_{13}, X_{15}\}; \\
\alpha_R(X^{(5)}) &= \frac{\text{card } \underline{RX}^{(5)}}{\text{card } \overline{RX}^{(5)}} = \frac{5}{11} = 0.45; \\
\rho_R(X^{(5)}) &= 1 - \alpha_R(X^{(5)}) = 0.55.
\end{aligned}$$

The obtained results indicate that the target sets $X^{(1)}$ and $X^{(4)}$ completely belongs to the union of the classes $E_1, E_2, E_4, E_6, E_8, E_9$ and E_2, E_4, E_6, E_7, E_9 ; $\alpha_R(X^{(1)}) = 1$, $\alpha_R(X^{(4)}) = 1$. The sets $X^{(2)}$, $X^{(3)}$ and $X^{(5)}$ cannot be classified with certainty because $\rho_R(X^{(2)}) = 0.58$, $\rho_R(X^{(3)}) = 0.64$ and $\rho_R(X^{(5)}) = 0.55$.

In such a situation, the mathematical apparatus of the rough set theory offers three strict rules for the classification of target sets that characterize transactions and their itemsets:

$$\begin{aligned}
X_i &\in E_j, \text{ if } X_i \in POS_R(E_j); \\
X_i &\notin E_j, \text{ if } X_i \in NEG_R(E_j); \\
X_i &\in E_j \text{ or } X_i \notin E_j, \text{ if } X_i \in BN_R(E_j).
\end{aligned}$$

It implies that for a reliable classification of transactions, only first and fourth decision rules can be used.

7 Conclusion

The presented approach is aimed at structuring of the generated transaction database in the process of forming a market basket using the rough set theory. The basis of this theory is the procedure for equivalence relations formation that is used to distinguish categories (classes) of transactions that are considered indiscernible within each category. When the transactions cannot be completely described by the obtained classes, specific approximations and estimates of their accuracy are introduced, which assess the degree of belonging or completeness of non-membership of such transactions in these classes.

Ultimately, a classified transaction database will increase the selectivity of the search for binary association rules.

References

1. Zayko, T. A., Oliynyk, A. A., Subbotin, S. A.: Association rules in data mining [in Russian]. Herald of the National Technical University "KhPI". Subject issue: Information Science and Modelling, vol. 39(1012), pp. 82-96. (2013).
2. Oliynyk, A. A., Zayko, T. A., Subbotin, S. A.: Method for synthesis of bases of numeric associative rules [in Russian]. Electronics and Informatics, vol. 2(61), pp. 61-66 (2013).
3. Kovalenko, I., Davydenko, Ye., Shved, A.: Formation of consistent groups of expert evidences based on dissimilarity measures in evidence theory. In: 14th International conference on Computer sciences and Information technologies (CSIT 2019), IEEE Press, Lviv, pp. 113-116 (2019). doi: 10.1109/STC-CSIT.2019.8929858
4. Shved, A., Kovalenko, I., Davydenko, Y.: Method of Detection the Consistent Subgroups of Expert Assessments in a Group Based on Measures of Dissimilarity in Evidence Theory. Advances in Intelligent Systems and Computing IV, Springer International Publishing, vol. 1080, pp. 36-53 (2020). doi: 10.1007/978-3-030-33695-0_4
5. Gorodetsky, V. I., Samoylov, V. V.: Association and casual rule mining using associative bayesian networks [in Russian]. Trudy SPIIRAN, vol. 9, pp.13-65 (2009).
6. Fisun, M., Horban, H.: Implementation of the information system of the association rules generation from OLAP-cubes in the post-relational DBMS caché. In: 11th International conference on Computer sciences and Information technologies (CSIT 2016), IEEE Press, Lviv, pp. 40-44 (2016). doi: 10.1109/STC-CSIT.2016.7589864
7. Billig, V. A., Korneeva, E. I., Syabro, N. A.: Association Rules. Compared Analysis of the Tools [in Russian]. Software Journal: Theory and Applications, vol. 2, pp.1-41 (2016). doi: 10.15827/2311-6749.16.2.2
8. Galkina, E. V.: The combined use of the decision tree method and associative analysis in management [in Russian]. International research journal, vol. 9(51), pp. 29-32 (2016). doi: 10.18454/IRJ.2016.51.095
9. Fisun, M., Horban, H., Dvoretzkyi, M.: Methods of Searching for Association Dependencies in Multidimensional Databases. In: 13th International conference on Computer sciences and Information technologies (CSIT 2018), IEEE Press, Lviv, pp. 88-93 (2018). doi: 10.1109/STC-CSIT.2018.8526737
10. Moldavskaya, A. V.: Method of forming multi-leveled sequential patterns [in Russian]. Programming problems, vol. 2-3, pp. 158-163. (2016).

11. Kovalenko, I., Davydenko, Y., Shved, A.: Searching for Pareto-optimal solutions. *Advances in Intelligent Systems and Computing IV*, Springer International Publishing, vol. 1080, pp. 121-138 (2020). doi: 10.1007/978-3-030-33695-0_10
12. Pawlak, Z.: *Rough sets theoretical aspects of reasoning about data*. Kluwer Academic Publishers, Boston; London (1991).
13. Uzga-Rebrovs, O.: Knowledge representing features in rough sets [in Russian]. In: 7th International Scientific and Practical Conference: Environment. Technology. Resources, vol. 2, pp. 169-175 (2009).
14. Skowron, A., Dutta, S.: Rough sets: past, present, and future. *Natural Computing*, 17, 4, pp. 855–876 (2018). doi: 10.1007/s11047-018-9700-3
15. Tripathy, H. K., Tripathy, B. K., Das, P. K.: An intelligent approach of rough set in knowledge discovery databases. *International Journal of Computer, Electrical, Automation, Control and Information Engineering*, vol. 1(11), pp. 3437-3440 (2007).
16. Weihua, X., Xiaoyan, Z.: Fuzziness in Covering Generalized Rough Sets. In: *Chinese Control Conference, Hunan*, pp. 386-390 (2007). doi: 10.1109/CHICC.2006.4347200
17. Wang, J., Peng, L.: Research on Expression of Rough Equality Sets. In: *IEEE Pacific-Asia Workshop on Computational Intelligence and Industrial Application, Wuhan*, pp. 337-341 (2008). doi: 10.1109/PACIIA.2008.289
18. Lin, H., Wang, Q., Lu, X., Li, H.: Hybrid multi-granulation rough sets of variable precision based on tolerance. In: *12th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), Zhangjiajie*, pp. 231-235 (2015). doi: 10.1109/FSKD.2015.7381945
19. Sun, B., Gong, Z.: Rough Fuzzy Sets in Generalized Approximation Space. In: *5th International Conference on Fuzzy Systems and Knowledge Discovery, Shandong*, pp. 416-420 (2008). doi: 10.1109/FSKD.2008.178
20. Wei-feng, D., Hai-ming, L., Yan, G., Dan, M.: Another kind of fuzzy rough sets. In: *IEEE International Conference on Granular Computing, Beijing*, vol. 1, pp. 145-148 (2005). doi: 10.1109/GRC.2005.1547254
21. Dai, J., Chen, W., Pan, Y., Sequent calculus system for rough sets based on rough Stone algebras. In: *IEEE International Conference on Granular Computing, Beijing*, vol. 2, pp. 423-426 (2005). doi: 10.1109/GRC.2005.1547326

Multi-Layer Model and Training Method for Information-Extreme Malware Traffic Detector

Viacheslav Moskalenko^[0000-0001-6275-980], Alona Moskalenko^[0000-0003-3443-3990],
Artur Shaikhov^[0000-0003-3277-0264], Mykola Zaretskyi^[0000-0001-9117-5604]

Sumy State University, Rimsky-Korsakov st., 2, Sumy, 40007, Ukraine
v.moskalenko@cs.sumdu.edu.ua, a.moskalenko@cs.sumdu.edu.ua

Abstract. Model-based on multilayer convolutional sparse coding feature extractor and information-extreme decision rules for malware traffic detection is presented in the paper. Growing sparse coding neural gas algorithms for unsupervised pre-training of the feature extractor are used. Random forest regression model as a student in knowledge distillation from sparse coding layers is proposed for speed up inference mode. Information-extreme learning method based on binary encoding with tree ensembles and class separation with radial basis function in binary Hamming space are proposed. Information-extreme classifier is characterized by low computational complexity and high generalization ability for small labeled training sets. Simulation results with an optimized model on test open datasets confirm the suitability of proposed algorithms for practical application.

Keywords: malware detection system, convolutional sparse coding network, growing neural gas, tree ensembles, random forest regression, information criterion, information-extreme machine learning.

1 Introduction

Existing malware traffic detection systems still do not provide high-reliability solutions, as there are a constant increase the number and variety of new sources of malware traffic and a small number of relevant labeled data [1, 2]. Thus, the use of hand-crafted features for the description of observations leads to a decline the informativeness of the features description and the effectiveness of learning of the decision rules of the malware traffic detection system [2, 3]. Therefore, the most promising approach to the synthesis of a features extractor is the use of ideas and methods of machine learning for the hierarchical (deep) representation of observations for unlabeled data [4, 5].

Conventional approaches to deep supervised machine learning require a significant amount of labeled training examples and computational resources [6, 7]. In addition, models trained with a supervisor based on gradient descent and its modifications are vulnerable to adversarial attacks, noise and data novelty. To increase the informativeness of the feature representation of observations, it is promising to use ideas and methods of sparse coding and unsupervised competitive learning [8, 9]. This allows to use the large volume of unlabeled data with maximum efficiency. Among the ways to

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

increase the generalization ability of the decision rules are considered ensemble algorithms, error correction codes and methods of class separation within the geometric approach. Also high speed of packet flow in modern networks require high productivity of traffic analysis algorithms. To reduce computational complexity of data analysis models, different methods of model pruning and knowledge distillation are used. However, models hybridization and integrated use of different methods bring some uncertainties to the final result, so the solution in this approach requires research and verification. In this case, information criteria are considered the best metrics for validation and verification of the result, because they directly characterize the reduction of uncertainty in decision-making and are less sensitive to outlier and imbalances in the data.

2 Formal Problem Statement

Let the CTU-Mixed and CTU-13 datasets are given data collections from the real network environment by CTU researchers from 2011 to 2015, which are formed as pcap-files [4, 5]. The first CTU-Mixed dataset can be used for training a feature extractor. The second CTU-13 dataset contains labeled flows and it could be used to train the decision rules for detecting malware network traffic.

It is necessary to build an informative feature extractor and reliable decision rules using labeled and unlabeled datasets through optimization of model parameters. In the process of training, it is necessary to maximize the information efficiency criterion of the malware traffic detector

$$\bar{E}^* = \frac{1}{M} \sum_{m=1}^M \max_{\{k\}} E_m^{(k)},$$

where $E_m^{(k)}$ is information efficiency criterion of recognition the class X_m^o on k -th step of training; $\{k\}$ – ordered set of training steps.

When the malware traffic detector functions in its inference mode, it is necessary to provide computational efficiency for high speed traffic.

3 Literature Review

Convolutional multi-layer neural networks allow forming an informative hierarchical features representation of input observations [6]. In addition, they have already shown high efficiency in solving problems of machine vision and analysis of time series [6], [7]. Meanwhile, supervised training requires a large amount of labeled data, the labeling of which may be expensive or inaccessible in a reasonable amount of time. The unsupervised training of convolutional networks is aimed at efficient use of unlabeled examples, which are usually available quite a lot. It is carried out based on an autoencoder or Restricted Boltzmann machine, which requires a large amount of training

data and long learning time to obtain an acceptable result [8]. In work [9] it is proposed to use alternative approach based on k-means cluster-analysis algorithm to speed up feature set training. However, k-means is characterized by slow convergence and sub-optimality of the results due to the hard-competitive nature of its learning scheme and the sensitivity to initial cluster initialization.

In work [10] is proposed a combination of the principles of neural gas and sparse coding for the feature set training on unlabeled data. Given approach is characterized by soft-competitive learning scheme that facilitates robust convergence to close to optimal features distributions over the training sample. At the same time, embedding of sparse coding methods can increase the immunity against interference and generalization ability of features representation. Also, it is a well-known fact that sparse representations of the input data are a crucial tool for combating adversarial attacks and the production of de-correlated features as a result of the explaining-away effect. However, the size of feature set is unknown beforehand and it is selected by the developer, which leads to increase the optimization time.

The required size of feature set in each layer of hierarchical representation is difficult to predict in advance, so the promising approach to feature set learning is to use the principles of growing neural gas, which automatically determines the required number of neurons (features) [11]. The presence of a mechanism for the adding of new neurons, as well as the removal of excessive old ones, makes the algorithm more flexible compared to the classical neuron gas, but it also has serious disadvantages. The small values of the period between the iterations of the generation of new neurons λ lead to the instability of the learning process and the distortion of the formed structures, as here observed the excessively frequent adding of new neurons. The high value of the period λ provides the expected effect, but at the same time it leads to a significant slowdown in the algorithm. However, in the works [11, 12] it was shown that achieving stability of learning could be done by setting the “radius of reach” of the neurons, which involves the replacement of the parameter λ on the threshold of maximum distance of the neuron from each points of the training set attributed to it. However, the mechanisms for updating neurons and assessing the remoteness of the points of the input space to the neurons have not yet been reviewed in order to adapt the learning process to the sparse coding of observations.

The main disadvantage of sparse coding in representation learning is the use of an iterative procedure during the inference which slows down the recognition process. One of the popular ways to accelerate models is to use the principles of knowledge distillation, where the redundant model acting as a teacher can be replaced by a light-weight model acting as a student [13]. The ensemble of decision trees is a flexible and computationally efficient model, which can potentially be used as a student model to approximate the sparse coder [14]. However, no such research had been conducted and the effectiveness of such an approach is unknown, which underscores the relevance of this issue.

In addition the decision rules are important components in the malware detection systems. As a rule, it represents a trainable classifier. At the same time, the effectiveness of training a classifier is often considered as a measure of the effectiveness of the feature extractor [5]. The most popular algorithm for classification analysis is the

method of support vector machine, where the training of decision rules takes place within the framework of a geometric approach by constructing linear separable hyper-surface in the secondary features space [15]. However, this algorithm requires a lot of hyper-parameters adjustments and its performance depends on the complexity of the kernel functions. In work [16], were proposed the construction of decision rules by adaptive binary encoding of the input features and the optimizing in information sense the radial-basis based separable hyper-surface in the Hamming binary space. Such a classifier has high operational efficiency, since it uses low computing complexity operations as comparison and logical XOR.

4 Model and Training Method for Malware Traffic Detector

The internal characteristics of the unit of traffic (packet stream or session) are best displayed in the front part of its bytes, which contains connection data and some content data. The process of converting a pcap-file into a training data set involves three main steps: the separation of traffic into discrete units, taking into account some granularity, clearing traffic by removing empty and duplicate units, forming training images. When dividing traffic into discrete units, one can consider the following granularities: TCP connection, flows, session, service, and host. In this paper, it is proposed to divide the incoming traffic into flows, where a number of packets have the same tuple of five elements: the source and destination IP address, the source and destination ports, the protocol number. In this case, the length of the stream is limited to 784 bytes, so longer streams are cropped, and shorter ones are supplemented by zero bytes. As a result, we have an image of 28x28 pixels, which will be delivered to the input of the feature extractor. The brightness of each pixel is normalized to the range [0, 1].

As a basis for building architecture of features extractor was used a convolutional network is known as LeNet-5 [5], the main modification of which relates to use the unfixed number of convolutional filters, the amount of which is determined during the layer-wise training. The pixel activation of each channel of features map is offered to calculate based on greedy-L0 Orthogonal Matching Pursuit algorithm (OMP) or L1-regularized least angle regression algorithm (LARS) with the function of ReLU activation [17]. In order to accelerate the model in the inference mode, it is possible to replace the computationally intensive search for sparse coefficients with a non-iterative approximating encoder (Figure 1). According to distillation knowledge principle, the training set for approximation encoder is formed from input of the layer and pseudo-labels from output of the layer. In this case, pseudo-labels are obtained by OMP or LARS algorithms.

It is proposed to implement sparse coding with OMP and LARS algorithms where stop criterion based on achievement of 30% non-zero entries in sparse code. A Local Contrast Normalization layer, placed after the sub-sampling layer, before the next layer, amplifies the informative features and weakens the rest of the pixels of the feature map.

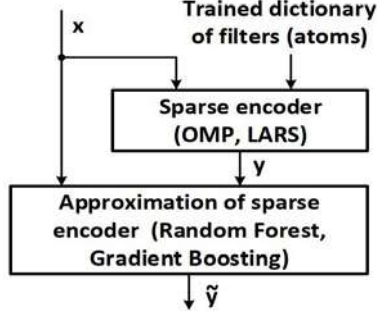


Fig. 1. Knowledge distillation diagram for each layer of feature extractor

The dataset for training of feature extraction layer is formed by decomposition of images or activation maps to patches. These patches are reshaped to 1D vectors, which put on the input of growing sparse coding neural gas algorithm, main steps of which are given below [16].

1. Initialization of the counter of training vectors $t := 0$.
2. Two initial nodes (neurons) w_a and w_b are assigned by random selection from the training set. Nodes w_a and w_b are connected by an edge whose age is zero. These nodes are considered non-fixed.
3. Selected from the dataset the following vector x , which is normalized to a unit length (L2-normalization).
4. Normalizing each base vector $w_k, k = \overline{1, M}$ to a unit length (L2-normalization).
5. Calculation of the similarity of the input vector x to the base vectors $w_{s_k} \in W$ for their sorting

$$-(w_{s_0}^T x)^2 \leq \dots \leq -(w_{s_k}^T x)^2 \leq \dots \leq -(w_{s_{M-1}}^T x)^2.$$

6. The closest node is selected w_{s_0} and the second closest to the node w_{s_1} .
7. Increase the age of all incident edges w_{s_0} by one.
8. If w_{s_0} is fixed, then should move to step 9, otherwise, to step 10.
9. If $(w_{s_0}^T x)^2 \geq \nu$, then proceed to step 12. Otherwise, should be added a new non-fixed neuron w_r to a point that coincides with the input vector $w_r = x$, also is adding a new edge, that connects w_r and w_{s_0} , then proceed to step 13.
10. The node w_{s_0} and its topological neighbors (the nodes connected to it by the edge) are displaced in the direction to the input vector x by the next formulas [10]

$$\Delta w_{s_0} = \varepsilon_b \eta_t y_0 (x - y_0 w_{s_0}), \quad y_0 := w_{s_0}^T x, .$$

$$\Delta w_{s_n} = \varepsilon_n \eta_t y_n (x - y_n w_{s_n}), \quad y_n := w_{s_n}^T x, .$$

$$0 < \varepsilon_b < 1, \quad 0 < \varepsilon_n < \varepsilon_b, \quad$$

$$\eta_t := \eta_0 (\eta_{final} / \eta_0)^{t/t_{max}},$$

where Δw_{s_0} , Δw_{s_n} – vectors of correction of weight of the neuron-winner and its topological neighbors, respectively; ε_b , ε_n – the constants of updated forces of weighting coefficients of the neuron-winner and its topological neighbors respectively; $\eta_0, \eta_t, \eta_{final}$ – initial, current and final learning rate respectively.

11. If $(w_{s_0}^T x)^2 \geq \nu$, note the neuron w_{s_0} as fixed.
12. If w_{s_0} and w_{s_1} are connected by edge, then its age is reset, otherwise a new edge with a zero age is formed between w_{s_0} and w_{s_1} .
13. All edges in the graph with the age more than a_{max} are removed. In the case that some nodes do not have incident edges (become isolated), they are also removed.
14. If $t < t_{max}$ then proceed to step 15, otherwise – increment of the counter of steps is $t := t + 1$ and then proceed to step 3.
15. If all neurons are fixed, the execution of the algorithm stops, otherwise proceed to step 3 and a new epoch of learning begins (repetition of the training set).

Features extractor can be fine-tuned based on the backpropagation algorithm with a temporary or permanent neural classifier at the model output [17]. Since in the conditions of nonstationarity the informativeness of features in advance cannot be known, the fine tuning is not provided in our algorithm. The purpose of the feature extractor is to disentangle explanatory factors.

The information-extreme classifier requires binary representation of the input signal to build error-correction decision rules. The ensemble of decision trees is a computationally effective method for inducing informative binary features of observations (Figure 2). Nodes of decision trees are numbered. Numbers of nonzero bits of resulting binary code correspond to the numbers of nodes through which the decision path lies [16].

Information-extreme classifier under inference mode make decision on belonging of input datapoint x with appropriate binary representation b to one class from set $\{X_z^o \mid z = \overline{1, Z}\}$ according to maximum value of membership function $\mu_z(b)$ through the expression $\arg \max_z \{\mu_z(b)\}$.

In this case membership function $\mu_z(b)$, the optimal container of which has support vector b_z^* with dimension N_2 and radius d_z^* , is derived from formula

$$\mu_z(b) = \exp\left(-\sum_{i=1}^{N_2} b_i \oplus b_{z,i}^* / d_z^*\right). \quad (1)$$

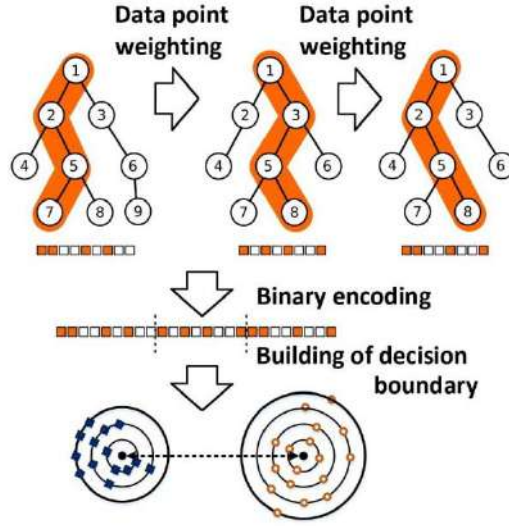


Fig. 2. Classifier Architecture

Let $D = \{x_j, y_j \mid j = \overline{1, n}\}$ be a training set, where n is size of dataset and y_j is label of j -th datapoint, which correspond to one class from set of classes $\{X_z^o \mid z = \overline{1, Z}\}$. In this case, classifier evaluates belonging of j -th datapoint x_j with N_1 features to one of the Z classes performs feature encoding using decision trees and decision rules constructed in radial basis of binary Hamming space. The training of information-extreme classifier is performed according to the following steps.

1. Initialize weight $w_j = 1/n$.
2. For $k = 1, \dots, K$ do
3. Bootstrap D_k from D using probability distribution $P(X = x_j) = w_j$.
4. Train decision tree T_k on D_k using entropy criterion to measure the quality of split.
5. Binary encoding of x_j datapoint from D using concatenation of results from $T_1, \dots, T_k$ trees. The output of this step is a binary matrix $\{b_{z,s,i} \mid i = \overline{1, N_2}; s = \overline{1, n_z}; z = \overline{1, Z}\}$, where N_2 is a number of induced binary features and n_z is a number of samples corresponded to class X_z^o . Hence the equality $n = \sum_z n_z$ condition is met.

6. Build information-extreme decision rules in radial basis of binary Hamming space and compute optimal information criterion:

$$E_z^* = \max_{\{d\}} E_z(d), \quad (2)$$

where $\{d\} = \{0, 1, \dots, \left(\sum_i b_{z,i} \oplus b_{c,i} - 1\right)\}$ is a set of concentric radiuses with center b_z (support vector) of data distribution in class X_z^o , which computed using rule

$$b_{z,i} = \begin{cases} 1, & \text{if } \frac{1}{n_z} \sum_{s=1}^{n_z} b_{z,s,i} > \frac{1}{Z} \sum_{c=1}^Z \frac{1}{n_c} \sum_{s=1}^{n_c} b_{c,s,i}; \\ 0, & \text{otherwise.} \end{cases} \quad (3)$$

where E_z – training efficiency criterion of decision rule for X_z^o class, which is computed as the normalized modification of the S. Kullback's information measure [16]:

$$E_z = \frac{1 - (\alpha_z + \beta_z)}{\log_2(2 + \varsigma) - \log_2 \varsigma} \cdot \log_2 \left[\frac{2 - (\alpha_z + \beta_z) + \varsigma}{(\alpha_z + \beta_z) + \varsigma} \right], \quad (4)$$

where α_z , β_z are the false-positive and false-negative rates of classification of input vectors as belonging to the X_z^o class; ς is any small non-negative number, introduced to avoid uncertainty when dividing by zero.

7. Test obtained information-extreme rules on dataset D and compute error rate for each sample from D . Under the inference mode, decision on belonging of datapoint b to one class from set $\{X_z^o \mid z = \overline{1, Z}\}$ is made according to maximum value of membership function $\mu_z(b)$ through the expression $\arg \max_z \{\mu_z(b)\}$. In this case membership function $\mu_z(b)$ of binary representation b of input datapoint x to X_z^o class, the optimal container of which has support vector b_z^* and radius d_z^* , is derived from formula (2).

8. Update $\{w_j\}$ proportional to errors on datapoint x_j :

$$w_j = 1 - \mu_{m'}(x_j), \quad m' = y_j;$$

$$w_j = \frac{w_j}{\sum_j w_j}.$$

9. If $|E_k^* - E_{k-1}^*| < \varepsilon$ and $k < K/2$ abort loop, where $\varepsilon = 0.001$.

Thus, the resulting model consists of several layers of tree ensemble with optimal in informational sense decision rules at the output.

5 Result and discussion

The training sample formed with CTU-Mixed for the training of the feature extractor contains 10,000 instances. To train the information-extreme classifier are formed by 1000 instances per class in the training and test datasets. In growing sparse coding neural gas algorithm were chosen the following parameters $\varepsilon_b = 0.5$, $\varepsilon_b = 0.05$, $a_{\max} = 100$, $\eta_0 = 1$ та $\eta_{\text{final}} = 0.01$. The parameter of the threshold of neuron fixation ν and the parameter of the maximum number of trees K of the classifier are adjusted by scrolling through the values. Table 1 shows the dependence of the number of neurons in the first M_1 and second M_2 layers of feature extractor, the criterion of the effectiveness of training averaged over the classes $\bar{E}$ and accuracy by the validation sampling of the parameter ν . In the tree ensembles, max depth is set to 5 and max features is set to $\sqrt{N_1}$.

Table 1. Dependence of information criteria and number of neurons from model parameters

ν	M_1	M_2	$\bar{E}$	Validation accuracy
0.10	15	11	0.106	74
0.15	17	13	0.138	77
0.20	23	13	0.138	77
0.25	25	13	0.138	77
0.30	27	15	0.149	78
0.35	27	15	0.220	83
0.40	33	17	0.255	85
0.45	34	22	0.255	85
0.50	40	25	0.366	90
0.55	49	31	0.459	93.0
0.60	66	43	0.466	93.2
0.65	70	45	0.501	94.1
0.70	99	45	0.550	95.2
0.75	145	57	0.554	95.3
0.80	161	120	0.591	96.1
0.85	220	147	0.603	95.4
0.90	322	238	0.611	95.0

The analysis of Table 1 shows that increasing the threshold ν leads to an increase in the number of neurons in the process of unsupervised training the features extractor. At the same time, increasing the threshold from 0.8 to 0.9 practically does not affect the accuracy of the decision rules. It means, that the value $\nu^* = 0.8$ is optimal and allows to form a more compact features representation (compression), meanwhile $\nu = 0.9$ allows to form a sparse representation based on overcomplete basis.

Knowledge distillation is implemented with Random Forest regression as student model, where the number of decision trees is limited to 150. The obtained model has equivalent accuracy. In this case, the inference time is reduced by 65 times.

Figure 3 shows a graph of maxima's changes of the information criterion (4) averaged in the set of classes in dependence on the number of decision trees in informa-

tion-extreme classifier with $v^* = 0.8$. In this case, the maximum number of trees is limited, $K=100$.

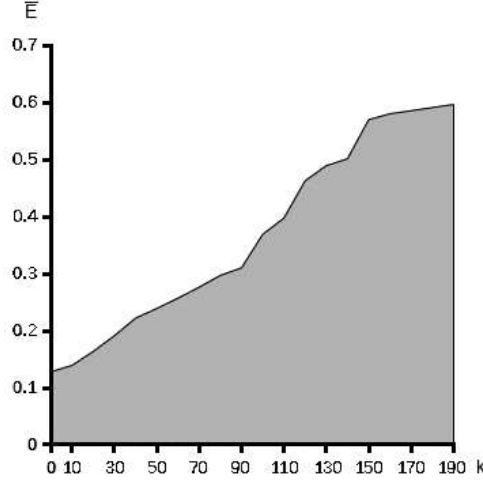


Fig. 3. A graph of the change of the average information criterion (4) in dependence from the number of decision trees in information-extreme classifier

The analysis of Figure 3 shows that the optimal value of the hyper parameter K^* is equal to 185. Further increase of the parameter K does not lead to an increase in the accuracy of the decision rules. At optimal parameters of the extractor and the classifier, the accuracy of detection of malware traffic is 96.1%. It indicates on the informative nature of the features descriptive of observation. Figure 4 shows the dependence of the information criterion (4) on the code radius of the container of each class.

The analysis of Figure 4 shows that the maximum values of information criterion of learning for the first and second classes are equal to $E_1^* = 0.590$ and $E_2^* = 0.597$, respectively, and the optimal values of radii of the corresponding containers of the classes of recognition $d_1^* = 26$, $d_2^* = 32$ (in code units). In this case, the inter-center Hamming distance is 65 indicating compactness of the feature vector distributions and the clarity of partition in the binary Hamming space.

Thus, the proposed training algorithm allows determining automatically the optimal number of neurons at each layer. At the same time, approximation of the sparse encoder by the non-iterative model, Random Forest regression, allowed accelerating the inference mode.

The results of simulation on data from CTU-Mixed and CTU-13 datasets show that obtained result is superior to result from [4] and [5] and it is acceptable for practical applications.

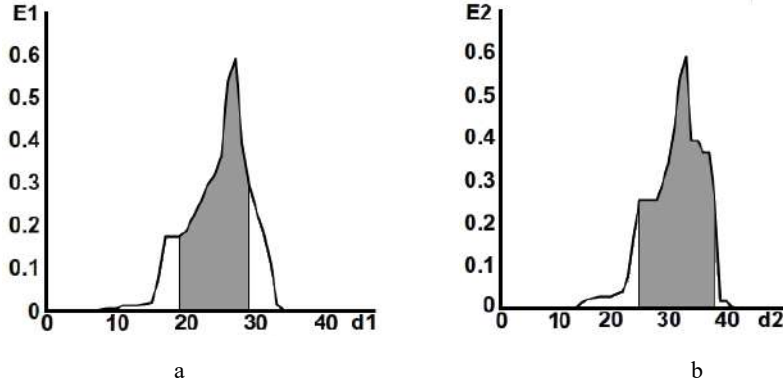


Fig. 4. Charts of dependency of the information criterion (4) on the radii of containers of classes: a – class of normal traffic; b – class of malware traffic

6 Conclusions

10. The scientific novelty of the obtained results is as follows:

- the algorithm of growing sparse coding neural gas is proposed for the first time, which allows unsupervised learning the optimal set of neurons for each layer of the convolution sparse coding model of feature extractor model;
- for the first time it was proposed to apply the principle of knowledge distillation to reduce computational costs in the algorithms of sparse coding through the application of approximation by the random forest model, which in the inference mode is non-iterative and computationally efficient;
- for the first time, an information-extreme algorithm of supervised learning is proposed for constructing the decision rules of the detector of malware network traffic.

11. The practical value of obtained results obtained for malware traffic detection systems is a developing a new learning method that effectively uses both labeled and unlabeled training sets. The results of simulation with using the CTU-Mixed and CTU-13 datasets confirm the effectiveness of the obtained decision rules in identifying the malware in test samples of traffic. In this case, the accuracy of the decision rules of the malware traffic detector is 96.1%.

7 Acknowledgment

The work was performed in the laboratory of intellectual systems of the computer science department at Sumy State University with the financial support of the Ministry of Education and Science of Ukraine in the framework of state budget scientific and research work of DR No. 0117U003934.

References

1. Skrzewski, M.: Flow Based Algorithm for Malware Traffic Detection. *Computer Networks*. 271-280 (2011).
2. Berkay Celik, Z., Walls, R., McDaniel, P., Swami, A.: Malware traffic detection using tamper resistant features. *MILCOM 2015 – 2015 IEEE Military Communications Conference*. (2015).
3. Iglesias, F., Zseby, T.: Analysis of network traffic features for anomaly detection. *Machine Learning*. 101, 59-84 (2014).
4. Yousefi-Azar, M., Varadharajan, V., Hamey, L., Tupakula, U.: Autoencoder-based feature learning for cyber security applications. *2017 International Joint Conference on Neural Networks (IJCNN)*. (2017).
5. Wei Wang, Ming Zhu, Xuwen Zeng, Xiaozhou Ye, Yiqiang Sheng: Malware traffic classification using convolutional neural network for representation learning. *2017 International Conference on Information Networking (ICOIN)*. (2017).
6. Szegedy, C., Wei Liu, Yangqing Jia, Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. (2015).
7. Zhao, B., Lu, H., Chen, S., Liu, J., Wu, D.: Convolutional neural networks for time series classification. *Journal of Systems Engineering and Electronics*. 28, 162-169 (2017).
8. Qiying Feng, Chen, C., Long Chen: Compressed auto-encoder building block for deep learning network. *2016 3rd International Conference on Informative and Cybernetics for Computational Social Systems (ICCSS)*. (2016).
9. Tang, J., Wang, D., Zhang, Z., He, L., Xin, J., Xu, Y.: Weed identification based on K-means feature learning combined with convolutional neural network. *Computers and Electronics in Agriculture*. 135, 63-70 (2017).
10. Labusch, K., Barth, E., Martinetz, T.: Sparse Coding Neural Gas: Learning of overcomplete data representations. *Neurocomputing*. 72, 1547-1555 (2009).
11. Mrazova, I., Kukacka, M.: Image Classification with Growing Neural Networks. *International Journal of Computer Theory and Engineering*. 422-427 (2013).
12. Palomo, E., Lopez-Rubio, E.: The Growing Hierarchical Neural Gas Self-Organizing Neural Network. *IEEE Transactions on Neural Networks and Learning Systems*. 1-10 (2016).
13. Li, H., Lin, S., Chen, C., Chiang, C.: Layer-Level Knowledge Distillation for Deep Neural Network Learning. *Applied Sciences*. 9, 1966 (2019).
14. Zhou, Y., Zhou, Z., Hooker, G.: Approximation Trees: Statistical Stability in Model Distillation, <https://arxiv.org/abs/1808.07573>. (2018).
15. Kim, S., Yu, Z., Kil, R., Lee, M.: Deep learning of support vector machines with class probability output networks. *Neural Networks*. 64, 19-28 (2015).
16. Moskalenko, V., Moskalenko, A., Korobov, A., Semashko, V.: The Model and Training Algorithm of Compact Drone Autonomous Visual Navigation System. *Data*. 4, 4 (2018).
17. Y. Gwon, M. Cha, H. T. Kung: Deep Sparse-coded Network (DSN). *2016 International Conference on Pattern Recognition*. (2016).

Analysis and Experimental Realization of the Logistic Map Using Arduino Pro Mini

Volodymyr Rusyn¹[0000-0001-6219-1031], Sergey Subbotin²[0000-0001-5814-8268], Aceng Sambas³[0000-0002-1623-0770]

¹Yuriy Fedkovych Chernivtsi National University, Kotsybinsky str., 2, Chernivtsi, 58012, Ukraine
rusyn_v@ukr.net

²Zaporizhzhia National Technical University, Zhukovsky str., 64, Zaporizhzhia, 69063, Ukraine
subbotin@zntu.edu.ua

³Universitas Muhammadiyah Tasikmalaya, Jl. Tamansari No. KM 2,5, Mulyasari, Kec. Tamansari, Tasikmalaya, Jawa Barat, 46196, Indonesia
acenx.bts@gmail.com

Abstract. In this paper, we used one-dimensional system that demonstrate chaotic behavior. Equation, analysis and system conditions are presented. Modern software environment LabView was used for analysis of the iteration series with different parameter r . Connection scheme, nominal components and programming code are also presented. For practical realization and visualizing of the logistic map we used Arduino Pro Mini and ten light-emitting diodes (LEDs) with ten resistors for each part of segment of the range $[0;1]$. The Arduino was connected to a computer through the USB port and programmed using a language similar to C++. Sketch was uploaded into Arduino using program software ArduinoIDE.

Keywords. Nonlinear, logistic map, Arduino

1 Introduction

Theory of chaos applied in various fields of research, such as economy [1-3], biology [4], mathematics [5-12], optics [13], robotics [14], memristors [15-20], electronics schemes [21-34], etc.

The logistic map is a one-dimensional discrete-time map that, despite its formal simplicity, exhibits an unexpected degree of complexity. Historically it has been one of the most important and paradigmatic systems during the early days of research on deterministic chaos. The map was popularized in a 1976 paper by the biologist Robert May [35] in part as a discrete-time demographic model analogous to the logistic equation first created by Pierre François Verhulst.

The logistic map is defined by the following equation:

$$x_{n+1} = rx_n(1-x_n) \quad (1)$$

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

The logistic map is a very simple system, which can produce chaotic behavior with the right values of the parameter r . Sensible values of r range from 0 to 4; also, the values of x range from 0 to 1.

Fig. 1 shows plot the behavior of the orbits for $r = 2.6$, $r = 3.3$, and $r = 3.9$. The software implementation of logistic map using LabView software environment shows in Fig. 2.

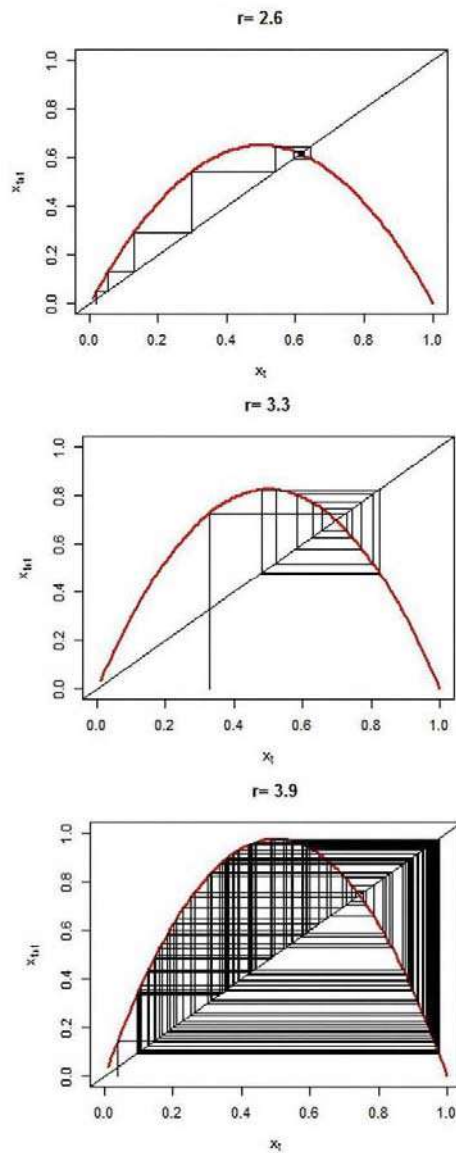


Fig. 1. The behavior of the orbits for different parameter r

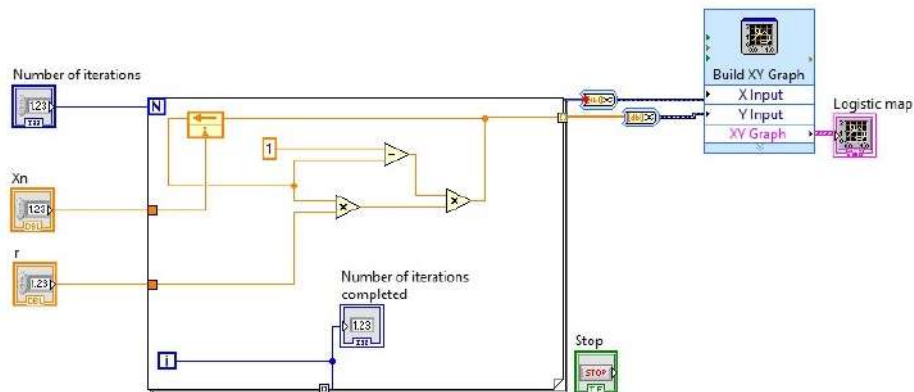


Fig. 2. Block scheme for realization of the logistic map

Fig. 3 and Fig. 4 shows computer modeling results for different initial conditions X_0 . Number of iterations equal 100. Results are presented for two cases: 1 – $X_0 = 0.4$ and 2 – $X_0 = 0.8$.

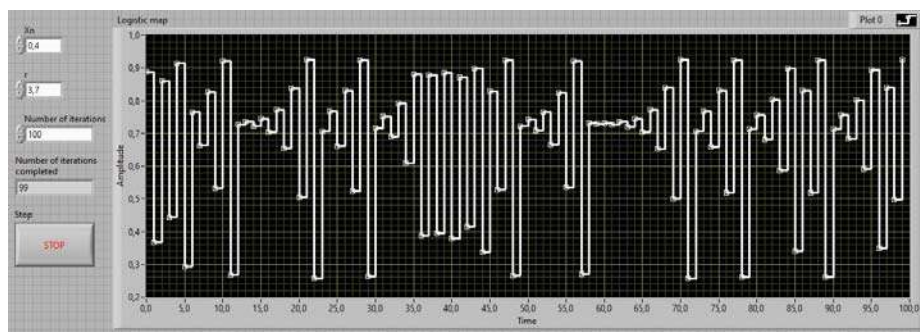


Fig. 3. Iteration series of logistic map for $X_0 = 0.4$

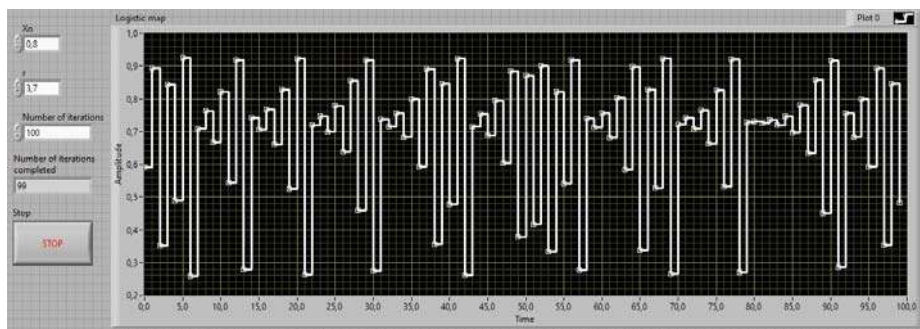


Fig. 4. Iteration series of logistic map for $X_0 = 0.8$

2 ARDUINO board

The top board is Arduino Pro Mini, the bottom – USB-UART / USB-TTL converter and some connectors (Fig. 5).



Fig. 5. Set of the Arduino board, converter and some connectors

Now about the Arduino Pro Mini board / platform. Structurally, it is a board with a microcontroller soldered on it, a RESET button, a power chip, and other peripherals that are not essential for us at this stage.

There are two versions of the Pro Mini: one running at 3.3V at 8 MHz, the other at 5V at 16 MHz. The board is based on the Atmega 168 or Atmega 328 microcontroller – the difference between them is the volume of the system-programmable Flash memory – 16 or 32 kB. This is the so-called "program memory", i.e. the memory in which the program will be recorded and its contents will not change during operation. Atmega is built on the so-called Harvard architecture in which the "program memory" and "data memory" are implemented separately, for greater speed and reliability. The

"data memory" is divided into 2 parts: operational SRAM, which for 168 and 328 is 1 KB, and a permanent EEPROM (Electrically Erasable Programmable Read-Only Memory) of 512 bytes, data from which is not "lost" when the power is turned off.

The connection scheme is quite simple and it's rather difficult to make a mistake (Fig. 6).

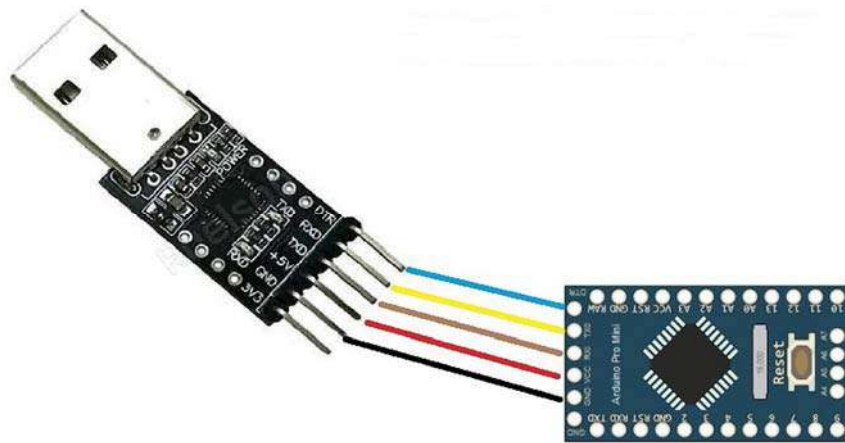


Fig. 6. The connection scheme for programming of the Arduino

3 Program realization

The Arduino integrated development environment is an environment in which an Arduino board can be programmed.

A written program or code is called SKETCH. In this work, the Arduino IDE is used as an environment in which the Arduino Pro Mini program is written, compiled and uploaded on the Arduino board as depicted in Fig. 7.

The Arduino can be connected to a computer through the USB port and programmed using a language similar to C++.

A simple program (sketch) used to operate the microcontroller and visualize the logistic map by looking at the array of blinking LEDs. A glowing LED represents an iteration of the one-dimensional map, and it is linked with a value in the interval $[0;1]$.

As example, the first LED from the left represents the first interval segment $[0;0.1)$, the second LED represents the segment $[0.1; 0.2)$, and so on. The last LED represents the final segment $[0.9; 1.0]$. When the microcontroller iterates the logistic map, a value belonging to one of these ten intervals is returned, and the corresponding LED turns on for 500 ms. This process is repeated infinitely, so we can observe the orbit followed by the logistic map by watching the blinking LEDs sequence.

The initial conditions for the logistic map can be changed and uploaded again into the microcontroller in any time. Thus, we can explore the behavior of the chaotic map when different values of the parameters r and X_0 are chosen.

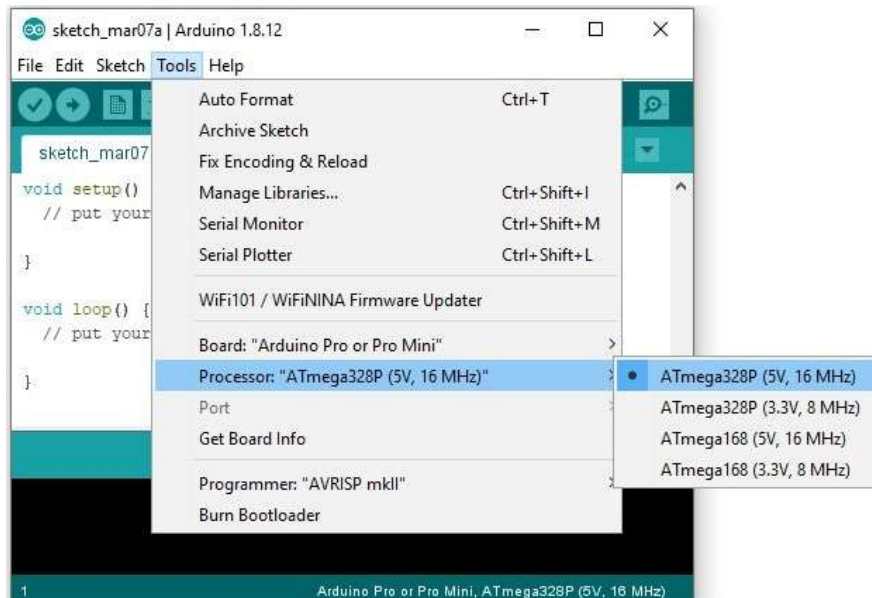


Fig. 7. Arduino IDE

Arduino programs require two mandatory functions: `setup()` and `loop()`. Any variable or constant defined outside these two functions is considered to be global. In the `setup()` function, we tell the microcontroller that there are ten light-emitting diodes connected to the digital pins and that they are intended to be turned on and off. In the `loop()` function, the logistic map is iterated, and the visualization process takes place as we observe the LEDs turning on and off, one after another, following the evolution of the nonlinear system.

The software to connect the LEDs using the following code:

```
// Logistic Map
// Select of the pin for each light-emitting diode (LED)
const int NbrLEDs = 10;
const int LEDpin[] = {4, 5, 6, 7, 8, 9, 10, 11, 12, 13};

const int wait = 500; // Wait for 500 milliseconds

// Parameters of Logistic Map
const double r = 3.8; // Logistic map constant

double X0 = 0.3; // Initial position
double X = X0; // X0 - first calculated point

// For initializes the LED pins use setup ()
void setup() {
```

```

    for (int i = 0; i < NbrLEDs; i++) {
        pinMode(LEDpin[i], OUTPUT);
    }
}

// For turn on/off LEDs and iterates the Logistic Map use
loop ()
void loop ()
{
    if (X < 0.1)
        blinkLED(LEDpin [0]) ;
    else if ((X >= 0.1 ) && (X < 0.2))
        blinkLED(LEDpin[1]);
    else if ((X >= 0.2) && (X < 0.3))
        blinkLED(LEDpin[2]);
    else if ((X >= 0.3) && (X < 0.4))
        blinkLED(LEDpin[3]);
    else if ((X >= 0.4) && (X < 0.5))
        blinkLED(LEDpin[4]);
    else if ((X >= 0.5) && (X < 0.6))
        blinkLED(LEDpin[5]);
    else if ((X >= 0.6) && (X < 0.7))
        blinkLED(LEDpin[6]);
    else if ((X >= 0.7) && (X < 0.8))
        blinkLED(LEDpin[7]);
    else if ((X >= 0.8) && (X < 0.9))
        blinkLED(LEDpin[8]);
    else
        blinkLED(LEDpin[9]);

    // Logistic Map function with iterates
    X0 = X;
    X = r * X0 * (1.0 - X0);
}

// Function for blinkLED
// turn on/off LEDs
void blinkLED(const int pin) {
    digitalWrite(pin, HIGH); // turn LED on
    delay (wait); // wait 500 milliseconds
    digitalWrite(pin, LOW); // turn LED off
}

```

4 Experimental realization

Fig. 8 shows connection scheme for experimental demonstration of chaotic behavior that realized one-dimensional system.

We connected an array of ten light-emitting diodes (HL1-HL10) to the digital pins (D4-D13) of the microcontroller using 220 Ohm resistors (R1-R10).

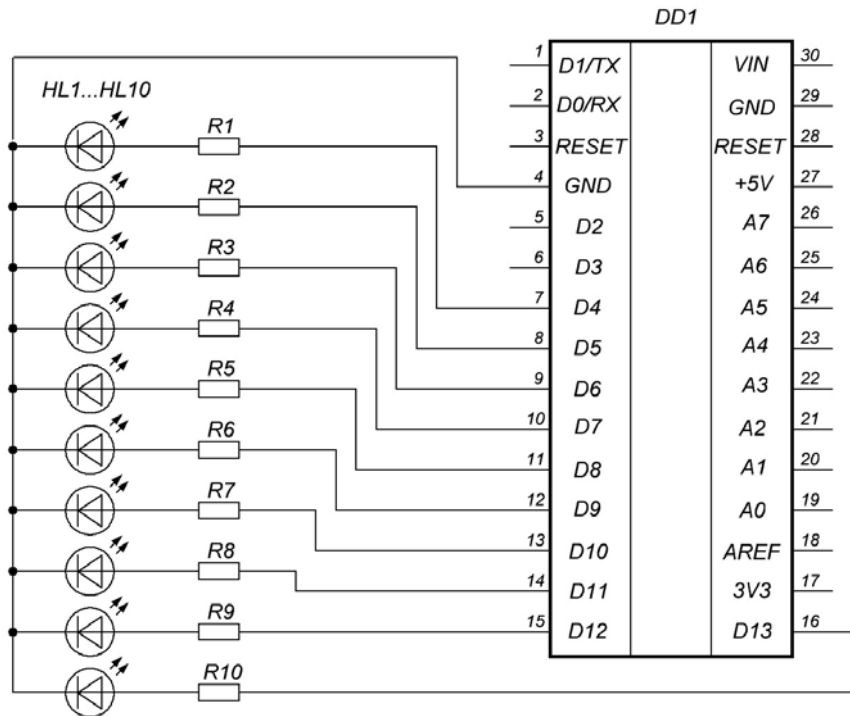


Fig. 8. Connection scheme for experimental demonstration of chaotic behavior

Fig. 9 shows experimental realization of logistic map for demonstration of chaotic behavior.

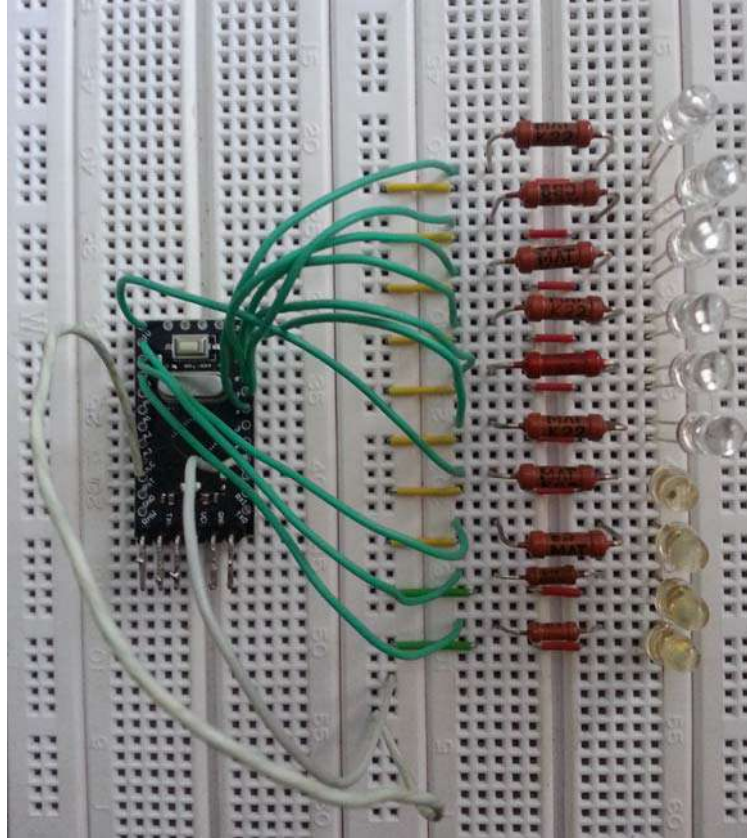


Fig. 9. Experimental realization for demonstration of chaotic behavior

5 Conclusions

The Arduino Pro Mini, ten light-emitting diodes (LEDs) and ten resistors for each part of segment of the range $[0;1]$ was used for practical realization and visualizing of the logistic map. It is one-dimensional system that demonstrate chaotic behavior.

Equation, system conditions and analysis of the iteration series with different parameter r are shown. For this analysis was used modern software environment Lab-View. Connection scheme, nominal components and programming code are also presented.

The Arduino was connected to a computer through the USB port and programmed using a language similar to C++. Sketch was uploaded into Arduino using program software ArduinoIDE.

References

1. Ramazanov, S., Tishkov, B.: Non-linear Forecasting of the State of a Socio-eco-oriented Innovative Economy in the Context of Systemic Crises. *CEUR Workshop Proceedings* **2422**, 181-193 (2019)
2. Pribylova, L.: Bifurcation routes to chaos in an extended Van der Pol's equation applied to economic models. *Electronic Journal of Differential Equations* **53**: 1-21 (2009)
3. Rusyn, V., Savko, O.: Modeling of Chaotic Behavior in the Economic Model. In 8th Chaotic Modeling and Simulation International Conference, Paris, 26-29 May 2015, pp. 705-712, (2015)
4. Hajnova, V., Pribylova, L.: Two-parameter bifurcations in LPA model. *Journal of Mathematical Biology*. **75**(5): 1235-1251 (2017)
5. Skiadas, C.H. Exact solutions of stochastic differential equations: Gompertz, generalized logistic and revised exponential. *Methodology and Computing in Applied Probability* **12**(2): 261-270 (2010)
6. Skiadas, C.H., Skiadas, C. Chaotic Modelling and Simulation: Analysis of Chaotic Models, Attractors and Forms. In: Taylor & Francis Group, pp. 1–345, LLC (2008)
7. Lingala, N., Namachchivaya, N.S., Pavlyukevich, I., Wedig, W.: Random perturbations of periodically driven nonlinear oscillators. *Procedia IUTAM* **19**, 91-100 (2016)
8. Lingala, N., Namachchivaya, N.S., Pavlyukevich, I.: Random perturbations of a periodically driven nonlinear oscillator: escape from a resonance zone. *Nonlinearity* **30**(4), 1376-1404 (2017)
9. Si, Y., Wang, J., Fečkan, M.: Controllability of linear and nonlinear systems governed by Stieltjes differential equations. *Applied Mathematics and Computation*, **376**, 125139 (2020)
10. Battelli, F., Fečkan, M.: On the Poincaré-Adronov-Melnikov method for the existence of grazing impact periodic solutions of differential equations. *Journal of Differential Equations*, **268**(7), 3725-3748 (2020)
11. Fečkan, M., Sathiyaraj, T., Wang, J.R.: Synchronization of butterfly fractional order chaotic system. *Mathematics*, **8**(3), 446 (2020)
12. Sher, M., Shah, K., Fečkan, M., Khan, R.A.: Qualitative analysis of multi-terms fractional order delay differential equations via the topological degree theory. *Mathematics*, **8**(2), 218 (2020)
13. Rusyn, V., Samila, A., Skiadas, Ch.: Computer modeling and practical realization of chaotic circuit with a light-emitting diode. In: Fourteenth International Conference on Correlation Optics, Chernivtsi, 16-19 September 2019, pp. 113690D (2019)
14. Milicka, P., Cízek, P., Faigl, J.: On Chaotic Oscillator-based Central Pattern Generator for Motion Control of Hexapod Walking Robot. *CEUR Workshop Proceedings* **1649**, 131-137 (2016)
15. Chua, L.: Memristor – the missing circuit element. *IEEE Trans. Circuit Theory*. **18**(5): 507-519 (1971)
16. Itoh, M., Chua, L.: Dynamics of Hamiltonian systems and memristor circuits. *International Journal of Bifurcation and Chaos*, **27**(2): 1730005 (2017)
17. Wang, Z.F., Shi, L., Wu, H., Helian, N., Chua, O.L.: Fractional memristor. *Applied Physics Letter* **111**, 243502 (2017)
18. Rusyn, V., Khrapko, S.: Memristor: modeling and research of information properties. *Springer Proceedings in Complexity*. Springer, pp. 229-238 (2019)
19. Yu, D., Zheng, C., Iu, H. H. C., Fernando, T., Chua, L. O.: A new circuit for emulating memristors using inductive coupling. *IEEE Access* **5**, 1284-1295 (2017)

20. Chua, L.: Memristor, Hodgkin–Huxley, and edge of chaos. *Nanotechnology* 24(38): 383001 (2013)
21. Vaidyanathan, S., Sambas, A., Kacar, S., & Çavuşoğlu, Ü.: A new three-dimensional chaotic system with a cloud-shaped curve of equilibrium points, its circuit implementation and sound encryption. *International Journal of Modelling, Identification and Control*, 30(3): 184-196 (2018)
22. Vaidyanathan S, Sambas A, Kacar S and Cavusoglu U.: A new finance chaotic system, its electronic circuit realization, passivity based synchronization and an application to voice encryption. *Nonlinear Engineering*, 8(1): 193-205 (2019)
23. Vaidyanathan, S., Azar, A. T., Rajagopal, K., Sambas, A., Kacar, S., & Çavusoglu, Ü.: A new hyperchaotic temperature fluctuations model, its circuit simulation, FPGA implementation and an application to image encryption. *IJSPM*, 13(3): 281-296 (2018)
24. Rusyn, V.: Modeling and Research Information Properties of Rucklidge Chaotic System Using LabView. In: 10th Chaotic Modeling and Simulation International Conference, Barcelona, 30 May – 2 June 2017, pp. 739-744, (2015)
25. Rusyn V., Kushnir M., Galameiko O.: Hyperchaotic Control by Thresholding Method. In: 11th Modern Problems of Radio Engineering, Telecommunications and Computer Science International Conference, TCSET 2012 - Proceedings, Lviv-Slavske, 21-24 February 2012, art. No. 6192785, p. 67 (2012)
26. Di Marco, M., Innocenti, G., Forti, M., Tesi, A.: Control Design for Targeting Dynamics of Memristor Murali-Lakshmanan-Chua Circuit. In 18th European Control Conference (ECC), pp. 4332-4337 (2019).
27. Vaidyanathan, S., Feki, M., Sambas, A., Lien C. H.: A new biological snap oscillator: its modelling, analysis, simulations and circuit design. *International Journal of Simulation and Process Modelling*, 13(5): 419-432 (2018)
28. Vaidyanathan, S., Azar, A. T., Sambas, A., Singh, S., Alain, K. S. T., Serrano, F. E.: A novel hyperchaotic system with adaptive control, synchronization, and circuit simulation. In: *Advances in System Dynamics and Control*, pp. 382-419, IGI Global (2018)
29. Vaidyanathan, S., Dolvis, L. G., Jacques, K., Lien, C. H., Sambas, A.: A new five-dimensional four-wing hyperchaotic system with hidden attractor, its electronic circuit realisation and synchronisation via integral sliding mode control. *International Journal of Modelling, Identification and Control*, 32(1): 30-45 (2019)
30. Vaidyanathan, S., Sambas, A., Zhang, S.: A new 4-D dynamical system exhibiting chaos with a line of rest points, its synchronization and circuit model. *Archives of Control Sciences*, 29(3): 485-506 (2019)
31. Rusyn, V., Skiadas, Ch.: Threshold Method for Control of Chaotic Oscillations. *Springer Proceedings in Complexity*. Springer, pp. 217-229 (2020)
32. Chua, L.: A zoo of strange attractors from the canonical Chua's circuits. In: *Proceedings of the IEEE 35th Midwest Symposium on Circuits and Systems*, Cat. No. 92CH3099-9, Vol. 2, pp. 916-926 (1992)
33. Rusyn, V., Mohamed, M.A., Purwandari, D., Mamat, M., Titaley, J., Pinontoan, B.: Chaotic and Controlling Regimes of a New Modified Chua's Generator. *Journal of Advanced Research in Dynamical and Control Systems*, 12(2), 556-561 (2020)
34. Rusyn, V., Sadli, M., Mamat, M., Mujiarto, Sanjaya Mada, W.S.: Computer Modelling of a New Simple Chaotic Generator. *Journal of Physics: Conference Series*, 1477, 022010 (2020)
35. May, Robert M.: Simple mathematical models with very complicated dynamics. *Nature*. 261(5560): 459-467 (1976)

A Method of Hidden Faults Opposition for FPGA-Based Components of Safety-Related Systems

Oleksandr Drozd¹[0000-0003-2191-6758], Vitaliy Romankevich²[0000-0003-4696-5935],
Alexei Romankevich³[0000-0001-5634-8469], Mykola Kuznietsov⁴[0000-0002-3043-5924],
Myroslav Drozd⁵[0000-0003-0770-6295]

^{1,4,5}Odessa National Polytechnic University, Ave. Shevchenko 1,
65044, Odessa, Ukraine

^{2,3}National Technical University of Ukraine “Igor Sikorsky Kyiv
Polytechnic Institute”, Victory avenue, 37, Kyiv, Ukraine

¹drozd@ukr.net, ^{2,3}romankev@scs.kpi.ua,

⁴koliaodessa@ukr.net, ⁵myroslav.drozd@opu.ua

Abstract. The paper is devoted to the problem of hidden faults, which is inherent in safety-related systems aimed at ensuring the functional safety of high-risk facilities to counter accidents. The problem of hidden faults is considered from the perspective of a resource-based approach as a problem of growth from a lower level of replication to the next level of diversification in the development of models, methods and means. Computer systems in critical applications have risen to the level of diversification in the division of the operating mode into normal and emergency, in the input data and structurally functional checkability, which for digital components have become different in these modes. Digital components continue to be traditionally stamped based on matrix structures that reflect the level of replication. The lag of the components from the development of the system creates a problem of hidden faults which can be accumulated during the normal mode and reduce the fault tolerance of the components and the functional safety of the system in emergency mode. We propose a method of counteracting hidden faults by raising components to the level of diversification in the promising field of FPGA designing. The proposed method uses the natural version redundancy inherent in the program code of the FPGA projects with LUT-oriented architecture. The method generates and selects versions of the program code, reducing many hidden faults of short circuits between neighboring inputs of LUT units. Possible hidden faults are eliminated by increasing the checkability of the FPGA project in normal mode and the trustworthiness of the results calculated in emergency mode.

Keywords: safety-related system, digital component, FPGA project, LUT unit, program code, hidden fault, resource-based approach, growth problem, checkability, trustworthiness, version redundancy

1 Introduction and Related Works

The most important direction in the development of information technologies is their improvement in the field of critical applications. This area, which is exceptional in nature, has already acquired a scale commensurate with the living space of mankind. Power grids and power plants, high-speed ground and air transport, hazardous chemical production, space technology and weapons are high-risk facilities [1, 2].

Risk assessment includes two factors related to the probability of an accident and the cost of its consequences. The factor of emergency consequences has a tendency of constant growth, which is caused by development of objects of high risk in the direction of their complication and increase of capacity, as well as increase of their number, density of location and proximity to densely populated areas [3, 4].

Only reducing the probability of accidents can counteract the growth of this factor. This mission is devoted to information technology implemented in safety-related systems, which, according to international standards, are aimed at "ensuring functional safety of both the system and the control object for preventing accidents and reducing the consequences if they occur" [5, 6].

The challenge in achieving functional safety "is to design the system in such a way as to prevent dangerous failures or to control them when they arise" [7]. This provision automatically includes hidden faults in the concept of "dangerous failures", as they pose a problem of their control for safety-related systems and their components [8, 9].

The fault control function is related to the development of the concept of checkability. At an early stage, this concept is known as testability or structural checkability of the digital circuit, characterizing it from the point of view of simplicity in the development of tests for fault detection [10, 11].

The structural nature of testability is due to the fact that it is completely determined by the structure of the digital circuit. The next stage in the development of checkability is related to on-line testing of digital circuits [12, 13], where checkability also shows dependence on input data, which characterizes it as structurally functional. The definition of self-testing circuits formulated in the theory of designing the totally self-checking circuits is known [14, 15]. Self-testing circuits are evaluated in checkability taking into account input data [16, 17].

Safety-related systems make a significant contribution to the development of the concept of checkability of the digital components by diversifying the operating mode into normal and emergency. Inputs are also included in the diversification process and become different in these modes. Then the structurally functional checkability of the digital circuit is converted into a double-mode, which differs in normal and emergency operation due to different input data. Such development of checkability converts harmless hidden faults of conventional computers working only in one operating mode, into dangerous ones. Indeed, a fault hidden throughout the operating mode will have no effect on the functionality of a regular computer. A hidden fault becomes a problem for safety-related systems because such faults can be accumulated in normal mode without causing errors under conditions of insufficient structurally functional checkability of the circuit, i.e. lack of necessary input data. In the emergency mode, the checkability is enriched with new input data and creates conditions for the manifestation of accumulated faults in the form of errors [18, 19].

The problem of hidden faults has a long history known from unsuccessful attempts to detect these faults using imitation modes aimed at recreating emergency conditions. The activation of such modes often provides for the shutdown of emergency protections, which has become one of the causes of the Chernobyl disaster.

In addition, history knows many examples of emergency consequences as a result of unauthorized activation of imitation modes because of human factor or due to the resulting fault [20, 21].

The presence of imitation modes creating a real danger to functional safety can be explained by two reasons:

- the high significance attached to hidden faults, which are feared more than emergency conditions created by imitation mode;
- lack of confidence in the fault tolerance of the solutions used, on which functional safety of critical systems and control objects is built [22-24].

The resource-based approach, which explores the integration of the artificial world created by human into the natural one, refers to the problem of hidden faults as a growth challenge. This approach identifies three levels in the development of models, methods and means: replication, diversification and self-sufficiency as a development goal. It shows the transition of safety-related systems to diversification and the backlog of their components, which continue to be stamped at a lower level of replication using matrix structures. Such classification of the problem of hidden faults determines the ways of its solution by raising the components to the system level [25, 26].

One of the most important directions in the development of digital components for critical systems is associated with FPGA designing. FPGA technologies are also a prime example of replication level dominance. FPGA chips contain Configurable Logic Blocks or Logical Elements, prepared iterative array multipliers, and chains for rapid carry propagation to add parallel codes, the libraries of IP-cores with matrix structures [27, 28].

However, FPGA refers to programmable hardware that raises stamped element matrices to the level of diversification by using the natural version redundancy in program code of FPGA projects with LUT-oriented architecture [29-31].

In this architecture, the computing process is organized using LUT units that are logic function generators. Their arguments arrive at the inputs of the LUT unit. The description of the logical function is stored in the memory of the LUT unit and written to this memory in the process of programming the FPGA project as program code. In the case of four inputs A, B, C and D, the memory of the LUT unit contains $2^4 = 16$ bits [32, 33].

Versions of the program code are created for each pair of the LUT units where output of the first LUT unit is connected to the input of the second one. The signal between the LUT units may be transmitted by a direct or inverse value using one of two versions of the program code. The inverse value at the output of the first LUT unit of the pair is provided by inverting the bits of its memory and changing its program code accordingly. The obtained inversion at the input of the second LUT unit of the pair is compensated by changing the program code with replacement bits of the memory [34].

This form of redundancy has been used to increase the trustworthiness of the calculated results by generating program code versions and selecting the best one from

the position of masking the faults between neighboring LUT unit inputs of the FPGA project [35, 36].

The selection of the version with the best structurally function checkability of the LUT units of the FPGA project in the normal mode of the safety-related system or trustworthiness of the results calculated in the emergency mode is proposed in [37].

A disadvantage of both solutions using version redundancy of program code is simulation of calculations on all normal and emergency mode inputs for each combination of versions generated by the LUT units of the FPGA project. The number of program code versions is defined as 2^Z , where Z is the number of first LUT units of all pairs.

For example, in the case of $Z = 30$ and $Z = 60$, simulation of calculations performed in the FPGA project is repeated for each input word of each mode more than 10^9 and 10^{18} times, respectively. The number of inputs is defined as 2^U , where U is the number of inputs of the simulated scheme of the FPGA project. For $U = 20$, the simulation must be repeated 10^{15} and 10^{24} times. Such a large number of iterations significantly limits the capabilities of the method in the processing of complex circuits. In addition, the choice of versions that increase the checkability of the FPGA project in normal mode or the trustworthiness of the results with the onset of emergency mode helps to reduce a set of the hidden faults, but in general is not the best solution.

We offer a method to counter hidden faults of the FPGA project taking into account the peculiarities of this kind of faults. The method allows maximum use of version redundancy of program code to reduce many possible hidden faults. Section 2 contains the main provisions of the proposed method. Section 3 describes the case study of the method using the FPGA project on example of the iterative array multiplier.

2 Main Provisions of the Suggested Method

The proposed method uses the synergy of several types of natural version redundancy. First of all, the method takes into account the natural version redundancy of safety-related systems, which is evident in their designing for operation in two essentially different modes: normal and emergency.

In addition, the method uses version redundancy of hidden fault elimination solutions. They pose a danger to fault tolerant decisions while meeting two conditions: accumulation of faults during normal mode and their manifestation in the form of errors in emergency mode. Therefore, the hidden fault is eliminated if at least one of the above conditions is excluded. Thus, the resistance to hidden faults can be achieved with the use of two versions of the solution, which consists in improving the checkability of LUT units and the trustworthiness of the results calculated on them, respectively.

The fault of the short circuit between the two neighboring inputs of the LUT unit also demonstrates natural version redundancy. It consists of two fault states: its masking or error manifestation in the case of the same signal values at the neighboring inputs of the LUT unit and otherwise, respectively. Fault masking increases the trust-

worthiness of the calculated results, and its manifestation in the form of error improves the checkability of the LUT unit.

Versions of the program code allow to control the input of the second LUT unit. They are assigned to this input directed or inverse value. Change version manages the state of the fault, showing it or masking for improving checkability of the LUT unit or the trustworthiness of the results. These improvements can be achieved concurrently, assigning them to different modes: normal and emergency, respectively.

The method generates and considers all versions of the program code that can be created for the second LUT unit of each pair. For example, if only two of the four LUT unit inputs connected to the outputs of the previous LUT units, then these LUT units will form two pairs and 4 versions.

The method distinguishes between three sets M_N , M_E and $M_{N\&E}$ of bits in the memory of each LUT unit: bits addressable in the normal, emergency, and in both modes, respectively. Faults which cause errors in bits of both M_N and $M_{N\&E}$ sets are not hidden, as can be detected in the normal mode. Therefore, the fault of the short circuit may be hidden, but if they cause errors only in bits of the M_E set and do not contain errors in bits, addressed in the normal mode.

The fault circuit between neighboring inputs of the LUT unit maintains proper access to its memory bits for identical values of these inputs and indicates the remaining bits in the values corresponding to zero values of the signals.

Examples of the effect of the short circuit faults on accessing the memory of the LUT unit are shown in Fig. 1.

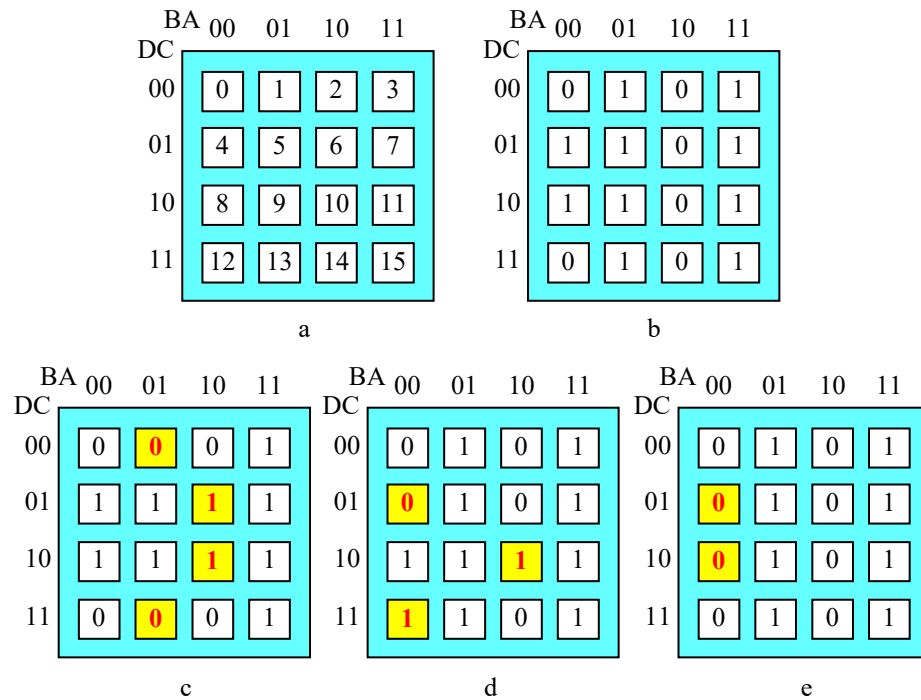


Fig. 1. Examples of LUT unit memory: numbers (a) and values (b) of bits and

memory in cases of shorts between A and B (c), B and C (d), C and D (e) inputs

The numbering of the LUT unit bits and the correct values of bits of memory, which is programmed with the $ABBA_{16}$ code, shown in Fig. 1, a and b. The memory of the LUT unit for cases of shorting inputs A and B, B and C, C and D is shown in Fig. 1, c, d, e, respectively.

Shorting the A and B inputs copies the values of bits BA 00 column into the memory array columns BA 01 and BA 10. BA 11 column bits retain their value. Shorting the B and C inputs copies the value of bits located at the intersection of column BA 00 and BA 01 with lines DC 00 and DC 10 into bits at the intersection of column BA 00 and BA 10 with a DC 01 and DC 11 lines as well as into bits at the intersection of column BA 10 and BA 11 with lines DC 00 and DC 10. The bits located at the intersection of columns BA 10 and BA 11 with lines DC 01 and DC 11 retain their values. Shorting the C and D inputs copies the values of bits of the DC 00 line into lines DC 01 and DC 10. Bits of the DC 11 lines retain their values. Erroneous bit values are highlighted in yellow.

The method performs the following steps:

Step 1: Determination of all second LUT units of the circuit and for each of them the set of all versions of program code.

Step 2. Simulation of calculations executed in FPGA project for all of the input data, i.e., U times, with determination of the $M_{NE} = M_N \cup M_{N\&E}$ and M_E bit sets in memory for each second LUT unit of the pair.

Step 3. Determination of the M_{NE} and M_E sets of bits in the memory for each version of each second LUT unit of the pair.

Step 4. Determination of all possible faults of the short circuit between neighboring inputs of each second LUT unit of the pair for cases where at least one of these inputs is connected to the output of the previous LUT unit.

Step 5. Determination of the program codes in the view, distorted under the influence of any faults. These program codes are generated for each version of each second LUT unit of pairs and compared with the correct program code versions. The erroneous bits detected in the sets M_{NE} and M_E , form the sets $M_{NE\ ER}$ and $M_{E\ ER}$, respectively.

Step 6. The FPGA project program code is generated using versions containing the minimum number of bits in $M_{E\ ER}$ sets with $M_{NE\ ER} = \emptyset$.

Step 7: The resulting program code is compared with the initial and least successful in the number of bits in the $M_{E\ ER}$ sets with $M_{NE\ ER} = \emptyset$ to evaluate the capabilities of the method.

The method improves the checkability of the FPGA project in the normal mode, and thus eliminates the hidden faults by choosing versions with sets of $M_{NE\ ER} \neq \emptyset$. In case $M_{NE\ ER} = \emptyset$, the method selects the versions with the lowest number of bits in the $M_{E\ ER}$ set to reduce the set of hidden faults that manifest themselves in an emergency mode.

Reducing many hidden faults, performed in both modes, aimed at improving the trustworthiness of the results calculated in the emergency mode. From this position, the trustworthiness of the FPGA project results with respect to hidden faults of the short circuit between neighboring inputs of LUT units can be estimated taking into

account erroneous memory bits addressed only in emergency mode in the case of $M_{NE\ ER} = \emptyset$.

The trustworthiness of the result read from the output of the LUT unit can be estimated as $T_{LUT} = (1 - (A_{E\ ER} / (3A_{E\ ER})) K_{NE\ ER}) \times 100\%$, where $A_{E\ ER}$ and A_E are the number of bits in the $M_{E\ ER}$ sets for all three types of short circuit and in the M_E set, respectively, $K_{NE\ ER} = 1$ if $M_{NE\ ER} = \emptyset$, and $K_{NE\ ER} = 0$ otherwise.

The trustworthiness of the FPGA project results can be estimated by the arithmetic average of the T_{LUT} values calculated for all LUT units.

The contribution that is made to the trustworthiness by the checkability of the LUT units in the normal mode can be estimated similarly taking into account the error memory bits of the $M_{E\ ER}$ set in the case of $M_{NE\ ER} \neq \emptyset$.

Comparison of the best solution with the initial project and the least successful version of the program code shows the effectiveness of the method in the specific example of FPGA design and the potential of the method, respectively.

3 Case Study of the Proposed Method

Experimental verification of the method was carried out using CAD Quartus Prime 18.1 Lite Edition on the example of a study of a 4-bit iterative array multiplier implemented in Intel Max 10 FPGA 10M50DAF672I7G [38, 39]. The digital circuit of the obtained FPGA project contains 8 inputs which are supplied with 4-bit operands, 30 LUT units with four inputs for performing the multiplication operation, and 8 product outputs. The digital circuit simulation was carried out using the program implementation of the method. The program was developed in the free Delphi 10 Seattle demo version [40].

As initial data, the program uses a description of a digital circuit with an indication of the connections of its inputs or outputs of previous LUT units to the inputs of each next LUT unit and the outputs of the circuit. In addition, the number of digital circuit inputs, the number of LUT units and their program codes are indicated.

The program presents the results of examining the digital circuit on the main panel, which is shown in Fig. 2.

The main panel is invoked by pressing the «Start» key and allows to complete the program on «Exit» command. The main panel allows to view the values of the memory bits for all LUT units operating at different threshold values S , dividing the input data of the normal and emergency modes.

The values of the factors smaller than the S threshold refer to the normal mode, and the rest to emergency one. The "S: 2 - 9" key determines 8 threshold values from 2 to 9. Each press of this key shifts the range of S values by one to the maximum: "S: 8 - 15" and then the value "S: 2 - 9". The "LUT # 22" key specifies the number 22 of the considered LUT unit. Clicking on this key allows to proceed to the LUT unit with the following number. The largest number is replaced by the first. Bits addressable in the normal and in emergency mode only, colored in green and yellow, respectively. Blue color indicates addressing in both modes.

The lower part of the panel shows the results of the proposed method for each value of the S threshold. Line "V" shows the decimal number of the best version of

the program code. The binary code $dcb a_2$ of the version number determines the inverted inputs of the second LUT unit by the unit values of the corresponding bits.

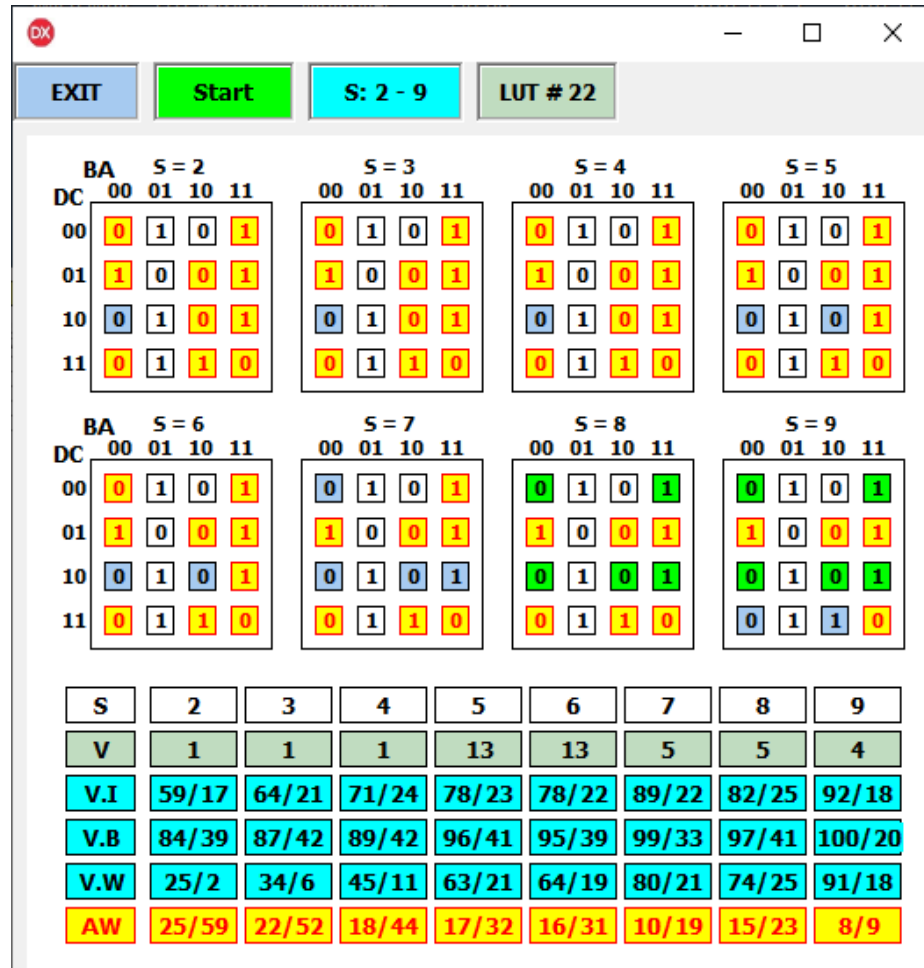


Fig. 2. The main panel of the program implementation of the suggested method

For example, version $13 = 1101_2$ means inverting inputs D, C and A. Number $0 = 0000_2$ indicates the preservation of the source program code [41].

The next three lines «V.I», «V.B» and «V.W» evaluate the trustworthiness of the calculated results to the initial, the best and the least successful version of the program code for FPGA project, respectively. The numerator includes the trustworthiness of the results calculated in the FPGA project, and the denominator contains the contribution that is made to the trustworthiness of the checkability of the LUT units in normal mode. The last line in the numerator and denominator shows the gain in the

trustworthiness of the best solution compared to the initial project and the least successful version of the program code, respectively.

Diagrams of the dependence in trustworthiness of the results on the S threshold for the best, initial and least successful solution are shown in Fig. 3.

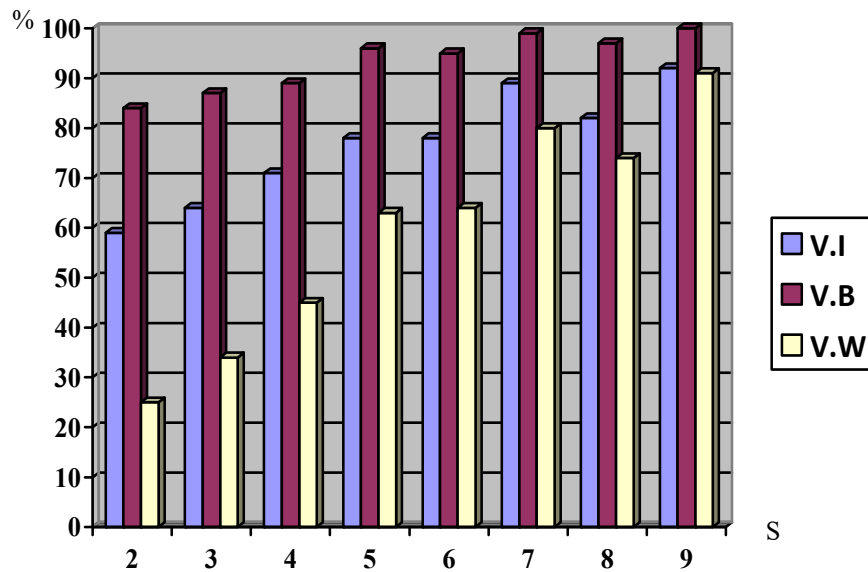


Fig. 3. Diagrams of the dependence in trustworthiness of the results on the S threshold for the best, initial and least successful solution

Diagrams show a tendency to increase trustworthiness of the results with an increase in the S threshold.

Therefore, it is important that the greatest gain in the best solution obtained by the proposed method is achieved for small threshold values that are typical for circuits operating in normal mode at a noise level. For $S = 2$ and $S = 3$, the trustworthiness of the results increases relative to the initial program code from 59% to 84% and from 64% to 87%, i.e. by 25% and 23%, respectively.

4 Conclusions

FPGA designing, which is a promising direction in the development of digital components for safety-related systems, opens up new possibilities for solving the hidden fault problem inherent in such systems. The programmability of FPGA projects allows to solve this problem as a growth challenge by raising the components to the level of diversification, where critical systems are located in operating modes, input data and digital circuit checkability.

The proposed method uses the version redundancy in the program code of FPGA projects with a LUT-oriented architecture to reduce the set of hidden faults in the short circuit of neighboring inputs of the LUT units.

Such faults distort the addressing of the memory bits of the LUT units. Memory bits read at the wrong address may have erroneous values that reduce the trustworthiness of the calculated results.

A hidden fault is dangerous with errors that are not detected in normal mode and distort the results in emergency one.

The method reduces the number of hidden faults in two ways. The first way is to search for each LUT unit the program code versions, showing a failure in the normal mode. The second way applies to LUT units for which there are no such version, and chooses the version with the least amount of erroneous values in bits, addressed only in emergency mode.

The advantage of the proposed method is its low complexity, limited not by the set of all versions of the FPGA project program code, but by the set of LUT units with mutually independent examination of their program codes.

Further research is planned in the direction of expanding the circle of problems which can be identified and solved as a growth challenge in relation to safety-related systems and their components based on the development and practical application of the concepts of checkability and natural version redundancy, in particular in FPGA designing.

References

1. International Atomic Energy Agency, Evaluation of the Status of National Nuclear Infrastructure Development, IAEA Nuclear Energy Series No. NG-T-3.2, IAEA, Vienna (2008)
2. Hiromoto, R. E., Sachenko, A., Kochan, V. et. al.: Mobile Ad Hoc Wireless Network for Pre- and Post-Emergency Situations in Nuclear Power Plant. In: WS 2014 - 2nd IEEE International Symposium on Wireless Systems within the Conferences on Intelligent Data Acquisition and Advanced Computing Systems, Offenburg, Germany, pp. 92-96 (2014) doi: 10.1109/IDAACS-SWS.2014.6954630
3. Ivanchenko, O., Kharchenko, V., Moroz, B. et. al.: Risk Assessment of Critical Energy Infrastructure Considering Physical and Cyber Assets: Methodology and Models. In: IDAACS 2018 - 10th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, Lviv, Ukraine, pp. 225-228 (2018) doi: 10.1109/IDAACS-SWS.2018.8525594
4. Smith, D. J.: Reliability, Maintainability and Risk. Practical Methods for Engineers, 9th Edition, Butterworth-Heinemann (2017)
5. International Atomic Energy Agency, On-line Monitoring for Improving Performance of Nuclear Power Plants, Part 2: Process and Component Condition Monitoring and Diagnostics, IAEA Nuclear Energy Series No. NP-T-1.2, IAEA, Vienna (2008)
6. Smith D., Simpson K.: The Safety Critical Systems Handbook, 4th Edition, Butterworth-Heinemann (2016)
7. International Electrotechnical Commission, Nuclear Power Plants: Instrumentation and Control for Systems Important to Safety – General Requirements for Systems, Rep. IEC 61513, IEC, Geneva (2001)

8. Efanov, D., Lykov, A., Osadchy, G.: Testing of relay-contact circuits of railway signalling and interlocking. In: EWDTs 2017 - IEEE East-West Design and Test Symposium, Novi Sad, Serbia, pp. 242-248 (2017) doi: 10.1109/EWDTs.2017.8110095
9. Drozd, O., Antoniuk, V., Nikul, V., Drozd, M.: Hidden faults in FPGA-built digital components of safety-related systems. In: TCSET 2018 - 14th International Conference "Modern problems of radio engineering, telecommunications and computer science, Lviv-Slavsko, Ukraine, pp. 805-809 (2018) doi: 10.1109/TCSET.2018.8336320
10. Hahanov, V., Litvinova, E., Obrizan, V., Gharibi, W.: Embedded method of SoC diagnosis. *Elektronika in Elektrotehn*, no. 8, 3-8 (2008)
11. Matrosova, A., Nikolaeva, E., Kudin, D., Singh, V.: PDF testability of the circuits derived by special covering ROBDDs with gates. In: EWDTs 2013 - IEEE East-West Design and Test Symposium, Rostov-on-Don, Russia, pp. 1-5 (2013) doi: 10.1109/EWDTs.2013.6673183
12. Coppad, D., Sokolov, D., Bystrov, A., Yakovlev, A.: Online Testing by Protocol Decomposition. In: IOLTS - 12th IEEE International On-Line Testing Symposium, Como, Italy, pp. 263-268 (2006) doi: 10.1109/IOLTS.2006.45
13. Drozd, A., Drozd, J., Antoshchuk, S., Nikul, V., Al-dhabi, M.: Objects and Methods of On-Line Testing: Main Requirements and Perspectives of Development. In: EWDTs 2016 - IEEE East-West Design & Test Symposium, Yerevan, Armenia, pp. 72-76 (2016) doi: 10.1109/EWDTs.2016.7807750
14. Anderson, D. A., Metzger, G.: Design of Totally Self-Checking Circuits for n-out-of-m Codes. *IEEE Trans. on Computers*, vol. C-22, 263-269 (1973) doi: 10.1109/T-C.1973.223705
15. Metra, C., Schiano, L., Favalli, M., Ricco, B.: Self-checking scheme for the on-line testing of power supply noise. In: DATE 2002 - Design, Automation and Test in Europe Conference, Paris, France, pp. 832-836 (2002) doi: 10.1109/DATE.2002.998395
16. Chakrabarty, K., Swaminathan S.: Built-in self-testing of high-performance circuits using twisted-ring counters. In: ISCAS 2000 - IEEE International Symposium on Circuits and Systems, Geneva, Switzerland (2000) doi: 10.1109/ISCAS.2000.857029
17. Kondratenko, Y.P., Kozlov, O.V., Topalov, A.M., Gerasin, O.S. Computerized system for remote level control with discrete self-testing. In: *CEUR Workshop Proceedings Open Access*, vol-1844, pp. 608-619 (2017) <http://ceur-ws.org/Vol-1844/10000608.pdf>
18. Drozd, A., Kharchenko, V., Antoshchuk, S., Sulima, J., Drozd, M.: Checkability of the digital components in safety-critical systems: problems and solutions. In: EWDTs 2011 - IEEE East-West Design & Test Symposium, Sevastopol, Ukraine, 2011, pp. 411-416 doi: 10.1109/EWDTs.2011.6116606
19. Drozd, A., Antoshchuk, S., Drozd, J. et. al.: Checkable FPGA Design: Energy Consumption, Throughput and Trustworthiness. In: *Green IT Engineering: Social, Business and Industrial Applications, Studies in Systems, Decision and Control*, vol. 171, Berlin, Heidelberg: Springer International Publishing, pp. 73-94 (2019) doi: 10.1007/978-3-030-00253-4_4
20. Gray, R.: The true toll of the Chernobyl disaster, *BBC Future* (2019) <https://www.bbc.com/future/article/20190725-will-we-ever-know-chernobyls-true-death-toll>
21. Gillis, D.: The Apocalypses that Might Have Been. [Online]. Available: <https://www.damninteresting.com/the-apocalypses-that-might-have-been/>.
22. Edstrom, J., Tilevich, E.: Reusable and Extensible Fault Tolerance for RESTful Applications. In: 11th International Conference on Trust, Security and Privacy in Computing and Communications, Liverpool, UK, pp. 737-744 (2012) doi: 10.1109/TrustCom.2012.244
23. Atamanyuk, I., Kondratenko, Y.: Computer's Analysis Method and Reliability Assessment of Fault-Tolerance Operation of Information Systems. In: *CEUR-WS*, vol. 1356, Lviv, Ukraine, pp. 507-522 (2015)
24. Romankevich, A., Feseniuk, A., Maidaniuk, I., Romankevich, V.: Fault-tolerant multiprocessor systems reliability estimation using statistical experiments with GL-models. In: *Advances in Intelligent Systems and Computing*, vol. 754, pp. 186-193 (2019)

25. Drozd, J., Drozd, A., Al-dhabi, M.: A resource approach to on-line testing of computing circuits. In: EWDTS 2015 - IEEE East-West Design & Test Symposium, Batumi, Georgia, pp. 276-281 (2015) doi: 10.1109/EWDTS.2015.7493122
26. Drozd, O., Kharchenko, V., Rucinski, A. et. al.: Development of Models in Resilient Computing. In: DESSERT 2019 - 10th IEEE International Conference on Dependable Systems, Services and Technologies, Leeds, UK, pp. 2-7 (2019) doi: 10.1109/DESSERT.2019.8770035
27. Tyurin, S.F., Grekov, A.V., Gromov, O.A.: The principle of recovery logic FPGA for critical applications by adapting to failures of logic elements. World Applied Sciences Journal, 328-332 (2013) doi: 10.5829/idosi.wasj.2013.26.03.13474
28. Jaechon Jung, Ibrahim Ahmed: Development of FPGA-based reactor trip functions using systems engineering approach, Nuclear Engineering and Technology, March 2016, pp. 2-11 (2016) doi: 10.1016/j.net.2016.02.011
29. Palagin, A., Opanasenko, V.: The implementation of extended arithmetic's on FPGA-based structures. In: IDAACS 2017 - 9th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, vol. 2, Bucharest, Romania, pp. 1014-1019 (2017) doi: 10.1109/IDAACS.2017.8095239
30. Chernov, S., Titov, S., Chernova, L. et. al.: Algorithm for the simplification of solution to discrete optimization problems. Eastern-European Journal of Enterprise Technologies 3 (4), 1-12 (2018) doi: <https://doi.org/10.15587/1729-4061.2018.133405>
31. Zashcholkina, K., Ivanova, O.: The control technology of integrity and legitimacy of LUT-oriented information object usage by self-recovering digital watermark. In: CEUR Workshop Proceedings, vol. 1356, pp. 498-506 (2015)
32. Cyclone II Architecture. Cyclone II Device Handbook Version 3.1.-Altera Corporation (2007) http://www.altera.com/literature/hb/cyc2/cyc2_cii51002.pdf
33. Toshinori, S.: Basic Knowledge to Understand FPGAs. In: Principles and Structures of FPGAs, H. Amano (edit), Springer, USA, New-York, pp. 1-22 (2018)
34. Zashcholkina, K., Ivanova, O.: LUT-object integrity monitoring methods based on low impact embedding of digital watermark. In: TCSET 2018 - International Conference "Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering, Lviv-Slavske, Ukraine, pp. 519-523 (2018) doi: 10.1109/TCSET.2018.8336255
35. Drozd, A., Drozd, M., Kuznetsov, M.: Use of Natural LUT Redundancy to Improve Trustworthiness of FPGA Design. In: CEUR Workshop Proceedings, vol. 1614, pp. 322-331 (2016)
36. Pleskacz, W., Jenihhin, M., Raik, J. et. al.: Hierarchical Analysis of Short Defects between Metal Lines in CMOS IC. In: 11th Euromicro Conference on Digital System Design Architectures, Methods and Tools, Parma, Italy, pp. 729-734 (2008) doi: 10.1109/DSD.2008.98
37. Drozd, A., Drozd, M., Martynyuk, O., Kuznetsov, M.: Improving of a Circuit Checkability and Trustworthiness of Data Processing Results in LUT-based FPGA Components of Safety-Related Systems. In: CEUR Workshop Proceedings, vol. 1844, pp. 654-661 (2017)
38. Intel Quartus Prime Standard Edition User Guide: Getting Started, <https://www.intel.com/content/dam/www/programmable/us/en/pdfs/literature/ug/ug-qps-getting-started.pdf>, last accessed 2019/03/20
39. Max 10 FPGA Device Architecture (2017), https://www.intel.com/content/dam/www/programmable/us/en/pdfs/literature/hb/max-10/m10_architecture.pdf, last accessed 2019/03/20
40. Delphi 10 Seattle: Embarcadero (2015) <https://www.embarcadero.com/ru/products/delphi/>
41. Drozd, O., Kuznetsov, M., Martynyuk, O., Drozd, M.: A method of the hidden faults elimination in FPGA projects for the critical applications. In: DESSERT 2018 - 9th IEEE International Conference on Dependable Systems, Services and Technologies, Kyiv, Ukraine, pp. 231-234 (2018) doi: 10.1109/DESSERT.2018.8409131

Docker Container based PaaS Cloud Computing Comprehensive Benchmarks using LAPACK

Dmitry Zaitsev¹[0000-0001-5698-7324], Piotr Luszczek²[0000-0002-0089-6965]

¹Odessa State Environmental University, 15 Lvivska Str., Odessa, 65016, Ukraine

daze@acm.org

²University of Tennessee's Knoxville, 1122 Volunteer Blvd, Knoxville, TN 37996, USA

luszczek@icl.utk.edu

Abstract. Platform as a Service (PaaS) cloud computing model becomes widespread implemented within Docker Containers. Docker uses operating system level virtualization to deliver software in packages called containers. Containers are isolated from one another and comprise all the required software, including operating system API, libraries and configuration files. With such advantageous integrity one can doubt on Docker performance. The present paper applies packet LAPACK, which is widely used for performance benchmarks of supercomputers, to collect and compare benchmarks of Docker on Linux Ubuntu and MS Windows platforms. After a brief overview of Docker and LAPACK, a series of Docker images containing LAPACK is created and run, abundant benchmarks obtained and represented in tabular and graphical form. From the final discussion, we conclude that Docker runs with nearly the same performance on both Linux and Windows platforms, the slowdown does not exceed some ten percent. Though Docker performance in Windows is essentially limited by the amount of RAM allocated to Docker Engine.

Keywords. Cloud computing, PaaS, Docker, benchmark, LAPACK, Top500

1 Introduction

A concept of cloud computing [1], widely spread recently, means on-demand availability of computer system resources. Peter Mell and Tim Grance from National Institute of Standards and Technology (NIST) define cloud computing as a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort or service provider interaction. Their cloud model is composed of five essential characteristics, three service models, and four deployment models. Five essential characteristics include: on-demand self-service, broad network access, resource pooling, rapid elasticity, and measured service. The service models include: Software as a Service (SaaS), Platform as a Service (PaaS), and Infrastructure as a

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

Service (IaaS). Deployment models are the following: Private cloud, Community cloud, Public cloud, and Hybrid cloud.

Among known implementations of PaaS service model, Docker Containers [2-4] are rapidly developing and attracting more and more customers in various application areas. Docker uses operating system level virtualization to deliver software in packages called containers. Containers are isolated from one another and contain all the required software, including operating system API, libraries and configuration files. Containers can communicate with each other through well-defined channels. All containers are run by a single operating-system kernel and are thus more lightweight than virtual machines [5-6]. The software that hosts the containers is called Docker Engine and is installed now on Linux, MS Windows, and Apple MacOS platforms. Works on Docker has started in 2010 and first released in 2013. Recently Docker wins millions of developers and customers offering a public repository of containers. It supports also parallel and distributed computing with Docker Swarm technology.

Developers of Docker claim that the application within a container runs quickly and reliably from one computing environment to another. On the one hand, it is utmost convenient tool encapsulating within a single container everything required to run an application – all dependencies including operating system API, libraries, configurations etc. On the other hand, it is efficient because it is light-weighted compared with the virtual machine concept. Moreover, it is ubiquitous, running everywhere where Docker Engine is installed and access to repository of containers is provided. Thus, it looks like a miracle that looks a bit of suspicious for IT professionals [6,7]. The design of research presented in the paper was either to prove it strictly and convincingly or to depose a legend.

In the present paper after a brief overview of Docker Container technology, we come to LAPACK software [8,9], widely applied for performance benchmarks including modern supercomputers [10,11]; then we meticulously create a series of containers to run LAPACK, compose one solid container for benchmarks, and upload it to Docker Hub repository; we run the LAPACK container on Ubuntu Linux and MS Windows 10 platforms installed on the same computer and also planning to supply additional information regarding benchmarks for Apple MacOS platform in future; finally we represent benchmarks graphically and discuss obtained results. A conclusion that Docker completely corresponds to the company claims showing rather good performance for any available platform accomplishes the paper.

2 Basics of Developing and Running Docker Containers

Docker [2-4] represents one of the most successful implementations of the PaaS Cloud Computing concept. A Docker image encapsulates an application together with its entire environment including libraries and operating system and runs on Docker Engine. Docker Engine can be started on Linux, Windows, MacOS and in future on other operating systems. Thus a certain independence of an image from operating environment is provided while it is claimed that Docker runs an image considerably faster than a virtual machine.

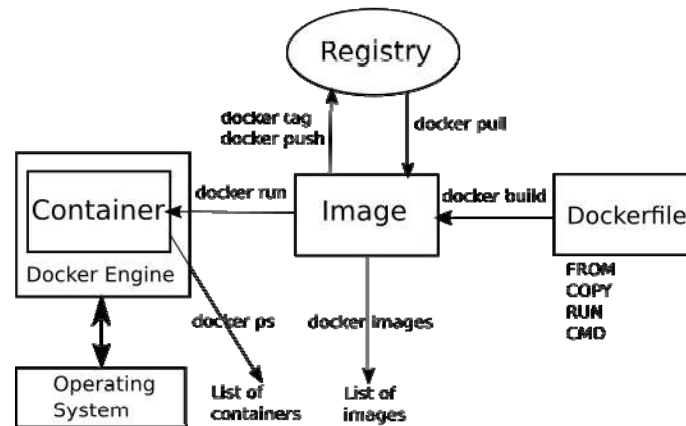


Fig. 1. Docker brief scheme of work.

Docker command line interface executes commands starting with “docker” prefix. To run an image, we use “docker run” command. For a quick start, we can either run *hello-world* image or use explicitly Linux *echo* command starting image *ubuntu*:

```
docker run hello-world
docker run ubuntu /bin/echo 'Hello world'
```

On processing *run* command, Docker creates and starts a new container from the image, downloading a new image if required. A container runs within Docker Engine which can be considered as a kind of thin virtual machine (Fig. 1); recently Docker Engine works within Linux, MS Windows, and MacOS. There are more than fifty Docker commands [2]. Among the most frequently used, we mention: “docker info” to display system-wide info; “docker images” to list images; “docker ps” to list containers (running images); “docker build” to build an image from a *Dockerfile* which represents a textual file specifying how the image should be built. For instance, we build and run image specified by the following file named *Dockerfile* and stored in the current directory:

```
cat Dockerfile
FROM ubuntu:latest
RUN /bin/bash
docker build -t u2 .
docker run -it u2
#
```

The option “-t” specifies the image name, and the option “-it” specifies an interactive mode of work; after starting, the image issues an invitation “#” waiting a bash command to process. A *Dockerfile* begins with “FROM” instruction that specifies parent image from which the current image is built. To use a file when building an image, we copy it into the image using “COPY” instruction. Instruction “RUN” speci-

fies commands which build a new layer of the current image. When an image is started, a command specified in “CMD” instruction is executed. Dockerfiles containing all the considered instructions are studied in Section 4.

On default, all the built images are stored locally within the current computer. It is rather convenient to store images in a cloud repository (registry), for instance <http://hub.docker.com>. For this purpose we create the corresponding account and enter it using “docker login” command. Local images are identified using 6 octets represented as a hex number, we can obtain identifies via “docker images” command. To push an image to a repository we tag it with a name using “docker tag” command and then push it using “docker push” command with specified name. When working on a new computer, we can pull an image from repository using “docker pull” command. Pushing and pulling images are studied in detail in Section 4.

3 LAPACK as de-facto Standard for Performance Benchmarks

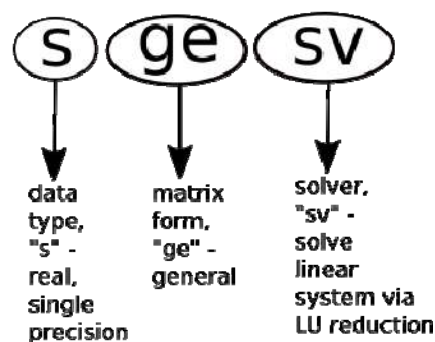
The LAPACK library contains a collection of numerical methods to solve problems arising from dense systems of linear equations assuming one of the following forms:

$$\begin{aligned} Ax &= b, \\ Ax &= \lambda x, \\ Ax &= \lambda Bx, \\ Av &= \sigma u. \end{aligned}$$

Less conceptually, the properties of matrices A and B dictate the specific of the algorithmic approach taken by the LAPACK solver. The overarching methodology involves the decomposition approach [12] that splits the algorithmic work into two distinct phases: factorization and substitution. The former is commonly the computationally-intensive step that consumes majority of execution time. The latter uses the output of the factorization to deliver the solution of the original problem at much lower cost than the first step. Moreover, the factorization results may be reused multiple times with different right-hand sides for much reduced computation time in many practical situations. Furthermore, numerical stability of using the factors is superior to using explicit inverse of the system matrix that may fail with overflows or division-by-zero even when the solve with the factors would still succeed to deliver a few digits of accuracy in the solution. The factorizations implemented in LAPACK include the LU, Cholesky, and QR factorizations that are often called one-sided factorizations because they apply numerical transformations to only one side of the original matrix and hence do not preserve the spectral properties. In contrast, the two-sided transformations do preserve the original matrix spectrum and are used in LAPACK’s Schur decomposition of a matrix. These two types of transformations present a different challenge to the modern hardware and thus may be used for a wide range of evaluations.

From the benchmarking standpoint, LAPACK offers a wide range of routines that reveal performance of characteristics of the tested hardware platform. At the basic

level there are BLAS subroutines that constitute the portable performance layer above the hardware that allows the rest of LAPACK to express more complex algorithms in an efficient manner. Using BLAS for benchmarking is an easy option to obtain trivial hardware metrics such as bandwidth and latency of the main memory by running some of the Level 1 and Level 2 BLAS subroutines. Moving up to Level 3 BLAS allows the user to test compute capability of the hardware platform both in terms of single-core performance as well as its scaling across all available sockets and cores contained therein. In practice, more complex measures of performance are obtained from LAPACK subroutines. One common example is the High Performance LINPACK benchmark [13] commonly referred to as HPL. The benchmark measure the time taken to solve a system of linear equations with a dense system matrix of an arbitrary size by using LU factorization with partial pivoting. This can be easily implemented by a single call to LAPACK's routine DGESV() and stands to show how LAPACK is an important performance tool. In fact, HPL has been used for decades for measuring efficiency and comparison the largest supercomputing machines in the world as is recorder by the TOP500 list [10]. Prior to proliferation of distributed memory machines, the shared memory supercomputers were dominating the list and, at that time, LAPACK's implementation of DGESV() was a perfect software implementation of HPL eligible for running and submitting TOP500 results. Later on, a distributed memory implementation of DGESV() was required and it had been provided by the ScaLAPACK library in the form of PDGESV() routine that not only computes locally on each of the distributed nodes but also coordinates the solution between the processes through an interconnect fabric, including Ethernet or Infini-Band, and a message passing library such as MPI. However, the detailed use and analysis of ScaLAPACK is outside of scope of this work due its complex software stack requirements and hardware complexities associated with running a modern distributed-memory cluster. These complex hardware-software interactions complicate evaluation of containers and might prevent from drawing proper conclusions about the associated overheads.



LAPACK packet is downloaded freely from its website [8] and installed according to the installation guide in some quarter of an hour, in its work it uses BLAS library

[9,13]. For instance, to solve a given system, we can use the following small program that initializes matrix A and vector of the right side b statically and then calls routine LAPACKE_sgesv to solve the corresponding system:

```
#include <stdio.h>
#include <lapacke.h>
#define size 3          // dimension of matrix A
int main()
{
    lapack_int n=size, nrhs=1, lda=size, ldb=1, info,
    ipiv[size];
    float A[size*size] = {2.0,5.0,8.0,
                          3.0,-6.0,9.0,
                          4.0,7.0,-1.0};
    float b[size] =      {13.0,25.0,-7.0};
    int i,j;
    info = LAPACKE_sgesv(LAPACK_ROW_MAJOR, n, nrhs, A, lda,
    ipiv, b, ldb);
    for (i=0; i<n; i++) printf("%f\n", b[i]); // print vec-
    tor x
}
```

Finally it prints the obtained vector of solutions which replaces the right side vector *b*. The program is written in C language and uses C interface of LAPACK called LAPACKE, all the corresponding routines have prefix “LAPACKE_”. LAPACKE only translates the call into the call of the corresponding LAPACK routine, sgesv in the considered example (a scheme of LAPACK routine names abbreviation is explained by Fig. 2). The program can be built and run using the following *Makefile*:

```
include ../../make.inc
all: lp_ex1
LIBRARIES = ../../$(LAPACKELIB) ../../$(LAPACKLIB)
$(BLASLIB)
lp_ex1: lp_ex1.o $(LIBRARIES)
    $(LOADER) $(LOADOPTS) -o $@ $^
    ./ $@
.c.o:
    $(CC) $(CFLAGS) -I. -I../include -c -o $@ $<
```

We assume that our example directory is situated at the same level as the standard LAPACKE example directory /lapack-3.8.0/LAPACKE/example. The computed vector of solutions can be checked by substitution into equations of the system:

```
daze@lion:~/lapack-3.8.0/LAPACKE/example$ make
gcc -O3 -I. -I../include -c -o lp_ex1.o lp_ex1.c
gfortran -o lp_ex1 lp_ex1.o ../../liblapack.a
../../liblapack.a ../../librefblas.a
```

```

./lp_ex1
-1.313167e+00
-1.000000e-01
1.800000e+00
daze@lion:~/lapack-3.8.0/LAPACKE/example$

```

Standard set of LAPACKE tests includes routines which solve a given linear system or implement computations by the least square method. At first, a system of a given size is generated with random elements and then it is solved, the obtained solutions printed. It uses two options: “-n” to specify the system size and “-nhrs” to specify the number of right-hand sides; for instance:

```

daze@lion:~/lapack-3.8.0/LAPACKE/example$
./xexample_DGESV_rowmajor -n 3
Entry Matrix A
  0.34  -0.11   0.28
  0.30   0.41  -0.30
 -0.16   0.27  -0.22
Right Rand Side b
  0.05
 -0.02
  0.13
LAPACKE_dgesv (row-major, high-level) Example Program Results
Solution
 -0.48
  1.19
  1.22
Details of LU factorization
  0.34  -0.11   0.28
  0.88   0.50  -0.55
 -0.48   0.43   0.15
Pivot indices
    1     2     3
daze@lion:~/lapack-3.8.0/LAPACKE/example$

```

One can check the obtained results and calculate the error, such intermediate results as LU-factorization and pivot indices are printed as well, though in the present study we are interested mainly in LAPACK running time to use it for benchmarks.

4 Install and Run LAPACK in Docker Containers

In the present section we create a Docker image to run LAPACK tests and also an additional image to solve the example equation from the previous section. We are going to compare multi-layer and solid images as well that is why we create a series

of images having the following structure of directories and files of the top directory *lpd*:

```
daze@lion:~$ ls -R lpd
lpd:
myubu1 myubu2 myubu3 myubu4 ubuntu-lapack
lpd/myubu1: Dockerfile
lpd/myubu2: Dockerfile lapack-3.8.0.tar.gz
lpd/myubu3: Dockerfile dt_example_DGESV_rowmajor.c
Makefile
lpd/myubu4: Dockerfile lp_ex1.c Makefile
lpd/ubuntu-lapack: Dockerfile lapack-3.8.0z.tar.gz
daze@lion:~$
```

We use a separate directory for a separate Docker image; besides the corresponding *Dockerfile* that specifies how to build the image, each directory contains the required software or other files which will be embedded into the image. We create the following directories and build the following images which we can use separately:

- myubu1 — recent Linux Ubuntu and essential developer tools;
- myubu2 — adds to *myubu1* LAPACK installed;
- myubu3 — adds to *myubu2* BLAS, LAPACKE, and LAPACKE examples installed;
- myubu4 — runs a program that solves the example linear system using LAPACK;
- ubuntu-lapack — a solid Docker image corresponding to *myubu3* (Fig. 3).

myubu3:	ubuntu-lapack:
LAPACKE	LAPACKE
myubu2:	LAPACK
LAPACK	Essential developer
myubu1:	Ubuntu
Essential developer	
Ubuntu	

Table 1 contains *scripts* and *Dockerfiles* to create the specified sequence of images. We can either run a script from file or type the corresponding commands in an explicit way as an input for Linux shell program.

Image *myubu1* starts from the latest image of *ubuntu* and adds such essential developer tools as basic compilers and *make* utility. We can use *myubu1* separately to develop programs in C, C++, Gfortran, and Python.

Image *myubu2* starts from the *myubu1* image and installs LAPACK on it. We can use *myubu2* separately to develop programs in Gfortran which call LAPACK functions or run LAPACK tests written in Gfortran.

Table 1. Commands and files to build and run images for benchmarks.

Image	File	Commands
<i>myubu1</i>	<i>Script</i>	docker image build -t myubu1 . docker run -it myubu1
	<i>Dockerfile</i>	FROM ubuntu:latest RUN apt-get update && apt-get install -y \ build-essential \ gfortran \ python RUN /bin/bash
<i>myubu2</i>	<i>Script</i>	docker image build -t myubu2 . docker run -it myubu2
	<i>Dockerfile</i>	FROM myubu1:latest COPY lapack-3.8.0.tar.gz . RUN tar -zxvf lapack-3.8.0.tar.gz RUN cp /lapack-3.8.0/make.inc.example /lapack-3.8.0/make.inc RUN ulimit -s unlimited && cd /lapack-3.8.0 && make RUN /bin/bash
<i>myubu3</i>	<i>Script</i>	docker image build -t myubu3 . docker run -it myubu3
	<i>Dockerfile</i>	FROM myubu2:latest COPY dt_example_DGESV_rowmajor.c /lapack- 3.8.0/LAPACKE/example COPY Makefile /lapack-3.8.0/LAPACKE/example RUN cd /lapack-3.8.0/CBLAS && make && cd /lapack-3.8.0/LAPACKE && make && cd /lapack- 3.8.0/LAPACKE/example && make RUN /bin/bash
<i>myubu4</i>	<i>Script</i>	docker image build -t myubu4 . docker run myubu4
	<i>Dockerfile</i>	FROM myubu3:latest

Image	File	Commands
		<pre> RUN mkdir /lapack-3.8.0/LAPACKE/myex COPY lp_ex1.c /lapack-3.8.0/LAPACKE/myex COPY Makefile /lapack-3.8.0/LAPACKE/myex RUN cd /lapack-3.8.0/LAPACKE/myex && make CMD /lapack-3.8.0/LAPACKE/myex/lp_ex1 </pre>
	<i>Script</i>	<pre> docker image build -t ubuntu-lapack . docker run -it ubuntu-lapack </pre>
<i>ubuntu-lapack</i>	<i>Dockerfile</i>	<pre> FROM ubuntu:latest COPY lapack-3.8.0z.tar.gz . RUN apt-get update && apt-get install -y apt-utils build-essential gfortran python && \ tar -zxvf lapack-3.8.0z.tar.gz && \ cp /lapack-3.8.0/make.inc.example /lapack-3.8.0/make.inc && \ ulimit -s unlimited && cd /lapack-3.8.0 && make && \ cd /lapack-3.8.0/CBLAS && make && cd /lapack-3.8.0/LAPACKE && make && cd /lapack- 3.8.0/LAPACKE/example && make RUN /bin/bash </pre>

Image *myubu3* starts from the *myubu2* image and installs BLAS, LAPACKE, and LAPACKE examples on it. We can use *myubu3* separately to develop programs in C which call LAPACKE functions or run LAPACKE tests written in C.

Image *myubu4* starts from the *myubu3* image and installs the example program *lp_ex1.c* and the corresponding *Makefile*. We use *myubu4* to solve the example linear system within Docker. When running image *myubu4* in Docker, the results coincide with the results obtained using LAPACK directly.

A solid image *ubuntu-lapack* contains all the tools and libraries installed in the sequence of images *myubu1*, *myubu2*, *myubu3* according to Fig. 3. It behaves in the same manner as *myubu3* though Docker manuals claim that a solid image runs faster. We will run this image on various platforms for the benchmarks. We push the final image *ubuntu-lapack* to our repository at <http://hub.docker.com> to make it public and use in Docker for Windows benchmarks:

```

docker tag ubuntu-lapack zsoftua/ubuntu-lapack
docker push zsoftua/ubuntu-lapack

```

When required, we can pull it from the repository:

```

docker pull zsoftua/ubuntu-lapack

```

We assume that in both cases above we are logged to a repository otherwise we can add a prefix with repository address to the image name to the left.

5 LAPACK-Docker Benchmarks in Linux, Windows, and MacOS

For the performance benchmarks, big systems are solved using LAPACKE, the output is redirected to NULL device. In the simple case, we can measure the running time using system utility *time* as follows:

```
daze@lion:~/lapack-3.8.0/LAPACKE/example$ time
./xexample_DGESV_rowmajor -n 5000 > /dev/null
real 0m38,057s
user 0m37,964s
sys 0m0,092s
daze@lion:~/lapack-3.8.0/LAPACKE/example$
```

Thus, the program running time for system of size 5000 is about 38,057 seconds. For more precise evaluation, test program *xexample_DGESV_rowmajor* is modified by adding code for measuring time and commenting all printing operators save error messages, we call the resulting program *dt_xexample_DGESV_rowmajor*:

```
daze@lion:~/lapack-3.8.0/LAPACKE/example$ time
./dt_xexample_DGESV_rowmajor -n 5000 > /dev/null
1 5000 25.868977s
real 0m26,654s
user 0m26,090s
sys 0m0,080s
daze@lion:~/lapack-3.8.0/LAPACKE/example$
```

The times obtained inside the program and by system utility *time* are very close with the difference less than 1 second. The essential difference with the previous listing is explained by the fact, the printing is commented in program *dt_xexample_DGESV_rowmajor*, thus we conclude that printing consumes about one third of time for a system of size 5000. Further we will use benchmark tests without printing results. As for the time measuring code, the following sketch program illustrates it:

```
#include <time.h>
#include <bits/time.h>
#include <sys/time.h>
double magma_wtime( void )
{
    struct timeval t;
    gettimeofday( &t, NULL );
    return t.tv_sec + t.tv_usec*1e-6;
}
```

```

...
double t1,t2;
t1=magma_wtime();
info = LAPACKE_dgesv(...);
t2=magma_wtime();
...
fprintf( stderr, "%d\t%d\t%d\t%fs\n", nrhs, n, t2-t1 );

```

To organize tests in a sequence, we compose the following tiny script:

```

for n in <list of time instants>;
do
    ./dt_xexample_DGESV_rowmajor -n $n;
done

```

For our benchmark tests, we use the same desktop computer Hare as in [14]: Intel Core i5 3.2GHz, 4 cores, RAM 8Gb. We compare the performance obtained directly in Linux with the performance obtained in Docker environment which runs both on Linux and MS Windows 10, besides we compare solid and many-layered images.

For instance, we run tests directly in Linux with:

```

daze@lion:~/lapack-3.8.0/LAPACKE/example$ for n in 1000
2000 3000 4000 5000 6000 7000 8000 9000 10000 11000 12000
13000 14000 15000; do ./dt_xexample_DGESV_rowmajor -n $n;
done

```

and we run tests in Docker with:

```

daze@lion:~/lpd/ubuntu-lapack$ docker run -it myubu3
root@24e7e0f0f7c0:/# cd /lapack-3.8.0/LAPACKE/example
root@24e7e0f0f7c0:/lapack-3.8.0/LAPACKE/example# for n in
1000 2000 3000 4000 5000 6000 7000 8000 9000 10000 11000
12000 13000 14000 15000; do ./dt_xexample_DGESV_rowmajor
-n $n; done

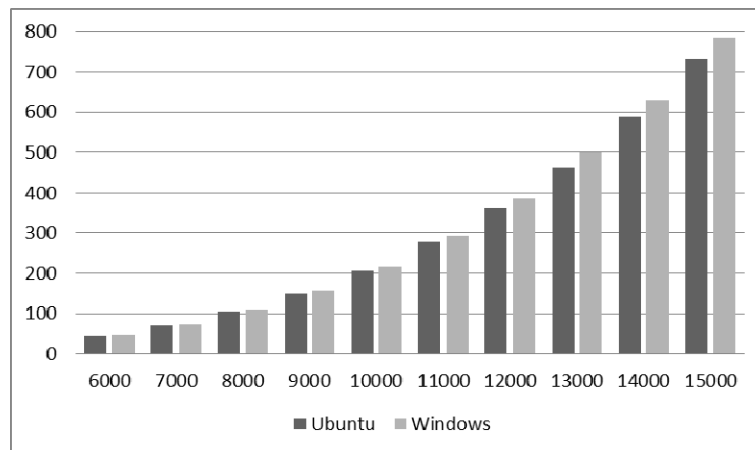
```

Basic obtained benchmarks are represented in Table 2. Matrix sizes from 1000 to 15000 are considered. We have four columns comparing a direct run of LAPACK on Ubuntu with running it within a Docker container either as multilayer or solid image, and finally, with running the solid Docker container within MS Windows 10. A brief vivid comparison is shown in Fig. 4.

As for the chosen range of the system size, it is limited by the RAM size when LAPACK goes out of memory. And as for Docker for Windows, the range depends on the amount of RAM allotted to Docker Engine. When Docker starts using virtual memory, the performance slows down considerably (Fig. 5).

Table 2. Comparing benchmarks of Docker using LAPACK.

Matrix size	LAPACK on Ubuntu (s)	Docker multi-layer image in Ubuntu (s)	Docker solid image in Ubuntu (s)	Docker solid image in MS Windows (s)
1000	0.230743	0.231323	0.233148	0.248542
2000	1.731192	1.729169	1.731458	1.814025
3000	5.706029	5.724043	5.785270	5.987914
4000	13.272057	13.306907	13.324464	13.912880
5000	25.860373	26.286293	25.965536	27.209460
6000	44.290466	44.523421	44.483824	46.647734
7000	70.136287	70.723309	70.605154	73.950816
8000	104.262183	105.543180	105.438383	110.346505
9000	148.151378	150.613408	150.461304	157.327095
10000	202.458930	206.776427	207.061478	217.058503
11000	271.838308	277.929413	278.195353	293.229381
12000	357.815169	361.973850	362.008640	387.046685
13000	462.599758	464.142452	463.041850	499.041814
14000	586.326644	586.678216	587.664414	631.014162
15000	729.811495	733.366776	731.759509	786.755366

**Fig. 4.** Comparing Docker performance in Linux (Ubuntu) and Windows.

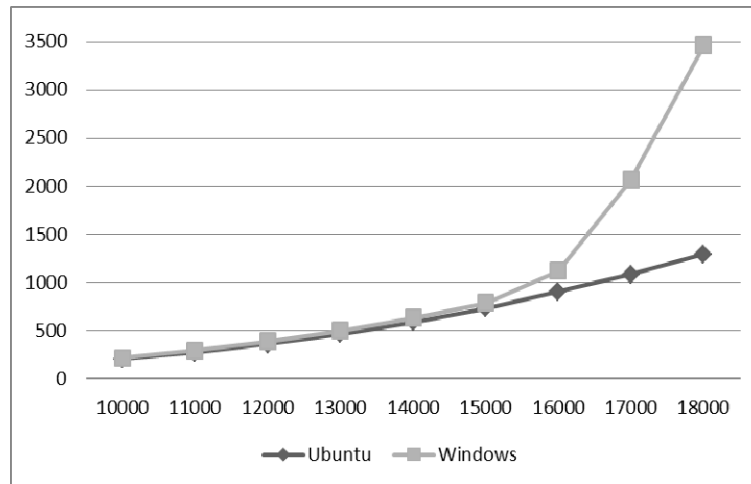


Fig. 5. Docker on Windows performance fall at exceeding RAM allocated to Docker Engine.

6 Final Discussions and Conclusions

Thus, using Docker yields very little slowdown (about one percent) in Ubuntu and little slowdown (about ten percent) in Windows that acknowledges that Docker platform is a perfect solution from performance point of view as well.

We should mention that Docker performance in Windows is considerably limited by the amount of RAM allocated to Docker Engine, a slowdown observed when active swapping within virtual memory starts.

Note that, the benchmarks have been collected for computations over real numbers. Recently computations over integer numbers become more significant for manifold applications to discrete event systems [15], fuzzy logic [16], cybersecurity, and artificial intelligence domains. For benchmarks over integer numbers, we can apply packet ParAd [14,17] recently developed based on clans composition theory [18].

7 References

1. Antonopoulos, N., Gillam, L. (ed.): Cloud Computing: Principles, Systems and Applications. Springer (2017).
2. Docker <https://www.docker.com/>
3. Ahmed, A., Pierre, G.: Docker Image Sharing in Distributed Fog Infrastructures. In: 2019 IEEE International Conference on Cloud Computing Technology and Science (CloudCom), Sydney, Australia, pp. 135-142 (2019). doi: 10.1109/CloudCom.2019.00030
4. Leon, D.: A Lightweight Container Middleware for Edge Cloud Architectures. In: Buyya, R., Srirama, S.N. (ed.): Fog and Edge Computing: Principles and Paradigm, Wiley, pp.145-170. (2019).

5. Zhang, P., Zhou, M., Wang, X.: An Intelligent Optimization Method for Optimal Virtual Machine Allocation in Cloud Data Centers. *IEEE Transactions on Automation Science and Engineering*. pp. 1-11 (2020). doi: 10.1109/TASE.2020.2975225
6. Lingayat, A., Badre, R.R., Gupta, A.K.: Performance Evaluation for Deploying Docker Containers On Baremetal and Virtual Machine. In: 2018 3rd International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, pp. 1019-1023 (2018).
7. Casalicchio, E., Perciballi, V.: Measuring Docker Performance: What a Mess!!! In: Proceedings of the 8th ACM/SPEC on International Conference on Performance Engineering Companion (ICPE). Association for Computing Machinery, New York, NY, USA, (2017).
8. LAPACK. <http://www.netlib.org/lapack/>
9. Angerson, E., et al: LAPACK: A portable linear algebra library for high-performance computers. In: Supercomputing '90: Proceedings of the 1990 ACM/IEEE Conference on Supercomputing, New York, NY, USA, pp. 2-11 (1990).
10. Top500. <http://top500.org>
11. Xie, G., and Xiao, Y.: How to Benchmark Supercomputers. In: 14th International Symposium on Distributed Computing and Applications for Business Engineering and Science (DCABES), Guiyang, pp. 364-367 (2015).
12. Stewart, G.W.: The decompositional approach to matrix computation. *Computing in Science & Engineering*. 2(1): 50-59 (2000).
13. Dongarra, J., Luszczek, P., Petitet, A.: The LINPACK Benchmark: Past, Present, and Future. *Concurrency and Computation: Practice and Experience*, 15(9): 803-820 (2003).
14. Zaitsev, D.A., Tomov, S., Dongarra, J.: Solving Linear Diophantine Systems on Parallel Architectures, *IEEE Transactions on Parallel and Distributed Systems*. 30(5): 1158–1169 (2019). doi: 10.1109/TPDS.2018.2873354
15. Zaitsev D.A.: Verification of Computing Grids with Special Edge Conditions by Infinite Petri Nets. *Automatic Control and Computer Sciences*. 47(7): 403–412 (2013). doi: 10.3103/S0146411613070262
16. Zaitsev, D.A., Sarbei, V.G., Sleptsov, A.I.: Synthesis of continuous-valued logic functions defined in tabular form. *Cybernetics and Systems Analysis*. 34(2): 190-195 (1998). doi: 10.1007/BF02742068
17. ParAd. <http://github.com/dazeorgacm/ParAd>
18. Zaitsev, D.A.: Sequential composition of linear systems' clans. *Information Sciences*. 363: 292–307 (2016). doi: 10.1016/j.ins.2016.02.016

A model of semantic web service in a distributed computer system

Nataliia Kulykovska¹[0000-0003-4691-5102], Stepan Skrupsky²[0000-0002-9437-9095],
Tetiana Diachuk³[0000-0002-2478-0588]

¹ "Zaporizhzhia Polytechnic" National University, Zhukovsky str., 64, Zaporizhzhia, 69063,
Ukraine

natalya.gontar@gmail.com

² "Zaporizhzhia Polytechnic" National University, Zhukovsky str., 64, Zaporizhzhia, 69063,
Ukraine

sskrupsky@gmail.com

³ "Zaporizhzhia Polytechnic" National University, Zhukovsky str., 64, Zaporizhzhia, 69063,
Ukraine

diacht@gmail.com

Abstract. One of the main tasks of any distributed computer system is the analysis of the properties of the obtained data, and their further use for logical reasoning. This is a difficult task due to the frequently encountered dynamics and heterogeneity of information. The aim of the work is to present a knowledge-based web service model. The web service integrates with the technologies of the semantic web in order to study the data, their processing and inference. The goal is achieved as follows: the meta-metamodel of the web service ontology and the metamodel of the semantic web service are structured using UML class diagrams and formalized using ALC description logic. The scientific novelty of the results is the presentation of a service description through an ontology. Through the introduction of ontologies, a transition is being made in data mining. In contrast to the application of standards: SOAP, WSDL, HTTP, the use of ontologies is argued that they contain structured information about the functional and non-functional characteristics of the service, and give flexibility to the description. When working with ontologies, there is a powerful logical apparatus for searching, combining and comparing ontologies. By introducing ontologies, knowledge of the web service is represented. The development of an ontology / web service pair gives an advantage over the SAWSDL, OWL-S standards, because the web service developer has the ability to create a more accurate semantic description of the service. We presented the language features and constructs of ontology of web service descriptions. Step by step, we created semantic description to describe web services. After learning how to use OWL to formally express web service semantics, we moved on to the issue of actually adding semantics to service descriptions. For the first time, formal models of service ontology and semantic web service are proposed.

Keywords: semantic web service, service oriented architecture, distributed systems, model, ontology, description logic

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

1 Introduction

Distributed computer systems (DS) based on a service-oriented architecture (SOA) are independent of development technologies and platforms, while applications running on one platform can call services running on other platforms in a standard way [1]. The main burden of performing computational operations with such an architecture rests with web services (WS) that solve all the problems of modeling the designed systems; client applications have only the simplest functions for preparing data and displaying simulation results.

Since WS can operate at a higher level of abstraction, analyzing and processing data types in a dynamic way, the individual software components are given the opportunity to interact more openly [2]. When using the universally described interfaces, it becomes possible to reuse software components, which reduces the complexity of the development of DS and correctly accumulate the data.

SOA is a heterogeneous environment, and it is developing towards creating a more structured set of solutions where WS should be represented in some unified way and equally detectable, able to communicate with other objects, and also be directly integrated with the Internet infrastructure and other services, independently from the functionality of the service [3].

Using semantic web technologies in the development of WS in the DS allows you to:

- go to the intellectualization of the description of XML-based interfaces and interactions, combining any type of application with another application, and providing freedom of change and development over time as long as the corresponding interface is supported;
- use a higher level of software abstraction and data description logic;
- take into account the weakness of the software, due to which the interaction between the applications of the service is not broken every time the design or implementation of a service changes;
- provide existing or legacy software service interface without changing the original applications;
- make decisions based on available data.

Key idea of semantic web services is to annotate web services with concepts which are defined in formal logic-based ontologies such that, from an artificial intelligence perspective, intelligent agents and service-based applications can actually reason on such formal service semantics. In contrast to web service descriptions, this may facilitate not only the semantic interoperability between services but their automated logic-based composition planning and a more precise service search [4].

The aim of the work is to present a knowledge-based web service model. The WS integrates with the technologies of the semantic web in order to study the data, their processing and inference. The scientific novelty of the results is the presentation of a service description through an ontology. Through the introduction of ontologies, a transition is being made in data mining.

2 Literature review

SOA is an approach for designing and developing DS [5]. By this approach, design, development, and implementation of DS are possible due to the web technology [6]. However, interoperability of services in SOA is not limited to web services [7, 8], the WS are the most suitable technology for successful SOA [9]. In DS based on SOA, information source and business functions can be converted into modular services units for control and management [10].

Web services are modular, self-describing, self-contained applications that are accessible over the Internet [11]. WS is a software component invocable over the web via an XML [12] message that follows the SOAP [13] standard. The component provides one or more operations for performing useful actions on behalf of the invoking client. These operations and the formats of the input and output messages are described using WSDL [14].

Description of services in a language neutral manner is vital for the widespread use of WS. For general usability, a service must be described and advertised. WSDL takes care of the description by providing a language to describe a service in enough detail to invoke any of its operations. Service providers describe their WS and advertise them in a universal registry called UDDI [15, 16]. This enables service requestors to search the registry and find services, which match their requirements. UDDI allows for the creation of registries that are accessible over the Web. A registry contains content from the WSDL descriptions as well as additional information such as data about the provider. Clients may use one or more registries to discover relevant services.

Prominent languages and formats for semantic service description are OWL-S (Web Ontology Language for Web Services) [17, 18, 19], WSML (Web Service Modeling Language) [20], the W3C standard SAWSDL (Semantic Annotations for WSDL and XML Schema) [21], USDL (Unified Service Description Language) [22, 23], Linked USDL [24], as well as the microformats hRESTS [25], SA-REST [26], and MicroWSMO. These description models mainly differ in their formal logic-based foundation and the possible extent of service annotation [27, 28, 29].

3 SOA

The SOA information model is registered with UDDI. Further, the client of the service can find it, call it, perform a certain task. The standards for exchanging information between services are SOAP, WSDL, UDDI, and a normal HTTP request. The SOAP and WSDL protocols provide a unified markup language for transmitted messages. It represents the functionality of a device connected to its managed resource. A user interface that provides a clear and standardized interface offers all the necessary functionalities for interacting with objects and related processes.

In order for SOA to enjoy greater success than its predecessors, it should consider the following attributes:

- scalable: The past solutions were not designed with the scale of the web in mind. SOA should work in a variety of settings, such as within an organization, between business partners and across the world [30];
- loosely-coupled: SOA is an evolution from tightly coupled systems to loosely coupled ones. Senders and receivers of a SOA should be independent of each other; the source can send the message independently of the target. Tight coupling is not suitable for SOA since it leads to monolithic and brittle distributed applications. Even trivial changes in one component lead to catastrophic breaks in function. Small changes in one application require matching changes in partner applications [31];
- interoperability: One party should be able to communicate with another party regardless of the machine they are running on;
- discovery: One party should be able to communicate with a second party selected from a set of competent candidates. Services need to be dynamically discoverable. This is accomplished through services such as a directory of service descriptions;
- abstraction: A SOA abstracts the underlying technology. Developers can concentrate on building services for business users rather than connecting systems and applications [32];
- standards: Interaction protocols must be standardized to ensure the widest interoperability among unrelated institutions. Contracts should also be standardized. Explicit contracts define what may be changed in an application without breaking the interaction. Furthermore, standards are the basis of interoperable contract selection and execution.

The current description of the interaction between WS is as follows [33]:

1. WS is created using some programming language.
2. WS is described by a WSDL document (normally generated by using a provided tool or by the built-in support of the development environment).
3. The service provider publishes the WS into the UDDI repository.
4. A web server hosts this WS by listening to HTTP traffic.
5. A client application (probably written in another programming language) searches the UDDI registry and discovers this service.
6. The client accesses the WSDL document, and a SOAP request message is generated based on the WSDL document.
7. The web server receives the SOAP request as part of a HTTP POST request, and it forwards this request to a WS request handler (a system-level application that is always running).
8. The WS request handler parses the SOAP message, invokes the right WS, and also creates the SOAP response. It finally sends the response to the web server.
9. The web server formulates a HTTP response, which includes the SOAP response, message and sends it back to the client.

4 Distributed knowledge-based systems

The semantic web should enable greater access not only to content but also to services on the web. Users and software agents should be able to discover, invoke, compose, and monitor web resources offering particular services and having particular properties, and should be able to do so with a high degree of automation if desired. Powerful tools should be enabled by service descriptions, across the web service lifecycle [34].

The semantic web concept introduces formal definitions called ontologies, which allows you to build models of heterogeneous objects in a domain, share knowledge and support the automation of the formulation of logical conclusions from this knowledge.

To create a holistic system for deploying distributed knowledge-based systems (DKBS) semantic web services in a specific field of activity, it is necessary to develop a formal ontological model of objects (web services) that are part of SOA.

By combining the key elements of SOA with knowledge-based systems, you can get a formal DKBS model. Figure 1 shows the structural model of DKBS. We single out the concept of the main artifact of the system through the Element class, its subclasses are such DKBS elements as a service, system, event, human actor, semantic service [35, 36].

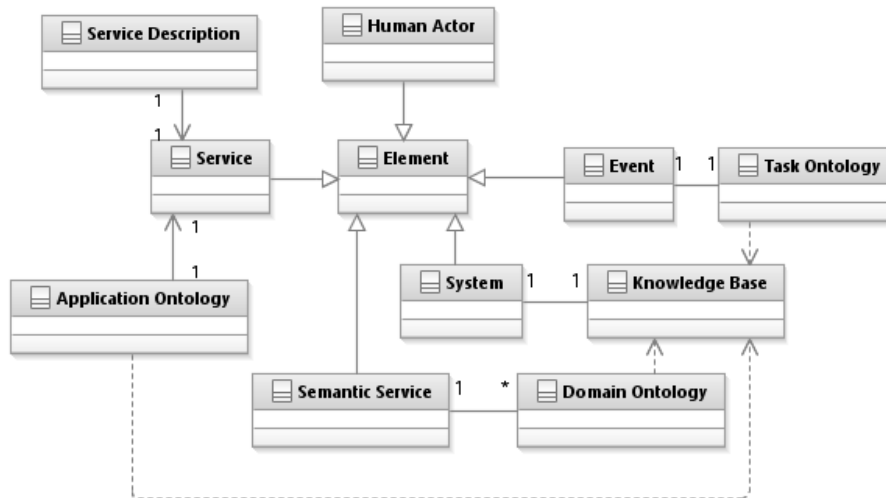


Fig. 1. Structural model of DKBS

The diagram shows that the service class consists of three elements: service description, application ontology, service interface. The semantic principle is reflected through such elements: domain ontology, application ontology, task ontology and knowledge base [37, 38].

A formal semantic service model should include:

1. Analysis of the structure of knowledge of the subject area of the DKBS, the main objects and relations:
 - (a) the use of least-logical and ontological methods for the formation of the terminology of the system domain and knowledge structure;
 - (b) structure of OWL ontology.
2. Repeated use of existing thesauruses, taxonomies, and ontologies of the DKBS domain:
 - (a) semantic search for relevant objects and analysis of means of representing knowledge and standards in the DKBS;
 - (b) a brief overview of relevant ontologies and other knowledge structures;
 - (c) integration of existing taxonomies and domain ontologies.
3. The architecture of methods for the automated improvement of the formal ontological model of the DKBS:
 - (a) the architecture of methods for the automated extraction of knowledge (terms and relationships) from natural language texts that relate to the DKBS;
 - (b) methods of automated linguistic processing of natural language texts;
 - (c) advanced OWL ontology.
4. Semantic search in the system based on the domain ontology:
 - (a) semantic search for objects;
 - (b) methods of semantic search for objects of the DKBS;
 - (c) methods of semantic search for SOA services;
 - (d) recommendations regarding the use of WS.

Search data for web services, their interactions, reviews, recommendations from service customers can be analyzed and converted into active knowledge, which allows us to better understand the physical world and create more value-added products and services.

DKBS is a synthesis of SOA and semantic technologies, which defines a service-oriented presentation of software and hardware components and a description of their formal semantics. Table 1 shows the architecture of the meta-modeling of the DKBS.

Table 1. The architecture of meta-modeling

Meta level	Simulation level
M3	Meta-metamodel / Meta-metamodel of service ontology
M2	Metamodel / Meta-model of service
M1	Data / Service
M0	Model / DKBS

The following should distinguish the basic concepts:

A web service is a software system identified by a unique web address (URL) with standardized interfaces.

Semantic web service (SWS) - complete elements of program logic with uniquely described semantics, accessible via the Internet and suitable for automated search, composition and execution, taking into account their semantics.

5 Semantic web service

A semantic meta-modeling of a WS is proposed (a meta-metamodel of a web service ontology, a meta-model of a WS) Thus, a new system object is formed - a SWS. In contrast to the application of standards: SOAP, WSDL, HTTP, the use of ontologies is argued that they contain structured information about the functional and non-functional characteristics of the service, and give flexibility to the description. When working with ontologies, there is a powerful logical apparatus for searching, combining and comparing ontologies. By introducing ontologies, knowledge is represented in the system. Developing an ontology / web service pair provides an advantage over SAWSDL, OWL-S, because a web service developer can create a more accurate description.

The meta-metamodel of the service ontology is shown in Figure 2, where Concepts are shown, i.e. units of information or data. Each concept has its own Logical Definition and a Textual Definition. In accordance with the concept, there may exist Instances, which, through the presentation of specific entities, connect abstract concepts with objects from the real world. In the same way, each Property of a concept has different Term - these are interpretation functions.

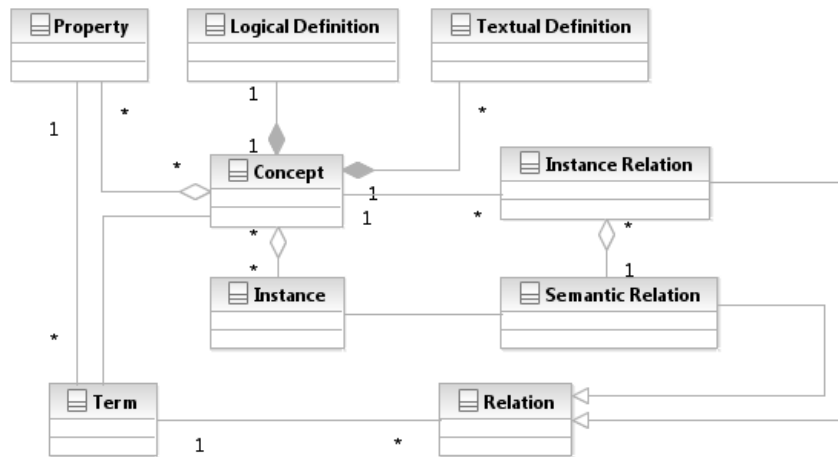


Fig. 2. Meta-metamodel of service ontology

Therefore, the meta-model of the service is its ontology, which describes all the characteristics of the service and its parameters (Fig. 3). A service primarily consists of Methods and Operations. The next component of the service is its Interface, which depends on already defined methods and operations. Binding is a component that provides the detailed information needed to access EndPoint. EndPoint is a specific service implementation where EndPoint associates a specific binding with a real address so that the service can be called.

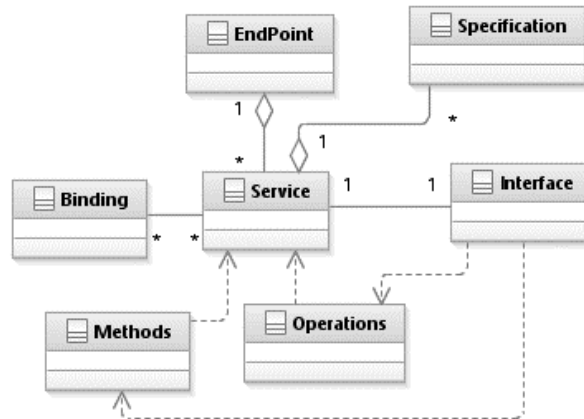


Fig. 3. Meta-model of the service

It is also important to understand the obvious difference between OWL and OWLS: OWL is an ontology language; it provides constructs and features we can use to create an ontology document. OWL-S, on the other hand, is just one such ontology created by using OWL.

Ontology is based on description logic (OWL-DL), which provides a platform for formal and machine-knowledge. To establish its syntax and semantics. Syntax, certain expressions (concepts, axioms, roles, etc.) are considered correctly constructed in this logic. How to interpret these expressions, i.e. gives them formal meaning.

Let $CN = \{A_1, \dots, A_m\}$ and $RN = \{R_1, \dots, R_n\}$ be finite nonempty sets of atomic concepts and atomic roles (also called concept names and role names). There are several dialects of descriptive logic that differ in expressive capabilities. The main operators of DL:

- $\top$ – concept «THING» ;
- $\perp$ – concept «NOTHING» ;
- $\cap$ – logical connective (conjunction);
- $\exists$ – the quantifier of existence;
- $\forall$ – quantifier value limitation;
- $\neg$ – logical connective (negation).

The syntax of DL, namely the ALC dialect, is a lot of concepts that are defined by an inductive definition:

- symbols $\top$ and $\perp$ - concepts (called truth and falsehood);
- every atomic concept A is a concept;
- if C is a concept, then $\neg C$ is a concept (called a complement to the concept);

- - if C and D are concepts, then $C \cap D$ and $C \cup D$ are concepts (intersection and union);
- - if C is a concept, and R is an atomic role, then $\exists R.C$ and $\forall R.C$ are concepts;
- - no other expressions are concepts.

In what follows, we will use a shorter notation to formulate the syntax. So, the syntax for the ALC logic concepts in this entry is as follows:

$$\perp \mid \top \mid \neg C \mid C \cap D \mid C \cup D \mid \exists R.C \mid \forall R.C \quad (1)$$

where A is an atomic concept, R is an atomic role, D are arbitrary concepts.

A complete description of the meta-metamodel of the service ontology in the formulations of the discription logic with the class diagram (Fig. 2) is given in Table 2.

Table 2. Meta-metamodel of service ontology in formulations of syntax description logic

Concept	Role
Property	isPartOf
LogicalDefinition	hasInstance
TextualDefinition	hasRelation
Concept	hasProperty
InstanceRelation	hasDefinition
Instance	hasObject
SemanticRelation	
Relation	
Term	

In this case, the values of the Concept column are atomic concepts, and the values of the Role column are atomic roles, then the concepts of ALC logic will be expressions:

$$\begin{aligned} & \text{Concept} \cap \text{Relation} \mid \text{Concept} \cap \text{Term} \mid \\ & \forall \text{hasPartOf}.\text{Concept} \mid \text{Property} \cap \exists \text{hasProperty}.\text{Concept} \end{aligned} \quad (2)$$

The semantics of logic is defined using the concept of interpretation. Interpretation is a pair $I = (\Delta, \cdot^I)$, consisting of a nonempty set Δ , called the domain of this interpretation and the interpretive function $\cdot^I$, which is:

- each atomic concept $A \in CN$ is an arbitrary subset of $A^I \subseteq \Delta$;
- each atomic role $R \in RN$ is an arbitrary subset of $R^I \subseteq \Delta \times \Delta$.

For a service ontology, domain interpretation Δ will be the set of all ontology definitions. Atomic concepts are comparable to the set of all existing rules and terms for describing the ontology. We interpret atomic roles as two-place relations connecting all knowledge with a concept. Then everything listed in Formula 2 acquires the following semantics: a concept with this interpretation means:

- $\text{Concept} \cap \text{Relation}$ - a lot of concepts of interconnection (i.e., relationship);
- $\text{Concept} \cap \text{Term}$ - many concepts of terms (term);
- $\forall \text{hasPartOf}.\text{Concept}$ - many concepts that are part of this concept;
- $\text{Property} \cap \exists \text{hasProperty}.\text{Concept}$ - many properties that are property of the concept.

A full description of the semantic web service model in the formulations of the description logic is given in the table 2.

Table 3. Metamodel of SWS in formulations of syntax description logic

Concept		Role	
EndPoint	ServiceName	hasParameters	hasInterface
Specification	Definition	hasInstance	hasDescription
Binding	Parameters	hasInput	hasSpecification
Methods	InputParameters	hasOutput	endPoint
Operation	OutputParametes	hasResult	hasOperation
Interface	ContactInformation	hasPrecondition	hasMethods
Service	Value	hasCode	hasName
Taxonomy	Code	isPartOf	
Result	ServiceClassifacation	hasPelation	
Text	ServiceParametersName	hasObject	
Name			

Thus, the formation of ALC logic concepts for the SWS model is as follows:

$$\begin{aligned} & \text{Service} \cap \text{Methods} \mid \text{Definiton} \cap \text{Text} \mid \\ & \forall \text{hasParameters}.\text{Service} \mid \text{Value} \cap \exists \text{hasParameters}.\text{Service} \end{aligned} \quad (3)$$

The semantic interpretation of formula 3 is as follows:

- $\text{Service} \cap \text{Methods}$ - many service methods;
- $\text{Definiton} \cap \text{Text}$ - many text descriptions;
- $\forall \text{hasParameters}.\text{Service}$ - many concepts that are service parameters;
- $\text{Value} \cap \exists \text{hasParameters}.\text{Service}$ - many values that are service parameters.

Thus, a semantic web service is a pair of ontology and web service, which are described above by the presented models. Each formal model of descriptive logic is easily ported to OWL.

6 Processing knowledge of semantic web service

Formal models of ontology and semantic web service are aimed at supporting the process of automated deployment of intelligent applications in heterogeneous environments and allows:

- hide the technological heterogeneity that is characteristic of many heterogeneous web services;
- hide the semantic heterogeneity inherent in the used heterogeneous domain ontologies in order to semantically annotate the data of semantic web services.

The following is the process of finding the web services the user needs. In this case, the user can specify both the functions or properties of the service itself, as well as the characteristics associated with the end result of the functioning of the service.

At the same time, the user can apply the terms from any ontologies available to him that describe the DKBS space and the corresponding subject area (with reference to the ontologies themselves). But in the description of the semantic web service, terms from other ontologies related to close domains can be used.

Therefore, the problem of comparing ontologies arises. In the general case, this problem is extremely complex and laborious, but if there are a number of restrictions on the comparable ontologies, it can be solved in an acceptable time.

For the intellectualization of data and the processing of knowledge stored in ontologies, logical inference blocks (reasoners) are used, which play a key role in knowledge-oriented intelligent systems.

There are many implementations of logical inference processors (reasoning engine) for OWL ontologies that differ in capabilities, applications, and quality of task execution. The generalized analysis allows us to divide them into three groups depending on the implementation method:

1. Table DL-processors. Traditionally, they were developed first to solve such problems. They have low performance, but are able to make conclusions on complex ontologies with many non-trivial constructions. This class includes resonators Pellet, RacerPro, FacT ++, as well as Hermit and SHER [41, 42, 43].
2. Disjunctive Datalog processors. They transform the ontology into a disjunctive Datalog program and use the deductive database technique and the resolution rule. Such processors have satisfactory performance when using some optimizations, but do not support certain OWL-designs, in particular cardinal restrictions and ratings. This group belongs to KAON2. [44].
3. Rules processors. Use rule processing systems for inference on ontologies. They have high performance, but can only process simple ontologies, devoid of many

important structures. Representatives of this group: Sesame / OWLIM, Jena, Owl-jesskb. [45].

There are two approaches to the implementation of inference: based on the rules (using the algorithms forwardchaining and / or backward-chaining) and based on the semantic scoreboard (semantic tableau). Based on the rules, Semantics, SDK and Owl原因 are implemented, and on the basis of the semantic scoreboard - Pellet.

Today, the Pellet system implements the most expressive descriptive logic using a high-performance logic inference (tableau-based algorithm), which is used to process ontologies described in the OWL DL language.

7 Conclusion

We presented the language features and constructs of ontology of WS descriptions. Step by step, we created semantic description to describe SWS. For instance, given that WSDL is also used to describe WS, understanding the relationship between WSDL and OWL is important for Semantic Web developers. We first built two models (structural and formal) to describe the ontology and web service. Examples of semantic interpretation of each model are given. Presented existing tools for further work with the obtained ontologies.

The advantages of using ontologies for WS:

1. Automatic discovery of SWS: Finding the desired service can be hard, especially when the service requester does not know of the existence of the service provided. However, to make SWS a real success, a way to discover the requested service should be provided; also, it has to be discovered automatically, with great accuracy and efficiency.
2. Automatic composition of the necessary services. Clearly, a software agent should be able to find all the necessary services and invoke them in the correct order to accomplish the business goal.

References

1. Ruzhi Xu, Peiguang Lin, Cheng Liu: Research on Distributed Knowledge Base System Architecture for Knowledge Sharing of Virtual Organization. Atlantis Press, pp. 349-357 (2010)
2. Fensel et. Al.: Implementing Semantic Web Services, Springer-Verlag Berlin Heidelberg (2008)
3. Portier B.: COA terminology overview, Part 1: Service, architecture, governance, and business terms (2007)
4. Klush, Matthias; Kapahnke, Patrick; Schulte, Stefan; Lecue, Freddy; Bernstein, Abraham (2016). Semantic web service search: a brief survey. Künstliche Intelligenz (KI), 30(2):139-147. DOI: <https://doi.org/10.1007/s13218-015-0415-7>
5. Zeppenfeld Klaus, Patrick Finger: SOA und WebServices, Springer-Verlag Berlin Heidelberg (2009)

6. Kulykovska, N. A.; Timenko, A., V; Ilyashenko, M. B.; Kirichek, G. G.: Distributed Knowledge Base System. PROBLEMELE ENERGETICII REGIONALE vol. 1-1 (40): 79-90 (2019)
7. Allemang D, Hendler J.: Semantic web for the working ontologist modeling in RDF, RDFS and OWL, Elsevier Inc. (2008)
8. Demirkan. H., Kauffman. R. J., Vayghan. J.A., Fill. H.G., Karagiannis. D., Maglio. P.: Service-oriented technology and management: Perspectives on research and practice for the coming decade. Electronic Commerce Research and Applications 7: 356–376 (2008)
9. Erl,T.: Service-Oriented Architecture Concepts , Technology , and Design ,Prentice Hall: 2-14 (2005)
10. Jack C.P. Cheng, Kincho H. Law, Hans Bjornsson, Albert Jones c, Ram Sriram: A service oriented framework for construction supply chain integration , Automation in Construction 19: 245–260 (2010).
11. Rudi Studer, Stephan Grimm, Andreas Abecker: Semantic Web Services, Springer-Verlag Berlin Heidelberg (2007)
12. Jeff Friesen: Java XML and JSON, Apress (2016)
13. Helen Hye-Young Paik, Angel Lagares Lemos, Moshe Chai Barukh, Boualem Benatallah, Aarthi Natarajan: Web Service Implementation and Composition Techniques, Springer International Publishing (2017)
14. Liang-Jie Zhang: Web Services, Springer-Verlag Berlin Heidelberg (2004)
15. James McGovern, Oliver Sims,nAshish Jain, Mark Little: Enterprise Service Oriented Architectures, Springer Netherlands (2006)
16. Klaus Zeppenfeld, Patrick Finger: SOA und WebServices, Springer-Verlag Berlin Heidelberg (2009)
17. Guang Chen, Tonghai Jiang, Meng Wang, Xinyu Tang, Wenfei Ji: Modeling and reasoning of IoT architecture in semantic ontology dimension. Computer Communications Volume 153: 580-594 (2020)
18. Matthias Klusch, Patrick Kapahnke, Stefan Schulte, Freddy Lecue & Abraham Bernstein: Semantic Web Service Search: A Brief Survey, KI - Künstliche Intelligenz vol. 30: 139–147 (2016)
19. Martin, D.; Burstein, M.; Hobbs, J.; Lassila, O.; McDermott, D.; McIlraith, S.; Narayanan, S.; Paolucci, M.; Parsia B.; Payne, T.; Sirin, E.; Srinivasan, N.; Sycara, K.: OWL-S: Semantic Markup for Web Services. (2004) www.w3.org/Submission/OWL-S/
20. Fensel, D.; Lausen, H.; Polleres, A: Enabling Semantic Web Services, Springer (2010)
21. Farrell, J.; Lausen, H.: Semantic Annotations for WSDL and XML Schema. (2007) www.w3.org/TR/sawsdl/
22. Kona, S.; Bansal, A.; Simon, L.; Mallya, A.; Gupta, G.; Hite, T.D.: USDL: A Service-Semantics Description Language for Automatic Service Discovery and Composition, Web Services Research (2009)
23. Lampe, U.; Schulte, S.: Self-Adaptive Semantic Matchmaking Using COV4SWS.KOM and LOG4SWS.KOM. In: B. Blake et al. (eds.) Semantic Web Services, Chapter 9, Springer (2012)
24. Pedrinaci, C.; Leidig, T.: Linked USDL Core (2011) www.linked-usdl.org/ns/usdl-core
25. Klusch, M.: The S3 Contest: Performance Evaluation of Semantic Service Matchmakers. In: Blake, M.B.; Cabral, L.; Koenig-Ries, B.; Kuester, U.; Martin, D. (Eds.): Semantic Web Services: Advancement through Evaluation; Springer (2012)
26. Gomadam, K.; Ranabahu, A.; Sheth, A.: SA-REST: Semantic Annotation of Web Resources. (2010) www.w3.org/Submission/2010/SUBM-SA-REST20100405/
27. Klusch, M.: Semantic Web Service Coordination. In: [56], Chapter 4. (2008)

28. Baader F. (editor), et al.: The Description Logic Handbook, Cambridge University Press (2003)
29. Gandon F.: Ontology engineering: A Survey and a Return of experience, INSTITUT DE RECHERCHE EN INFORMATIQUE ET AUTOMATIQUE (2002)
30. Wooldridge, M.J., Jennings, N.R.: Intelligent Agents: Theory and Practice. Knowledge Engineering Review 10(2): 115–152 (1995)
31. Gontar N. A., Kudermetov R. K.: Working out ontology of systems engineering of space system, Radio Electronics, Computer Science, Control 2: 131-137 (2011)
32. Kirichek G., Harkusha V., Timenko A., Kulykovska N. System for detecting network anomalies using a hybrid of an uncontrolled and controlled neural network. CS&SE@SW 2019: 2nd Student Workshop on Computer Science and Software Engineering : 138 – 148 (2019)
33. Cardoso J, Hepp M., Lytras M. (Eds.): The Semantic Web. Real-World Applications from Industry(2008)
34. Cardoso Jorge and Amit P. Sheth.: Semantic Web Services, Processes and Applications (2006)
35. Kulykovska N.A, Timenko A.V.: A Structure of Semantic Service in a Distributed Knowledge Based System. Computer Modeling and Intelligent Systems: Proceedings of the Second International Workshop on Computer Modeling and Intelligent Systems (CMIS-2019), In: CEUR Workshop Proceedings 2353, 533-544 (2019)
36. Kirichek, G., Tymoshenko, V., Rudkovskyi, O., Hrushko, S.: Decentralized System for Run Services. In: Second International Workshop on Computer Modeling and Intelligent Systems (CMIS-2019). In: CEUR Workshop Proceedings 2353, 860–872 (2019)
37. Allemang D.:P Semantic web for the working ontologist. Modeling in RDF, RDFS and OWL, Morgan Kaufmann Publishers (2008).
38. Yu, Liyang.: Introduction to Semantic Web and Semantic Web services, Springer (2007)
39. Leuf, Bo.: Technology Analyst, Sweden. The Semantic Web: crafting infrastructure for agency (2006)
40. Medjahed B. and Bouguettaya A.: Service Composition for the Semantic Web, Springer Science+Business Media (2011)
41. Cary Pennington: Introduction to Web Services, Computer Science (2007) DOI:10.4018/978-1-59904-045-5.ch007
42. Kishore, Rajiv, Ramesh, Ram (Eds.): Ontologies. A Handbook of Principles, Concepts and Applications in Information Systems, Springer US (2007)
43. Wang, H., Horridge, M., Rector, A.L., Drummond, N., Seidenberg, J.: Debugging OWL-DL Ontologies: A Heuristic Approach. vol. 3729: 745–757 (2005)
44. Ebrahimipour, Vahid, Yacout, Soumaya (Eds.): Ontology Modeling in Physical Asset Integrity Management, Springer International Publishing (2015)
45. Tamma, V., Dragoni, M., Goncalves, R.S., Ławrynowicz, A. (Eds.): Ontology Engineering. 12th International Experiences and Directions Workshop on OWL, OWLED 2015 (2015)

Autonomous Distributed Systems and their Simulation

Valeriy Kolesnikov^[0000-0002-1991-3614]

Sumy State University, Sumy, Ukraine

v.kolesnikov@cs.sumdu.edu.ua

Abstract. The development and actual testing of autonomous distributed systems is difficult, long, and expensive. Simulators can save time, the development and testing efforts, as well as bring the overall costs down. This paper describes an approach to designing and testing autonomous distributed systems. The design of such systems is based on organization model. For testing such systems, we developed DiSy-Sim simulator. Specifically, DiSy-Sim simulator simplifies testing of communication between system components by providing statistics on messages being used in communication. We demonstrate the use of the simulator for a swarm system of autonomous drones used for disaster relief aid delivery. DiSy-Sim simulator is applicable for a variety of distributed systems.

Keywords: autonomous distributed systems, swarms, simulation, organization model, DiSy-Sim, software engineering

1 Introduction

Autonomous distributed systems have found their application in many areas. For example, such systems can enable autonomous drones to perform a task together as a team. A swarm of such drones can work towards a common goal. Each drone in a swarm is autonomous, but drones can communicate among themselves in a distributed manner.

A study of autonomous systems has become important especially in recent years as such systems can be used, for instance, for humanitarian aid and disaster relief, search and rescue and search and track of a target. To this end, swarm based autonomous systems should be self-directing, with reliable communication among its parts, resilient to various interferences and self-adjusting in the face of failures of some of its parts.

It is a difficult task to develop autonomous distributed systems that are robust, flexible and fault tolerant. This paper describes how such systems can be designed utilizing an organization model approach and tested with the help of DiSy-Sim simulator. These efforts are a part of a bigger ecosystem for design, development, verification, validation, and deployment of autonomous distributed systems that the author is working toward.

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

The approach used in this study is the continuation of our efforts to model autonomous distributed systems utilizing the organization model. This approach provides a framework for defining goals that the organization exists to achieve, entities to achieve these goals and relationships among the entities, and finds its motivation in the Multiagent Systems Engineering methodology (MaSE) [1,2] to derive goals from the system description, and then design roles, capabilities and agents to achieve the organization goals.

For testing an autonomous distributed system with DiSy-Sim simulator we utilize the example of a swarm of autonomous drones for disaster relief aid delivery.

The remainder of this paper is organized as following: in section 2, we describe related work to designing and simulating autonomous distributed systems; next, we present an organization model approach to design autonomous distributed systems; we follow with an example of a swarm of autonomous drones for disaster relief aid delivery and present an organization instance for this example; next, we present DiSy-Sim simulator and discuss its application to the example of a swarm of autonomous drones for disaster relief aid delivery; finally, we end with conclusion and directions for our future work.

2 Related Work

We chose designing and implementing autonomous distributed systems using organization-based approach because research shows that: 1) organizational design is easier to understand as it is close to human organizations [3], 2) organization-based systems tend to be efficient and can efficiently adapt to changing conditions [3,4], 3) organizations are flexible [3,4], and 4) best designs utilizing organization-based approach are application and situation dependent [5]. Organizations can be centralized or decentralized. In a distributed scenario, decentralized organizations tend to exhibit higher performance and reliability [1].

For testing purposes, there are many simulators available that range from general in purpose to specific contexts. A performance comparison of recent network simulators is given in [6]. In our previous research [7-9] we studied and used NS2, NS3 [10] and J-Sim [11] simulators extensively.

NS3 is a discrete-event network simulator for internet systems. It is built using C++ and Python and provides scripting capability. The following are the key strong features of NS3: topology definition, model development, node and link configuration, execution environment with logging capability, performance analysis and graphical visualization. Four main drawbacks for us in using NS2 or NS3 are: 1) these are network simulators (they don't capture application layer communications), 2) they cannot simulate the movement of application entities over time, 3) they are written in C++ and Python (the language of choice for our ecosystem for design, development, verification, validation and deployment of autonomous distributed systems is Java) and 4) we want our simulator to be integrated in our ecosystem briefly described below.

J-Sim simulator solves two of the issues. First, it is written in Java. Also, it can be used in almost any system, in which object states change discretely and therefore capture application layer communications. The basic building blocks of J-Sim simulations are processes and queues. Whereas processes are active, queues and other elements of the simulation are passive. The simulation is executed step by step. During each step, only one process is given an opportunity to run. J-Sim provides a rich set of features, such as simulation of Dijkstra semaphores, simulation of message passing (blocking and non-blocking send and receive operations, symmetric, asymmetric, and indirect communication), independent random number generators that can be initialized by user-defined seed and therefore guarantee repeatability of experiments and console mode and batch and interactive GUI modes [11]. Still, J-Sim cannot simulate movements of application entities over time. Integration with InDiGO framework remains the main hurdle. Therefore, we opted to create our own simulator and add additional features valuable for InDiGO framework and for our ecosystem in general.

Of interest for us is a framework or a whole ecosystem for the ease of developing, testing, and deploying autonomous distributed systems. Several attempts have been done at creating and formalizing systems or frameworks to capture the concept of teamwork within an organization [12-16]. Although valuable for design of autonomous distributed systems, they fail to provide a wide accepted use.

In our previous work we proposed InDiGO, an infrastructure for the development of distributed systems [7]. InDiGO allows design of generic but customizable algorithms and provides tools to customize such algorithms for distributed applications [7,17]. We demonstrated advantages of using such a framework for distributed systems development [7-9] and plan to merge this research into InDiGO framework.

3 Organization Model

To ease the design of autonomous systems, we propose an organization model (O). This model describes the structure of the organization and combines all entities of the autonomous system and the relationships among them. Fig. 1 shows organization model using standard UML notations.

Organization model O contains a set of goals (G) that the system is trying to achieve, a set of roles (R) that achieve the stated goals, a set of agents (AG) that are capable of playing specific roles, a set of capabilities (C) that agents possess and that are required by the roles to achieve the goals, a set of assignments (A) that the organization performs to assign agents to play specific roles to achieve specific goals and a set of policies (P) that constrain the assignments.

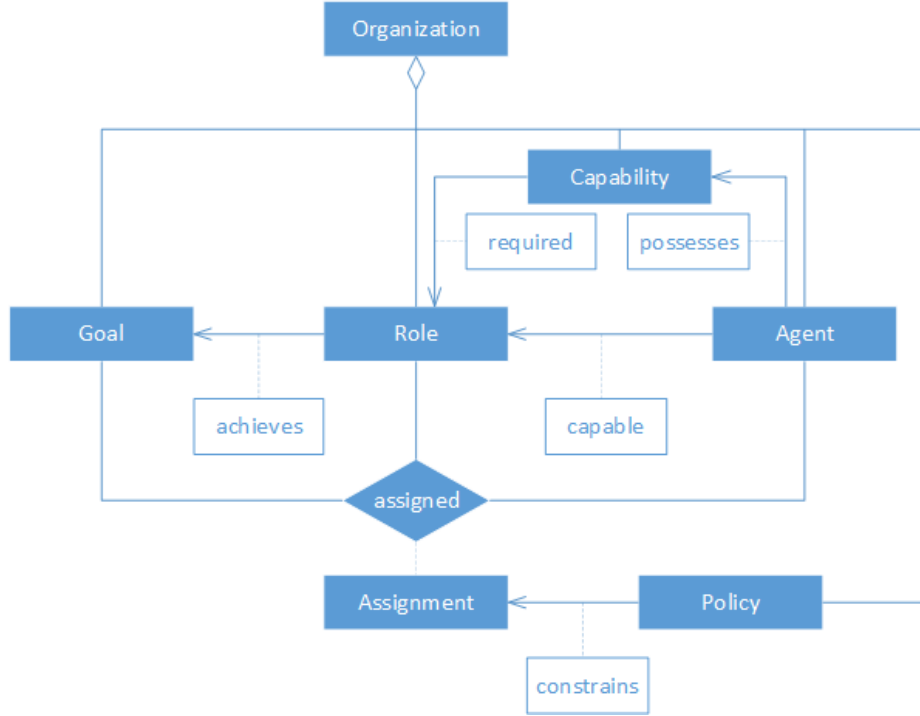


Fig. 1. Organization model

The organization model also describes relationships that exist among entities of the autonomous system. Among them are the following relationships: achieves relationship that exists between roles and goals, capable relationship that exists between agents and roles, possesses relationship that exists between agents and capabilities, required relationship that exists between capabilities and roles, constrains relationship that exists between policies and assignments and assigned relationship that exists between agents, roles and goals.

Formally, organization model O is a tuple.

$$O = \langle G, R, AG, C, A, P, \text{achieves}, \text{capable}, \text{possesses}, \text{required}, \text{constrains}, \text{assigned} \rangle \quad (1)$$

where

$$\text{achieves}: R \times G \rightarrow \text{Boolean} \quad (2)$$

$$\text{capable}: AG \times R \rightarrow \text{Boolean} \quad (3)$$

$$\text{possesses}: AG \times C \rightarrow \text{Boolean} \quad (4)$$

$$\text{required}: C \times R \rightarrow \text{Boolean} \quad (5)$$

$$\text{constrains: } P \times A \rightarrow \text{Boolean} \quad (6)$$

$$\text{assigned: } AG \times R \times G \rightarrow \text{Boolean} \quad (7)$$

Every autonomous system is built to achieve certain goals. In our model each goal has a definition. All goals are placed in a set G . Goals can be critical or noncritical. Function *critical* determines whether a goal is critical. At this point we do not consider various relationships among goals, for example, such as subgoals, parent and child goals, conjunctive and disjunctive goals and precedence between goals and leave this to our future work.

$$\text{critical: } G \rightarrow \text{Boolean} \quad (8)$$

The model is said to be complete if all the goals can be achieved. The model is said to be incomplete if at least one goal cannot be achieved. It is certainly desirable for a model to be complete and in some cases it is absolutely necessary. There are non-critical situations though, in which a model can be incomplete but sufficient for a system. We will research such scenarios in our future work.

A goal is directly achieved by a role or a set of roles that each agent is capable of playing. Therefore, each goal is indirectly achieved by an agent or a group of agents through the roles, which those agents are capable of playing. An agent can be capable of playing several roles depending on the inherent capabilities that the agent possesses. Each role requires one or several capabilities that agent/s possess.

Completeness of the model is validated through assignments. Assignments are used to capture information about which agents play which roles to achieve which goals. Assignments are constrained by the organization policies that dictate which assignments are allowed and which are not.

Along with general constraints that each system must satisfy, an organization might impose additional constraints through policies. An example of such a policy constraint could be one that states that certain roles must be achieved by several agents to provide redundancy. This redundancy might be important for agent's capabilities that degrade under certain conditions, such as capabilities that depend on a battery life for autonomous drones.

4 Organization Instance

In this section we present examples of what various entities of an organization model could be in a case of a system of autonomous drones and then apply organization model approach to designing a swarm system of autonomous drones that are used for disaster relief aid delivery.

Examples of agents could be autonomous drones, parts of autonomous drones or whole groups or swarms of autonomous drones.

Examples of capabilities are data acquisition capabilities, data access capabilities, data manipulation capabilities, data transmission capabilities, computational capabilities, sensor capabilities, GPS capabilities, etc. During the system life span capabilities can degrade or improve. For example, sensors usually degrade over time and some-

times their accuracy becomes an issue. Another example of a degrading capability, albeit for possibly a short period of time, would be a limited vision capability during a cloudy or rainy weather. On the other hand, learning algorithms can improve a certain capability. In these cases, the system reorganization could be advisable or necessary.

Each organization has a set of policies that constrain the assignment of agents to roles to goals. These policies could be general in nature or very specific to the autonomous system. An example of a general policy could be a policy that states that each role must be played by at least one agent. An example of a system specific policy could be a policy that states that no fewer than three agents ought to be involved in working on a specific goal.

4.1 A Swarm System of Autonomous Drones Used for Disaster Relief Aid Delivery

In this section we apply the organization model approach to designing a swarm system of autonomous drones that are used for disaster relief aid delivery. This example is intentionally kept to a minimum for simplicity and clarity of presentation.

Defining Goals. The purpose of this system is to deliver disaster relief aid to the disaster area. Therefore, we define the following two goals for this system:

- G1 – deliver disaster relief aid.
- G2 – navigate to site.

Defining Roles. Goals are achieved by roles that various agents play. For simplicity we define only two roles that will be directly mapped to achieve the two goals stated above. In general, an organization can have many roles that achieve various goals.

- R1 – carrier.
- R2 – navigator.

Defining Capabilities. Each role requires a set of capabilities that an agent must possess in order to be capable to play the role. For this system we define four capabilities and list them below:

- C1 – carrying capability.
- C2 – navigation capability.
- C3 – data processing capability.
- C4 – communication capability.

Defining Agents. We also define seven agents with various capabilities and list them below:

- A1 - A4 – carrier agents.
- A5 – navigator agent.
- A6 – A7 – data processing agents.

Organization Instance. After we define goals, roles, capabilities and agents, the organization algorithms make assignments taking into consideration policy constraints to create an organization instance that would achieve all the stated goals. We call such an organization instance a system. The organization also produces information to the user whether the system is complete. If a complete system cannot be formed, the organization tries to create a viable system. In all cases, initial system has to be valid.

An organization instance for our example is shown in Fig. 2. It is both complete and valid. It is complete because all goals are achieved. It is also valid because all organization policies are satisfied.

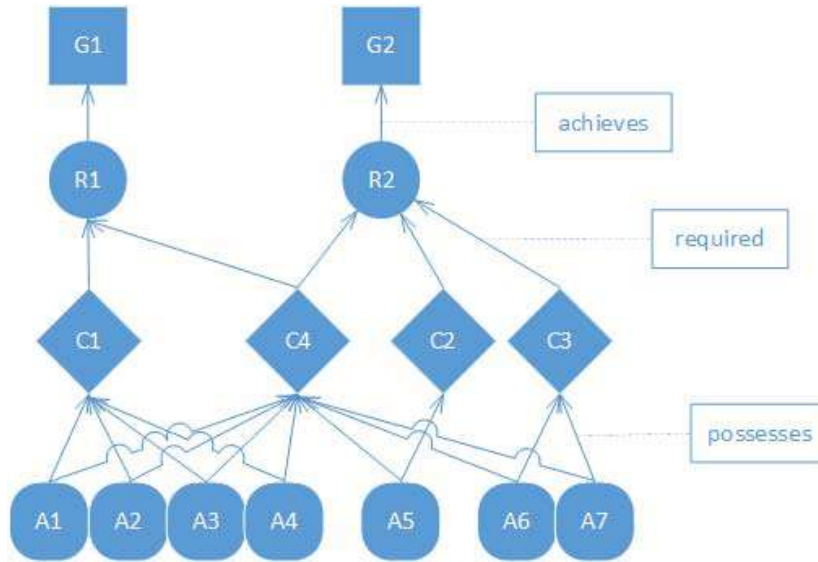


Fig. 2. Organization instance

5 Simulator

We developed a simulator DiSy-Sim (*Distributed Systems Simulator*) to test message passing algorithms in middleware and application layers of autonomous distributed systems, as well as to examine the work of a system in real time. The statistics allows distributed systems developers and testers to measure the efficiency of their message passing algorithms and compare various optimization techniques. The ability to see the work of a system in real time allows developers and testers to bring their development cost down and improve the overall quality. The end users of actual physical systems can use the simulator to make decisions on the quantity and the characteristics of the system entities to achieve the goals for the system.

5.1 Simulator goals

We set the following goals for the creation of DiSy-Sim simulator:

1. Ability to model message passing in a distributed system of autonomous entities.
2. Ability to provide statistics in terms of number of messages used by the algorithms and application itself; statistics should be easily available:
 - (a) for the whole system,
 - (b) for a particular algorithm,
 - (c) for a set of algorithms,
 - (d) for application layer,
 - (e) for a single entity,
 - (f) for a group of entities,
 - (g) for an entity type.
3. Simulation of the system in real time.
4. Graphical representation of messages over time.
5. Logging.

5.2 Simulator components

DiSy-Sim simulator was written in Java and tested on a Windows platform. Its main control window contains 3 panels as can be seen in Fig. 3: system settings/system view panel, simulation view panel and graphical stats panel. The panels can be resized within the main window. Each panel can also be shown in a separate window. Simulation view and graphical stats panels can be duplicated to show different parts of the system/statistics. A separate window for a single system entity can be opened by selecting the entity from either the simulation view panel or system view panel. Multiple windows for multiple entities can be opened concurrently.

5.3 Simulating organizational instance in DiSy-Sim

Below in Fig. 3 is a screenshot from the simulation in DiSy-Sim of the organizational instance described above.

Seven agents A1-A7 are moving to their destination. Communication protocols are based on the agents' roles and the message statistics is shown in the graphical statistics panel. The whole system consists of seven agents as can be seen from the system view panel. The agents' arrows show the direction, in which the drones are moving. The system successfully logged messages used by all the entities in the process of system execution.

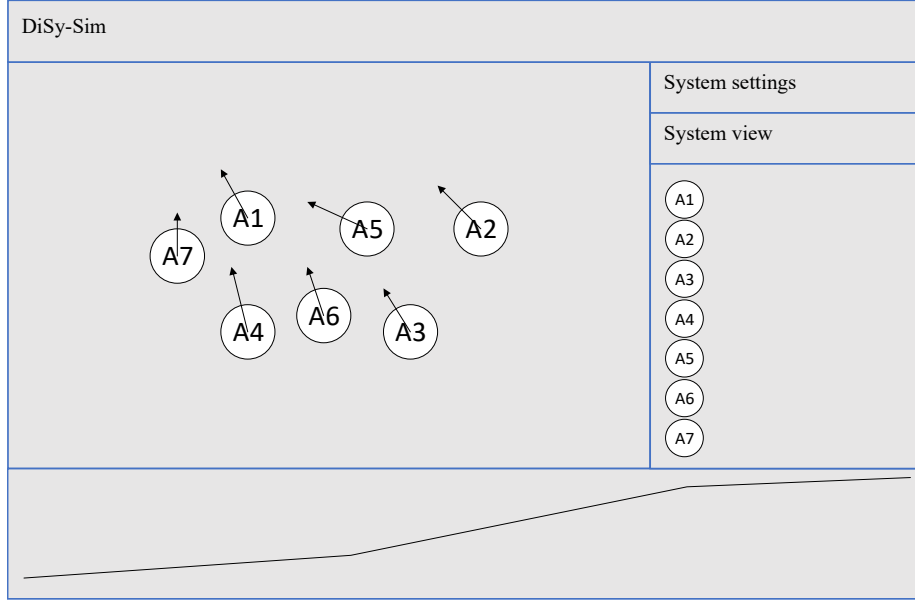


Fig. 3. DiSy-Sim simulator main window

To check the correctness of message calculation, we tested the system on a simple communication protocol that consisted of one navigator drone sending a broadcast message to all other drones during every time increment and other drones responding with an acknowledgement message sent back to the navigator drone upon receiving the navigation message. We tested the system using 100-time increments. The system correctly calculated the number of messages as can be seen in Fig. 4.

We also simulated this scenario with a different number of drones in a swarm. The system correctly calculated the number of messages in each test runs as can be seen from Fig. 5.

5.4 Applicable contexts

The DiSy-Sim simulator was designed for autonomous distributed systems, specifically for swarms of drones modeled using organizational approach. However, the simulator was designed for a general case of autonomous component based distributed systems and is applicable for such distributed systems as: sensor networks, distributed applications running on fixed computers/devices and distributed applications running on moving devices (ex., mobile phones), to name a few. We will demonstrate it with case studies in our future work.

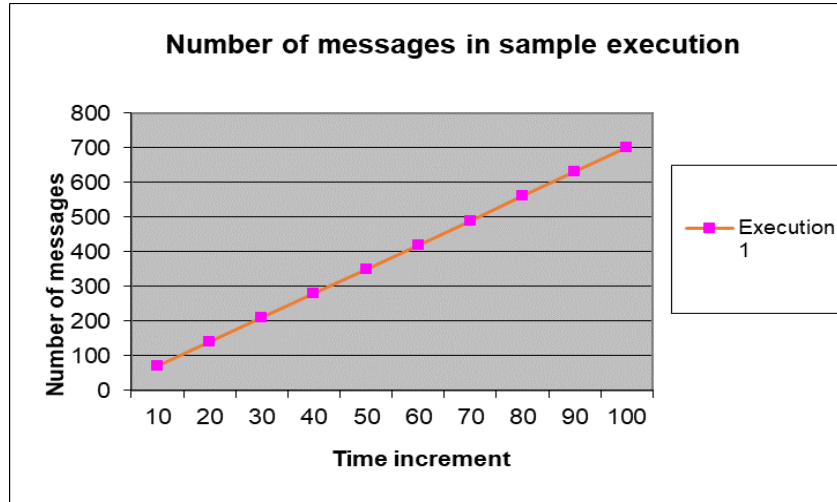


Fig. 4. Number of messages in sample execution

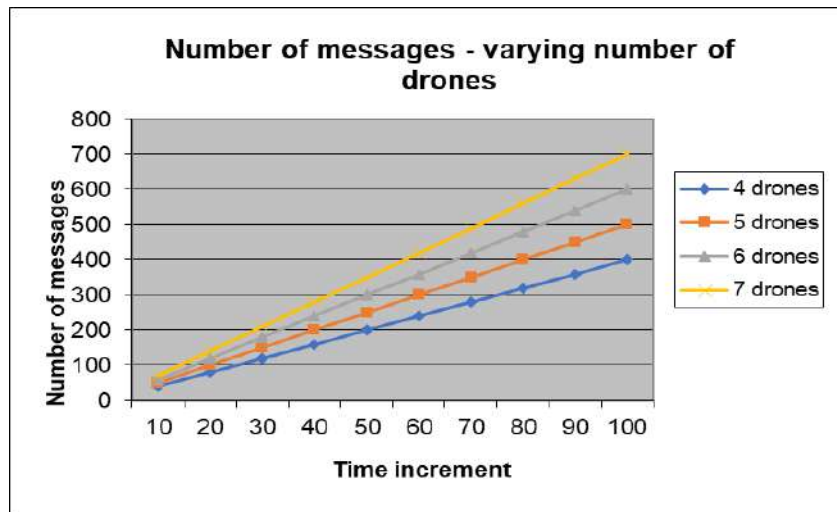


Fig. 5. Number of messages – varying number of drones

6 Conclusions and Future Work

In this paper, we presented an approach to designing and testing autonomous distributed systems. The design of such systems is based on organization model. This ap-

proach simplifies modeling of such systems and allows for better understanding of the system overall.

For testing autonomous distributed systems, we developed DiSy-Sim simulator. DiSy-Sim simulator simplifies testing of communication between system components by providing statistics on messages being used in communication.

The approach to designing and testing autonomous distributed systems presented in this paper is applicable for a variety of distributed systems. We demonstrated it for a swarm system of autonomous drones used for disaster relief aid delivery. Even though the example of an organizational instance for a swarm system of autonomous drones used for disaster relief aid delivery is simplified, it still shows the benefits of using such an approach.

In future work, we plan to enhance the organization model with additional features to account for failures in the system and the reorganization to take place.

We will add additional features to DiSy-Sim simulator, such as: single entity view with stats, replay capability, various terrains/obstacles for moving entities and failure simulation (loss of messages, failure of entity/equipment, battery power, etc.). We will apply DiSy-Sim simulator to other contexts, for instance, for sensor networks, distributed applications running on fixed computers/devices and distributed applications running on moving devices (ex., mobile phones).

We also plan to merge organization-based approach to modelling autonomous distributed systems into InDiGO framework and provide tools to extract optimization information from the model for middleware distributed algorithms of a distributed systems. We also have plans to research a possibility of expressing communication protocols through the organization-based modelling so that optimizations for communication in a distributed system could be realized based on the information from a design stage of a distributed system.

References

1. DeLoach, S.A., Wood, M.F., Sparkman, C.H.: Multiagent Systems Engineering. *The International Journal of Software Engineering and Knowledge Engineering* 11(3): 231–258 (2001)
2. DeLoach, S.A., Wood, M.F.: Developing Multiagent Systems with agentTool. In: *Agent Theories Architectures and Languages, 7th International Workshop*. LNCS, vol. 1986, Springer Verlag, Berlin (2001)
3. Carley, K.M.: Computational and Mathematical Organization Theory: Perspective and Directions. *Computational and Mathematical Organization Theory* 1(1): 39–56 (1995)
4. Carley, K.M.: Organizational Adaptation. *Annals of Operations Research* 75: 25–47 (1998)
5. Lawrence, P.R., Lorsch, J.W.: Organization and Environment: Managing Differentiation and Integration. *Journal of Artificial Intelligence Research* 7: 83–124 (1997)
6. Weingartner, E., Lehn, H., Wehrle, K.: A Performance Comparison of Recent Network Simulators, In: *2009 IEEE International Conference on Communications*, pp. 1–5. Dresden (2009)

7. Kolesnikov, V., Singh, G.: InDiGO: An infrastructure for optimization of distributed algorithms. In: 7th International Symposium on Parallel and Distributed Computing, pp. 401–408 (2008)
8. Kolesnikov, V., Singh, G.: Utilizing model checking for automated optimization information discovery in InDiGO. In: 8th International Symposium on Parallel and Distributed Computing, pp. 91–98 (2009)
9. Kolesnikov, V.: Realizing optimization opportunities for distributed applications in the middleware layer by utilizing InDiGO framework. In: 9th International Symposium on Parallel and Distributed Computing, pp. 85–92 (2010)
10. Kumar, A.R., Rao, S.V., Goswami, D.: NS3 Simulator for a Study of Data Center Networks, In: 12th IEEE International Symposium on Parallel and Distributed Computing, pp. 224–231 (2013)
11. Sobeih, A., Hou, J., Kung, L., Li, N., Zhang, H.: J-Sim: a simulation and emulation environment for wireless sensor networks. *Wireless Communications* 13: 104–119 (2006)
12. Cohen, P.R., Levesque, H.J.: Intention is Choice with Commitment. *Artificial Intelligence* 42(3) (1990)
13. Jennings, N.R.: Commitments and Conventions: The Foundation of Coordination in Multi-agent Systems. *The Knowledge Engineering Review* 8(3): 223–250 (1993)
14. Jennings, N.R.: Controlling Cooperative Problem Solving in Industrial Multi-Agent Systems Using Joint Intentions. *Artificial Intelligence* 75(2): 195–240 (1995)
15. Grosz, B., Kraus, S.: Collaborative Plans for Complex Group Action. *Artificial Intelligence* 86(2): 269–357 (1996)
16. Kinny, D., Ljungberg, M., Rao, A.S., Sonenberg, E., Tidhar, G., Werner, E.: Planned Team Activity. *Artificial Social Systems* 830: 226–256. LNAI, Springer-Verlag (1992)
17. Singh, G., Kolesnikov, V., Das, S.: Methodologies for optimization of distributed algorithms and middleware. In: The 22nd IEEE International Parallel and Distributed Processing Symposium (2008)

Promoting Training of Multi-Agent Systems

Petro Kravets¹[0000-0001-8569-423X], Vasyl Lytvyn²[0000-0002-9676-0180], Victoria Vysotska³[0000-0001-6417-3689], Yevhen Burov⁴[0000-0001-8653-1520]

Lviv Polytechnic National University, Lviv, Ukraine

Petro.O.Kravets@lpnu.ua¹, Vasyl.V.Lytvyn@lpnu.ua²,
Victoria.A.Vysotska@lpnu.ua³, Yevhen.V.Burov@lpnu.ua⁴

Abstract. The problem of incentive training of multi-agent systems in the game formulation for collective decision making under uncertainty is considered. Methods of incentive training do not require a mathematical model of the environment and enable decision making directly in the training process. Markov model of stochastic game is constructed and the criteria for its solution are formulated. An iterative Q-method for solving a stochastic game based on the numerical identification of a characteristic function of a dynamic system in space of state-action is described. Players' current gains are determined by the method of randomization of payment Q-matrix elements. Mixed player strategies are calculated using the Boltzmann method. Pure strategies are determined on the basis of discrete random distributions given by mixed player strategies. The algorithm for stochastic game solving is developed and results of computer implementation of game Q-method are analyzed.

Keywords – Multi-Agent System, Stochastic Game, Promotional Training, Q-method

1 Introduction

The functioning of most modern information systems (IS) is based on rigidly programmed algorithms. Unforeseen environmental influences in such systems may impair the stability of operating modes, which can lead to various types of emergency situations. To prevent critical states, distributed IS software must consist of interoperable standalone modules, be intelligent, flexible, and capable of independently monitoring environmental changes and making timely and appropriate decisions. Otherwise, such systems should be built on the principles of an agent-oriented methodology [1 – 10]. An IS agent is a standalone software module with elements of artificial intelligence, capable of making decisions on its own, interacting with the environment, other agents, and people as they accomplish the task. IS agents interact within the computer network. A population of computer network agents who solve a common problem is called a multi-agent system (MAS).

The operation of the MAS is usually carried out in the context of a priori uncertainty about the state of the decision-making environment and the actions of other

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

agents. In this regard, agent's behavior strategies must be adaptive at the expense of agents' ability to learn [11]. Among the methods of learning under uncertainty, incentive-based methods have gained practical appeal [12, 13] because they do not require a mathematical model of the environment and provide decision-making power directly in the learning process. The mechanisms of reflexive behavior of living organisms with developed nervous system are the basis of the stimulating training. An effective method of incentive learning is Markov Q-learning [14], which performs numerical identification of the characteristic function of a dynamic system in state-space. As a characteristic function is usually the function of the total expected reward agent.

Compared to single agent systems, the structure, operation, and research of multi-agent Q-learning methods are much more complicated [15]. Due to the collective interaction of agents, the stationary environment is transformed into a non-stationary class. The change in the state of the environment and the value of the benefits of each agent depend on the actions of the other agents. Generally, in an MAS, an agent cannot achieve a maximum gain equal to that of a single agent system. The optimal pay-offs of agents must be balanced and meet the criteria of benefit, fairness, balance. Thus, instead of the criterion of scalar maximization of the benefits of a single-agent system, the criteria of vector maximization of MAS winnings are introduced, for example, Nash equilibrium, Pareto optimality or other [16].

Provided the use of methods of Q-learning of MAS, an iterative construction of a system of characteristic Q-functions in the state-action space takes place, and the increment of the elements of these functions is carried out in the direction of achieving their collective equilibrium. To build the MAS, it is necessary to carry out preliminary studies on the basis of adequate mathematical models that will allow to study the dynamics of the system under uncertainty, to build strategies for the behavior of agents that provide optimal technical and economic parameters of the system functioning. Given the peculiarities of the subject area, namely, multi-agency, uncertainty of decision-making environment, antagonism or competitiveness of goals, communicativeness, coordination of actions, adaptability of agent behavior strategies, we use the mathematical apparatus of stochastic game theory [17 – 29]. Solving a stochastic game is to find the strategies of agents that maximize their winnings so as to ensure a certain collective balance of interests for all players. The search for optimal strategies for players in uncertainty will be performed on the basis of the promotional training method.

The purpose of the work is to construct an iterative method of incentive learning to solve the stochastic MAS game in uncertainty. To achieve this goal, it is necessary to develop a model of multi-agent stochastic game, to determine the criteria of collective equilibrium, the method and algorithm for solving the game problem.

2 The Mathematical Model of Stochastic Game

The stochastic game is determined by the tuple:

$$(S, p, A_i, r^i | i \in I), I = \{1, 2, \dots, L\},$$

where $S = \{s_1, \dots, s_M\}$ is set of all states of the environment, $p : S \times A \rightarrow \Delta(S)$ is system state change function defined in the space of probability distributions $\Delta(S)$ on the plural S , $A_i = (a_i(1), \dots, a_i(N_i))$ is multiple actions or pure strategies i -agent, $A = \times_{i \in I} A_i$ is set of combined agent actions, $r^i : S \times A \rightarrow R$ is reward function i -agent, I is multiple agents, L is number of agents, M is number of states, N_i is number of strategies i -agent.

In the general case of multiple actions $A_i = A_i(s) \quad \forall i \in I$ and combined actions $A = A(s)$ may depend on the state of the environment $s \in S$.

We adopt a Markov model [30, 31] of the dynamics of states of a system in which the probability of change of states p depends only on the current state of the environment and the current actions of the agents:

$$p(s_{t+1} = s' | (s_t, b_t), \tau = 0, 1, 2, \dots, t) = p(s_{t+1} = s' | s_t, b_t),$$

where $b_t \in A$ is combined action at time t .

At each point in time, the environment is in one of the states $s \in S$ and agents choose actions independently $a_i \in A_i$. After the implementation of the combined option $a = (a_1, \dots, a_L) \in A$ agents get random winnings r^i (otherwise is incentives or reinforcements), and the environment changes its state according to the probability distribution $p(s, a)$ with values per segment $[0, 1]$:

$$\sum_{s' \in S} p(s' | s, a) = 1.$$

The agent implements actions based on a mixed strategy $\pi_i : S \rightarrow A_i$, which determines the likelihood of action $a_i \in A_i$ in every state of the environment $\forall s \in S$.

Distribution $\pi_i \in \Pi_i$ takes the value on a unit simplex [11]

$$\Pi_i = \left\{ \pi \left| \sum_{a_i \in A_i} \pi(s, a_i) = 1, \pi(s, a_i) \geq 0 \right. \right\}.$$

If $\pi_i(s, a_i) \in \{0, 1\}$, then the agent determines the choices of solutions. Let the total payoff of each agent be determined by the function of discounted total payoffs:

$$\Upsilon_i = \sum_{t=0}^{\infty} \gamma^t r_t^i, \quad (1)$$

where $\gamma \in (0, 1]$ is discount option.

The goal i -agent is to maximize function (1) by formulating an effective strategy π_i :

$$V_{\pi}^i(s) = E_{\pi}[r_i | s_0 = s] \rightarrow \max_{\pi_i}, \quad \forall i \in I, \quad (2)$$

where $\pi = (\pi_1, \dots, \pi_L)$; E is symbol of mathematical expectation.

Stochastic game resolution is about defining agent behavior strategies π_i^* ($\forall i \in I$), which ensure fulfilment of one of the conditions of collective optimality, for example:

- 1) Nash equilibrium: $V^i(s, \pi_1^*, \pi_2^*, \dots, \pi_L^*) \geq V^i(s, \pi_1^*, \pi_2^*, \dots, \pi_{i-1}^*, \pi_i, \pi_{i+1}^*, \dots, \pi_L^*)$;
- 2) Pareto optimality: $V^i(s, \pi^*) \geq V^i(s, \pi)$.

3 Learning of Stochastic Game

Calculation $V_{\pi}^i(s)$ can be performed in a recursive form known in the literature as the Bellman equation [12 – 14]. Given (1), we obtain after simple transformations:

$$\begin{aligned} V_{\pi}(s | s_t = s) &= E(r_t) + \gamma \sum_{k=0}^{\infty} \gamma^k E(r_{t+k+1}) = E(r_t) + \gamma W_{\pi}(s_{t+1}) = \\ &= r(s, \pi(s)) + \gamma \sum_{s' \in S} p(s' | s, \pi(s)) V_{\pi}(s') \end{aligned} \quad (3)$$

where s' is probable future states of the system.

The agent's goal is to find a strategy π^* that maximizes function (2) for all states of the environment:

$$\forall \pi \forall s \in S \quad V^{\pi^*}(s) \geq V^{\pi}(s).$$

Since the choice of options is made by chance, it is for the purpose of comparing the effectiveness of actions when the system is in a state $s \in S$, current gains are useful to obtain from (2). For this purpose is specially built Q is average payoff feature that determines the cost of the action – the total payoff of the agent in the state s chose action a :

$$Q_{\pi}(s, a) = E_{\pi}[R | s_0 = s, a_0 = a]. \quad (4)$$

Expression (4) defines a tabular function of the values of the action options a in the states s .

Similarly to (3) we obtain:

$$Q_{\pi}(s, a) = r(s, a) + \gamma \sum_{s' \in S} p(s' | s, a) V_{\pi}(s'). \quad (5)$$

Adherence to the Bellman principle of optimality (5) ensures the optimal gain of the agent from the current state achieved $s \in S$ at all future times. Applying this principle to all states ensures a global optimal solution.

For optimal function selection strategies π^* for each state $s \in S$ we will get:

$$V_{\pi^*}(s) = \max_{a \in A} \left[r(s, a) + \gamma \sum_{s' \in S} p(s' | s, a) V_{\pi^*}(s') \right]. \quad (6)$$

From (6) one can obtain the optimal function of choosing strategies

$$\pi^*(s) = \arg \max_{a \in A} Q_{\pi^*}(s, a). \quad (7)$$

Optimization (7) can be performed by dynamic programming methods [30].

By analogy to single agent training, we define the payoff matrix i -players with current and future winnings in the direction of movement to the optimal collective state in space $S \times A$:

$$Q^i_*(s, a_1, \dots, a_L) = r^i(s, a_1, \dots, a_L) + \gamma \sum_{s' \in S} p(s' | s, a_1, \dots, a_L) \cdot V^i(s', \pi_1^*, \dots, \pi_L^*),$$

where $Q^i_*(s, a_1, \dots, a_L)$ is total discounted gain i -player provided the players select the action $(a_1, \dots, a_L)$ in the state s according to the optimal strategy of the game $\pi^* = (\pi_1^*, \dots, \pi_L^*)$.

In the conditions of a priori uncertainty of the transition probabilities between the states of the system $p(s, a_1, \dots, a_L)$ and winnings features $r^i(s, a_1, \dots, a_L)$ an iterative method is used to calculate the elements of the payoff matrix Q -learning [31]:

$$Q^i_{t+1}(s, a_1, \dots, a_L) = (1 - \alpha_t) Q^i_t(s, a_1, \dots, a_L) + \alpha_t [r^i_t + \gamma V^i_t(s_{t+1})] \quad (8)$$

where $\alpha_t \in (0, 1)$ is training option; $V^i_t(s_{t+1})$ is the operator of the cost of the system state in the direction of the optimal collective solution.

The type of the operator $V^i_t(s_{t+1})$ is determined by the condition of collective equilibrium, for example:

$$\begin{aligned} V^i_t(s_{t+1}) &= MM(Q^i_t(s_{t+1})) \text{ is maximin equilibrium;} \\ V^i_t(s_{t+1}) &= NE(Q^i_t(s_{t+1})) \text{ is Nash equilibrium;} \\ V^i_t(s_{t+1}) &= BR(Q^i_t(s_{t+1})) \text{ is best agent response;} \\ V^i_t(s_{t+1}) &= CE(Q^i_t(s_{t+1})) \text{ is correlated equilibrium;} \\ V^i_t(s_{t+1}) &= PE(Q^i_t(s_{t+1})) \text{ is Pareto optimality.} \end{aligned}$$

The above list may be supplemented by other already known and new equilibrium states that will determine the target aspect of the functioning of a distributed dynamic system. Method (8) can be applied to decipher a single agent game with nature as a partial case N -agent stochastic game if $I = \{i\}, |I| = 1, A = A_i$, when $V_t^i(s_{t+1}) = \max_{b \in A_i} Q_t^i(s', b)$, where $s' = s_{t+1}$.

The Maximin Equilibrium (MM) takes place in the game of two agents with zero sum of their payoff functions:

$$V^1(s_{t+1}) = \max_{\pi_1 \in \Pi_1} \min_{a_2 \in A_2} \sum_{a_1 \in A_1} \pi_1(s', a_1) Q_t(s', a_1, a_2) = -V^2(s_{t+1}) .$$

Nash Equilibrium (NE) is determined by the independent distribution of strategies by players who choose their own strategies independently of the choice of other agents. In a Nash equilibrium situation in mixed strategies $\pi^{NE}(s) = (\pi_1^{NE}(s), \dots, \pi_L^{NE}(s))$ it is not profitable for each agent to deviate from its own optimal strategy $\pi_i^{NE}(s)$, if other agents stick to the equilibrium point [31]:

$$\sum_{a \in A} Q_t^i(s', a) \pi_i^{NE}(a_i) \prod_{j \neq i} \pi_j^{NE}(s', a_j) \geq \sum_{a \in A} Q_t^i(s', a) \tilde{\pi}_i(a_i) \prod_{j \neq i} \pi_j^{NE}(s', a_j), \quad (9)$$

where $a = (a_1, \dots, a_L)$; $\pi_i^{NE}, \tilde{\pi}_i \in \Pi_i$.

Method (8) ensures that condition (9) is satisfied when the current value of the system state value operator is determined at point π^{NE} Nash equilibrium:

$$NE(Q_t^i(s_{t+1})) = \sum_{a \in A} Q_t^i(s', a) \prod_{j=1}^L \pi_j^{NE}(s', a_j) .$$

The set of NE equilibrium points in mixed strategies is a convex compact and can be calculated by linear programming methods (for bi-matrix games) or by solving a system of polylinear equations that determine the complementary rigidity condition:

$$\begin{aligned} \sum_{a_{-i} \in A_{-i}} Q_t^i(s, a_{-i}, a_i) \prod_{j \neq i} \pi_j(s, a_j) &= \sum_{a \in A} Q_t^i(s, a) \prod_{j=1}^L \pi_j(s, a_j), \quad \forall i \in I, \quad \forall a_i \in A_i, \quad \forall s \in S, \\ \pi_i(s, a_i) &> 0, \quad \sum_{a_i \in A_i} \pi_i(s, a_i) &= 1 . \end{aligned}$$

Unlike the Nash equilibrium, the Best Response (BR) method generates an optimal agent strategy in response to the actions of all other agents. The corresponding system state value operator in method (8) has the form [32]:

$$BR(Q_t^i(s_{t+1})) = \max_{\pi_i} \left(\sum_{a \in A} Q_t^i(s', a) \prod_{j=1}^L \pi_j(s', a_j) \right).$$

Correlated Equilibrium (CE) generalizes the Nash equilibrium by allowing players' strategies to depend. For this purpose, there is an arbitrator in the collective decision-making system, which according to the generalized distribution $\sigma \in \Delta(A)$

($\sum_{a \in A} \sigma(a) = 1$, $A = \times_{i=1}^L A_i$) recommends that players choose to take actions that form a combined option $a = (a_1, \dots, a_L)$. Player with a number i only receives component information combined option $a \in A$. This signal is perceived i -player as an optional offer to take action a_i . Each player secretly and independently chooses at a moment's time t action option a_i , possibly different from the proposed variant, and receives a current payoff $r^i(s_t, a_t)$, which is a function of the current state of the system s_t and the combined option $a_t \in A$. The environment then moves to a new state s_{t+1} according to the probability distribution $p(s_{t+1} | s_t, a_t)$ and the process is repeated at time $t+1$.

Correlated equilibrium is determined by the united distribution of player strategies $\sigma \in \Delta(A)$, when each agent does not have the motivation to deviate unilaterally [33]:

$$\sum_{a_{-i} \in A_{-i}} \sigma^{CE}(a_{-i} | a_i) Q^i[s', (a_{-i}, a_i)] \geq \sum_{a_{-i} \in A_{-i}} \sigma^{CE}(a_{-i} | a_i) Q^i[s', (a_{-i}, \tilde{a}_i)],$$

where $A_{-i} = \times_{j=1, j \neq i}^L A_j$, $A = A_{-i} \times A_i$, $a_{-i} \in A_{-i}$, $a = (a_{-i}, a_i) \in A$, $a_i, \tilde{a}_i \in A_i$,

$$\sigma^{CE}(a_i) = \sum_{a_{-i} \in A_{-i}} \sigma^{CE}(a_{-i}, a_i), \sigma^{CE}(a_{-i} | a_i) = \sigma^{CE}(a_{-i}, a_i) / \sigma^{CE}(a_i), \sigma^{CE}(a_i) > 0.$$

All Nash equilibrium points are correlated equilibrium points. If $\forall i \in I$ $\sigma^{CE}(a_{-i} | a_i) = \sigma^{CE}(a_{-i} | \tilde{a}_i)$, $\forall a_i, \tilde{a}_i \in A_i$, $\forall a_{-i} \in A_{-i}$, $\forall \sigma^{CE}(a_i), \sigma^{CE}(\tilde{a}_i) > 0$, then correlated equilibrium is also Nash equilibrium. To solve the game by method (8), the cost operator $V_t^i(s_{t+1})$ the state of the system is determined by the point σ^{CE} correlated equilibrium: $CE(Q_t^i(s_{t+1})) = \sum_{a \in A} \sigma^{CE}(a) Q_t^i(s', a)$. The set of CE equilibrium

points is non-empty, convex and compact and can be effectively calculated using linear programming methods. In the case of maximizing total player winnings, the task of linear programming is to find σ^{CE} can be formulated as follows:

$$\sum_{a \in A} \sigma(a) \sum_{i=1}^L Q^i(s, a) \rightarrow \max_{\sigma}, \quad \sum_{a_{-i} \in A_{-i}} \sigma(a_{-i}, a_i) (Q^i[s', (a_{-i}, a_i)] - Q^i[s', (a_{-i}, \tilde{a}_i)]) \geq 0,$$

$$\forall i \in I, \forall a_i \in A_i, \forall \tilde{a}_i \in A_i, \forall s \in S, \sigma(a) > 0 \quad \forall a \in A, \sum_{a \in A} \sigma(a) = 1.$$

Based on the distribution $\sigma(a) \quad \forall a \in A$ players' own strategies are determined $\pi_i \quad \forall i \in I$. There are various options for switching from σ to π_i depending on the type of strategies and players' level of awareness. For example, pure strategies determine the maximum value of the operator $CE(Q^i(s))$:

$$a_i = \arg \max_{\sigma} CE(Q^i(s)).$$

Taking into account that $\sum_{a_i \in A_i} \sigma(a_i) = 1$, it can be assumed that mixed strategies

$$\pi_i(a_i) = \sigma(a_i) = \sum_{a_{-i} \in A_{-i}} \sigma(a_{-i}, a_i), \text{ or:}$$

$$\pi_i(a_i) = \sum_{a_{-i} \in A_{-i}} \sigma(a_{-i}) Q^i(s, a_{-i}, a_i) / CE(Q^i(s)).$$

Pareto Equilibrium (PE) optimality occurs in the Common-Interest Markov Game, where the payoff matrices are the same for all players $Q_t^i(s, a) = Q_t^j(s, a) \quad \forall i, j \in I$, $\forall s \in S, \forall a \in A$ [33]. A game with different payoff matrices can be turned into a game of shared interests by a convolution

$$PE(Q_t(s_{t+1})) = \sum_{k=1}^L \lambda_k \sum_{a \in A} Q_t^k(s', a) \prod_{j=1}^L \pi_j^{PE}(s', a_j),$$

where $\lambda_j > 0 \quad (j = 1..L)$.

The search for the game's PE solution is done independently by agent strategies, similar to the search for a NE solution. A multi-agent game is optimal for Pareto if there is no common player strategy that improves the winnings of all players: $Q_t^i(s, \pi^{PE}) \geq Q_t^i(s, \pi)$. Pareto-optimal mixed strategies $\pi^{PE}(s) = (\pi_1^{PE}(s), \dots, \pi_L^{PE}(s))$ can be obtained by maximizing the convolution of concave (up) payoff functions:

$$\sum_{k=1}^L \lambda_k \sum_{a \in A} Q_t^k(s', a) \prod_{j=1}^L \pi_j(s', a_j) \rightarrow \max_{\pi}.$$

To calculate optimal collective strategies $\pi^*(s) = (\pi_1^*(s), \dots, \pi_L^*(s))$ (NE, CE, PE) agent with number i need to know Q -functions of all agents: $Q(s) = (Q_1^1(s), \dots, Q_1^L(s))$. In the absence of such information, each agent should evaluate the value Q -functions in the learning process. For this i -agent monitors the current gains of other agents and modifies their estimates Q -functions according to (8). To ensure that method (8) converges to one of the points of collective equilib-

rium, it is necessary to impose a limit on the rate of change of its adjustable parameters. The general limitations are as follows [11, 14, 34]:

$$\sum_{t=0}^{\infty} \alpha_t = \infty, \quad \sum_{t=0}^{\infty} \alpha_t^2 < \infty, \quad (10)$$

where $\alpha_t = t^{-\kappa}$ ($\kappa > 0$) is monotonically decreasing positive sequences of real values.

4 Stochastic Game Solving Algorithm

Step 1. Set the start time $t = 0$; the initial values of the payoff matrices

$Q_t^i(s, a_1, \dots, a_L) = \varepsilon \quad \forall s \in S, \quad \forall a_i \in A_i, \quad \forall i \in I$, where $0 < \varepsilon < 1$ is small positive value; the value of the gain discount parameter $\gamma \in (0, 1]$; the initial state of the system s_0 .

Step 2. Perform a random selection of agent actions $a = (a_1, \dots, a_L)$ based on strategies $\pi = (\pi_1, \dots, \pi_L)$. The value of the strategies can be calculated from the current estimates of the payoff matrices $\forall i \in I$:

$$\pi_i(s, a_i) = \sum_{a_{-i} \in A_{-i}} Q_t^i(s, a_{-i}, a_i) / \sum_{a \in A} Q_t^i(s, a), \quad \forall a_i \in A_i.$$

Step 3. Get current agent payouts $r_t = (r_t^1, \dots, r_t^L)$.

Step 4. Determine the new state of the system $s_{t+1} = s_t(a_1, \dots, a_L)$.

Step 5. Calculate function $V_t^i(s_{t+1})$ according to the specific condition of collective equilibrium (NE, CE, PE).

Step 6. Modify the payoff matrix $Q_{t+1} = (Q_{t+1}^i(s_t, a_t) | i = 1..L)$ according to (8).

Step 7. If $\|Q_{t+1}^i - Q_t^i\| < \varepsilon \quad \forall i = 1..L$, then ask $t := t + 1$ and go to step 2.

Step 8. Print the calculated values of the payoff matrices $Q(s) = (Q^1(s), \dots, Q^L(s))$ and strategies $\pi(s) = (\pi_1(s), \dots, \pi_L(s)) \quad \forall s \in S$. End of algorithm.

5 The Results of Computer Simulation

Let's solve the stochastic game of two agents with two pure strategies in a two-state environment. The matrices of the average payoffs of such a game are given in Table 1.

Table 1. Player gains matrix

States	Strategies	The first player		The second player	
	$\times$	$\pi_2(s_1, a_2[1])$	$\pi_2(s_1, a_2[2])$	$\pi_2(s_1, a_2[1])$	$\pi_2(s_1, a_2[2])$
s_1	$\pi_1(s_1, a_1[1])$	0.4	0.1	0.9	0.2
	$\pi_1(s_1, a_1[2])$	0.1	0.9	0.2	0.9
	$\times$	$\pi_2(s_2, a_2[1])$	$\pi_2(s_2, a_2[2])$	$\pi_2(s_2, a_2[1])$	$\pi_2(s_2, a_2[2])$
s_2	$\pi_1(s_2, a_1[1])$	0.4	0.6	0.5	0.2
	$\pi_1(s_2, a_1[2])$	0.6	0.8	0.6	0.7

Under uncertainty, the elements of the average payoff matrix $\left[v^i(s, a) \right]_{\substack{\forall s \in S \\ \forall a \in A}}$ a priori unknown and available for observation in the form of random current values

$$r^i(s, a) = \text{Normal}(v^i(s, a), d^i(s, a)),$$

distributed by normal law with mathematical expectation $v^i(s, a)$ and dispersion $d^i(s, a)$. Normally distributed random variables are obtained by summing twelve evenly distributed random numbers $\omega \in [0, 1]$:

$$r^i(s, a) = v^i(s, a) + \sqrt{d^i(s, a)} \left(\sum_{j=1}^{12} \omega_j - 6 \right), \quad (11)$$

where $d^i(s, a) = d > 0 \forall s \in S, \forall a \in A$. If at time t the system was in a state of disrepair $s \in S$, then after implementing pure strategies $a = (a_1, \dots, a_L)$, where $a \in A = \times_{i=1}^L A^i$, agents receive current winnings $r_t^i(s, a)$, calculated according to (11).

After receiving current winnings, each agent lists the corresponding item Q -matrix according to the algorithm modified for uncertainty conditions BR :

$$Q_{t+1}^i(s, a_1, \dots, a_L) = (1 - \alpha_t) Q_t^i(s, a_1, \dots, a_L) + \alpha_t (r_t^i + \gamma \max_{a_i} Q_t^i(s, a_1, \dots, a_L)). \quad (12)$$

Based on Q -matrices the current values of mixed strategies are calculated using the Boltzmann method:

$$\pi_i(a_i(k) | s) = e^{Q_i^*(s, a_i(k))/T} / \sum_{j=1}^{N_i} e^{Q_i^*(s, a_i(j))/T}, \quad k = 1..N_i, \quad (13)$$

where $Q_i^*(s, a_i(k)) = \max_{a_{-i}} r^i(a_{-i}, a_i(k))$, $a_{-i} \in A_{-i}$, $A_{-i} = \prod_{j=1, j \neq i}^L A^j$, $T > 0$ is temperature coefficient.

Vector elements of mixed strategies π_i define a discrete distribution by which the values of random pure strategies are determined i -agent at the next time:

$$a_i(s) = \left\{ A^i(s, k) \middle| k = \arg \left(\min_k \sum_{j=1}^k \pi_i(s, a_i(j)) > \omega \right), k = 1..N_i \right\}, \forall s \in S, \forall i \in I, \quad (14)$$

where $\omega \in [0, 1]$ is random variable with uniform distribution.

The change of states of the dynamic system is determined by the discrete distribution $p(s' | s, a) = p \quad \forall s \in S, \forall a \in A$:

$$s = \left\{ S(k) \middle| k = \arg \left(\min_k \sum_{j=1}^k p(j) > \omega \right), k = 1..M \right\}. \quad (15)$$

Let the states change $s \in S$ system is implemented with equal probabilities $p(s' | s, a) = (|S|^{-1}, k = 1..M) \quad \forall s \in S, \forall a \in A$, that is $p(s' | s, a) = (0.5; 0.5)$ for $|S| = 2$. Agents' average payoffs are calculated taking into account the transition probabilities of the medium from one state to another: $V^i = \sum_{s \in S} p(s) V^i(s)$, $i = 1..L$,

where $V^i(s) = \sum_{a \in A} v^i(s, a) \prod_{j=1}^L \pi_j(s, a)$ is average agent gain in the state $s \in S$.

The trajectories of changing agent strategies within a unit simplex and the appearance of average payoff functions $V^i(s)$, which correspond to the data of the table 1, is shown in Fig. 1 and Fig. 2.

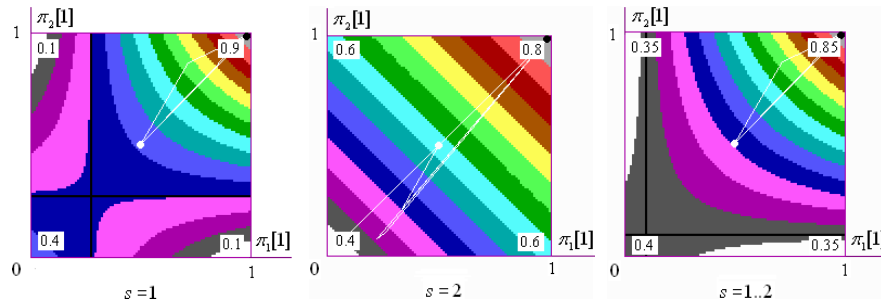


Fig. 1. First agent average payoff functions

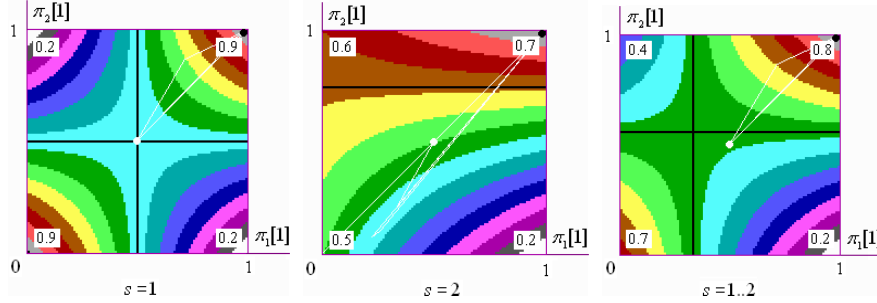


Fig. 2. Second agent average payoff functions

The BR method (12 – 15) ensures that stochastic play is solved at the vertices of a unit simplex with the maximum value of the mean gain function. The percentage of options for achieving optimal game resolution depends on the absolute difference between the two largest consecutive values of the average payoff features.

The convergence of the method is estimated by the error of fulfillment of the complementary slackness condition [35], weighted by mixed strategies:

$$\Delta = L^{-1} \sum_{i \in I} \|\pi_i - \tilde{\pi}_i\|^2, \text{ where } \tilde{\pi}_i = \text{diag}(\pi_i) \nabla V^i / V^i; \text{diag}(\pi_i) \text{ is diagonal square}$$

matrix of order N_i , formed from vector elements π_i ; $\nabla V^i = (V^i[j])_{j=1..N_i}$ is vector

median payoff function for fixed net strategies i -player; $V^i = \sum_{j=1}^{N_i} V^i[j] \pi_i[j]$ is

average payoff function i -player; $\|\cdot\|$ is Euclidean vector norm. The complementary slackness condition characterizes the gameplay in Nash mixed strategies. The weighted condition additionally takes into account the game's solutions in pure strategies. Average payoff function graphs Υ and norms for deviating mixed strategies from their target values Δ filed in Fig. 3.

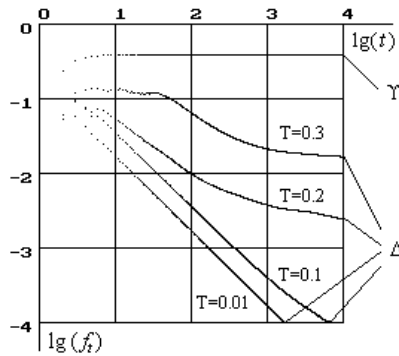


Fig. 3. Characteristics of the convergence of the game Q -method

The gains in Fig. 3 are averaged over the number of players: $\bar{Y} = L^{-1} \sum_{i=1}^L Y_i$, where $Y_i \geq 0$ is discounted current gains i -player. Deviation schedule drop Δ mixed strategies from their target indicates the convergence of the game Q -method.

The value of the temperature coefficient T has a significant impact on the convergence of the game method. The rate of convergence is determined by the rapid decline of the function graph Δ , which can be estimated by the value of the acute angle of linear approximation of the function graph Δ with the time axis. With the growth T rate of convergence of game Q -method decreases.

6 Conclusions

The promotional training method (8) considered in the deterministic version requires the knowledge of each agent Q -functions of all other agents. These functions are used by agents to identify strategies that provide method dynamics toward the points of collective equilibrium. Value Q -functions can be obtained by exchanging information between agents. If integrated information about Q -functions are not available to the agent, then he must determine their value independently in the learning process, observing the current benefits of other agents and performing evaluations Q -functions according to (8). If such observations are not possible, the agents may perform reflective assessments Q -functions of other agents.

Another method of constructing incentive training algorithms for agents under uncertainty is to apply the stochastic approximation method to the corresponding collective equilibrium condition.

The practical use of game-based promotional training methods requires their prior analysis to determine the conditions for convergence to a state of collective equilibrium. Such studies are based on the evaluation of sequences of random variables, which characterize the current deviations of players' strategies from their optimal values.

The rate of convergence of the game method of Q -learning is determined by the parameters α_t and T . Parameter α_t must satisfy the general conditions of stochastic approximation (10). The value of the parameter T depends on the absolute values of the elements of Q -matrices. It is experimentally established that for the given matrices of average payoffs the convergence of the game Q -method is provided at $T \in (0, 0.2]$ in the parameter value range $\alpha_t = t^{-\kappa}$, $\kappa \in (0, 1]$. The highest convergence rate of the game method of promotional learning is achieved at $T = 10^{-2}$.

References

1. Weiss, G.: Multiagent Systems. Second Edition. The MIT Press. (2013)

2. Lee, Y., Chong, Q.: Multi-agent Systems Support for Community-Based Learning. *Interacting with Computers*, 15(1), 33 – 55. (2003)
3. Kravets, P.: The Methodology of Multi-Agent Systems: a Modern State and Future Trends. In: *Proceedings of the International Conference on computer science and information technologies – CSIT'2006*, 125 – 127. (2006)
4. Dignum, F., Bradshaw, J., Silverman, B.G., Doesburg, W.: *Agent for Games and Simulations: Trends in Techniques, Concepts and Design*. Springer. (2009)
5. Kravets, P.: The Control Agent with Fuzzy Logic. In: *Perspective Technologies and Methods in MEMS Design, MEMSTECH*, 40 – 41. (2010)
6. Scerri, P., Vincent, R., Mailler, R.T.: *Coordination of Large-Scale Multiagent Systems*. Springer. (2010)
7. Byrski, A., Kisiel-Dorohinicki, M.: *Evolutionary Multi-Agent Systems: From Inspirations to Applications*. Springer. (2017)
8. Radley, N.: *Multi-Agent Systems – Modeling, Control, Programming, Simulations and Applications*. Scitus Academics LLC. (2017)
9. Yang, S., Xu, J.-X., Li, X., Shen, D. : *Iterative Learning Control for Multi-Agent Systems Coordination*. Wiley-IEEE Press. (2017)
10. Sun, Z.: *Cooperative Coordination and Formation Control for Multi-Agent Systems*. Springer. (2018)
11. Nazin, A. V., Poznyak, A. S.: *Adaptive Choice of Variants: Recurrence Algorithms* (in russian). Moscow: Science. (1986)
12. Kaelbling, L., Littman, M.L., Moore, A.W.: Reinforcement learning: A survey. In: *Journal of Artificial Intelligence Research*, 4, 237 – 285. (1996)
13. Sutton, R.S., Barto, A.G.: *Reinforcement Learning: An Introduction*. MIT Press. (1998)
14. Watkins, C.J.C.H., Dayan, P.: Q-Learning. In: *Machine Learning*, Kluwer Academic Publishers, Boston, 8, 279 – 292. (1992)
15. Ummels, M.: *Stochastic Multiplayer Games: Theory and Algorithms*. Amsterdam University Press. (2014)
16. Ungureanu, V.: *Pareto-Nash-Stackelberg Game and Control Theory: Intelligent Paradigms and Applications*. Springer. (2018)
17. Zheng, J., Cai, Y., Wu, Y., Shen, X.: Dynamic Computation Offloading for Mobile Cloud Computing: A Stochastic Game-Theoretic Approach. In: *IEEE Transaction on Mobile Computing*, 18(4), 771 – 786. (2018)
18. Chen, B.-S.: *Stochastic Game Strategies and their Applications*. CRC Press. (2019)
19. Kravets, P., Pasichnyk, V., Kunanets, N., Veretennikova, N. Game Method of Event Synchronization in Multiagent Systems. In: *Advances in Intelligent Systems and Computing (AISC), ICCSEEA 2019 – Proceedings*, 938, 378 – 387. (2019)
20. Kravets, P., Burov, Y., Lytvyn, V., Vysotska V.: Gaming Method of Ontology Clusterization. In: *Webology*, 16(1), 55 – 76. (2019)
21. Gao, J., You, F.: A Stochastic Game Theoretic Framework for Decentralized Optimization of Multi-Stakeholder Supply Chain Under Uncertainty. In: *Compute & Chemical Engineering*, 122, 31 – 46. (2019)
22. Lalropuia, K. C., Gupta, V.: Modeling Cyber-Physical Attacks Based on Stochastic Game and Markov Processes. In: *Reliability Engineering & System Safety*, 181, 28 – 37. (2019)
23. Lozovanu, D. Pure and Mixed Stationary Nash Equilibria for Average Stochastic Positional Games. In: *Frontiers of Dynamic Games*, 131 – 155. (2019)
24. Garrec, T. Communicating Zero-Sum Product Stochastic Games. In: *Journal of Mathematical Analysis and Applications*, 477(1), 60 – 84. (2019)

25. Kloosterman, A. Cooperation in Stochastic Games: a Prisoner's Dilemma Experiment. In: *Experimental Economics*, 1 – 21. (2019)
26. Lin, X., Adams, S. C., Beling, P. A. Multi-Agent Inverse Reinforcement Learning for Certain General-Sum Stochastic Games. In: *Journal of Artificial Intelligence Research*, 66, 473 – 502. (2019)
27. Saldi, N., Basar, T., Raginsky, M. Approximate Nash Equilibria in Partially Observed Stochastic Games with Mean-Field Interactions. In: *Mathematics of Operations Research*, 44(3), 1006 – 1033. (2019)
28. Yamamoto, Y. Stochastic Games with Hidden States. In: *Theoretical Economics*, 14, 1115 – 1167. (2019)
29. Fudenberg, D., Levine, D.K.: *The Theory of Learning in Games*. Cambridge. (1998)
30. Puterman, M. L.: *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley & Sons, New York. (2005)
31. Hu, J., Wellman, M. P., Nash Q-learning for general-sum stochastic games. In: *Machine Learning Research*, 4, 1039 – 1069. (2003)
32. Weinberg, M., Rosenschein, J.S.: Best-Response Multiagent Learning in Non-Stationary Environments. In: *AAMAS'04*, New York, USA. (2004)
33. Greenwald, A., Hall, K.: Correlated Q-learning. In: *Proceedings of the Twentieth International Conference on Machine Learning*, 242 – 249. (2003)
34. Kushner, H., Yin, G. G.: *Stochastic Approximation and Recursive Algorithms and Applications*. Springer Science & Business Media. (2013)
35. Neogy, S. K., Bapat, R. B., Dubey, D.: *Mathematical Programming and Game Theory*. Springer. (2018)

Information-measuring technologies in the metrological support of heat flux measurements

Vitalii Babak^[0000-0002-9066-4307], Svitlana Kovtun^[0000-0002-6596-3460],
Oleg Dekusha^[0000-0003-3836-0485]

Institute of Engineering Thermophysics of NAS of Ukraine, Marii Kapnist Str. 2a, 03057 Kyiv,
Ukraine
vdoe@ukr.net, KovtunSI@nas.gov.ua, ODekusha@nas.gov.ua

Abstract. Heat flux measurements are importance for energy, construction, space research, medicine, etc. The range of measured values is from several units to hundreds of thousands W/m^2 . The use of appropriate heat flux sensors makes it possible to determine heat flux local in space and time, to measure the components of the heat flow and coefficient of heat exchange and coefficient of heat transfer during complex heat exchange. For their correct application and obtaining reliable measurement results, it is necessary to provide periodically monitoring of metrological characteristics of heat flux sensors.

This paper considers the information-measuring technology of metrological support of heat flux measuring, which allows you to adapt the basic principles of metrology of thermal quantities to specific spatial and temporal arguments of the implementation of the heat flux measuring process. The concept of modular construction of the metrological complex is proposed, the features of which are the separation of the measurement range into parts using various methods of generating and transmitting thermal energy and a single module for recording and processing measurement information. Such an approach made it possible to expand boundaries of the measurement range while ensuring the established uncertainty values of the measurement results. The information-measuring system for implement this concept is proposed. It consists of structurally completed modules of thermal blocks, in which a certain principle of thermal energy generation is implemented, and a set of standard devices for maintaining and monitoring the parameters of the given system operation modes. The software module of the developed system allows you to register and process measurement information.

The created system allows studies of the conversion coefficient of heat flux sensors with varying temperature values. The experimental results of the temperature dependences of the conversion coefficient of a series of heat flux sensors with different sensitivity are also presenting in paper.

Keywords: heat flux density, heat flux sensor, metrological provision, measurement, information-measuring technology.

1 Introduction

The informative parameters that are recorded in the research and modernization of energy-intensive technologies, and the introduction of new energy-efficient materials are temperature and heat flow. The second characterizes the process of thermal energy propagation in the object of study, its heat exchange with the environment or other objects, as well as the intensity of the processes that take place.

In particular, in the construction industry, the quality of building materials is monitored by measuring the surface heat flux, control of local and transmission heat losses of enclosing structures of buildings [1...5]; in medicine – diagnostics of various diseases in traumatology, oncology, etc., determining the thermal comfort of the body [6, 7]; in space research determines the thermal state of objects [8]; in thermal power engineering - control of local heat losses and determining the quality of thermal insulation, control of integral heat losses of energy objects during operation, etc. [9...11].

For solve the problems of the building industry and medicine, it is important to measure small values of the heat flux density in the range $(1 \div 1\,000) \text{ W/m}^2$, in aero space research $(100 \div 20\,000) \text{ W/m}^2$, for control of energy facilities is actual to measure of the heat flux up to $200\,000 \text{ W/m}^2$. In all these cases, the value of the surface heat flux is an informative characteristic, which is measured using appropriate sensors in control and diagnostics systems and as a sensitive element of thermophysical devices.

At present, the issue of ensuring the uniformity of measurements of surface heat flux at the state level is being addressed in several countries of world. The metrological provision of heat flux in the United States deals the National Institute of Standards and Technology (NIST). Calibration of measuring instruments for heat flux is by radiation method with determination of the unit of measurement by the Stefan-Boltzmann law [12, 13]. The National Organization of the Netherlands TNO is carried out of calibration of heat flux measuring instruments by the conductive method with determination of the unit of measurement according to the Fourier law [14]. The reproduction of a unit of measurement in Russia is carried out by the conductive method of determining the heat flux density by the Fourier law [15]. In Ukraine the metrological provision of heat flux is based on the radiation method with determining the unit of measure by the Stefan-Boltzmann law [16]. In sum, unit size reproduction in the world is implemented in the range of values from 10 to 100 thousand W/m^2 .

Therefore, it should be noted that today metrology is provided for measuring the surface density of the heat flux in the range, the boundaries of which are orders of magnitude narrower than required. In addition, the high-intensity heat flux is reproduced mainly by models of black body, the structures of which today have a diameter of the output diaphragm of the order of 20...80 mm [17...19]. This places restrictions on the calibration of heat flow sensors larger than the diameter of the power source diaphragm. In this regard, the issue of ensuring the unity of measurements of heat flux is becoming a priority.

The aim of the work is to improve metrological support for heat flux measurements. Important here is the realization of the possibility of carrying out calibrations of measuring instruments of different types, expanding the range of measurements and minimizing resource costs.

2 Theoretical basis and Methodology of surface heat flux measurement

Unlike classical approach, which is mainly focused on static measurements, increasing their sensitivity and precision by improving the elemental base and developing physical and technical principles of measurement, trends in the modern metrology require an increase in the measurement automation, which means is to provide not only the measurement of physical quantities, but also their physical principles, carrying out the necessary calculations, displaying the obtained results in the required form and ting them to destination or storage.

In the modern realization of measurements to the first place comes information-measuring technologies, which are a set of measurement methods, hardware and software, integrated for the purpose of collecting, processing, storing, displaying and using measurement information in the form of measurement results with a certain level of uncertainty for the benefit of consumers.

The information-measuring technology of surface heat flux measurements metrological support, the constituent elements of which is presented in Fig. 1. Technology based on the justified use of different physical principles of reproducing a unit of surface heat flux density and transmitting its size in different parts of the extended dynamic range while providing a set level of uncertainty of the measurement results.

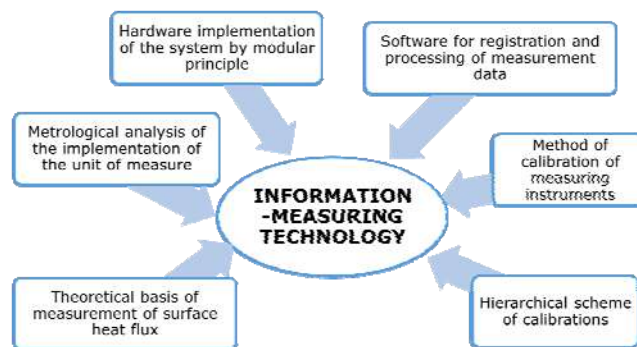


Fig. 1. Components of information-measuring technology for metrological support for measurements of surface heat flux

The theoretical basis of surface heat flux measurements in a wide range is based on the development of the basic elements of metrology of thermal quantities, in particular, the theoretical substantiation of methods and means of reproduction and transmission of the measurement unit, improvement of the hardware base for measuring instruments calibration and development of appropriate methodological tools. Fig. 2 are shown proposed methodology for ensuring the unity of measurements of the surface heat flux, which allows us to formulate the main perspective directions for the development of metrological support.

The methodology is based on three directions of development of scientific and applied foundations of ensuring uniformity of measurements, namely: theoretical, hardware and software and methodological support. Each of these directions involves solving a number of specific problems that take into account the peculiarities of the subject area of application of the measurements of the surface heat flux density, the methods and means of applied measurement, their characteristics.

In the process of reproduction and transmission of the unit of the heat flux on the surface of the test instrument, the field of the heat flux induced by the heat source is formed. The mechanism of formation of the field of heat flux is determined by the following main phenomena of heat transfer: thermal conductivity, radiation and convection, which describe the way of transfer of kinetic energy of particles (molecules, atoms, electrons, photons, etc.) during their thermal motion. The intensity and direction of energy transfer are determined by the heat flux vector.

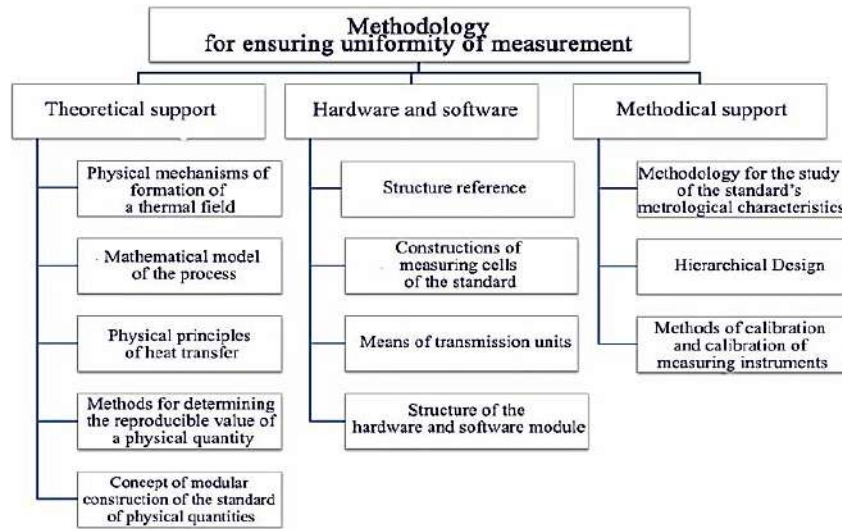


Fig. 2. Methodology for ensuring unity of heat flux measurement

Surface heat flux density is an energy flow that is transferred in the process of heat exchange through a unit of isothermal surface area over a period of time dt , and, irrespective of the method of forming thermal energy, in the general form can be determined by:

$$q(\mathbf{r}, \gamma_i, \tau) = F^{-1} \cdot \frac{dQ(\mathbf{r}, \gamma_i, \tau)}{dt} + \xi(\mathbf{r}, \gamma_i, \tau, F, Q), \quad (1)$$

where F – surface area of heat exchange; Q – heat flow through a unit of surface area of heat transfer; γ – input values to be measured to determine the magnitude of the surface heat flux density; $\xi(\mathbf{r}, \gamma_i, \tau, F, Q)$ – a random field that determines the effect of a set of factors on the formation of a unit of measure.

Equation (1) determines the method of obtaining the value of the surface heat flux, taking into account the action of various factors of influence, each member of which determines a set of input quantities and the physical effects that accompany the measurement process.

The implementation of the measurement unit depends on the precision of the measurement of the input quantities and the intensity of the action of a set of factors, each of which is characterized by such an indicator of accuracy as uncertainty. Given that each input value and each impact factor contains a certain uncertainty, the uncertainties of their totality have an effect on the total uncertainty of the results of the reproduction of the unit of surface heat flux density.

The summary measurement uncertainty is formed from the standard uncertainties of the components included in the equation for determining the surface flux density (1), and assuming that all the components of the uncertainty in (2) are uncorrelated, defined as:

$$u(q) = \sqrt{\sum_{i=1}^n k_i^2 u(\gamma_i)^2} , \quad (2)$$

where $u(q)$ – the total uncertainty of the implementation of the unit of measurement over a wide range; k_i – the coefficients of influence of the i -th component of determining the surface heat flux density; $u(\gamma_i)$ – the standard uncertainty of the individual components of determining the surface flux density.

The analysis of the world experience of the measurement unit realization [20...25] of the surface heat flux with the formation of heat energy by both radiation and conductive method is established that there is no single method of formation of thermal energy that would allow to reproduce the unit of measurement in a wide dynamic range without deterioration of such indices as uncertainty of measurement result .

For this reason it was proposed to divide the dynamic range by the heat flux density into parts, in each of which the reproduction of a unit of measurement occurs using different physical principles. Based on this, the concept of constructing a surface heat density standard on a modular principle was developed.

Therefore, to ensure the uniformity of measurements in the present conditions, the standard of surface heat flux should be implemented in the form of information-measuring system [26], which is a set of functionally-integrated measuring, control, computing and other auxiliary technical means formed to obtain measurement information, its transformation, processing, visualization and documentation. This principle ensures the technical and information-interoperability of the modules, simplifies maintenance, improves the accuracy and reliability of the system and allows to extend the functionality.

3 Implementation of conceptual model

The implementation of the proposed conceptual model of the standard of surface heat flux density involves different physical principles for the generation and transmission of thermal energy. To ensure uniformity of measurements across the range $1 \dots 50 \text{ W/m}^2$. It is proposed to use the model of absolutely black body as a source of thermal energy and to transmit the unit of measurement to the heat flux sensors in a radiation way. Determining the value of the surface heat flux reproduced in this way is carried out based on the Stefan-Boltzmann law.

Implementation of the unit of measurement in the range $2 \cdot 10^1 \dots 2 \cdot 10^4 \text{ W/m}^2$ made by the method of hot guarded plate by forming a heat flow using an electric heater, while the transfer of the unit size is carried out in a conductive way and the use of heat flux sensors. Determination of the magnitude of the reproduced value of the surface heat flux density is based on the Joule-Lenz law.

In the range $1 \cdot 10^4 \dots 2 \cdot 10^5 \text{ W/m}^2$ the formation of heat flow is proposed to be carried out by means of an emitter, whereby the unit size is transmitted in a radiation way. Both heat flux sensors and thermal radiation receivers can serve as transmitters.

The modular principle of construction of the information-measuring system allows to form a standard of surface heat flux density in a form structurally completed modules of thermal blocks in which the certain principle of formation of thermal energy is realized, and a set of standard devices for maintaining and controlling the parameters of the given modes of operation of the system. In Fig. 3 presented a scheme for constructing a standard modular principle.

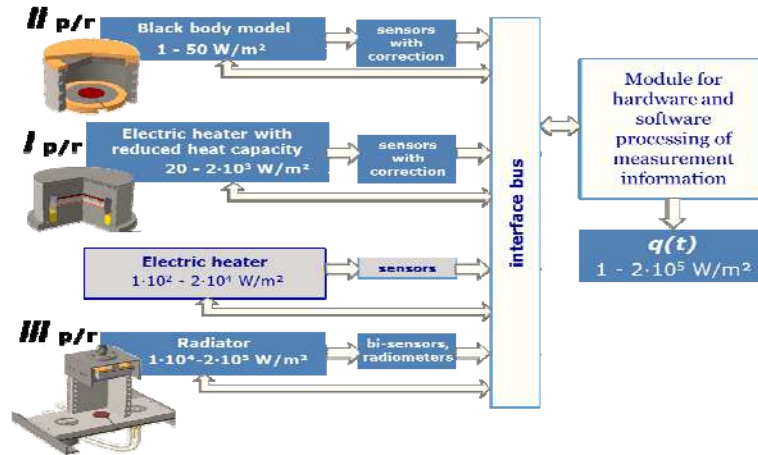


Fig. 3. Scheme of modular construction of the standard

According to this scheme, thermal blocks are developed in each part of the range, which creates conditions for reproducing the unit of surface heat flux density and transmitting its size to the appropriate measuring instruments.

I part of range. The thermal block contains an isothermal source of thermal energy to create first-order boundary conditions on the surface of the measuring instru-

ment (the heat flux sensor under test). The measurement unit is transmitted using only conductive method. In this case, the heat flux sensor has the same geometric dimensions in the radial direction as the working surface of the heat source.

The conductive method of using an electric heater is a significant advantage of reproducing a unit of surface heat flux density and transmitting its size. The most significant of these are the high accuracy of the task and the maintenance at a certain level of electrical quantities involved in the formation of heat flow, and the possibility of implementing the unit of measurement over a wide temperature range. The heat flux is formed on the main heater due to the Joule effect when passing electric current through it. Electric current is supplied from the power supply unit. To determine the value of the input power of the heater included consistently unambiguous measure of electrical resistance. Thus, power is determined by measuring the voltage at the contacts of the primary heater and at the potential terminals of a single measure of electrical resistance.

Structural model of formation of measurement results uncertainty by the conductive method in Fig. 4 is presented as a cause and effect Isikawa diagram.

The elements of the diagram are the input factors and quantities that directly measure to determine the surface heat flux density, as well as the physical effects that have an impact on the end result - the accuracy of the unit size reproduction.

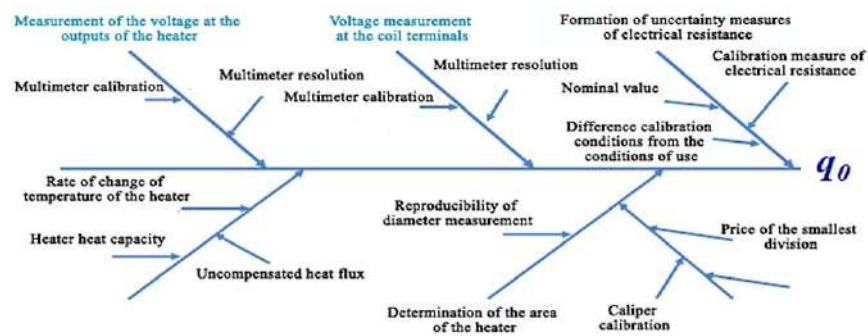


Fig. 4. Scheme of the formation of heat flux measurement results uncertainty by conductive method

It is established that the main sources of uncertainty of the measurement result are:

Measurement of voltage at the inputs of the heater by voltmeter; calibration of the voltmeter in the range of measured values of voltage; digit voltmeter; measuring the voltage at the inputs of a measure of electrical resistance by a voltmeter; calibration of the voltmeter in the range of measured voltage values at the terminals of the measure of electrical resistance; calibration of the measure of electrical resistance (when determining the thermal power of the electric current supplied to the heater).

Measuring the diameter of the heater with a caliper; calibration of the caliper (when determining the heater area). The heat losses are caused by temperature drift, heat capacity of the heater and uncompensated heat flow as additional sources.

The components of the heat flux reproduction uncertainty have been estimated. Calculations based on the results of the study of metrological characteristics of the

standard heat flux sensors and passport data on the standard devices. The estimation results for the three values of the surface heat flux density are presented in Table 1. To compare the calculated results, the estimates are given as a percentage of the contribution of each component of uncertainty from the action of the respective source to the total uncertainty of the measurement results taken as 100%.

As we can see, with the decrease in the value of the reproduced value of the surface density of the heat flux, the influence of the heat capacity of the main heater increases, and already by the values $q = 10 \text{ W/m}^2$ has a dominant influence.

Table 1. Estimates of the constituent sources contribution of uncertainty to the summary uncertainty

Sources of uncertainty	Percentage in summary uncertainty		
	10 W/m ²	100 W/m ²	2·10 ⁴ W/m ²
Measuring the diameter of the main heater	3	9	20
Measurement of voltage at the inputs of the main heater	1	4	11
Measuring the voltage at the inputs of the resistance coil	1	8	18
Calibration of the measure of electrical resistance	3	14	30
Uncompensated heat flow	6	23	2
Thermal losses due to the heat capacity of the main heater	86	41	18

Reducing the heat capacity of the main heater can be achieved by changing its thickness. However, this method is limited by the technological processes of manufacturing the heater and its finite size.

II part of range. Another way to reduce this effect is to avoid direct contact between the surfaces of the heat source and the sensor being tested, which is possible when using the radiation method of heat flow formation. Therefore, the second part of the range requires the creation of a thermal block with a low-intensity thermal radiation source. This source can be a model of a completely black body. The most productive and versatile for the reproduction of the surface heat flux density by the radiation method is a scheme for the construction of a thermal block with the use of a measuring cell formed by the flat designs of the heat source and heat sink on which the sensor under study is mounted and a side screen [27].

The scheme that implements the radiation method of reproducing the normalized value of thermal energy, assumes the presence in the device of a closed space formed by two diffusely radiating surfaces.

Determination of the surface heat flux density in the radiation method of supplying thermal energy is carried out on the basis of the Stefan-Boltzmann law. Therefore, the formation of heat flux occurs by maintaining the temperature difference between the heat source and the heat sink, taking into account the energy emitted by the heat transfer surfaces, their relative location and the thermo-radiation characteristics.

Fig. 5 shows a constructive model of the uncertainty formation of the result of surface heat density reproduction by the radiation method.

From Fig. 5 it follows that the main factors of influence on the uniform distribution of the heat flow field in the absolute method of measurement include:

- contribution of uncontrolled convective or convective-conductive components into the heat flow;
- the ratio between the dimensions of the cavity of the radiator: the diameter of the heat-sensing surface of the heat sink and the height of the screen;
- the nature of the distribution of local angular coefficients on the surface of the heat sink on which the sensor is located;
- the ratio of the values of the radiator temperature and the temperature control of the heat sink;
- thermo-radiation characteristics of heat exchange surfaces.

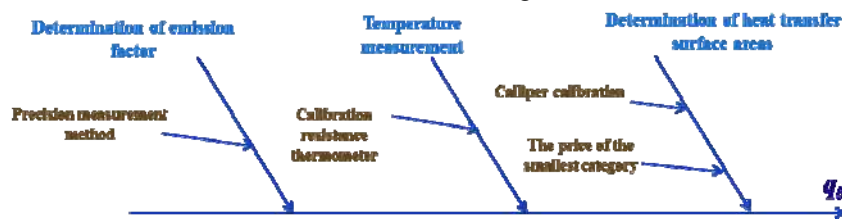


Fig. 5. Scheme of the formation of measurement results uncertainty by radiation method

The thermal energy source is made in the form of a flat model of a black body. A feature of this scheme is the use of a side black screen, thermostated at the same temperature as the heat sink. This allows to ensure almost complete absence of the convective component in the location of the investigated sensor, which contributes to the formation of a uniform field of flow of thermal radiation on the surface of the heat sink under conditions of low heat exchange. In this case, the design of the measuring cell of the heat block provides for the possibility of placing heat flow sensors in it without rigid requirements to the radial size of the means of transmitting the unit of measurement, but there are limitations regarding their inertia.

III part of range. The heat block should contain a source of high-intensity thermal energy, which causes certain structural features of the construction of the measuring cell. At high-intensity heat exchange, which is observed in the reproduction of heat flow in the range of values $(1 \cdot 10^3 - 2 \cdot 10^5) \text{ W/m}^2$, the use of the conductive method is also limited. In this case, the limitation is the heat resistance of the measuring cell materials and the test instrument. Therefore, in accordance with the structure shown in Fig. 3, in the third part of the range, it is proposed to produce a high-intensity heat flux by radiation.

The formation of high-intensity heat flux is possible by using a heat source with a temperature on the surface of the radiation body of at least 650... 1100 K. This characteristic is possessed by high-temperature models of absolutely black body and incandescent lamps.

The main sources of uncertainty in this case are the uneven distribution of the heat flux density entering the heat sink, as well as the contribution of the convective component to the resulting heat flux. Comparison of the characteristics of thermal emitters

[28], taking into account the redistribution of energy transfer methods, revealed the advantages of quartz halogen incandescent lamps, which have the lowest percentage of convection heat transfer, which has a significant effect on the formation of a homogeneous heat flux field on the heat sensing. For this part of the range, it is possible to calibrate contact measuring instruments that have adequate thermal stability and thermally independent sensitivity.

Creating a module of the third part of the range, the thermal block of which is essentially a measuring comparator, opens the way to the implementation of verification of standard means of contact and contactless measurements.

4 Hardware and software module of Information-measuring system

Implementation of the hardware part of the heat flux density standard surface is presented on Fig. 6. It is a set of functionally integrated three thermal and electronic units, ancillary equipment and a personal computer with appropriate software.



SPECIFICATIONS

Measurement range, W/m^2 :

I part of range $1 \div 50$;

II part of range $2 \cdot 10^1 \div 2 \cdot 10^3$;

III part of range $1 \cdot 10^4 \div 2 \cdot 10^5$

Extended uncertainty, %:

I part of range 0,6;

II part of range 0,6;

III part of range 0,7

Fig. 6. Implementation of the hardware part of the heat flux density standard

The thermal blocks provide the implementation of methods of reproduction and transmission of the unit of measurement of the surface heat flux density over a wide dynamic range. To convey the unit size in a specific part of the range, the heat flux sensors or thermal radiation receivers are placed in the appropriate measuring cell and provide the necessary fixed operating modes for temperature and surface heat flux density.

The regulation of the necessary modes of operation of the standard in terms of temperature and density of heat flow, registration, transmission, processing and archiving of measurement information is carried out using the hardware and software module of the system, the structure of which is presented in Fig. 7.

The hardware part of the hardware and software module of the system consists of task modules and maintenance of measurement modes and modules of primary data registration.

The measurement and maintenance modes module consists of four units that perform the functions of setting and controlling the thermal modes of the system and the corresponding boundary conditions.

The unit of reference and maintenance of the reference temperature is realized by means of the temperature controller of the reference couplings of the thermocouples, which controls the thermostat. The scheme is implemented with a pre-amplifier signal of the resistance thermometer and regulator-meter.

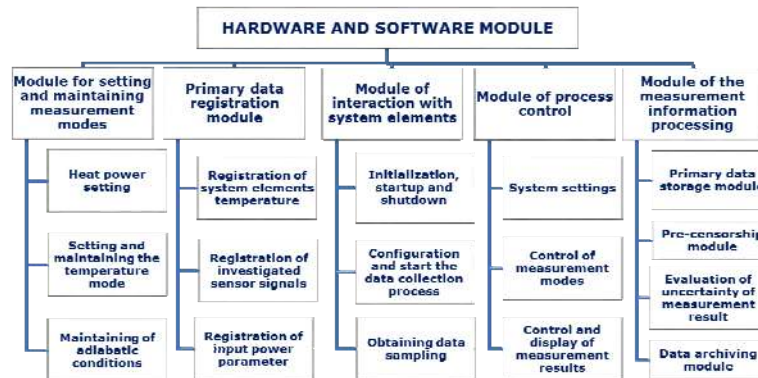


Fig. 7. Hardware software module of the system

The unit for maintaining adiabatic conditions is built on a controller that is designed to control the power supply to the guard heater. Due to the possibility of regulating the power of the guard heater in two ways in the control system is a two-position signal switch and two pre signal amplifiers, respectively, differential thermocouple and auxiliary sensor thermocouple. To set and maintain the temperature, a heat sink temperature controller and a protective screen temperature controller are used.

The heat sink temperature controller is used to control the heat output of the heat sink assembly according to the resistance thermometer.

The protective screen temperature controller is designed to regulate the temperature of the thermal screen according to the readings of the resistance thermometer built into the screen.

To regulate the temperature of the couplings of thermocouples, the heater, the heat sink and the protective screen, four standard RT-0102 model regulators are used, and Pt1000 thermocouples are used as resistance thermometers.

The thermal power task unit is implemented based on a standardized stable voltage source.

The module of registration of primary data consists of blocks of registration of temperature of elements of system, registration of parameters of power, registration of signals of the investigated sensors. ICP DAS I-7018 modules have been used to provide temperature recording of system elements and to register the signals of the sensors under study:

- 8-channel ADC with a bit rate of 16 bits and a conversion frequency of 10 Hz;
- dynamic range adjustment and calibration;

- support for RS-485 industrial trunk interface and addressing, which makes it possible to create a measuring network.

The power parameter recording unit is based on the Picotest precision digital multimeter M3500A model with a relative measurement error of 0.0015 / 0.0004, in%, in the voltage range from 0 V to 10 V and 0.002/ 0.0006, in %, in the range from 10 V to 100V, Fluke 8846A digital multimeter with an error of 0.0037/0.0035 within ± 100 mV in voltage measurement mode and uniquely measured electrical resistance P310 of 0.1 Ohm class of 0.01.

The system software is made in a modular structure, which made it possible to connect and integrate into the main program code of previously created dynamic libraries.

The developed software includes modules that are functionally divided into three groups: module of interaction with system elements, module of process control, module of the measurement information processing.

The system interaction module provides software coordination and data exchange with hardware by sequentially calling functions that meet different tasks:

- initial setup and initialization of the interface
- search for ID and initialization of individual I-7018 modules in the RS-485 network;
- adjusting the dynamic range and frequency of the I-7018 ADC module polls;
- Setting the poll cycle time, response timeout, data type, measurement channel poll queue that can be saved and downloaded from configuration files.

The resulting data sets are transmitted to the software modules that implement the control and processing functions of the measurement information - these are the modules for controlling the process of the system operation and processing of the measurement information.

The measurement information processing unit consists of a pre-censorship module of the measurement data samples, a primary data storage module, and a data archive module.

The pre-censorship module of measurement data is intended to avoid cases of storage of data with errors as a result of transmission through RS-232/485 and exclusion of measurement data with excess errors. The first problem is solved by converting the received ANSI string to a number, if at this stage an error occurs, the request is transmitted to the process control module. To solve the second problem, statistical criteria are used, which allow us to calculate marginal sample values.

The primary data storage module stores the input values of the signals used for further processing in the form of matrices, which allows to improve the processing efficiency of the data and the system as a whole, as well as normalize the input data to bring it to a predetermined range of values.

The data-archiving module is responsible for storing input values of signals which allows to not lose data in case of hardware or software failure. In addition to archiving, the operator commands automatic storage of data.

The process control module of the system provides general control of the modes of operation, including in the interactive mode, displaying the results of measurements using 2D graphs and tables

5 Experimental studies

With the use of the information-measuring system, experimental studies of the temperature dependence of the sensitivity of a number of heat flux sensors, thermocouples of which are made of copel and chromel thermoelectrode wires with nickel coating, with the variation of the ratio of the cross-sectional area temperature (30 ÷ 200) °C. In Fig. 8 the results of the experimental studies in the form of a dependence of the conversion factor of a series of heat flow sensors are presented.

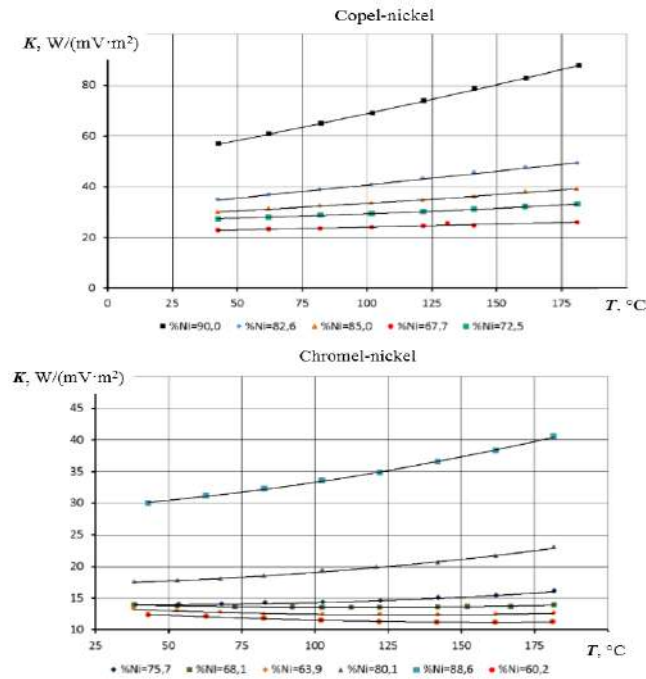


Fig. 8. Temperature dependences of the conversion factor of heat flow sensors with variation of the reduced cross-sectional area of the thermoelectrode

At each point in the range, five series repeated measurements were conducted. Expanded uncertainty of measurement results is less than 1.8 % at a coverage coefficient 2. As a result of such studies, it becomes possible not only to determine the metrological characteristics of the heat flux sensors, but also to select heat flux sensors whose characteristics will be optimal for the conditions of their subsequent operation.

6 Conclusions

The proposed information and measurement technology combines theoretical basis, software, proposed hardware for the implementation of the standard modular principle, its metrological analysis and the scheme of transferring the size of the unit

of measurement, techniques for the study of metrological characteristics of the measuring instruments, and allows to adapt the main provisions of metrology specific spatial and temporal arguments for the implementation of the process of measuring the surface heat flux density.

A mathematical model is proposed that illustrates the relationship between the reproduction unit of the unit of measurement and individual components of the uncertainty that affects the accuracy of forming the unit of measurement for different heat transfer processes. The uncertainty of the measurement of the surface density of the heat flux and the main factors of influence has been studied using a cause and effect Isikawa diagram.

On the basis of the conducted researches the perspective directions of development of metrological support of measurements of heat flow are substantiated, in particular, realization of the conceptual model of the standard of surface heat flux density were applied the different physical principles of generation and transmission of thermal energy. It allows to provide measurement of in a wide range of heat flux values.

Hardware and software are presented that implement the concept of modular construction of the surface heat flux density standard in the range of values from 1 W/m² to 200 000 W/m²

References

1. Biddulph, P., Gori, V., Elwell, C. A., Scott, C., Rye, C., Lowe, R., Oreszczyn, T.: Inferring the thermal resistance and effective thermal mass of a wall using frequent temperature and heat flux measurements. *Energy and Buildings*. Vol. 78: pp. 10–16 (2014). doi:10.1016/j.enbuild.2014.04.004
2. Ricci, R., Galatioto, A., Besalduch, L. A., Desogus, G., Di Pilla, L.: Building wall heat capacity measurement through flux sensors. *J. sustain. dev. energy water environ syst.*, 7(1) pp. 44-56 (2019). doi:10.13044/j.sdewes.d6.0234.
3. EN 12667: Thermal performance of building materials and products – Determination of thermal resistance by means of guarded hot plate and heat flow meter methods – Products of high and medium thermal resistance. (2001).
4. Dekusha, L., Kovtun, S., & Dekusha, O.: Heat Flux Control in Non-Stationary Conditions for Industry Applications. *Proceedings of 2nd Ukraine Conference on Electrical and Computer Engineering (UKRCON)*, Lviv, July 2019, IEEE: pp. 601-605 (2019). doi: 10.1109/UKRCON.2019.8879847
5. Babak, V., Dekusha, O., Kovtun, S., Ivanov, S.: Information-measuring system for monitoring thermal resistance. *CEUR Workshop Proceedings*, Vol. 2387: pp. 102-110 (2019). <http://ceur-ws.org/Vol-2387/20190102.pdf>
6. Liu, J., Xu, L.X.: Boundary information based diagnostics on the thermal state of the biological bodies. *Int. Journal Heat Mass Transfer*. Vol. 43, pp. 2827-2839 (2000).
7. Miyakawa, M., Bolomey, J.C.: *Non-invasive thermometry of the human body*. CRC Press Inc. (1996)
8. Lin, T.C.: Influence of Laminar Boundary-Layer Transition on Entry Vehicle Designs. *Journal of Spacecraft and Rockets*. Vol. 45 (2), pp. 165-175 (2008).
9. Pontt, J.: Diagnostics of Insulation Condition and Risk Evaluation of Electrical Machines Employed in Mining Application. *IEEE Industry Applications Society Annual Meeting. IEEE Conferences IAS 2009: 1-3* (2009)

10. Jussila H.K., Mityakov, A.V., Sapozhnikov, S.Z., Mityakov, V.Y., Pyrhönen, J.: Local Heat Flux Measurement in a Permanent Magnet Motor at No Load. *Transactions on Industrial Electronics*. Vol. 60 (11), IEEE: pp. 4852 – 4860 (2013).
11. Howey, D.A., Childs, P.R. N., Holmes, A.S.: Air-gap convection in rotating electrical machines. *Transactions on Industrial Electronics*. Vol. 59 (3), IEEE: pp. 1367–1375 (2012).
12. Tsai, B.K., Gibson, C.E., Murthy, A.V. & others: NIST MEASUREMENT SERVICES: Heat-Flux Sensor Calibration. NIST Special Publication 250-65 (2004).
13. Heat Flux Meters: Calibration. Nordtest method. NT FIRE 050. NORTEST, Helsinki (1995).
14. Cherepanov, V.Ya.: Gosudarstvennyiy pervichnyiy etalon edinitsyi poverhnostnoy plotnosti teplovogo potoka. *Doklad Federalnomu Agentstvu po tekhnicheskomu regulirovaniyu i metrologii (FGUP «SNIIM»)*, Novosibirsk (2007).
15. ASTM C1130-07:2012. Standard Practice for Calibrating Thin Heat Flux Transducers. ASTM International (2012).
16. Nazarenko, L.A., Polevoi, V.I., Bondarenko, L.I.: Derzhavnii spetsialnii etalon odinitsi energetichnoyi osvitchenosti nekogerentnym vyrominyuvannym. *Ukrainskii metrologichnii jurnal*. Vol. 1: pp. 31–36 (1995).
17. Zarr, R., Carvajal, S., & Filliben, J.: Sensitivity Analysis of Factors Affecting the Calibration of Heat-Flow-Meter Apparatus. *Journal of Testing and Evaluation*. V. 47: 20170588 (2019). doi:10.1520/JTE20170588.
18. Murthy, A.V., Tsai, B.K., Saunders, R.D.: Facility for calibrating heat flux sensors at NIST: an overview. *Proc. ASME Heat Transfer Division*, New York: ASME, Vol. 3: pp. 159-164 (1997).
19. Olsson, S.: Calibration of Radiant Heat Flux Meters-The Development of Water Cooled Aperture for Use with Black Body Cavities. SP Report 1991:58, Nordtest Project 873-90, Swedish National Testing and Research Institute, Sweden (1991).
20. Turzó-András, E., Magyarlaki, T., Németh, S., Kovács, T.: Intercomparison of heat flux sensors. Report of EUROMET Project No 426, Hungarian Trade Licensing Office, Thermometry Department (2010).
21. Turzo-Andras, E., Blokland, H., Hammerschmidt, U. et al.: Round-Robin Test of Heat Flux Sensors. *International Journal of Thermophysics*. Vol. 32: pp. 2708–2715 (2011). doi:10.1007/s10765-011-1078-8.
22. Hammerschmidt, U.: Guarded Hot Plate (GHP) Method: Uncertainty Assessment. *International Journal of Thermophysics*, Vol. 23 (6), (2002).
23. Anderson, P., Wetterlund, I.: Uncertainty in Heat Flux Calibrations performed according to NT FIRE 050. Nordtest Project № 1525-01, SP Swedish National Testing and Research Institute. SP Report 2001:34.
24. Murthy, A.V., Tsai, B.K., Saunders R.D.: Radiative Calibration of Heat-Flux Sensors at NIST: Facilities and Techniques. *Journal of Research of the National Institute of Standards and Technology*. Vol. 105 (2), pp. 293-305 (2000).
25. Osadchy, S.M., Potapov, B.G.: The state standard of the unit of density of the radiation heat flux in the range from 5 to 2500 kW/m². *Almanac of Modern Metrology*, Vol. 12: pp 65-72 (2017).
26. Babak, V., Babak, S., Yeremenko, V.: *Teoretichni osnovy informatsiyno-vimiryvalnyh system: Pidruchnyk*, NAU, Kyiv (2017).
27. Kovtun, S., Dekusha, L., & Vorobiov, L.: Features of reproduction the heat flux density unit of measurement by radiation method. *The scientific heritage*, P.1 (22): pp 51-55 (2018).
28. Ayzenberg Yu.: *Spravochnaya kniga po svetotekhnike*, Znack, Moscou (2006).

Inter-domain routing method under normalized Quality of Service based on hierarchical coordination

Oleksandr Lemeshko¹[0000-0002-0609-6520], Maryna Yevdokymenko¹[0000-0002-7391-3068],
Zhengbing Hu²[0000-0002-6140-3351] and Oleksandra Yeremenko¹[0000-0003-3721-8188]

¹ Kharkiv National University of Radio Electronics, Kharkiv, 14 Nauky Ave., Ukraine

² Central China Normal University, Wuhan, Hubei, P.R.China

oleksandr.lemeshko.ua@ieee.org, maryna.yevdokymenko@ieee.org,
hzb@mail.ccnu.edu.cn, oleksandra.yeremenko.ua@ieee.org

Abstract. The inter-domain routing method under normalized Quality of Service (QoS) based on hierarchical coordination in a Software-Defined Network (SD-WAN, Hybrid SD-WAN) is proposed in the paper. The novelty of the method is that the routing solutions are aimed at ensuring the normalized QoS in terms of average transmission rate and average end-to-end packet delay. The method is based on the use of a decomposition flow-based routing model, which consists of the inter-domain routing interaction and ensuring the normalized QoS conditions that obtained during tensor network modeling. Therefore, the inter-domain QoS routing problem was presented as an optimization problem with a quadratic optimality criterion. A number of numerical examples confirmed the efficiency and effectiveness of the proposed method in terms of providing the normalized QoS within the finite number of iterations. Reducing the number of such iterations helps to decrease the amount of service traffic transferred between routers and SDN controllers at different levels, as well as minimize the time for solving the inter-domain QoS-routing task.

Keywords: Method, Inter-domain routing, Hierarchical coordination, Quality of Service, End-to-end delay, SDN.

1 Introduction

Providing a specified level of Quality of Service for user requests has been and remains the primary purpose of the functioning of modern infocommunication networks (ICNs), in which various technological means and protocols of distribution and reservation of network resources, traffic management, etc. are involved. At the same time, the problem of QoS provision is exacerbated in the conditions of considerable territorial distribution and heterogeneity of ICNs, which significantly influence the scalability of traffic management solutions.

An effective way to increase scalability is to use multi-domain Software-Defined Networks (SDN) when a significant list of traffic management functions is translated

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

to many network operating system controllers, which in turn are built into a clear functional hierarchy [1-6].

Assigning a separate controller for each network domain contributes to a significant reduction in the amount of service traffic circulating on the network, the size of the routing tables, and improving the timeliness of traffic management tasks solutions, among which routing is of an important consideration. On the other hand, full implementation of these advantages requires a radical modernization of models, methods, and protocols of routing with their functional adaptation to the features of the hierarchical architecture of SDN-controllers and the multidomain structure of ICNs.

2 Overview of existing solutions regarding hierarchical routing

Currently the most well-known protocols that implement the principles of hierarchical multi-domain routing are considered to be OSPF and Integrated IS-IS used in IP networks, as well as the somewhat outdated PNNI in the ATM network [4-6]. However, these protocols still use Dijkstra's combinatorial algorithm to calculate routes, which does not take into account the hierarchical features of network construction. It should also be added that QoS requirements in the same protocols are organized indirectly only through routing metrics, which are usually related to the bandwidth of ICN links and paths. Implementing such an approach does help improve the overall QoS level on the network, but does not guarantee the numerical values of the end-to-end QoS indicators, such as the average packet delay for a particular flow.

A number of solutions have been proposed in scientific papers on hierarchical routing [7-14], which are mainly represented by flow-based models and optimization methods of calculation. In articles [7, 8, 12-14], such solutions are adapted to the features of SDN architectures. An important feature of the results obtained in [9-11] is that they are based on the use of the provisions, principles, and postulates of the theory of hierarchical multilevel systems [15, 16]. First, these solutions impose several hierarchical levels of decision-making on routing in ICNs. Second, they imply a decomposition representation of a network mathematical model, which can be described, for example, by a system of differential equations [17]. Third, a mandatory component of hierarchical routing is the coordination process, which is implemented by the top level in relation to the routing decisions of the lower levels.

From the point of view of QoS routing implementation the approach [18-24] is particularly noteworthy since is based on the use of tensor research methodology, within which it was possible to obtain in an analytical form the conditions for ensuring the Quality of Service on a number of indicators – bandwidth, average delay, probability of packet loss. Therefore, in this paper, we will propose the solution to the problem of hierarchical coordination of inter-domain routing in the Software-Defined infocommunication network with the provision of normalized QoS, that is when the requirements for the level of end-to-end QoS are set for each domain in the form of corresponding norms. The presented solution is a further development and integration of the results obtained in [9-11, 23, 24].

3 Decomposition model of inter-domain QoS routing in Software-Defined infocommunication network

Suppose that the network structure is represented as an oriented graph $G = (R, E)$, where R is the set of vertices that model the routers and E is the set of graph edges that describes the communication links of the network. In the general case, a packet flow is generated when providing a particular information service on the network. We denote the number of flows circulating in the network by K , then $|K| = \tilde{K}$ is the power of the set, which quantitatively characterizes the total number of flows in the ICN. For each k th flow ($k \in K$), its average packet rate (intensity) λ_{req}^k is known, which is measured in packets per second (1/s) and defines the requirements for the ICN bandwidth allocated to that flow.

When developing a decomposition model of inter-domain routing, suppose that the network consists of N interconnected subnets – domains. Then let each p th individual domain in the ICN be described by the graph G subgraph $G^p = (R^p, E^p)$, where $R^p = \{R_i^p; i = \overline{1, m_p}\}$ is the set of routers of the p th domain, $E^p = \{E_{i,j}^p; i, j = \overline{1, m_p}, i \neq j\}$ is the set of links connecting the routers of the p th domain, m_p and n_p are the total numbers of the routers and communication links in the p th domain respectively.

During network decomposition, the boundary between domains passed through network routers, as implemented, for example, in OSPF [10]:

$$R^p \cap R^q \neq \emptyset \text{ and } E^p \cap E^q = \emptyset, \quad p \neq q, \quad (1)$$

that is, some network routers may belong to several adjacent domains at a time. We also define for each p th domain a set of border routers B^p ($B^p \in R^p$). In turn, the entire set of the p th domain border routers can be divided into two subsets: $B_{in}^{p,k}$ is the subset of the border routers through which the packets of the k th flow income into the p th domain; $B_{out}^{p,k}$ is the subset of the border routers through which the packets of the k th flow outgo from the p th domain. For each communication link $E_{i,j}^p$, we denote its bandwidth $\phi_{i,j}^p$, which is measured in packets per second (1/s).

As a result of solving the hierarchical inter-domain routing task for each p th domain, it is necessary to calculate the routing variables $x_{i,j}^{p,k}$ that characterize the fraction of the k th packet flow transmitted in the link $E_{i,j}^p \in E^p$. Then, for each p th domain router, the k th flow conservation conditions must be met to ensure the connectivity of intradomain sections of the inter-domain routes.

If the p th domain is transit for the k th packet flow, then such conditions take the form:

$$\left\{ \begin{array}{l} \sum_{B_{in}^{p,k} \in B^p} \left(\sum_{E_{i,j}^p \in E^p} x_{i,j}^{p,k} - \sum_{E_{j,i}^p \in E^p} x_{j,i}^{p,k} \right) = 1; \\ \sum_{E_{i,j}^p \in E^p} x_{i,j}^{p,k} - \sum_{E_{j,i}^p \in E^p} x_{j,i}^{p,k} = 0; \\ \sum_{B_{out}^{p,k} \in B^p} \left(\sum_{E_{i,j}^p \in E^p} x_{i,j}^{p,k} - \sum_{E_{j,i}^p \in E^p} x_{j,i}^{p,k} \right) = -1. \end{array} \right. \quad (2)$$

The system of equations (2) must be met separately for each k th packet flow. The first condition of the system (2) covers all the border routers through which the k th flow arrives at the p th domain. The second condition in (2) is introduced for internal p th domain routers that are transit for the k th packet flow. The third condition must be satisfied for all border routers through which the k th flow outgoes from the p th domain.

If the k th packet flow arrives at a network through the p th domain, and its source is the router R_i^p , for example, then for this network the first condition of the system (2) will be somewhat simplified and will look like: $\sum_{E_{i,j}^p \in E^p} x_{i,j}^{p,k} = 1$. The rest of the

equations in the system (2) will remain unchanged. When the router R_i^p from the p th domain is the k th packet flow receiver, only the last equation of system (2) will be simplified and will look as follows: $\sum_{E_{j,i}^p \in E^p} x_{j,i}^{p,k} = 1$.

In addition, to prevent links congestion on the p th network domain, it is important to fulfill the following conditions:

$$\sum_{k \in K} \lambda_{req}^k x_{i,j}^{p,k} \leq \phi_{i,j}^p, \quad p = \overline{1, N}. \quad (3)$$

Let us denote as $\lambda_{i,j}^{p,k} = \sum_{k \in K} \lambda_{req}^k x_{i,j}^{p,k}$ as the average intensity of the k th packet flow transmitted in the link $E_{i,j}^p \in E^p$. Then, when implementing multipath routing, the routing variables are imposed by the constraints:

$$0 \leq x_{i,j}^{p,k} \leq 1. \quad (4)$$

Variables (4) are the coordinates of the routing vectors $\vec{x}_p^k$ that set the result of the solution for the k th flow routing problem in the p th domain. During distributed calculation of vectors $\vec{x}_p^k$ across each p th domain, it is important to ensure the structural and functional connectivity of inter-domain routes that are traverse multiple routers of different domains. In order to ensure inter-domain connectivity, the model (1)-(4) introduces the following inter-domain interaction conditions [10]:

$$C_{p,q}^k \vec{x}_p^k = C_{q,p}^k \vec{x}_q^k, \quad p, q = \overline{1, N}, \quad p \neq q, \quad k \in K, \quad (5)$$

where $C_{p,q}^k$ is the interaction matrix of the p th and q th domains of the size $m_{p,q} \times m_x^{p,k}$; $m_{p,q}$ is the number of routers through which the border between the p th and q th domains passes; $m_x^{p,k}$ is the number of coordinates $x_{i,j}^{p,k}$ of the $\vec{x}_p^k$.

Thus for the network structure shown in Fig. 1 that consists of two domains, the boundary passes through two routers. In the designations of the first domain (Fig. 1) the border routers are R_3^1 and R_4^1 , while in the designations of the second the border routers are R_1^2 and R_2^2 . In the gaps of communication links, their bandwidth is shown (1/s).

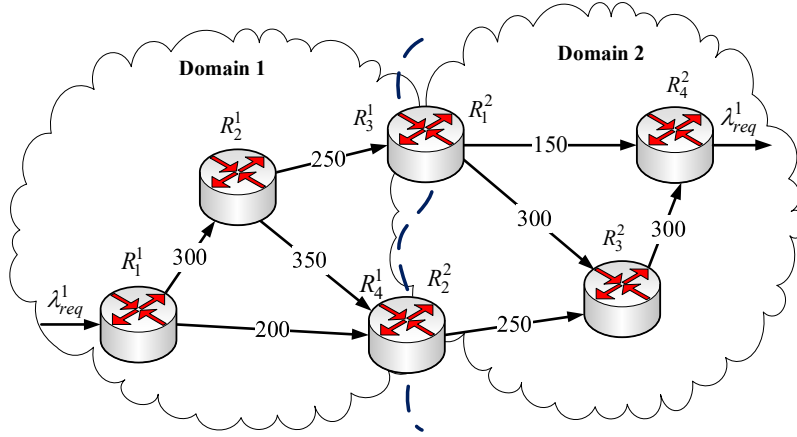


Fig. 1. An example of the ICN structure to be investigated.

4 Terms of providing normalized Quality of Service over domains in the infocommunication network

For each network flow, its transmission rate, average end-to-end delay, and packet loss probability, certain limits are set with respect to their boundary (minimum and/or maximum) values, which determine the level of customer QoS. In case of support of multidomain ICN architecture, the provision of set values of the end-to-end QoS indicators is often realized on the basis of their previous normalization. Let us introduce the following notation:

- τ_{req}^k is the requirement for the boundary value of the average end-to-end delay of the k th packet flow in the ICN, which is measured between the incoming router of the source domain and the outgoing receiver router of the k th packet flow;
- $\tau_{req}^{p,k}$ is the normalized requirement for the average delay of packets of the k th flow in the ICN p th domain ($\tau^{p,k}$).

Thus, for each flow $k \in K$, the requirements τ_{req}^k for the average end-to-end delay of packets are distributed (normalized) in some way between separate domains with the following conditions:

$$\sum_{p=1}^N \tau_{req}^{p,k} \leq \tau_{req}^k \text{ and } \tau^{p,k} \leq \tau_{req}^{p,k}. \quad (6)$$

Within the framework of this study, it is considered that τ_{req}^k and $\tau_{req}^{p,k}$ are predetermined. This raises the problem of formulating the conditions of normalized QoS on the network in terms of calculating expressions for $\tau^{p,k}$. To obtain such conditions it is necessary to use the functionality of tensor analysis of networks. The results presented in [18-25] allowed obtaining analytical expressions for calculating the values of end-to-end QoS indicators, which were evaluated and analyzed between a pair of individual routers. For the case considered in this study, in the structure of an arbitrary network domain, multiple routers may simultaneously belong to each of the sets $B_{in}^{p,k}$ and/or $B_{out}^{p,k}$, which is a special characteristic of transit domains. The following methodology is proposed to take into account the features of the multidomain ICN architecture.

1. During the k th flow routing for each p th domain the pair of routers R_{in}^p and R_{out}^p is determined, between which $\tau^{p,k}$ will be calculated. In the domain that served as the source of the k th packet flow, the R_{in}^p was the router through which the k th flow arrived into the ICN. For the domain that served as the k th packet flow

receiver, the R_{out}^p was the router through which the k th flow outwent from the ICN.

2. Additional imaginary routers R_{in}^p and R_{out}^p , which become adjacent for routers from $B_{in}^{p,k}$ and $B_{out}^{p,k}$ sets respectively through imaginary communication links, are introduced to the ICN structure at the boundaries of the p th domain that interacts with other domains through two or more routers. If the boundary between the p th and q th domains passes through several routers, as shown, for example, in Fig. 2, then the imaginary routers R_{in}^p and R_{out}^p coincide. The introduction of imaginary routers is conditioned by the fact that value $\tau^{p,k}$ can now be evaluated and analyzed already between a pair of routers R_{in}^p and R_{out}^p on the basis of the approach proposed in [10, 18]. In order to prevent impact of packet delays $\tau^{p,k}$ in imaginary communication links that connect real border routers to imaginary ones, their bandwidths must go to $+\infty$ during the calculations.
3. With the aim of further geometrization of the ICN structure, the continuous numbering of the communication links in the p th domain is adopted. For this purpose, the set of network links belonging to the p th domain is denoted as $V^p = \{v_z^p, z = 1, \tilde{n}_p\}$, where $\tilde{n}_p$ and $\tilde{m}_p$ are the number of communication links and routers respectively in the p th domain taking into account the introduced imaginary elements of the network. Thus, for an example of the structure of ICN shown in Fig. 2, $\tilde{n}_1 = \tilde{n}_2 = 6$ and $\tilde{m}_1 = \tilde{m}_2 = 5$. Such an increase in the number of communication links and routers leads to an increase in the number of variables $x_{i,j}^{p,k}$ with the condition (2) being replaced by the following flow conservation conditions on the p th domain routers:

$$\left\{ \begin{array}{l} \sum_{E_{i,j}^p \in E^p} x_{i,j}^{p,k} - \sum_{E_{j,i}^p \in E^p} x_{j,i}^{p,k} = 1, \text{ if } R_i^p = R_{in}^p; \\ \sum_{E_{i,j}^p \in E^p} x_{i,j}^{p,k} - \sum_{E_{j,i}^p \in E^p} x_{j,i}^{p,k} = 0, \text{ if } R_i^p \neq R_{in}^p, R_{out}^p; \\ \sum_{E_{i,j}^p \in E^p} x_{i,j}^{p,k} - \sum_{E_{j,i}^p \in E^p} x_{j,i}^{p,k} = -1, \text{ if } R_i^p = R_{out}^p. \end{array} \right. \quad (7)$$

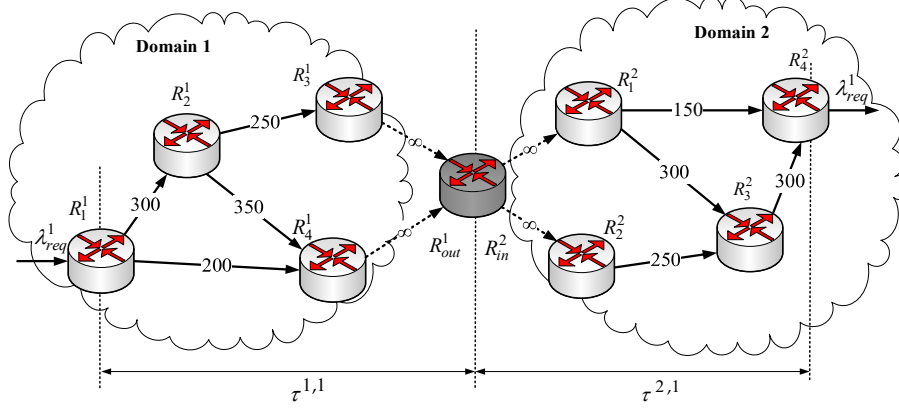


Fig. 2. The principle of introducing imaginary routers and links into the network structure.

According to the methodology of tensor modeling of ICN [18, 19, 23, 24], the domain structure determines the anisotropic space formed by coordinate paths. Network edges (links), circuits, interpolator paths, and node pairs can act as coordinate paths where the network poles are the routers R_{in}^p and R_{out}^p . The dimension of this space is determined by the total number of edges in the network and is equal to $n_p^\sim$ [18, 24]. From all possible interpolator (namely, end-to-end from the source to the destination) paths in the p th domain, we choose κ_p linearly independent $\{\gamma_i, i = \overline{1, \kappa_p}\}$. Whereas the set of internal node pairs is represented by the set $\{\varepsilon_j, j = \overline{1, \vartheta_p}\}$. These sets form the basis of the $n_p^\sim$ -dimensional space of the network structure:

$$\kappa_p^\sim = n_p^\sim - m_p^\sim + 2; \quad \vartheta_p^\sim = m_p^\sim - 2; \quad n_p^\sim = \kappa_p^\sim + \vartheta_p^\sim. \quad (8)$$

When routing the k th flow in the selected space, the p th domain can be represented by a mixed bivalent tensor $Q = T \otimes \Lambda$, where $\otimes$ is the tensor multiplication operator and the components of the tensor are a univalent covariance tensor of average packet delays T and a univalent contravariant tensor of the average intensities of flows Λ in the coordinate paths of the selected domain. Further, the index p in all tensor quantities and their projections will be omitted for greater clarity of information perception, because it is always a question of only physical quantities that are associated with the p th domain. The tensor can be written in the index form:

$$q_j^i = \tau_j \lambda^i, \quad (i, j = \overline{1, n_p^\sim}), \quad (9)$$

where τ_j is the average packet delay along the j th coordinate path (s); λ^i is the average packet flow intensity along the i th coordinate path (1/s).

In the selected $n_p^\sim$ -dimensional space, the tensor (9) will be set in one of two coordinate systems. The first is the coordinate system (CS) of the network edges $\{v_k, k = \overline{1, n_p^\sim}\}$, the second one is the CS of linearly independent interpolator paths $\{\gamma_i, i = \overline{1, \kappa_p}\}$ and internal node pairs $\{\varepsilon_j, j = \overline{1, \vartheta_p}\}$, the projections of the tensor in which will be denoted by the index $\gamma\varepsilon$.

In the case of modeling the operation of the network routers interface by the queuing system M/M/1, the coordinates of the projection of the metric tensor G in the basis of edges G_v represented by diagonal elements of the matrix will be determined by the expression:

$$g_v^{zz} = \lambda_v^z (\varphi_z - B_v^z), \quad (10)$$

where λ_v^z is the intensity of the k th flow in the z th communication link when using the continuous link numbering; B_v^z is the intensity of the aggregated flow in the z th communication link, which is defined as follows:

$$\lambda_v^z = \lambda_{req}^k x_{i,j}^{p,k}, \quad B_v^z = \sum_{k \in K} \lambda_{req}^k x_{i,j}^{p,k} \quad \text{under } v_z^p = E_{i,j}^p. \quad (11)$$

Projections of the twice contravariant metric tensor G when changing the CS of its consideration are transformed by the law

$$G_{\gamma\varepsilon} = A^t G_v A, \quad (12)$$

where $G_{\gamma\varepsilon}$ is the projection of the metric tensor in the CS of interpolator paths and internal node pairs; A is the $n_p^\sim \times n_p^\sim$ matrix of covariant coordinate transformation in the transition from the CS of interpolator paths and internal node pairs to the basis of edges; $[\cdot]^t$ is the operation of the matrix transposition.

As shown in [23, 24], the matrix $G_{\gamma\varepsilon}$ can be represented as

$$G_{\gamma\varepsilon} = \left\| \begin{array}{c|c} G_{\gamma\varepsilon}^{(1)} & G_{\gamma\varepsilon}^{(2)} \\ \hline G_{\gamma\varepsilon}^{(3)} & G_{\gamma\varepsilon}^{(4)} \end{array} \right\|, \quad (13)$$

where $G_{\gamma\varepsilon}^{(1)}(t)$ is the square $\kappa_p^\sim \times \kappa_p^\sim$ submatrix; $G_{\gamma\varepsilon}^{(2)}(t)$ is the square $\vartheta_p^\sim \times \vartheta_p^\sim$ submatrix; $G_{\gamma\varepsilon}^{(3)}(t)$ is the $\kappa_p^\sim \times \vartheta_p^\sim$ submatrix; $G_{\gamma\varepsilon}^{(4)}(t)$ is the $\vartheta_p^\sim \times \kappa_p^\sim$ submatrix.

Then, as shown in [22, 23], the conditions for ensuring the normalized QoS in the p th domain when routing the k th flow (8) take the form:

$$\Lambda_\gamma \leq \left(G_{\gamma\epsilon}^{(1)} - G_{\gamma\epsilon}^{(2)} \left[G_{\gamma\epsilon}^{(4)} \right]^{-1} G_{\gamma\epsilon}^{(3)} \right) T_\gamma^{req}, \quad (14)$$

where $[\cdot]^{-1}$ is the matrix inversion operation; Λ_γ is the $\kappa_p^\sim$ -dimensional vector of flow intensities in the interpolator paths of the selected domain with coordinates λ_γ^i that are connected by the following condition

$$\sum_{i=1}^{\kappa} \lambda_\gamma^i = \lambda_{req}^k; \quad (15)$$

T_γ^{req} is the $\kappa_p^\sim$ -dimensional vector of average packet delays in the interpolator paths of the selected domain, each of which coordinates τ_i^γ corresponds to the condition

$$\tau_i^\gamma = \tau_{req}^{p,k}, \quad i = \overline{1, \kappa_p^\sim}. \quad (16)$$

5 Method of hierarchical coordination inter-domain routing in Software-Defined infocommunication network with provision of normalized QoS

The hierarchical coordination inter-domain routing method will be based on solving an optimization problem for the calculation of vectors of routing variables $\vec{x}_p^k$ ($p = \overline{1, N}, k \in K$) subject to constraints (3)-(7), (14)-(16) by using the following optimality criterion:

$$\min F, \quad F = \sum_{p \in N} \sum_{k \in K} (\vec{x}_p^k)^t H_p^k \vec{x}_p^k, \quad (17)$$

where H_p^k is the diagonal matrix of routing metrics of links of the p th domain.

The goal coordination principle [9-11, 15, 16] will be used to solve the optimization problem (17). Then, moving to the problem at the unconditional extremum, it is necessary to maximize by $\vec{\mu}$ the Lagrangian of the form:

$$L = \sum_{p=1}^N \sum_{k \in K} (\vec{x}_p^k)^t H_p^k \vec{x}_p^k + \sum_{p=1}^N \sum_{q=1}^N \sum_{\substack{k \in K \\ q \neq p}} (\vec{\mu}_{p,q}^k)^t (C_{p,q}^k \vec{x}_p^k - C_{q,p}^k \vec{x}_q^k), \quad (18)$$

where $\bar{\mu}$ is the vector of Lagrange multipliers; $\bar{\mu}_{p,q}$ are subvectors of the vector $\bar{\mu}$ assigned to each of the vector-matrix domain interaction conditions (5).

Given that within the principle of goal coordination, the Lagrange multiplier vectors $\bar{\mu}$ are calculated at the upper level and for the lower level are values known, the expression (18) can be represented in the following decomposition form: $L = \sum_{p=1}^N L_p$,

where

$$L_p = \sum_{k \in K} (\bar{x}_p^k)^t H_p^k \bar{x}_p^k + \sum_{\substack{q=1 \\ p \neq q}}^N \sum_{k \in K_p^+} (\bar{\mu}_{p,q}^k)^t C_{p,q}^k \bar{x}_p^k - \sum_{\substack{q=1 \\ p \neq q}}^N \sum_{k \in K_p^-} (\bar{\mu}_{q,p}^k)^t C_{p,q}^k \bar{x}_p^k, \quad (19)$$

where K_p^+ is the subset of flows incoming to the p th domain from other domains;

K_p^- is the subset of flows outgoing from the p th domain ($K_p^+, K_p^- \subset K$).

Within the framework of the proposed method, the general problem of hierarchical coordination of inter-domain routing is formulated as a two-level optimization problem:

1. At the lower level, SDN controllers of the domains calculate the routing variables represented by vectors $\bar{x}_p^k$ during the minimization of Lagrangians (19) under constraints (3), (4), (7), and (14)-(16). The results of the calculations are sent to the top level, namely to the SDN controller of the network.
2. At the top level, the SDN controller of the network coordinates the lower-level solutions to ensure that the conditions (5) are met by modifying the Lagrange multiplier vectors:

$$\bar{\mu}_{p,q}^k(a+1) = \bar{\mu}_{p,q}^k(a) + \nabla \bar{\mu}_{p,q}^k, \quad \nabla \bar{\mu}_{p,q}^k(x) \Big|_{x=x^*} = C_{p,q} \bar{x}_p^k - C_{q,p} \bar{x}_q^k, \quad (20)$$

where a is the iteration number; $\nabla \bar{\mu}_{p,q}^k$ is the gradient of the function (19).

3. The modified values of the Lagrange multiplier vectors $\bar{\mu}_{p,q}^k$ are transmitted to the lower level for the calculation of new routing vectors $\bar{x}_p^k$. The calculation process becomes iterative. Inter-domain route connectivity will be ensured when the gradient values (20) approach zero.

From the technological point of view, minimizing the number of iterations of the procedure (20), when obtaining the desired optimal solution, aims to reduce the amount of service traffic transmitted between hierarchical levels about the results of

calculations at each iteration, and to decrease the total time of solving the problem of inter-domain routing in the ICN as a whole [9-11].

6 Investigation of the proposed method of hierarchical coordination inter-domain QoS routing

Let us investigate the proposed method of hierarchical coordination inter-domain routing in the ICN in order to confirm its functionality, adequacy, and efficiency of the obtained calculation results. In the framework of the numerical example, let us analyze the peculiarities of the solutions to the problem of hierarchical coordination inter-domain QoS routing for the variant of the infocommunication network structure shown in Fig. 1. As an example, consider a single-flow case when, in the course of a study, the requirements for the QoS level in a multidomain network were given by the following parameters:

$$\lambda_{req}^1 = 350 \text{ 1/s and } \tau_{req}^1 = 100 \text{ ms, } \tau_{req}^{1,1} = 40 \text{ ms, } \tau_{req}^{2,1} = 60 \text{ ms.} \quad (21)$$

Then Fig. 3 presents a solution to the problem of inter-domain QoS routing prior to the start of the coordination procedure (20). In Fig. 3, the link breaks show the following data (top to bottom): packet flow intensity, bandwidth, and average packet delay in this link.

The characteristic feature of the obtained solution (Fig. 3) is that the conditions for providing the normalized QoS (6), (21) are fulfilled: the maximum end-to-end delay in the first domain was 38.5 ms and in the second domain it was 48.6 ms. However, the conditions for inter-domain interaction (5) were not met. Inter-domain route connectivity was ensured after the third iteration of the coordination procedure (20).

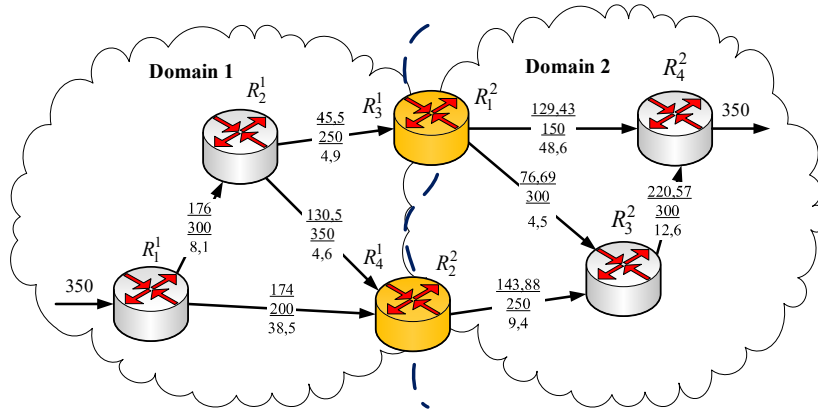


Fig. 3. The initial solution to the problem of inter-domain QoS routing under requirements (21).

The coordinated solution to the problem of inter-domain QoS routing is presented in Fig. 4, within which the specified normalized values of the average packet delays

in each of the domains (21) were provided: in the first domain the maximum end-to-end delay was 38.5 ms, and in the second domain it was 58.3 ms.

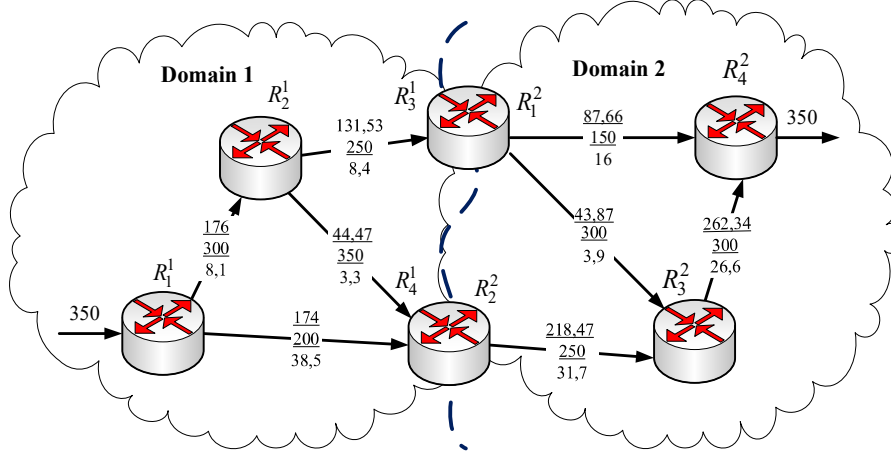


Fig. 4. The final (coordinated) solution to the problem of inter-domain QoS routing under requirements (21).

7 Conclusion

The paper proposes a method of hierarchical coordination inter-domain routing in SDN, which is a further development of the solutions presented in [9-11, 23, 24]. The novelty of the method is that the routing solutions obtained with it are aimed not only at increasing the scalability of ICN, but also at ensuring the normalized QoS in terms of average transmission rate and end-to-end average packet delay.

The proposed method is based on the use of the decomposition flow-based model of inter-domain routing (3)-(7). The model was supplemented by the conditions for the provision of normalized QoS (14)-(16), which were formulated in an analytical form on the basis of tensor modeling of the ICN (8)-(13).

Within the proposed method, the problem of inter-domain QoS routing was presented in the optimization form with a quadratic optimality criterion (17). The goal coordination principle was used to solve the optimization problem. During the study of the method, its functionality and efficiency were confirmed in terms of ensuring the normalized QoS. It was found experimentally that the method converged to the optimal solution for the finite number of iterations (20). For the structure of the network that was selected as the test (Fig. 1), the number of iterations of the coordination procedure (20) with the proper setting of the gradient search did not exceed three iterations. Reducing the number of such iterations helps to decrease the amount of service traffic that is transmitted across the network between routers and SDN controllers at different levels, as well as minimizing the overall time for solving the inter-domain QoS routing task.

The prospect of further research in this area is that the QoS metrics, which each domain should provide, are not set statically, for example, on the SDN controller, but can be redistributed dynamically between domains with the fulfillment of the conditions (6) in accordance with their structure, capacity, and utilization.

References

1. Wibowo, F.X., Gregory, M.A., Ahmed, K. and Gomez, K.M.: Multi-domain software defined networking: research status and challenges. *Journal of Network and Computer Applications*, 87, 32–45 (2017). <https://doi.org/10.1016/j.jnca.2017.03.004>
2. Katsalis, K., Rofoee, B., Landi, G., Riera, J.F., Kousias, K., Anastasopoulos, M., Kiraly, L., Tzanakaki, A., Korakis, T.: Implementation experience in multi-domain SDN: Challenges, consolidation and future directions. *Computer Networks*, 129, 142–158 (2017). <https://doi.org/10.1016/j.comnet.2017.09.005>
3. Bliat, O., Ben Mamoun, M., Benaini, R.: An overview on SDN architectures with multiple controllers. *Journal of Computer Networks and Communications*, 2, 1–8 (2016). <https://doi.org/10.1155/2016/9396525>
4. Medhi, D., Ramasamy, K.: *Network routing: algorithms, protocols, and architectures*. Morgan Kaufmann (2017)
5. Misra, S., Goswami, S.: *Network Routing: Fundamentals, Applications, and Emerging Technologies*. Wiley (2017)
6. Szigeti, T., Zacks, D., Falkner, M., Arena, S.: *Cisco Digital Network Architecture: Intent-based Networking for the Enterprise*. Cisco Press (2018)
7. Wójcik, R., Domżał, J., Duliński, Z., Rzym, G., Kamisiński, A., Gawłowicz, P., Jurkiewicz, P., Rzaśa, J., Stankiewicz, R., Wajda, K.: A survey on methods to provide interdomain multi-path transmissions. *Computer Networks*, 108, 233–259 (2016). <https://doi.org/10.1016/j.comnet.2016.08.028>
8. Eun, J., Jung, H.: August. The implementation of domain routing protocol in hierarchical domain network model. In: 2015 17th Asia-Pacific Network Operations and Management Symposium (APNOMS), pp. 396–399. IEEE (2015). <https://doi.org/10.1109/APNOMS.2015.7275350>
9. Lemesko, O., Nevzorova, O., Hailan, A.M.: February. Hierarchical method of routing and resource allocation in DiffServ-TE network. In: 2018 14th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering (TCSET), pp. 1014–1018. IEEE (2018). <https://doi.org/10.1109/TCSET.2018.8336366>
10. Lemesko, O., Yermenko, O., Nevzorova, O.: Hierarchical method of inter-area fast re-routing. *Transport and Telecommunication Journal*, 18(2), 155–167 (2017). <https://doi.org/10.1515/tjt-2017-0015>
11. Nevzorova, Y.S., Arous, K.M., Salakh, M.T.R.: Method for hierarchical coordinated multi-cast routing in a telecommunication network. *Telecommunications and Radio Engineering*, 75(13), 1137–1151. (2016). <https://doi.org/10.1615/TelecomRadEng.v75.i13.10>
12. White, R., Tantsura, J.E.: *Navigating Network Complexity: Next-generation routing with SDN, service virtualization, and service chaining*. Addison-Wesley Professional (2015)
13. Lin, S.C., Akyildiz, I.F., Wang, P., Luo, M.: QoS-aware adaptive routing in multi-layer hierarchical software defined networks: A reinforcement learning approach. In: 2016 IEEE International Conference on Services Computing (SCC), pp. 25–33. IEEE (2016). <https://doi.org/10.1109/SCC.2016.12>

14. Amin, R., Reisslein, M., Shah, N.: Hybrid SDN networks: A survey of existing approaches. *IEEE Communications Surveys & Tutorials*, 20(4), 3259–3306. IEEE (2018). <https://doi.org/10.1109/COMST.2018.2837161>
15. Mesarovic, M.D., Macko, D., Takahara, Y.: *Theory of hierarchical multilevel systems*. Elsevier (2000)
16. Singh, M.G., Titli, A.: *Systems: decomposition optimisation and control*. Pergamon (1978)
17. Segall, A.: The modeling of adaptive routing in data-communication networks. *IEEE Transactions on Communications*, 25(1), 85–95 (1977)
18. Leshchko, A.V., Evseeva, O.Y., Garkusha, S.V.: Research on tensor model of multipath routing in telecommunication network with support of service quality by greater number of indices. *Telecommunications and Radio Engineering*, 73(15), 1339–1360 (2014). <https://doi.org/10.1615/TelecomRadEng.v73.i15.30>
19. Leshchko, O., Yevdokymenko, M., Naors, Y.: Development of the tensor model of multipath QoS-routing in an infocommunication network with providing the required quality rating. *Eastern-European Journal of Enterprise Technologies*, 5 (2), 40–46 (2018). <https://doi.org/10.15587/1729-4061.2018.141989>
20. Strelkovskaya, I., Solovskaya, I., Grygoryeva, T., Paskalenko, S.: The solution to the problem of the QoS characteristics definition for self-similar traffic serviced by the W/M/1 QS. In: 2016 Third International Scientific-Practical Conference Problems of Infocommunications Science and Technology (PIC S&T), pp. 40–42. IEEE (2016). <https://doi.org/10.1109/INFOCOMMST.2016.7905330>
21. Strelkovskaya, I., Solovskaya, I., Makoganiuk, A.: Optimization of QoS characteristics of self-similar traffic. In: 2017 4th International Scientific-Practical Conference Problems of Infocommunications. Science and Technology (PIC S&T), pp. 497–500. IEEE (2017). <https://doi.org/10.1109/INFOCOMMST.2017.8246447>
22. Strelkovskaya, I.V., Solovskaya, I.N.: Tensor model of multiservice network with different classes of traffic service. *Radioelectron. Commun. Syst.* 56, 296–303 (2013). <https://doi.org/10.3103/S0735272713060058>
23. Leshchko, O.V., Yermenko, O.S., Hailan, A.M.: QoS solution of traffic management based on the dynamic tensor model in the coordinate system of interpolating paths and internal node pairs. In: 2016 International Conference Radio Electronics & Info Communications (UkrMiCo), pp. 1–6. IEEE (2016). <https://doi.org/10.1109/UkrMiCo.2016.7739625>
24. Yermenko, O.: Development of the dynamic tensor model for traffic management in a telecommunication network with the support of different classes of service. *Eastern-European Journal of Enterprise Technologies*. 6(9), 12–19 (2016). <https://doi.org/10.15587/1729-4061.2016.85602>

Experimental research of optimizing the Apache Spark tuning: RDD vs Data Frames

Sergii Minukhin¹[0000-0002-9314-3750], Maksym Novikov¹[0000-0002-8978-044X], Natalia Brynza¹[0000-0002-0229-2874], Dmytro Sitnikov²[0000-0003-1240-7900]

¹Simon Kuznets Kharkiv National University of Economics, 9a Nauka Ave., Kharkiv 61166 Ukraine

minukhin.sv@gmail.com, maksym.novikov@hneu.net,
natalia.brynza@hneu.net

²Harkiv National University of Radio Electronics, 14 Nauka Ave., Kharkiv 61166, Ukraine
dmytro.sytnikov@nure.ua

Abstract. In this paper results and analysis of experimental research for determining the effectiveness of changing the parameters (as compared to standard values) of tuning Apache Spark for minimizing application execution time have been presented. The structure of a test dataset has been developed using RDD and Data Frames, based on which it is possible to create during a minimal time text files with a size greater than 4 GB having properties (characteristics) set up for testing. A peculiarity of test data is the fact that they often reflect basic properties of real world problems. The investigation includes 2 stages: at the first stage a comparative analysis of RDD and Data Frames is carried out for the standard settings of Apache Spark; at the second stage experiments for different sizes of an input test dataset for assessing the influence of parallelism levels, a block size in HDFS and the parameter `spark.sql.shuffle.partitions` in Spark Data Frames have been conducted. The obtained results substantiate the influence of the `spark.sql.shuffle.partitions` value on the test task execution performance. For this parameter ranges and change trends have been found. Also, levels of parallelism that maximally influence the execution time have been determined. It has been proven that for certain sizes of input test files the size of an HDFS block can be set up by default. Results of computational experiments have been demonstrated in tables and graphs. They confirm the effectiveness of the suggested changes to the Apache Spark settings as compared with the standard ones for different sizes of tested files.

Keywords. Apache Spark, Resilient Distributed Dataset, Data Frames, HDFS, shuffling, level of parallelism, data processing, data set, application, execution time.

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

1 Introduction and related works

Spark is a general-purpose, fault-tolerant, and fast cluster computing platform for processing big data. Apache Spark is a popular platform for data analytics that has been used by many organizations (Yahoo!, eBay, Baidu, Netflix [1]) owing to its in-memory computing framework. The Apache Spark framework is open source. It offers a number of tuning capabilities and can be customized. Its configuration has a 100 parameters that can be easily set up and changed by the user and tailored according to the cluster and application he uses [2]. The main types and description of Apache Spark parameters are presented in Table 1 [2, 3].

Table 1. The main types and description of Apache Spark parameters

Parameter type	Description
Application Properties	Related to basic properties of an application, for example, the application name, the CPU and the memory resources that are supposed to be allocated to the coordinating process (driver), the memory allocated for all executor processes that perform computations
Runtime Environment	Settings for environment (class paths, java options and logging)
The level of parallelism	It relates to the number of tasks in which each Resilient Distributed Dataset (RDD) is divided into parts. It should be set in a way that the resources of a cluster are completely utilized
Shuffle Behavior	Such parameters represent the shuffling mechanism of Spark, and they include methods of shuffling, buffer settings, sizes, memory selected for shuffling
Execution Behavior	Fractions of memory, number of execution cores and the level of parallelism
Scheduling	Related to the scheduling mode and define, in particular, the maximum number of CPU cores that will be used
Networking	Related to timeout options, ports, optimal values for network retries, heartbeat pauses
Compression and Serialization	These parameters define if compression is applied or not, which compression codec should be used, parameters of the codec, the serializer to be used and its buffer options
Spark UI	These parameters are mainly associated with UI event-logging

The existing methodological approaches to optimizing settings (resetting standard values) of the Spark parameters allow increasing computational performance for solving problems related to various fields are considered in [1-3]. However, owing to

substantial differences in applications and their different performance (with respect to input data characteristics), developing a generalized algorithm that works effectively for any application is difficult [1]. Thus it is important to improve the execution performance of Spark for different application types, which can be done with the help of: optimizing the physical execution plan of a Spark job; efficiently scheduling parts of a Spark job on cluster nodes; and selecting the right configuration for the cluster, such as the number of machines (processors) and the resources available on each machine (RAM, I/O devices, network channel, HDD or SSD) [4]. Apache Spark applications are executed in many steps, where each step includes multiple tasks running parallelly on multiple worker nodes. Resilient Distributed Dataset (RDD) [4, 5] provides interfaces for data transformations and parallelization. These RDDs are distributed among cluster nodes. There are two types of operations on RDDs in Spark: transformations, which convert an RDD to another RDD, and actions, which operate on an RDD for producing a final result. This allows creating a job DAG (Directed Acyclic Graph) of transformations before calling an action, and therefore, optimization can be carried out on this DAG prior to execution [2, 6, 8]. For each action triggered inside of a Spark application, the DAG scheduler develops an execution plan to complete it. The conception of this execution plan supposes assembling as many transformations with narrow dependencies as possible into stages [2, 5, 6]. So, The Spark DAG schedules each stage for execution [2, 8]. The RDD interface also provides the possibility of caching data in memory in a way allowing them to be read at next iterations of the job without I/O latency [7, 8].

Spark SQL [8, 9] is a module that is built on top of the Spark core engine in order to process structured or semi-structured data. Spark SQL introduces a novel extensible optimizer called Catalyst [9]. Catalyst makes it easy to add data sources, optimization rules, and data types. There is a possibility of repartitioning data in RDDs. It is possible to do with the help of a specific transformation, such as `groupByKey` and `sortByKey`, which leads to a new RDD, where data are differently partitioned across machines. Such a redistribution of data is called data shuffling. An API Data-Frame has been created to perform relational operations on data such as `select`, `filter`, and `join` [2, 9]. Data Frame in Spark is a distributed data collection organized in the form of named columns. Conceptually, this is equivalent to a table in a relational database or data frame in R/Python, but with more powerful optimization. Data frames can be created from different sources – files with structured data, Hive tables, external databases or existing RDD [2, 9]. Data Frames in Spark automatically optimize the execution of queries with the help of the query optimizer Catalyst. At the beginning of calculations in Data Frame Catalyst compiles operations that have been used for building Data Frame in a physical execution plan and then generates byte-code JVM for the plans that are more optimized than a handwritten code [8, 9]. Unlike RDDs, Data Frames normally tracks their schema and support different relational operations, which leads to a more optimized execution. Data Frames can be constructed from tables in the available Big Data infrastructure or from existing RDDs [9]. Data Frames can be managed with direct SQL queries and also with the help of the DataFrame DSL (domain-specific language), where various relational operators and transformers can be used, such as `Where` and `GroupBy`. Any Data Frame can also be considered as RDD of Row objects, which allow users to call

RDD of Row objects, which allow users to call procedural Spark APIs such as map [9, 10].

Big Data Benchmarking Requirements include the capability of processing data of different types, e.g., unstructured, semi-structured, structured data, and different sources [9, 11, 13]. Big data workloads selected in the benchmark suite should correspond to the diversity of application scenarios. Some researchers have used 3 benchmark applications: sort-by-key, shuffling and k-means (e.g., part of the HiBench benchmark [11-13]). The above applications have been chosen because they can be viewed as representatives of a variety of applications under the condition that they cover both cpu- and shuffling-intensive cases (Sort-by-key is both computation- and shuffling-intensive) [2, 12, 13]. A great number of transformations require data shuffling across the cluster, which includes Join, ReduceByKey, GroupByKey, Cogroup. In this paper such an approach serves as a basis for the development of a test data generator.

Profiling Spark applications. Many researches divide all methods for tuning Apache Spark into several classes. A brief description of these is as follows:

- using pairs of parameters on benchmarking applications and the application of a graph algorithm to build complex candidate configurations [2, 6];
- using arbitrary combinations of the parameters [2, 3, 8];
- tuning parameters with the help of Machine Learning methods [6-8].

The work [2] considers Spark's shuffling optimization using two approaches. The first one is optimization via columnar compression, which unfortunately does not lead to any significant improvement of performance. The second one applies file consolidation during shuffling. The shuffle phase is all-to-all communication instrument and it can potentially introduce network, disk, memory and CPU scheduling overheads. A lot of transformations require data shuffling across the cluster, including Join, ReduceByKey, GroupByKey, Cogroup. For optimizing shuffle performance the following approaches are used: 1) Emulating Spark behavior by merging intermediate files; 2) Creating large shuffle files [14-16].

The main operations associated with shuffling are represented in Table 2 [15].

Table 2. Shuffle operations

Configuration	Description	Tuning advice
Spark.shuffle.file.buffer	This parameter is used to set the buffer size of the bufferedOutputStream of the shuffle write task.	If the available memory resources are sufficient, it can increase the size of this parameter (such as 64k), so as to reduce the number of times the disc file overflows during the shuffle write process, which can reduce the number of disc IO times and improve performance

Spark.reducer.maxSizeInFlight	This parameter is used to set the buffer size of the shuffle read task, and this buffer decides how much data can be drawn at a time.	If the available memory resources are sufficient, we can increase the size of the parameters (such as 96 MB), thereby reducing the number of times the data is pulled, which can reduce the number of network transmissions and improve performance
Spark.shuffle.io.maxRetries	Shuffle read task from the shuffle write task where the node is pulling their own data, if the network due	For those jobs that contain a very time-consuming shuffle operation, it is recommended to increase the maximum number of retries (for example, 60 times) to avoid data failure due to factors such as the full gc of the JVM or network instability.
Spark.shuffle.io.retryWait	This parameter represent the retry interval for each retry of the data).	It is recommended to increase the interval length (e.g. 60s) to increase the stability of the shuffle operation
Spark.shuffle.memoryFraction	This parameter represent the memory size of the Executor memory, which is assigned to the shuffle read task for the aggregation operation.	This parameter is explained in resource parameter tuning. If the memory is sufficient, and rarely use the persistence operation, it is recommended to increase this ratio, to shuffle read the aggregation operation of more memory, in order to avoid the lack of memory caused by the frequent process of reading writing disk.
Spark.shuffle.sort.bypassMergeThreshold	When ShuffleManager is SortShuffleManager, if the number of shuffle read tasks is less than this threshold (default is 200), shuffle write process will not be sorted, but directly in accordance with the way to optimize the HashShuffleManager to write data.	When you use SortShuffleManager, if we do not need to sort operation, it is recommended that this parameter will be larger, greater than the number of shuffle read task.

The contributions of this paper are as follows:

- a technique for generating test files of different sizes with properties defining requirements to test data processing has been developed;

- a comparative performance analysis of using Resilient Distributed Dataset and Data Frame in the case of standard Spark settings has been carried out; the influence of input files of different sizes has been investigated;
- the analysis of the Data Frame technology performance under simultaneously changing settings (tuning) related to parallelism levels and shuffle partitions (Data Frames) for input files of different sizes has been conducted.

2 Problem statement

The problem considered in this paper is the optimization of distributed computations, namely, experimental determining potentially most important tuning parameters for Apache Spark for minimizing the application execution time in the Apache Spark cluster based on using new Data Frame technologies for working with data as compared with the technology Resilient Distributed Dataset; obtaining quantitative estimates for the application of the API Data Frame technology for different settings (tuning) of the Spark parameters and experimental determining their values providing minimum execution time.

3 Experiments Environment setup and Data Set generation

When conducting experiments, the following software and hardware for the work of the Apache Spark cluster have been used:

4 work stations (1 master, 3 workers) with the following characteristics: CPU: Intel Core i5-7500 CPU @ 3.40 GHz; number of cores: 4 (without HyperThreading);

L1 cache: 256 KB; L2 cache: 1 MB; L3 cache: 6 MB; RAM: 8 GB DDR4 @ 2400 MHz; HDD: 500 GB (1 TB); interface: SATA; computer network – Ethernet; network adapter: Realtek RTL8411@1 GBit/s; OS: Ubuntu Linux 16.0 LTS. For distributed computations on the cluster the following versions of Apache Spark and Apache Hadoop and Spark deployment modes have been used: Master – Spark 3.0.0 (Preview 2), Name Node – Hadoop (HDFS) – 3.2.1, YARN (Fair), client mode; workers – Spark 3.0.0 (Preview 2), Data Node – Hadoop (HDFS) 3.2.1, YARN (Fair), client mode. For testing a dataset containing information on movies from the site <https://datasets.imdbws.com> has been used (a description of the dataset is presented on the resource <https://www.imdb.com/interfaces>). The dataset consists of several files linked with a key and contains columns with different data types, such as string, numerical, Boolean and date. The following dataset files have been used for testing: name.basics; title.basics; title.akas; title.episode; title.principals; title.ratings.

The size of the created basic test file is about 4 GB. In order that we can obtain results that can be compared and analyzed, datasets of higher volumes have been generated (Big Data) based on the main dataset. The following method has been used for generating these datasets: the original files have been copied and linked with each other (merged) and the key fields have been changed so that records are considered unique after the application of the aggregation, sorting and joining by key operations. Thus, test data in the range from 8 GB to 16 GB have been obtained. The operations

that have been used for a load imitation have been identical for RDD and Data Frames, namely, filter, join, groupBy, sortBy, and the aggregation functions count and sum. For each file in the dataset the operations of filtering have been applied, then the operations of sorting and aggregations, and after that, the linking by key with another file has been carried out. Thus, the operations that require both computer intensity (processors) and those where input-output speed is a key factor have been applied. The testing has been carried out in two modes: a) using RDD and b) using Data Frames.

For the file name.basics the following filters have been applied: deathDate != null and primaryProfession != miscellaneous. Their implementation through the DataFrames API is as follows:

```
names_df.filter(f.col('deathYear').isNotNull())
.filter(f.array_contains(f.split(f.col('primaryProfession'), ','), 'miscellaneous'))
```

and for RDD API:

```
names_rdd.filter(lambda r: r['deathYear'] is not
None).filter(lambda r: False if r['primaryProfession'] is
None else 'miscellaneous' not in r['primaryProfession']).
```

For the file title.akas the following filters have been applied: isOriginalTitle = '1' and region = 'US'. Their implementation through the DataFrames API is the following:

```
akas_df.filter(f.col('isOriginalTitle') ==
'1').filter(f.col('region') == 'US')
```

and for the RDD API:

```
akas_rdd.filter(lambda r: r['isOriginalTitle'] ==
'1').filter(lambda r: r['region'] == 'US').
```

For the file title.episode the following filter has been applied: episodeNumber > 10. Its implementation through the DataFrames API is as follows:

```
episodes_df.filter(f.col('episodeNumber').cast('int') >
10))
```

and for the RDD API:

```
episodes_rdd.filter(lambda r: False if r['episodeNumber']
is None else int(r['episodeNumber']) > 10).
```

For the file title.principals the following filters have been used: category != 'self' and category != 'cinematographer' and ordering <= 3. Their implementation through the DataFrames API is as follows:

```
principals_df.filter((f.col('category') != 'self') &
(f.col('category') !=
'cinematographer')).filter(f.col('ordering').cast('int')
<= 3)
and for the RDD API:
```

```
principals_rdd.filter(lambda r: r['category'] != 'self'
and r['category'] != 'cinematographer').filter(lambda r:
int(r['ordering']) <= 3).
```

For the file title.ratings the filters averageRating > 5.0 and numVotes > 10 000 have been applied. Their implementation through the DataFrames API is the following:

```
ratings_df.filter(f.col('averageRating') >
5.0).filter(f.col('numVotes').cast('int') > 10_000)
```

and for the RDD API:

```
ratings_rdd.filter(lambda r: float(r['averageRating']) >
5.0).filter(lambda r: int(r['numVotes']) > 10_000).
```

For the file title.basics the filters 'Comedy' in genres and titleType != 'short' have been used. Their implementation through the DataFrames API is as follows:

```
titles_df.filter(f.array_contains(f.split(f.col('genres'),
','), 'Comedy') & (f.col('titleType') != 'short'))
```

and for the RDD API:

```
titles_rdd.filter(lambda r: False if r['genres'] is None
else 'Comedy' in r['genres'].split(',')).filter(lambda r:
r['titleType'] != 'short').
```

After the filtration, the sorting and arrogation operations have been applied. For the file title.principals the operation of sorting by vlues of the column 'tconst' has been used, which is implemented through the Data Frames API: principals_df.orderBy('tconst'); in the case of the RDD API: principals_rdd.sortBy(lambda r: r['tconst']). For the file title.basics the operation of grouping by genres has been applied, which is implemented through the Data Frames API:

```
titles_df.withColumn('genres',
f.split(f.col('genres').getItem(0))).groupBy('genres').ag
g(f.count(f.lit(1))) и для RDD API: titles_rdd.groupBy(lambda r: None
if r['genres'] is None else
r['genres'].split(',')[0]).map(lambda r: (r[0],
len(r[1]))).
```

The last operation is a superposition of several files by key, which is implemented through the DataFrames API:

```
titles_df.join(akas_df, f.col('tconst') ==
f.col('titleId')).join(ratings_df,
on=['tconst']).join(epsisodes_df, on=['tconst'])
```

and the RDD API:

```
titles_rdd.join(akas_rdd).join(ratings_rdd).join(epsisodes
_rdd).
```

All the above operations can be divided into two categories:

CPU intensive (mainly associated with filtering or computing values in the framework of one row);

I/O (operations associated with data exchange between cluster nodes, for example, aggregation or sorting).

The considered operations used for generating test data with different properties are represented in Table 3.

Table 3. Operations for generated data

Operation	CPU intensive	I/O intensive
name.basics	filters deathDate != null and primaryProfession != miscellaneous	superposition with another dataframe by key 'nconst'
title.akas	filters isOriginalTitle = '1' and region = 'US'	superposition with another dataframe by key 'tconst'
title.episode	filters episodeNumber > 10	superposition with another dataframe by key 'tconst'
title.principals	filters category != 'self' and category != 'cinematographer' и ordering <= 3	sorting by values of the column 'tconst'
title.ratings	filters averageRating > 5.0 and numVotes > 10 000	superposition with another dataframe by key 'tconst'
title.basics	filters 'Comedy' in genres and titleType != 'short'	operations of grouping by genre in the column 'genres' superposition with other data frames by the key 'titleId'

4 RDD vs Data Frame

For conducting experiments and comparative analysis the following setting for Apache Spark parameter tuning have been used (Table 4).

Table 4. Apache Spark settings

Parameters	Description	Default	Rules of thumb	Values
spark.driver.cores	Number of cores to use for the driver process	1	1-4	1
spark.driver.memory	Amount of memory to use for the driver process	1 GB	1-8 GB	1 GB
spark.executor.cores	The number of cores to use on each executor	All the available cores on the worker	1-4	3
spark.executor.memory	Amount of memory to use per executor process	1 GB	1-8 GB	6 GB
spark.io.compression.codec	The codec used to compress internal data	lz4	lz4/lzf	lz4
spark.reducer.maxSizeInFlight	Maximum size of map outputs to fetch simultaneously from each reduce task	48 MB	24-48 MB	48 MB
spark.shuffle.compress	Whether to compress map output files	true	true/false	true
spark.default.parallelism	Default number of partitions in RDDs returned by transformations	Number of cores on the local machine	9-72	9-72
spark.sql.shuffle.partitions	Number of partitions to use when shuffling data for joins or aggregations	200	50-200	50-200
spark.sql.shuffle.buffer	Size of the in-memory buffer for each shuffle file output stream, in KB unless otherwise specified. These buffers reduce the number of disk seeks and system calls made in creating	32	16-64	64

	intermediate shuffle files			
input data size	Total size of input data files	–	4-16 GB	4-16 GB
hdfs.blocksize	Size of blocks in which files are stored in hdfs	128 MB	64-256 MB	64-256 MB

The conducted experimental research has shown the absence of influence of changing standard values for some tuning Spark parameters, which are shown in Table 4, namely, `spark.shuffle.compress`, `spark.reducer.maxSizeInFlight`, `spark.driver.memory` and `spark.io.compression.codec` on the execution time. The value of the parameter `spark.sql.shuffle.buffer` (64 MB) has been obtained based on multiple experimental results for RDD and Data Frames. Therefore the presented methodology has been built on the investigation of the parameters input data size, `spark.default.parallelism`, `spark.sql.shuffle.partitions`, and `hdfs.blocksize` on the task execution time. The experimental research included two stages: at the first stage a comparison of test results for input files of 4GB – 16GB has been carried out for different Spark settings in the cases RDD and Data Frame; at the second stage the influence of parallelism levels and the parameter `spark.sql.shuffle.partitions` on the execution time for files with different sizes for different values of the HDFS data block has been investigated.

As a performance metric, the application execution time (*Execution Time*) has been used. It can be calculated according to formula [17]:

$$Execution\ Time = Finish\ Time - Start\ Time ,$$

where *Finish Time* is the application completion time; *Start Time* is the application start time.

Stage 1.

The experiments have been carried out for different sizes of the HDFS data block (by default 128 MB). The results obtained for the test task execution time in the cases of RDD and Data Frames are shown in Table 5.

Table 5. Task execution time for RDD and Data Frames for different values of input files sizes and levels of parallelism

Type	Input file size, GB	Level of parallelism	Execution time
RDD	4	9	8 min. 35 sec.
	4	36	7 min. 35 sec.
	4	72	6 min. 59 sec.
	8	9	38 min. 45 sec.
	8	36	17 min. 28 sec.
	8	72	14 min. 51 sec.
	16	9	49 min. 23 sec.

	16	36	40 min. 45 sec.
	16	72	31 min.28 sec.
Data Frame	4	9	1 min. 37 sec.
	4	36	1 min. 53 sec.
	4	72	1 min. 33 sec.
	8	9	2 min.54 sec.
	8	36	2 min. 58 sec.
	8	72	3 min. 29 sec.
	16	9	5 min. 3 sec.
	16	36	5 min.19 sec.
	16	72	5 min. 13 sec.

The analysis of results shown in Table 5 allows us to come to a conclusion that the performance of computations with the help of Data Frame is significantly higher than with the using of RDD (in the range of 600%-700%), and it changes with changing the level of parallelism as follows: it decreases for RDD in the range of 15%-230% and it increases for Data Frames. The latter result shows the necessity of investigating the influence of other factors (parameters) used in Data Frames on the performance of computations.

Stage 2.

At this stage the combined influence of changing the parameter settings Level of Parallelism and spark.sql.shuffle.partitions on the application execution time has been investigated. A peculiarity of the suggested approach is this. The parameter spark.sql.shuffle.partitions decreases (it is equal to 200 by default) in the range from 50 to 200. The results of the conducted computing experiments are shown in Table 6.

Table 6. Task execution time for Data Frames for different values of input files sizes, levels of parallelism, and spark.sql.shuffle.partitions

Level of Parallelism	spark.sql.shuffle.partitions	Execution Time
Input File size=8 GB, Block size=128 MB		
9	200	2 min.54 sec.
	150	2 min.45 sec.
	100	2 min.52 sec.
	50	2 min.40 sec.
36	200	2 min.58 sec.
	150	2 min.48 sec.
	100	2 min.52 sec.
	50	2 min.43 sec.
72	200	3 min.29 sec.
	150	2 min.48 sec.
	100	2 min.43 sec.
	50	2 min.45 sec.
Input File size=8 GB, Block size=64 MB		

9	200	3 min. 7 sec.
	150	2 min. 49 sec.
	100	2 min. 35 sec.
	50	2 min. 48 sec.
36	200	2 min. 29 sec.
	150	2 min. 28 sec.
	100	2 min. 24 sec.
	50	2 min. 21 sec.
72	200	2 min. 21 sec.
	150	2 min. 21 sec.
	100	2 min. 18 sec.
	50	2 min. 19 sec.
Input File size=4 GB, Block size=128 MB		
9	200	1 min. 37 sec.
	150	1 min. 38 sec.
	100	1 min. 34 sec.
	50	1 min. 35 sec.
36	200	1 min. 53 sec.
	150	1 min. 31 sec.
	100	1 min. 32 sec.
	50	1 min. 28 sec.
72	200	1 min. 33 sec.
	150	1 min. 32 sec.
	100	1 min. 30 sec.
	50	1 min. 30 sec.
Input File size=4 GB, Block size=64 MB		
9	200	1 min. 58 sec.
	150	1 min. 44 sec.
	100	1 min. 41 sec.
	50	1 min. 36 sec.
36	200	1 min. 34 sec.
	150	1 min. 34 sec.
	100	1 min. 31 sec.
	50	1 min. 31 sec.
72	200	1 min. 36 sec.
	150	1 min. 34 sec.
	100	1 min. 36 sec.
	50	1 min. 33sec.
Input File size=16 GB, Block size=128 MB		
9	200	5 min. 3 sec.
	150	5 min. 2 sec.
	100	5 min. 7 sec.
	50	4 min. 59 sec.

36	200	5 min. 19 sec.
	150	5 min. 13 sec.
	100	5 min. 9 sec.
	50	5 min. 11 sec.
72	200	5 min. 13 sec.
	150	5 min. 11 sec.
	100	5 min. 11 sec.
	50	5 min. 11 sec.

The obtained results show that for relatively small test files (4 GB and 8 GB) values of the parameter `spark.sql.shuffle.partitions` have the maximum influence on the execution time; at that the optimal parallelism level is from 36 to 72 and the minimum execution time is stable when `spark.sql.shuffle.partitions` equals 50. We should note that for the same files the optimal HDFS data block size is 64 MB or 128 MB, i.e. it practically does not have any influence on the target variable. For a 16 GB file the same settings do not bring substantial results. Besides, the most important result of the conducted research and computational experiments is the fact that in all experiments the default value of `spark.sql.shuffle.partitions` leads to worse results.

The graphs for the dependence of the test task execution time on parallelism level and the parameter `spark.sql.shuffle.partitions` are shown in Fig. 1, 2 (the following notation is used: IFS – Input File Size, BS – Block Size – block size in HDFS).

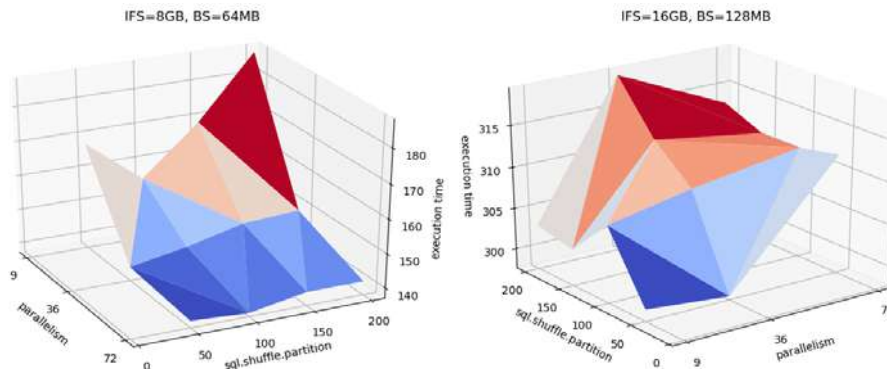


Fig. 1. Graphs for the dependence of the test task execution time on different values of BS for IFS=8 GB.

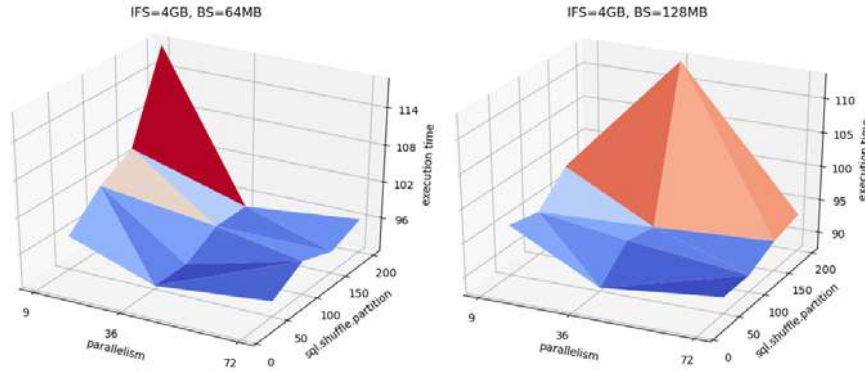


Fig. 2. Graphs for the dependence of the test task execution time on different values of BS for IFS=4 GB.

On graphs 1, 2, results of conducted execution time calculations are visualized. These results should be interpreted as follows. The minimum execution time values are shown in blue and its gradient, the maximum ones are shown in red and its gradient. The results confirm the fact (see Table 6) that in general, for small size files minimum execution time values are determined when the parallelism levels are 36 or 72, and the parameter `sql.spark.shuffle.partitions` equals 50.

5 Conclusions

A new technology for working with data in Apache Spark – API Data Frames – has been investigated. A comparative analysis of performance for RDDs and Data Frames in Apache Spark has been conducted on test data with properties defined by different data volumes. Software for generating test data with pre-defined characteristics has been developed and studied, taking into consideration CPU and I/O intensity. A comparative analysis of the technologies RDD and Data Frames has been carried out, which has shown a significant advantage of Data Frames and the influence of the parallelism level as the most important factor of increasing performance of computations for RDD for the standard Spark settings. By conducting a series of experiments, it has been obtained that the most important parameter for the Data Frame technology that influences the execution time is `spark.sql.shuffle.partitions`. Its changes in the range from 50 (experimentally considered to be the minimum value) to 200 (default value) have shown that for input files of small sizes the default value gives the worst results, and increasing the performance of computations is determined by decreasing the values of this parameter. The investigation of the parallelism level influence has allowed determining optimal (in the context of settings for minimizing execution time) values for this parameter - 36 and 72. For big size files decreasing the value `spark.sql.shuffle.partitions` and increasing the parallelism level within the selected ranges does not practically lead to a decrease in execution time. This fact probably means that when an input file size increases, it is necessary to increase the

spark.sql.shuffle.partitions value. At the same time, it is quite difficult to assess the influence of the parallelism level and directions of its change.

The conducted research has demonstrated the topicality of solving Big Data problems in the framework in-memory Apache Spark [18-20], for which the optimization of settings differs from that for small files. Using well-known benchmarkings for assessing distributed computations performance [11, 13, 18, 19] does not always allow determining an optimal set and values of the Spark tuning parameters for RDDs and Data Frames in the package and stream modes. Therefore, there are unsolved problems related to investigating conditions for minimizing execution time for big size files and saving properties of files generated synthetically, according to the considered technique – forming big size files by cloning small files.

In the future we are planning to test the above technologies for data with volumes higher than 50 GB, maybe 100 GB, under the condition of the scalability of the Apache Spark cluster. An important research direction is also associated with investigating the performance of the considered technologies when they are used for methods and tests in machine learning [19-21].

References

1. Wang, K., Khan, M. M. H., Nguyen, N., Gokhale, S.: A Model Driven Approach towards Improving the Performance of Apache Spark Applications. In: 2019 IEEE International Symposium on Performance Analysis of Systems and Software (ISPASS), 24-26 March 2019, pp. 233-242 (2019) doi: 10.1109/ISPASS.2019.00036.
2. Gounaris, A., Torres, J.: A Methodology for Spark Parameter Tuning. Big Data Research, vol. 11, pp. 22-32 (2018) <https://doi.org/10.1016/j.bdr.2017.05.001>.
3. Petridis, P., Gounaris, A., Torres, J.: Spark Parameter Tuning via Trial-and-Error. In: Advances in Big Data: Proceedings of the 2nd INNS Conference on Big Data. INNS 2016. Advances in Intelligent Systems and Computing, vol. 529, pp.226-237 (2016) https://doi.org/10.1007/978-3-319-47898-2_24.
4. Zaharia, M., Chowdhury, M., Franklin, M.J., Shenker, S., Stoica, I.: Spark: cluster computing with working sets. In: Proceedings of the 2nd USENIX Conference on Hot Topics in Cloud Computing (HotCloud), pp. 1-7 (2010) <https://www.icsi.berkeley.edu/icsi/node/5074>.
5. Zaharia, M., Chowdhury, M., Das, T., Dave, A., Ma, J., McCauly, Franklin, M. J., Shenker, S., Stoica, I.: Resilient distributed datasets: a fault-tolerant abstraction for in-memory cluster computing. In: Proceedings of the 9th USENIX conference on Networked Systems Design and Implementation. USENIX Association, pp. 15-28 (2012) doi/10.5555/2228298.2228301.
6. Wang, G., Xu, J., He, B.: A Novel Method for Tuning Configuration Parameters of Spark Based on Machine Learning. In: 2016 IEEE 18th International Conference on High Performance Computing and Communications; IEEE 14th International Conference on Smart City; IEEE 2nd International Conference on Data Science and Systems (HPCC/SmartCity/DSS), Sydney, NSW, pp. 586-593 (2016) doi: 10.1109/HPCC-SmartCity-DSS.2016.0088.
7. Choi, I. S., Yang, W., Kee, Y.: Early experience with optimizing I/O performance using high-performance SSDs for in-memory cluster computing. In: IEEE International Confer-

ence on Big Data (Big Data), pp. 1073-1083 (2015) doi: 10.1109/BigData.2015.7363861.

8. Mustafa, S., Elghandour, I., Ismail, M. A.: A Machine Learning Approach for Predicting Execution Time of Spark Jobs. *Alexandria Engineering Journal*, vol. 57, issue 4, pp. 3767–3778 (2018) <https://doi.org/10.1016/j.aej.2018.03.006>.
9. Armbrust, M., Xin, R.S., Lian, C., Huai, Y., Liu, D., Bradley, J. K., Meng, X., Kaftan, T., Franklin, Michael J., Ghodsi, A., Zaharia, M.: Spark SQL: Relational Data Processing in Spark. In: *SIGMOD '15: Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data*, pp. 1383–1394 (2015) <https://doi.org/10.1145/2723372.2742797>.
10. Zaharia, M., Xin, R., Wendell, P., Das, T., Armbrust, M., Dave, A., Meng, X., Rosen, J., Venkataraman, S., Franklin, M., Ghodsi, A., Gonzalez, J., Shenker, S., Stoica, I.: Apache Spark: a unified engine for big data processing. *Communications of the ACM*, vol. 59, no. 11, pp. 56-65 (2016) <https://doi.org/10.1145/2934664>.
11. Han, R., John, L. K., Zhan, J.: Benchmarking Big Data Systems: A Review. *IEEE Transactions on Services Computing*, vol. 11, issue 3, pp. 580-597 (2018) doi: 10.1109/TSC.2017.2730882.
12. Clemente-Castello, F., Nicolae, B., Mayo, R., Fernandez, J.: Performance Model of MapReduce Iterative Applications for Hybrid Cloud Bursting. *IEEE Transactions on Parallel & Distributed Systems*, vol. 29, issue 8, pp. 1794-1807 (2018) doi: 10.1109/TPDS.2018.2802932.
13. Huang, S., Huang, J., Dai, J., Xie, T., Huang, B.: The HiBench benchmark suite: Characterization of the MapReduce-based data analysis. In: *2010 IEEE 26th International Conference on Data Engineering Workshops (ICDEW 2010)*, Long Beach, CA, pp. 41-51 (2010) doi: 10.1109/ICDEW.2010.5452747.
14. Davidson, A., Or, A.: Optimizing shuffle performance in spark. University of California, Berkeley Department of Electrical Engineering and Computer Science (2013) <https://www.semanticscholar.org/paper/Optimizing-Shuffle-Performance-in-Spark-Davidson/d746505bad055c357fa50d394d15eb380a3f1ad3>.
15. Nirali, R., Shyam, D.: Shuffle Performance in Apache Spark. *International Journal of Engineering Research & Technology (IJERT)*, vol. 4, issue 02, pp.177-180 (2015).
16. Vandana, Velamuri, M., Narendra, K. P.: Shuffle phase optimization in spark. In: *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, Udupi, pp. 1028-1034 (2017) doi: 10.1109/ICACCI.2017.8125977.
17. Web UI, available at: <https://spark.apache.org/docs/3.0.0-preview/web-ui.html#jobs-tabN>, last accessed 2020/04/20.
18. Li, M., Tan, J., Wang, Y. et al.: SPARKBENCH: a spark benchmarking suite characterizing large-scale in-memory data analytics. *Cluster Computing*, vol. 20, no. 3, pp. 2575–2589 (2017) <https://doi.org/10.1007/s10586-016-0723-1>.
19. Rodrigues, J., Vasconcelos, G.: Big Data Machine Learning Benchmark on Spark. *IEEE Dataport* (2019) <http://dx.doi.org/10.21227/t8bg-yc46>.
20. Shi, S., Wang, Q., Xu, P., Chu, X.: Benchmarking State-of-the-Art Deep Learning Software Tools. In: *7th International Conference on Cloud Computing and Big Data (CCBD)*, Macau, pp. 99-104 (2016).
21. Liu, Y., Zhang, H., Zeng, L., Wu, W., Zhang, C.: MLbench: benchmarking machine learning services against human experts. *Proceedings of the VLDB Endowment*, vol. 11, no. 10, pp. 1220–1232 (2018) doi:<https://doi.org/10.14778/3231751.3231770>.

Models of Information Processing Optimization for Technical Interoperability in a Network of Distributed Situational Centers

Oleksii Kovalenko¹[0000-0002-9639-3544], Vitalii Visnevsky¹[0000-0003-2204-0487], Vladimir Kosolapov¹[0000-0002-5264-7872]

¹ Institute of Mathematical Machines and Systems Problems NAS of Ukraine,
Academician Glushkov Av. 42, Kyiv, 03187, Ukraine
Kovalenko.O.E@nas.gov.ua, vit.vizual@gmail.com,
KosolapovVL1953@gmail.com

Abstract. The problems and directions of development of distributed information services of situational systems are considered. A method for assessment the interoperability of distributed situational centers is proposed. The urgency of creating a network of distributed situational centers based on convergent technological solutions is substantiated. Models and algorithms for providing technical interoperability between participants of the process of situational management in the network of distributed situational centers have been developed. The potential for further improvement of situational management systems for strategic planning based on the management (control) of the capabilities of the target systems is shown.

Keywords: system of situational management, network of distributed situational centers, convergence of information technologies, interoperability optimization, system capability

1 Introduction

An urgent issue for the development of information technologies and systems is the task of creating a Network of Distributed Situational Centers (NDSCs) based on convergent technological solutions. In this regard, it is of particular importance to develop information services for information and analytical systems based on NDSCs to support and make decisions at different levels of state governance. Developing and implementing into practice the system of situational centers (SC) for solving public administration tasks is an integral part of the creation of effective governance systems for different levels of the country's economic complex.

Standard ISO/IEC 10746 [1] defines the essential concepts necessary to specify reference model of open distributed processing (RM-ODP) systems from five prescribed viewpoints. It provides a well-developed framework for the structuring of specifications for large-scale, distributed systems.

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

The framework for system specification provided by ISO/IEC 10746 has four fundamental elements:

- An object modeling approach to system specification;
- The specification of a system in terms of separate but interrelated viewpoint specifications;
- The definition of a system infrastructure providing distribution transparencies for system applications;
- A framework for assessing system conformance.

The RM-ODP framework provides five generic and complementary viewpoints on the system and its environment:

- The enterprise viewpoint, which focuses on the purpose, scope and policies for the system. It describes the business requirements and how to meet them;
- The information viewpoint, which focuses on the semantics of the information and the information processing performed. It describes the information managed by the system and the structure and content type of the supporting data;
- The computational viewpoint, which enables distribution through functional decomposition on the system into objects, which interact at interfaces. It describes the functionality provided by the system and its functional decomposition;
- The engineering viewpoint, which focuses on the mechanisms and functions required to support distributed interactions between objects in the system. It describes the distribution of processing performed by the system to manage the information and provide the functionality;
- The technology viewpoint, which focuses on the choice of technology of the system. It describes the technologies chosen to provide the processing, functionality and presentation of information.

In context of ISO/IEC 10746 standard interoperability notion is defined as capability of objects to collaborate, that is, the capability mutually to communicate information in order to exchange events, proposals, requests, results, commitments and flows [2].

2 Formal statement of the problem

When implementing the information-analytical process of information processing during the alternatives preparation and decision-making in NDSC, there is rising the task to administrate the sequence of information processing and control flows such that the system of the SC associated with one same problem could provides the necessary level of interoperability in the processes of information, analytical and expert servicing for participants of situation management processes (PSMP) at a minimum cost of time. The PSMP set includes experts, analysts, decision-makers, and other involved stakeholders and sites. The information and analytical process is formed in the form of an application that determines the need for information resources and analytical and expert capabilities and time of service of each SC. The convergent

approach involves the composition of heterogeneous components in a single system [3], ensuring compliance with 12 basic interoperability principles [4]: 1) subsidiarity and proportionality; 2) openness; 3) transparency; 4) reusability; 5) technological neutrality and data partability; 6) user-centricity; 7) inclusion and accessibility; 8) security and privacy; 9) multilingualism; 10) administrative simplification; 11) preservation of information; 12) assessment of effectiveness and efficiency.

These basic principles are divided into the following 4 categories:

- Principle setting the context for governance actions on interoperability (1st principle);
- Core interoperability principles (2nd to 5th principles);
- Principles related to generic user needs and expectations (6th to 9th principles);
- Foundation principles for cooperation among public administrations (10th to 12th).

A heterogeneous network of SCs is considered. Let M be the set of users (PSMP) of NDSC; n is the number of SCs (in our case, $|M| > n$, that is, multi-user and content-different nodes of monitoring and analysis of the information being processed (economic, environmental, military, etc.); c_k is estimation of the k -th user information needs ($k \in M$); R_i is the bandwidth estimation of the i -th SC of the NDSC; P_{ij} – the flow of information redistribution among i -th and j -th SCs ($i, j=1,2,\dots,n$) in the network; q is the maximum permissible information flow among two SCs.

We introduce a variable

$$x_{ki} = \begin{cases} 1, & \text{if the } k\text{-th user is served by the } i\text{-th SC of the NDSC, } i=1,\dots,n \\ 0, & \text{otherwise} \end{cases}.$$

The formalized view of the model for actual loading H_i of the i -th SC ($i=1,2,\dots,n$) in NDSC during of user's servicing in NDSC is as follows:

$$H_i = \max_i \sum_{k \in M} c_k x_{ki} \rightarrow \min, \quad (1)$$

$$\sum_{k \in M} c_k x_{ki} \leq R_i, i=1,2,\dots,n, \quad (2)$$

$$\frac{1}{M} \sum_{i=1}^n x_{ki} \leq 1, k \in M, \quad (3)$$

$$P_{ij}(x_{ki}, x_{kj}) \leq q; k \in M; i \neq j; i, j=1,2,\dots,n. \quad (4)$$

Condition (2) expresses the limitation on bandwidth information of the i -th SC, condition (3) is a requirement of obligatory service of all users (PSMP) and belonging of the user to the service group of local SC, condition (4) regulates the service flows between users (PSMP) of NDSC.

The problem in statement of (1) - (4) is the problem of zero-one integer programming (Boolean integer programming). To solve problem (1) - (4), an algorithm is proposed that implements the search method with local optimization [5], which al-

allows one to obtain a solution of problem (1) - (4) with acceptable accuracy for practical operation of the NDSC.

3 Literature review

The problem of efficient distribution of tasks in automated computer systems (ACS) network was considered in other works by Glushkov [6], Morozov [7] and others. NDSC is a development and extension of the concept of ACS network to date. One of the key mechanisms for the effective functioning of NDSC is to ensure interoperability among its constituents.

According to the definition of interoperability in accordance with ISO/IEC 24765 [2] and the interoperability framework (EIF) adopted at the EU [4], the basic level of interoperability is technical interoperability (Fig. 1).

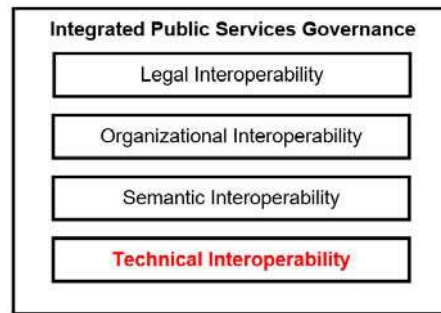


Fig. 1. Four levels of interoperability

On the basis of the EIF [4] the European Interoperability Reference Architecture (EIRA) was developed [8]. Various aspects of interoperability have been considered in [9-13].

4 Models of optimization of technical interoperability

4.1 Model of optimization of loading of NDSC

The paper proposes models and algorithms for providing technical interoperability between PSMPs in NDSC. Consider the problem of PSMP service in a distributed system, which can consist of several local (regional) SCs and aimed at solving various content problems. As a rule, when organizing the interaction of many SCs, appropriate expert assessments and calculations of their loading are conducted in order to serve the needs of the PSMP in a timely manner. However, over time, uploads can change significantly (for example, through the introduction of new tasks and management information flows), and NDSC will not be able to meet PSMP requirements at the right time and volume. The challenge is to redistribute the load of NDSC in order to ensure a uniform load on organizational, human and technical and informa-

tion resources for timely and qualitative decision-making in SCs that cater to PSMPs at various levels.

The outputs for solving problem (1) - (4) are formed in the form of matrix A . The rows of the matrix represent the estimation of information needs of PSMPs served by the NDSC. The columns of matrix A are the actual loading of the NDSC, i.e. H_i ($i=1,2,\dots,n$).

The algorithm for solving the problem of customer service in NDSC is presented in the form of the following five steps.

Step 1. If all the columns of matrix A are revised, then proceeds to step 5. Otherwise go to the next step. In the analysis, matrix A is searched for the column with the maximum sum of rows, that is, with the maximum load of a specific SC. Let such a column be the column i . The next step is to fix an area of three columns i , $i+1$, $i-1$ ($i=1,2,\dots,n$). If $i+1$ or $i-1$ do not exist (for finite elements), then the column i neighborhood will be truncated.

Step 2. In the i -th column there is a non-zero element $c_k x_{ki}$ and the possibility of shifting this element to the right or left according to conditions (2) - (4) is checked. If such shifts are possible, they are executed. To remember the movements of elements of matrix A in the next step of the algorithm (in order to determine possible movements in the next step) is an auxiliary matrix B , which elements are numbers from the set $(-1, 0, 1)$. Initial values of matrix elements are equal 0. When moving the element $c_k x_{ki}$ to the location of the element $c_k x_{ki-1}$, the e_{ki} element of the matrix B takes a value of -1 , when moving the $c_k x_{ki}$ to the location $c_k x_{ki+1}$ the e_{ki} element takes the value $+1$, otherwise the value of the e_{ki} element remains equal to 0. If the following iterations of the algorithm necessitate the displacement of some element $-c_k x_{ki-1}$ of matrix A in place of the element $c_k x_{ki}$, and the element $-e_{ki}$ of matrix B has a value of -1 , as well as the need to move the element $-c_k x_{ki+1}$ of matrix A to the location of the element $c_k x_{ki}$, and element $-e_{ki}$ of matrix B is $+1$, then these movements are not made and go to row $k+1$ of matrix A .

Step 3. In this step, check which of the three options (original and two received) gives the best result by criterion (1). If the shift was possible only to the left or only to the right, then both variants are subject to verification – initial and obtained. If, at the m -th iteration, it is found that $H_i^m = H_i^{m-1}$, then the process is completed and H^m is taken as an approximate solution for column i . The corresponding changes are made to matrix A , matrix B and vector $H = \{H_i\}$ ($i=1,2,\dots,n$). Thus, the i -th column is excluded from consideration and proceeds to step 1 of the algorithm. Otherwise, step 4 is performed.

Step 4. Go to row $k+1$ of matrix A and step 2 of the algorithm. If all rows was reviewed, then the process ends, the value of H_i^k is taken as an approximate solution for the problem with resolved column i and the corresponding changes are made to matrix A , matrix B and vector $H = \{H_i\}$ ($i=1,2,\dots,n$). Thus, the i -th column is excluded from consideration and step 1 of the algorithm is performed.

Step 5. After all the columns of matrix A have been revised, the solution of problem (1) - (4) ends. A new pin has been obtained for users in the local SC.

The above algorithm for solving problem (1) - (4) of servicing the PSMP in the NDSC is finite due to the finite number of SCs in the network.

The results for the 9-node loading redistribution algorithm for the first 5 iterations are presented in Fig. 2. The graph number in the diagram corresponds to the iteration number. The horizontal axis shows the numbers of the NDSC nodes, along the vertical axis the volume of nominal information flow for a specific decision making on a given problem, which should be processed within the regulated time limit.

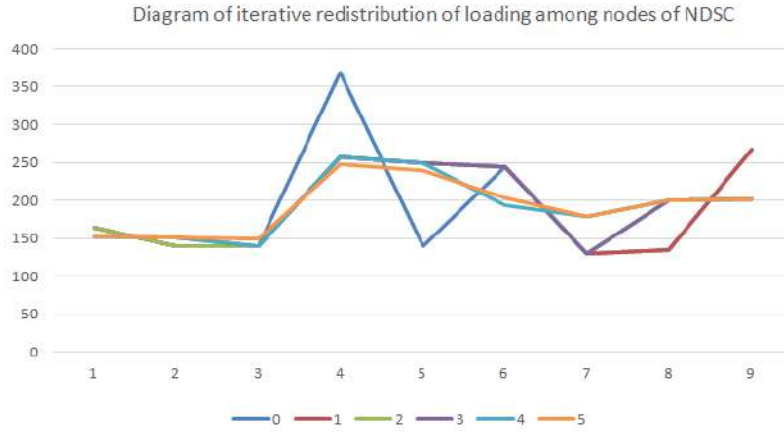


Fig. 2. Results for 5 iterations of loading redistribution for NDSC with 9 nodes

Thus, the algorithm for solving the problem (1) - (4) of optimization of loading for NDSC based on zero-one programming allows to increase qualitatively the level of information service of users in the NDSC and thereby to increase the efficiency of its functioning. This example is a prototype of PSMP servicing in NDSC by different criteria and experts assessments in the relevant monitoring and analysis areas stored in distributed data warehouses.

4.2 Model of minimization of servicing time in NDSC

When implementing the information-analytic process of information processing for downloading the corresponding SC, the task of managing the sequence of processing of information and control flows arises, so that the complex of SCs provides the maximum convenience of information, analytical and expert service to users (PSMP) with the minimum time consuming [7]. Such a problem is dual to the primal problem discussed above. The information-analytical process is formed as request that defines the need for information resources and analytical and expert capacities and time of service in each local SC. Such a task is formulated as follows.

Consider the common distributed information resource of the NDSC, which exists in n SCs $A_1, A_2, \dots, A_n$, which can provide various information, analytical, expert or forecast (foresight) services. The SC system executes some request for information, analytical or expert service for PSMP. An request is considered serviced if it is serviced by each local system SC sequentially, starting with SC of A_1 . Suppose that the time of service of an application in the i -th SC is an independent random variable with

a distribution function $F^i(t)$ and a distribution density $\varphi_i(t) = F^i(t)$. The system works like this. The request is serviced at SC A_1 as soon as it arrives. At the time of arriving of request to SC A_2 there are two possible cases:

1) the application was serviced by SC A_1 earlier than SC A_2 was involved and is awaiting response from SC A_2 . This pays some fine for a delay of request servicing;

2) the application was serviced by SC A_1 later than the request was sent to SC A_2 , that is, SC A_2 is awaiting the completion of servicing of the request in SC A_1 . In this case a fine is paid for a delay of SC A_2 . In this way, interaction between other SCs is performed.

Obviously, the fine f_i is a random variable. It is necessary to determine such moments $x_1, x_2, \dots, x_n$ of switching on of SCs $A_2, A_3, \dots, A_n$, that the mathematical expectation of the total fine (penalty) equal to the sum of fines paid due to the discrepancy of moments x_i of switching on of the i -th SC with the moment η_{i-1} of termination of service of the request in the i -th SC was minimal.

Let α be the unit cost of idle time of request and let β be the unit cost of idle time of Situational Center. Then the penalties related to the discrepancy between moments x_i of switching on and moments η_{i-1} of termination of service of the application in $(i-1)$ -th SC are determined as follows:

$$f_1(\eta_1, x_2) = \begin{cases} \alpha(x_2 - \eta_1), \eta_1 \leq x_2, \\ \beta(\eta_1 - x_2), \eta_1 \geq x_2, \end{cases} \quad (5)$$

where $\eta_1 = \tau_1$,

$$f_2(\eta_2, x_3) = \begin{cases} \alpha(x_3 - \eta_2), \eta_2 \leq x_3, \\ \beta(\eta_2 - x_3), \eta_2 \geq x_3, \end{cases} \quad (6)$$

where $\eta_2 = \max(\eta_1, x_2) + \tau_2$,

$$f_{n-1}(\eta_{n-1}, x_n) = \begin{cases} \alpha(x_n - \eta_{n-1}), \eta_{n-1} \leq x_n, \\ \beta(\eta_{n-1} - x_n), \eta_{n-1} \geq x_n, \end{cases} \quad (7)$$

where $\eta_{n-1} = \max(\eta_{n-2}, x_{n-1}) + \tau_{n-1}$,

$$Mf_1(\eta_1, x_2) = (\alpha + \beta) \int_0^{x_2} F(u_1) du_1 + \beta(M_{\tau_1} - x_2), \quad (8)$$

$$\begin{aligned} Mf_2(\eta_2, x_3) = & (\alpha + \beta) \int_{x_2}^{x_3} \left(\int_{x_2}^{u_1} F(u_1) \varphi_2(u_2 - u_1) du_1 + \beta(M_{\tau_1} - x_2) \right) du_2 + \\ & + \beta \left(M_{\tau_1} + M_{\tau_2} + \int_0^{x_2} F(u_1) du_1 - x_3 \right) \end{aligned} \quad (9)$$

$$Mf_3(\eta_3, x_4) = (\alpha + \beta) \int_{x_3}^{x_4} \left(\int_{x_3}^{u_3} \left(\int_{x_2}^{u_2} F(u_1) \varphi_2(u_2 - u_1) du_1 \right) \varphi_3(u_3 - u_2) du_2 \right) du_3 + \\ + \beta \left(M_{\tau_1} + M_{\tau_2} + M_{\tau_3} + \int_0^{x_2} F(u_1) du_1 + \int_{x_2}^{x_3} \left(\int_{x_2}^{u_2} F(u_1) \varphi_2(u_2 - u_1) du_1 \right) du_2 - x_4 \right), \quad (10)$$

$$Mf_{n-1}(\eta_{n-1}, x_n) = (\alpha + \beta) \int_{x_{n-1}}^{x_n} \left(\dots \left(\int_{x_2}^{u_2} F(u_1) \varphi_2(u_2 - u_1) du_1 \right) \dots \right) du_{n-1} + \\ + \beta \left(M_{\tau_1} + \dots + M_{\tau_n} + \int_0^{x_2} F(u_1) du_1 + \int_{x_2}^{x_{n-2}} \left(\dots \left(\int_{x_2}^{u_2} F(u_1) \varphi_2(u_2 - u_1) du_1 \right) \dots \right) du_{n-2} - x_n \right). \quad (11)$$

The total fine is equal

$$F(x) = \sum_{i=2} Mf_{i-1}(\eta_{i-1}, x_i) \quad (12)$$

The use of the classical approach to minimize the function $F(x)$ is due to the need to calculate multiple integrals and the value of the laws of distribution $F_i(t)$, which is not always possible. However, these difficulties can be eliminated by using methods of stochastic programming [14], in terms of which we formulate the problem. The problem is to find the minimum of function

$$F(x) = Mf(\eta, x) \quad (13)$$

under condition

$$x_{i+1} \geq x_i \geq 0, i = 2, 3, \dots, n, \quad (14)$$

where

$$f(\eta, x) = \alpha \sum_{i \in I} (x_i - \eta_{i-1}) + \beta \sum_{j \in J} (\eta_{j-1} - x_j) \quad (15)$$

moreover

$$\eta_i = \max(\eta_{i-1}, x_i) + \tau_i, i = 2, 3, \dots, n.$$

We note that function (15) is convex with respect to x_i for fixed η_{i-1} .

Problems (13) - (15) are a special case of the stochastic programming problem. To solve them, a stochastic quasi-gradient projection method [14] can be used, which is described by the procedure

$$x_i^{s+1} = \pi_x(x^s - \rho_s W_i^s), s = 0, 1, \dots, i; i = 2, 3, \dots, n, \quad (16)$$

where s is the iteration number; X is the projection area intersected by the constraints of (14); $\pi_x(\bullet)$ is an operator that maps each point $x \in R^\pi$ to a point $\pi_x(x) \in X$ so that $\|x - \pi_x(x)\|^2 \leq \|y - x\|^2$ for any $y \in X$; ρ_s is a step multiplier; W_i^s is a random vector such that $M(W_i^s | x_i^0, x_i^1, \dots, x_i^s) = F_x^A(x_i^s)$, $s = 0, 1, \dots, i$; $i = 2, \dots, n$, where $F_x^A(x_i^s)$ is the gradient (generalized gradient) of the function $F(x)$. The sequence $\{x_i^s\}$, $s = 0, 1, \dots, i$, conditioned by procedure (16), converges with probability 1 to the solution of problems (13), (14).

The computational scheme for solving the problem is formalized as follows. Let x^s be given on the s -th iteration (initial x^0 is given). The following steps are required to obtain the $(s + 1)$ -th approximation.

Step 1. Observe the implementation of x_i^s of the value x_i .

Step 2. Calculate the vector W_i^s .

Step 3. As a result of performing a design operation on the area $X = \{x_i : x_i \geq x_{i-1} \geq 0\}$, the following approximation is determined as follows:

$$x_i^{s+1} = \max(0, \max(x_{i-1}^s; x_i^s - \rho_s W_i^s)), s = 0, 1, \dots, i; i = 2, 3, \dots, n.$$

The above algorithm is easy to implement and does not depend on the type of law of distribution of random service time of the PSMP application.

It is convenient to use the value of the smoothed mean

$$Z^{s+1} = Z^s + 1/W^s (f(\eta^s, x^s) - Z^s),$$

which is an estimate of the mathematical expectation of a random function $f(\eta^s, x^s)$.

The algorithm defined for managing the load on the SC is thus an effective tool for managing the NDSC in order to provide maximum convenience in the implementation of information, analytical and expert services to PSMP with minimization of information processing time.

5 Conclusion

Proposed models allow to solve the problems of technical interoperability in terms of ensuring efficient loading and minimizing the time of servicing applications in distributed information systems by resolving the primal and dual problems of mathematical programming.

To resolve the primal problem, the use of the zero-one integer programming method is justified and an algorithm for its implementation is proposed. To resolve the dual problem, the use of stochastic programming method is justified and an algorithm for its realization is proposed. The algorithm defined for managing the load on the SCs is thus an effective tool for managing the NDSC in order to provide maximum

convenience in the implementation of information, analytical and expert services to PSMP with minimization of information processing time.

Proposed models of technical interoperability support and algorithms for managing the information and analytical environment of NDSC create opportunities for further improvement of systems for strategic planning based on the capabilities management of complex situational systems.

References

1. ISO/IEC 10746-2:2009 – Information technology — Open distributed processing — Reference model: Foundations. <https://www.iso.org/ru/standard/55723.html>
2. ISO/IEC 24765:2017 – Systems and Software Engineering – Vocabulary. <https://www.iso.org/standard/71952.html>
3. Kovalenko, O. E.: Compositional convergence of information technologies in systems of situational management based on the knowledge domain model. *Mathematical modeling in economy*. **2** (15), 39-44. (2019). doi: 10.35350/2409-8876-2019-15-2-40-45
4. New European Interoperability Framework (EIF). ISA² - Interoperability solutions for public administrations, businesses and citizens. https://ec.europa.eu/isa2/eif_en https://ec.europa.eu/isa2/sites/isa/files/eif_brochure_final.pdf
5. Zabrodskiy, V.A., Skurikhin, V.I.: Optimization of the functioning of the enterprise ACS. *Vyshcha Shkola, Kiev* (1978) (In Russian)
6. Glushkov, V.: Introduction to ACS, 2-nd ed. *Tekhnika, Kiev* (1972) (In Russian)
7. Morozov, A.O., Kosolapov, V.L., Otrishko, A.V.: Means of information service of objects of intellectual property. *Mathematical Machines and Systems*. **3**: 136-147 (2002). (In Ukrainian)
8. European Interoperability Reference Architecture (EIRA). ISA² - Interoperability solutions for public administrations, businesses and citizens. <https://joinup.ec.europa.eu/asset/eia/description>
9. Koulou, A., Zemzami, M., El Hami, N., Elmir, A., Hmina, N.: Optimization in collaborative information systems for an enhanced interoperability network. *Int. J. Simul. Multidisci. Des. Optim.* **11**, 2 (2020) doi: 10.1051/smdo/2019021
10. Leal, G., Guedria, W., Panetto H.: A semi-automated system for interoperability assessment: an ontology- based approach. *Enterprise Information Systems*. **14** (3), 308-333. (2020). doi: 10.1080/17517575.2019.1678767
11. Guédria, W., Naudet, Y., Chen, D.: Maturity Model for Enterprise Interoperability. *Enterprise Information Systems* **9** (1), 1–28. (2015). doi:10.1080/17517575.2013.805246
12. Gabriel da Silva Serapião Leal, Wided Guédria, Hervé Panetto: Interoperability Assessment: A Systematic Literature Review. *Computers in Industry*. **106**, 111-132. (April 2019). doi:10.1016/j.compind.2019.01.002
13. Khisro, J., Sundberg, H.: Enterprise interoperability development in multi relation collaborations: Success factors from the Danish electricity market, *Enterprise Information Systems*, (2018). doi: 10.1080/17517575.2018.1528633
14. Yermolyev, Yu.M.: Methods of stochastic programming. *Nauka, Moscow* (1976). (In Russian)

Method of Monitoring of Young Athletes' Physical State Indicators Based on Wearable Devices Usage

Anzhelika Parkhomenko^[0000-0002-6008-1610], Anastasiia Volynska^[0000-0002-6008-1610],
Yaroslav Zalyubovskiy^[0000-0002-6847-8778], Andriy Parkhomenko^[0000-0001-8265-0530], and
Marina Kalinina^[0000-0002-7327-0285]
National University «Zaporizhzhia Polytechnic», 64, Zhukovskogo str., Zaporizhzhia, 69063,
Ukraine
parhom@zntu.edu.ua, stasja.volynska@gmail.com, zalubovskiy.
ja@kliongroup.com, andriy2872073@gmail.com, shinkari73@gmail.com

Abstract. The task of monitoring of human physical state during sports training is considered in this paper. The possibilities of popular on the market wearable devices have been analysed for its solving. It is shown that the main restriction on the usage of the considered devices is the absence of the function of data collecting for the next processing and recommendation formation. Therefore, this work is devoted to the research and development of the method of monitoring the indicators of young athletes' physical state based on the usage of inexpensive fitness bracelet and developed web application.

Keywords: indicators of physical state, wearable devices, fitness bracelet, heart rate, anaerobic threshold, dynamic monitoring, Web Bluetooth API, web application.

1 Introduction

The formation of a professional athlete is a long process that requires constant systematic training related to physical and psychological stress. For this process to be fruitful and not harm the athlete, the coach must be able to properly plan the training program. It is especially important to monitor the health status of young athletes as very intense physical activity can lead to irreversible consequences. Therefore, the implementation of the opportunity for the coach to effectively monitor the indicators of the physical state of juniors during training is an urgent task [1, 2].

In general, an active lifestyle and dynamic monitoring of own physical status are becoming increasingly popular today. Accordingly, every year on the market more and more wearable devices appear that help to monitor indicators of human physical state. Bill Gates cited the usage of wearable devices for monitoring cardiac activity, as one of the most important technologies in 2019 [3]. Different modern heart rate sensors, fitness bracelets, smart watches and even wireless Bluetooth earbuds are presented on the market that can be connected with a smartphone for cardio data transmission [4]. Such companies as Apple, AliveCor, Withings have presented wearable devices with electrocardiogram (ECG) function, which provide accuracy close to the accuracy of medical equipment [5]. Another example is low-cost clinical-

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

grade AliveCor KardiaMobile ECG Monitor. This is a portable electrocardiograph with two electrodes that can be attached to the back of iOS and Android smartphones [6]. Typically, all these devices are focused on individual usage and the data are usually just displayed locally for self-monitoring. They do not provide centralized data collection and storage for processing and generating recommendations for the following physical activities planning. In some cases, where data collection and visualization in the cloud service is provided, nevertheless, the data is available only to the owners of the devices for self-monitoring.

The goals of the work are research and practical implementation of the method of obtaining and processing information on the indicators of the physical state of the juniors directly in dynamics during sport training.

2 State-of-the-Art

Studies have shown that the main task of a sports coach is to adapt the standard training system to the individual capabilities and characteristics of a particular athlete, as well as setting real training goals and current tasks.

In the management of sports training, we can distinguish such directions as:

- The control of sport activities (accounting and analysis of the work done).
- The level of athlete's readiness monitoring (biological, mental and technical).
- Generalization and analysis of monitoring data to make changes in the planning of sports activities.

For example, in the framework of the official program of sports training of children and juniors, the Squash Federation of Ukraine uses a mandatory individual card of the athlete, but in paper form [7]. The record is maintained from the very beginning of the child's training and is further performed at each new stage of preparation. The monitoring is performed in two directions: medical examination and control of physical state indicators.

In the study of the functionality of existing software products in this area, the following was revealed. Most of the existing systems provide standard accounting functions based on the schedule, occupancy of halls, coaches and the mark of attendance by the client. Official web sites of sports associations and sports clubs generally have informational character.

An example [8] is the fitness club training accounting system, which is presented as a mobile application for the Android and iOS platforms. Its main functions are customer registration; season ticket purchase; training attendance records; recording for group classes; data about club staff; information about training plans. The application does not monitor the state of health during training, but it is possible to enter anthropological indicators (such as weight, waist size, etc.) and thus to monitor their changes.

Work [9] shows the results of the development of web applications for the Squash Federation of Ukraine. It provides the following functions: user registration; weekly planning of the training schedule for each junior; viewing of schedule; distribution of messages about changes in the schedule; confirmation of participation in the training;

assessment by the coach of the athlete for each type of exercise and the overall assessment for the training and report generation.

The authors of [10] have shown the results of the study on the interdependence of the psychological state and recovery period of athletes depending on several parameters, such as the training load, heart rate, recovery of heart rate and heart rate variability. Data processing was carried out using LISREL (linear structural relations) which is a proprietary statistical software package. However, data collection was carried out by traditional methods, without the usage of specialized software tools.

The paper [11] analyzes the objective and subjective indicators of the athletes' physical state. Objective indicators were taken into account both at rest (for example, blood counts, heart rate), and during exercise (for example, oxygen consumption, heart rate reaction). They were compared with subjective indicators (e.g. mood, perceived stress). All indicators were also evaluated by their response to acute and chronic training load. An electronic search was conducted using the existing data from different databases, so the results are not suitable for a particular group of kids and kind of sport.

The authors [12] suggested monitoring the weekly training load and assessing it based on the health rating scale. The described tools are based on the use of Microsoft Excel also provides for the collection of information and manual data entry for subsequent analysis and providing feedback to the athlete to manage stress or fatigue and prevent possible consequences.

Thus, the investigation has revealed that the existing approaches and software systems are not focused on the control of the health of a person who is engaged in the gym. However, monitoring indicators of the athlete's physical state during a particular training session and subsequent analysis of these indicators are the most informative for assessing the success of the training plan as a whole. Such accounting is useful for both amateurs and professional athletes, especially for beginners, including kids.

3 Analysis of Wearable Devices and Applications for Monitoring a Person's Physical State

In addition to professional medical equipment various wearable devices are actively used today to monitor the indicators of the person's physical state. They are fitness bracelets and smart watches.

Modern fitness bracelets can be conditionally divided into three groups: trackers, smart bracelets and health bracelets [13]. Trackers are highly specialized gadgets that support the functions of a heart rate monitoring, pedometer, alarm clock and tracking sleep phases. Smart bracelets are advanced tracker models. In addition to fitness functions, they have an indication of calls and messages, reminders about classes, GPS and a compass. They are equipped with a display and look like a "smart" watch. The most popular manufacturers are Jawbone, Polar, Fitbit, Garmin, Samsung, Xiaomi. Health bracelets are designed for professional usage in medicine or sports. The main differences of these devices are the built-in blood pressure monitor, as well as sensors for measuring the level of oxygen in the blood. GSKIN E11 fitness

bracelet is designed both for monitoring sports performance and for monitoring the health status of elder people. In addition to common functions such as measuring heart rate, steps and calories, this fitness bracelet provides measuring of blood pressure, oxygen level in the blood and building an ECG. During the work budget fitness bracelet Xiaomi Mi Band 2 was examined. This device performs all the basic functions of a standard fitness bracelet, such as measuring heart rate, calculating steps taken. Additionally, it is equipped with an OLED screen with which the user can view the number of steps taken per day and measure the pulse without using a smartphone.

Smart watches are computerized devices that act as a watch, fitness tracker, switch and control panels. They can perform as:

- Applications for a smartphone (broadcast calls, messages, transmit voice commands, launch applications on the phone, control other types of smart technology, etc.).
- Independent gadgets (make calls, send messages, take pictures, play music, etc.).

Depending on the operating system, a smart watch can be the following types: Android (e.g. Garmin Fenix, SMA S3); Android Wear (e.g. Asus ZenWatch, Huawei Watch, Motorola 360); IOS (Apple Watch); Tizen (Samsung Gear Sport, Samsung Gear S3). Almost all models can be used to monitor the physical state: support the training program, track the route and determine the heart rate and the number of steps based on the usage of a heartbeat sensor and accelerometer. The watch can synchronize with other computers or smartphones with the usage of wireless data transfer technology. This type of equipment includes Samsung's Galaxy Watch Active. Using special motion sensors and built-in GPS, this watch automatically monitors up to seven types of activity, as well as the user's heart rate. It sends instant notifications when anomalies are detected.

The software for fitness bracelets is a specialized utility that allows to send various information from fitness bracelets to smartphones using the wireless connection. After taking measurements, the fitness bracelet is synchronized with the smartphone and transfers the data for processing to the application. In addition, some programs allow communication between people who use these bracelets. Almost all manufacturers of fitness bracelets provide their own unique application that works only with devices from this manufacturer.

For example, Mi Fit is a mobile application for fitness gadgets manufactured by Xiaomi, including fitness bracelets. The user enters his biometric data (gender, age, blood type, weight, height), sets a goal that he seeks to achieve and synchronizes his fitness bracelet with the smartphone. The application tracks the number of steps taken over a certain period of time, calculates the number of calories burned. The application has two modes of usage: daily mode, when only the number of steps is calculated, and a training mode. In training mode, the user determines the type of activity and the application, at certain intervals, measures the pulse and determines the level of anaerobic threshold (AT). After training, a general report is displayed in graphs and statistics. The application is designed for operating systems such as Android and iOS.

Google Fit is the activity monitoring app designed for both smartphones and smart watches, supported by the Android platform. It is a universal application and can sync

with Android Wear, Xiaomi Mi, Runkeeper, Sleep as Android, Nike +, Lifesum, MyFitnessPal, Basis, Strava, Withings. In addition, the application is synchronized with other programs that track your health status. All data is stored on the server and the user can see it not only through the application but also through his personal Google account. The advantage of this software is that the data are obtained from many synchronized with smartphone devices and from sensors that are located on the smartphone.

Thus, fitness bracelets can be used as an inexpensive solution for dynamic monitoring of athletes' physical state indicators. However, the urgent task is to develop a web application for centralized data collecting from such devices and processing for further usage by coaches in training planning.

4 Investigation of Variants of Monitoring System Realization on the Basis of Fitness Bracelet

Xiaomi Mi Band 2 fitness bracelet was chosen as a device for data collection, due to its low price and at the same time the availability of basic functionality for monitoring athletes' health indicators.

The manufacturer does not provide an open Application Programming Interface (API) either for the bracelet or for its proprietary application. The fitness bracelet is a Bluetooth Low Energy (BLE) client-device and it transmits data via Bluetooth. To obtain data, we needed a device that would act as a server. This role can be performed either by a personal computer (PC) or a smartphone.

The implementation of the technology for obtaining data transmitted by the device was an important task of the project. Several variants have been considered:

- To write own mobile application that would synchronize with the device and exchange data.
- To receive data from another application through the API.
- To receive data without the usage of a smartphone but with a PC.

The first option was to synchronize the bracelet with a smartphone. For this purpose, an Android application was developed on the basis of the library for working with BLE devices. This application transfers data to the web application server. The first step in the interaction of the application with the BLE device is to connect to the Generic Attribute Profile (GATT) server on the device. For this aim, the `connectGatt()` method is used. The application connects to the GATT server located on the BLE device and returns an instance of `BluetoothGatt`, which can be used to perform GATT client operations later. The Android application is a GATT client. `BluetoothGattCallback` is used to obtain the results by the client, such as the status of the connection and any additional GATT client operations. As soon as the application connects to the GATT server and discovers the services, it can read and write attributes, where it is supported. In practice, the application hasn't connected to the bracelet all the time. Sometimes the device hasn't appeared at all during the search. When the connection could be established, other imperfections of the used library

were revealed, in particular, the characteristics and descriptors could not be recognized automatically.

Another option was to receive data through the universal Google Fit application. Data from the application is automatically synchronized with the user's Google account and stored on the server. The documentation points out the API with which it is possible to receive and to add data to this server. This was the easiest variant since data can be obtained directly to a previously developed web-based application for training squash players and there is no need to create additional software. During the experiment, it was revealed that it was possible to get from the Google server only the data that has been sent there from our application. The data obtained by it from a fitness bracelet was inaccessible. Thus, we can conclude that this API is necessary only for the developers of applications for fitness bracelets who want to synchronize data from their devices with Google accounts of their users.

The next phase of research was to consider a new Web Bluetooth technology. The first version of this specification allows client-side websites with a central role to connect to GATT servers through BR/EDR (Basic Rate/ Enhanced Data Rate) or BLE (Bluetooth Low Energy) connection. This means that Web Bluetooth allows to control any BLE device, including a smart bracelet, directly from a PC or a smartphone without the installation of a special application. In addition, such a solution will be cross-platform. Web Bluetooth technology is now supported by all popular browsers of the latest versions: Chrome, Samsung Internet v6.4, Opera, Servo, Firefox, Microsoft Edge, Safari. Although the solution is cross-platform, restrictions on operating systems exist. The technology is supported on the following operating systems: Android 6.0 Marshmallow+, iOS X Yosemite+, Linux Kernel 3.19+ and BlueZ 5.41+, Windows 10. It is necessary to know the list of characteristics and descriptors for working with a BLE device. We got these parameters using the free configuration software nRF Connect for BLE devices. This app connects to any BLE device, detects and analyses services and features. Table 1 contains UUID (Universally Unique Identifier) characteristics of Mi Band bracelet and Fig. 1 shows results of its scanning.

Table 1. UUID characteristics of the device

Name	UUID
Hardware service (HRDW)	0000fee0-0000-1000-8000-00805f9b34fb
Heart Monitor Service (HMS)	0000180d-0000-1000-8000-00805f9b34fb
Heart Rate Measure Characteristic (HRM)	00002a37-0000-1000-8000-00805f9b34fb
Heart Monitor Control Characteristic (HMC)	00002a39-0000-1000-8000-00805f9b34fb
Sensor Characteristic (SENS)	00000001-0000-3512-2118-0009af100700

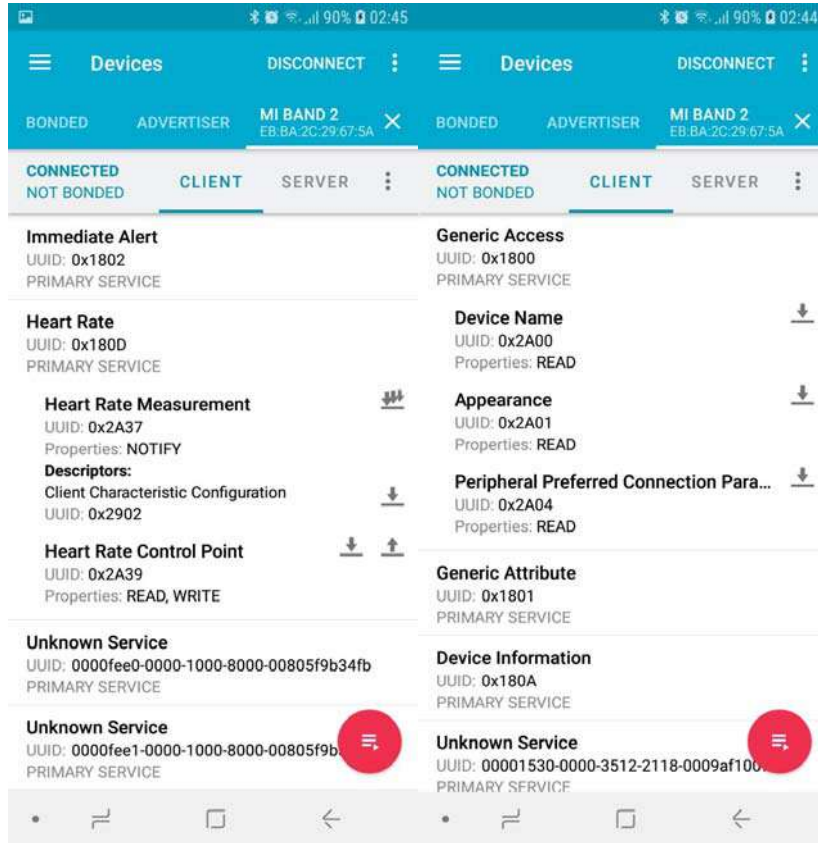


Fig. 1. The results of fitness bracelet Mi Band 2 scanning with nRF Connect

Web Bluetooth provides an API for development in the JavaScript language. The Internet also has an example of a working application for searching and connecting to a fitness bracelet. This application stably connected to the device and took heart rate measurements. Therefore, as a result of the study, Web Bluetooth technology was chosen for the development.

5 Model and Algorithm of Athletes' Physical State Indicators Identification

Existing models of training effects development use three groups of key variables. They are quantified training load, quantified athletic performance and variables that describe the state and dynamic characteristics of fatigue and fitness [14]. Training load is a combination of intensity, duration and frequency of training. In this work, the TRIMP (TRaining IMPulse) method is used for assessing training intensity, since it takes into account the combination of the external load and the physiological

response of the athletes' body [15]. The TRIMP based model can be presented in the form (1):

$$M = \langle t, x, y, z \rangle \quad (1)$$

where t is a training volume, measured in minutes, x is a set of indicators related to heart rate (HR) or lactate response, y is a set of indicators related to maximum HR or oxygen consumption; z is a set of indicators related to ventilation or anaerobic threshold (AT).

In accordance with the proposed algorithm (Fig. 2), such indicators as the maximum HR during training, the pulse zones and their duration, the level of metabolism or AT and the pulse zone in accordance with the AT are calculated.

Stage 1. Assessment of the intensity of physical activity. Determining the basic values of the maximum HR depending on the age (Table 2) and comparing them with current ones during training. Typical data for juniors are shown in Table 2.

Stage 2. Calculation of training intensity zones, which is based on a percentage of the maximum HR (Table 3).

Stage 3. Calculation of the AT level based on special tests to obtain the most accurate data on training zones. If there is no test data, it is possible to use the theoretical data from Table 4.

Stage 4. Evaluation of levels of training zones according to Table 5.

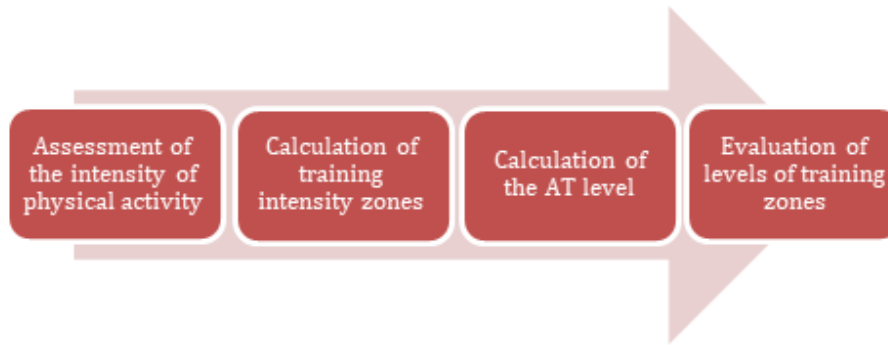


Fig. 2. Algorithm of athletes' physical state indicators identification

Table 2. Typical value of maximum HR for juniors

Age	Boys	Girls
11	200	199
12	199	198
13	198	197
14	197	196
15	197	195
16	195	195
17	194	194

Table 3. Calculation of heart rate zones for HR = 197

Zone	% from MHR	Heart rate	Feeling of exercise
1	<66%	<110,2	Very easy
2	66%-73%	110,2 – 121,9	Easy
3	73%-84%	121,9 – 140,2	Normal
4	84%-91%	140,2 – 151,9	Hard
5	>91%	>151,9	Very hard

Table 4. Theoretical calculation of the AT level

Age	HR
9-11	179±3
11-14	161±5
14-17	145±7

Table 5. Calculation of heart rate zones for AT = 161

Zone	Description	% from AT	HR	Feeling of exercise
1	Restoration	<69%	<111	Very easy
2	Endurance	69%-84%	111-135	Easy
3	Tempo	84%-95%	135-152	Normal
4	AT	95%-106%	152-170	Hard
5	Maximum oxygen consumption	>106%	>170	Very hard

6 Development of the Web Application for Dynamic Monitoring Indicators of the Athletes' Physical State

Functional requirements for the application are pulse reading; synchronization with the web application; determination of indicators of athlete's physical state during training; report generation. Potential users of the system are coaches, juniors and parents. The program is a part of a previously developed web application that has a three-tier architecture (Fig. 3). It includes the server level, as well as the database with which the application server level works. A client program connects to the application server. Fitness bracelet connects to the client program using the Web Bluetooth API. HTML, CSS, JavaScript, PHP languages were used for the application development. A cross-platform integrated development environment PhpStorm from JetBrains has been also used.

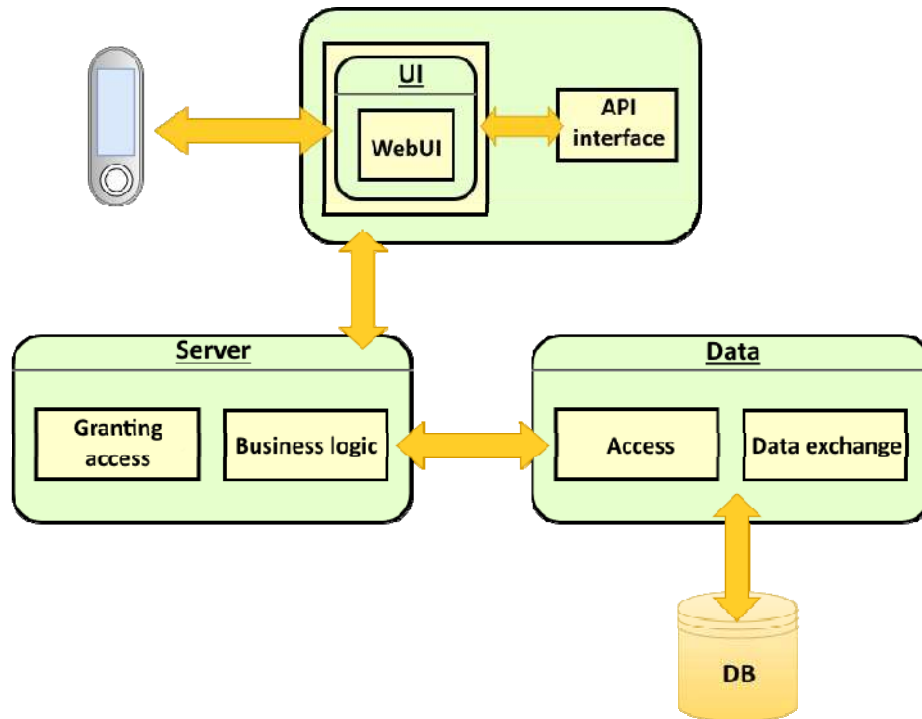


Fig. 3. Web application architecture

Conventionally, the process of functioning of the system is divided into two stages: connecting to the device and processing the received data. User interaction with the application is necessary only at the beginning and at the end of the process.

At the beginning, the user selects the necessary device and connects to it. After that, the application via Bluetooth receives data from the fitness bracelet. Heart rate and time data are stored in an array. Further, the program calculates the pulse zones, finds the maximum heart rate for the entire training period. All received data is displayed in the form of a graph and a table. After that, the received data is transmitted to the server, where it is entered into the database.

Fitness tracker connection is implemented in JavaScript and it is located in the `miband.js` file.

In order to connect the device to the web application, the `navigator.bluetooth.requestDevice` method is used. This method accepts an array of filters that return only devices that match the specified Bluetooth GATT functions or device names. As noted above, the Mi Band 2 device has a Heart Rate feature with `UUID = 0x180d`. After executing this code, the browser will display a window with the devices found near it. The functions used for working with the fitness tracker are shown in Table 6.

Table 6. Description of miband.js functions

Function name	Description
device.gatt.connect()	Connecting to a device
getPrimaryService(0x180d)	Getting main service with UUID = 0x180d - heart rate reading
getCharacteristic('00001531-0000-3512-2118-0009af100700')	Getting pulse measurement characteristics
characteristic.startNotifications	Tracking data changes
device.gatt.disconnect()	Disconnect from device

During training, an array is formed with the pulse and measurement time data. Data processing occurs after the device is turned off. Processing of the received data occurs in the miband.js file using the function and CharacteristicValue (data []). An array with the received data is passed to the function, there, using the standard switch () operator, the data is divided into 5 pulse zones and the time of athlete's stay in this or that zone is calculated. The maximum heart rate is found. The array is not transferred to the server using AJAX technology.

This technology enables the "background" exchange of browser data with a web server. As a result, when updating data, the website does not restart completely, and web applications are faster and more convenient. Below there is a function for transmitting the data to the server.

```
$.ajax({url: "test.blade.php",type: "POST", data:
({source: data, len: data.lenght()}) dataType: "int",
beforeSend: function(){ console.log("Please wait..."); },
success: function(d){ console.log ("Success!"); });
test.blade.php - the file to which the data is sent;
POST - type of data transmission;
source: data, len: data.lenght() - the transmitted
array and its length;
dataType: "int" - data type;
beforeSend:function(){ console.log("Please
wait...");}, - when sending data to the console, a message
is displayed "Please wait..."
function(d){console.log ("Success!"); } - if the
sending was successful, a message is displayed in the
console "Success!".
```

Views contain HTML code that the application passes to the user. Views are in the resources/views directory.

The Blade Templating Engine template structure was used for this work. It uses special tags and functions to separate view logic and application code. In order to separate the main part of the layout from the content, the content is highlighted with @ section marks (section name). Layout of page templates is done with the usage of

layers. A layer is implemented with the <div> tag. Layers are independent of each other. One layer may contain another.

Fig. 4 shows a web page with basic information about indicators of an athlete's physical state during training. The data collected by system can be effectively used by coach for improving the effectiveness of training for optimal development of athletes.

For individual planning of athletes' training loads a special environment can be realized based on the fuzzy reasoning model for decision support system [15-16].

For the recommendations formation about the individual combination of intensity, duration and frequency of training for different juniors, knowledge-based recommendation method can be realized that were presented in [17-18].

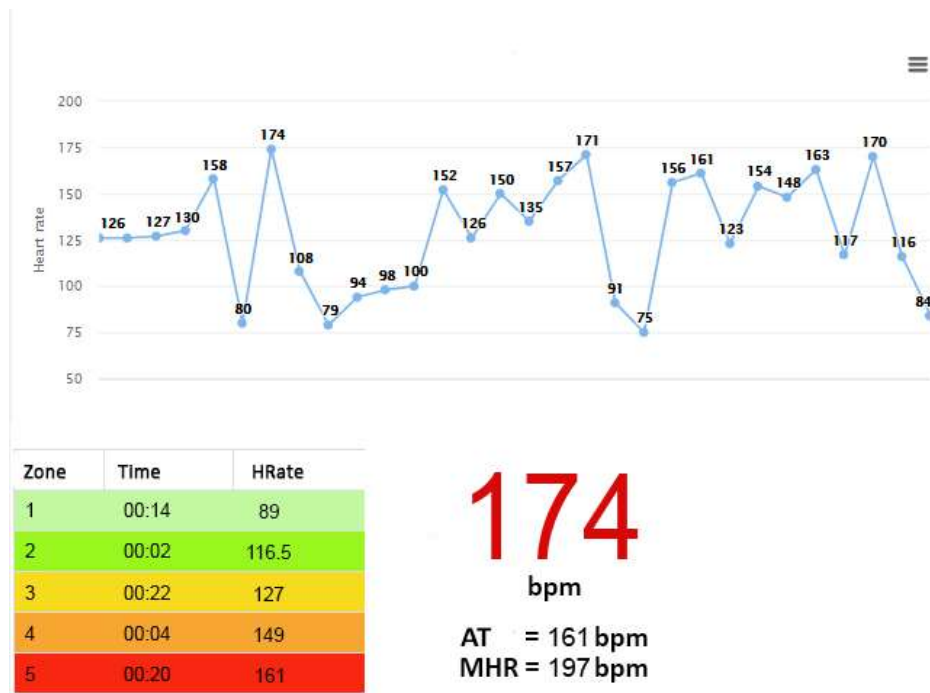


Fig. 4. Form with indicators based on heart rate data

7 Conclusion

Information regarding the health, level of physical development, biological maturation of athletes is evaluated in the framework of an in-depth medical examination by the medical specialists. Nevertheless, dynamic monitoring of the main indicators of athletes' physical state is very important for trainers, as it allows assessing the real state and the level of athlete's functional capabilities.

This work proposes a method for monitoring athletes' state during training based on fuzzy sets of variables. Basic indicators have been identified, which can be judged by the physical loads in training and relative to the athlete's physical development.

Various methods of obtaining data from a fitness bracelet to a web application have been analyzed and the best way to solve the task was found. A web application has been developed that became a part of the management system for training squash athletes.

The scientific novelty of the work lies in the fact that the method of monitoring indicators of athletes' physical state has been further developed through the implementation of the method of reading data from a fitness bracelet with the usage of Web Bluetooth technology.

The practical value of the work is that the implemented method makes it easy to obtain data for monitoring the physical state of athletes directly during training.

In the future, the application requires the implementation of the ability to receive data from multiple devices simultaneously.

Acknowledgements. This work is carried out partly with the support of Erasmus + KA2 project BIOART «Innovative Multidisciplinary Curriculum in Artificial Implants for Bio-Engineering BSc/MSc Degrees» (586114 – EPP – 1 – 2017 – 1 – ES - EPPKA2 – CBHE – JP)

References

1. Stöggl, T.L., Sperlich, B.: Training intensity, volume and recovery distribution among elite and recreational endurance athletes. *Frontiers Media SA* (2019)
2. Lysov, L.P. Methodical planning of the training program: scientific guide. SPb (2016)
3. How we'll invent the future, by Bill Gates. MIT Technology Review. <https://www.technologyreview.com/lists/technologies/2019/>
4. Morris, R.: The 6 best heart rate monitors of 2020. <https://www.verywellfit.com/best-heart-rate-monitor-4157709>
5. 7 of the best fitness trackers for monitoring heart rate. <https://mashable.com/roundup/best-fitness-trackers-smart-watch-heart-rate/>
6. Mobile health moves forward: FDA approves AliveCor's Heart Monitor for the iPhone. <https://techcrunch.com/2012/12/04/mobile-health-moves-forward-fda-approves-alivecors-heart-monitor-for-the-iphone/>
7. Children's squash development program. <http://squash.ua/programma-podgotovki-detej/>
8. Mobile application for the Colosseum fitness club. <https://cmsmagazine.ru/journal/cases-4339/>
9. Parkhomenko, A.V., Pozdnyakov, O.A., Volynska, A.V., Ilyin, E.M.: Automated system of squash training planning and accounting. In: The annual scientific-practical conference of teachers, scientists, young scientists, graduate students and students «Science Week-2018», Zaporizhzhya, Ukraine, 1100 (2018) (in Ukrainian)
10. Nicolas, M., Vacher, P., Martinet, G., Mourot L.: Monitoring stress and recovery states: Structural and external stages of the short version of the RESTQ sport in elite swimmers before championships. *J. of Sport and Health Science*. **8** (1), 77-88 (2019)
11. Saw, A.E., Main, L.C., Gatin, P.B.: Monitoring the athlete training response: subjective self-reported measures trump commonly used objective measures: a systematic review. *Br J Sports Med*. **50** (5), 281–291 (2016)
12. McFarland, M., Bird, S.P.: A wellness monitoring tool for youth athletes. *J. of Australian Strength and Conditioning*. **22** (4), 22-26 (2014)
13. Fitness bracelet against smart watches. What to choose? https://kfc.ua/ru/blog/fitnes_braslet_protiv_smart_chas__chto_zhe_vybrat.html

14. Borresen J., Lambert M.I.: The quantification of training load, the training response and the effect on performance. *J. Sports Medicine*. **39** (9),779-795 (2009)
15. Taha, T., Thomas, S.: Systems modelling of the relationship between training and performance. *Sports Medicine*. **33** (14),1061-1073 (2003)
16. Ling, S. H., Nuryani, Nguyen, H. T.: Evolved fuzzy reasoning model for hypoglycaemic detection. In: The annual international conference of the IEEE engineering in medicine and biology, 4662-4665 (2010)
17. Parkhomenko, A., Gladkova, O., Parkhomenko, A.: Recommendation system as a user-oriented service for the remote and virtual labs selecting. *Advances in Intelligent Systems and Computing*. **917**, 600-610 (2019)
18. Subbotin, S., Gladkova, O., Parkhomenko, A.: Knowledge-based recommendation system for embedded systems platform-oriented design. In: The Proceedings of the XIII International scientific and technical conference on computer science and information technologies, Lviv, Ukraine, 368-373 (2018).

Computer Modeling of Microstructures with Probabilistic Cellular Automata Method Using Different Nucleation Rate Functions

Oleksii Vodka ^[0000-0002-4462-9869]

Department of Dynamics and Strength of Machines National Technical University “Kharkiv Polytechnic Institute”, Kharkiv, Ukraine
Oleksii.Vodka@gmail.com

Abstract. Computer modelling widely uses cellular automata method for simulation of the microstructure of various materials. This work deals with the modification of cellular automata method – probabilistic cellular automata. The software for the synthetic generation of microstructures of polycrystalline materials has been developed. Using the software, the coefficient of the form, the normalized grain area, the scale factor and the angle of grain orientation are determined. For the obtained data, statistical processing and probability density functions have been obtained. According to statistical parameters, the comparison of the obtained results with the parameters of the microstructure of pure iron and H62 copper alloy has been performed. Modelling has been performed with different nucleation rate models.

Keywords: microstructure, probability cellular automata, microstructure quantitative characteristics, nucleation rate

1 Introduction

Computer simulation is a powerful tool in reproducing microstructures of different materials. It is the use of computer simulation to reproduce various microstructures quickly and efficiently. For practical purposes, it is important to reproduce microstructures that are statistically equivalent to real materials. They have the greatest influence on the mechanical properties of materials.

One common method for modelling the microstructure of polycrystalline materials is the method of cellular automata. Cellular automata began to be used in the middle of the XX century [1–3]. The term "cellular automaton" means a set of dependent elements with given states and rules. These rules determine the states of these elements and dependencies between them vary in time. Time and states are discrete. The use of the described models for the formal modelling of self-reproductive organisms was first proposed in the work of von Neumann [2]. Elements of cellular automata are proposed to be represented by one-dimensional, two-dimensional or multidimensional infinite rectangular tables. The state of the element varies depending on its state and state of its closest neighbours.

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

Cell-based models are widely used to predict behaviour in various industries. Cellular automata method widely uses in cryptography [4, 5], image processing [6–9], biology [10–12], physics [13, 14]. All these works cover a lot of possibilities of using the method of cellular automata. However, one of the most successful applications of the method is in the field of material science and engineering [15]. A lot of modifications of the method have been developed for the simulation of crystallization, recrystallization and evolution of the internal structure of materials [15–25]. Papers [16, 18, 19] deals with recrystallization in metals and alloys. In [17] authors reconstructs deformed microstructure using a cellular automaton. Irregular cellular automaton modification introduced in [20]. Microstructure and grain structure evolution during recrystallization [21, 23], additive manufacturing [22] and welding process [25] are also modelled.

In most works, the method of cellular automata is used to synthetically reproduce the microstructure of materials. However, the mentioned works almost do not investigate the influence of cellular automaton parameters on the quality of the created microstructures. Also, to increase the possibilities of modelling processes with stochastic nature, the modification of the method has been developed. Probabilistic method of cellular automata also uses in material internal structure simulation [16, 17, 19]. There are other ways to material modelling [26–28] and simulation planning [29, 30].

It should be noted work [17]. In the work partially investigated the distribution of grain size. However, the simulation is performed on a hexagonal grid. Thus, almost completely absent works devoted to the study of the quality of synthetically created microstructures and their statistical correspondence with real microstructures.

The scientific novelty of this paper is to develop a modification of the method of cellular automata, which uses probabilistic neighbours, as well as to determine the most successful parameters of the method by comparing the results with real microstructures. The practical value of developing such a modification of the method is to improve the quality of synthetic microstructures that can be used to model the mechanical behaviour of materials in multilevel modelling.

2 Problem statement

It is proposed to develop software that allows simulating the microstructure of metals using the probabilistic method of cellular automata. To do this the following tasks have been formulated:

- To study the developed modifications of the method of cellular automata for the simulation of microstructures;
- To develop software that allows simulating the microstructure of metals according to randomly assigned centres of grains and various of nucleation rate functions;
- Test the work of the software, determine the statistical characteristics of the size, shape and orientation of the grains;
- Compare the results with real microstructures.

3 Modelling with cellular automata

The cellular automaton K from a mathematical point of view [31–33] is an ordered set of four components:

$$K = \langle z^d, \nu, A, \varphi \rangle \quad (1)$$

where z^d is set of d -dimensional vectors with integer coordinates (cell space);

$\nu = \{\nu_i \mid \nu_i = (x_{i1}, \dots, x_{di}), \exists \nu_i = 0\}, i = 1, \dots, m$ is a finite set of power m of vectors with zero vector (cell neighbourhood template);

A is the finite set of power k states of the cell with a dedicated state of rest $\emptyset$ (the alphabet of the cellular automaton);

φ is the local function of the transitions, defined in the discrete-time, which changes the states of the cell, which is a zero element in the template, depending on the state of the cells that form the neighbourhood pattern $\varphi: A^m \rightarrow A$; with $\varphi(\emptyset, \emptyset, \dots, \emptyset) = \emptyset$. The state of all cells at the time t creates the current configuration $c^t: z^d \rightarrow A$.

All cells form a cellular automaton grid. Grids can be different types, differing in size and shape of cells. Each cell is a finite automaton which states are determined by the states of neighbouring cells and its states.

Cellular machines, in general, are characterized by the following properties:

- Changing the values of all cells occurs simultaneously after calculating the new state of each grid cell.
- The grid is homogeneous. It is impossible to distinguish any two places on the grid over the landscape.
- Interactions are local. Only the surrounding cells (usually neighbouring ones) can affect this cell.
- The set of states of the cell is finite.

In a two-dimensional (planar) case, the grid is implemented by a two-dimensional array. Each cell has eight neighbours. To eliminate boundary effects, the grid can be wrapped in a torus. It allows to use the following ratio for all automaton cells:

$$a'_{i,j} = \varphi(a_{i,j}, a_{i-1,j}, a_{i-1,j+1}, a_{i,j+1}, a_{i+1,j+1}, a_{i+1,j}, a_{i+1,j-1}, a_{i,j-1}, a_{i-1,j-1}) \quad (2)$$

In this work, for modelling the crystallization process, the cell can be in two states: the melt and the crystal.

To move from the deterministic to the probabilistic method of cellular automata, the following idea is used. Rules for switching from one state become non-deterministic. Thus, the transition from state to state occurs with a certain probability. Figure 1 shows typical transfer rules ϕ for the probabilistic method.

The application of this probabilistic approach allows creating more complicated rules of transition from state to state and reducing the effect of discreteness on the results of the algorithm.

From a physical point of view, the crystallization process should be uniform in all directions. Due to local fluctuations and temperature gradients, the uniformity of the chemical composition of the melt, crystallization occurs nonuniformly in different directions. To simulate this effect, it is proposed to set the crystallization rate in the form of an ellipse. The ellipse radii correspond to the crystallization rate in the corresponding directions. To determine the transition probability, an ellipse has been projected onto field cells. The shaded area indicates the probability of the transition of neighbouring cells to a solid-state. A more detailed description of the algorithm is given in [34].

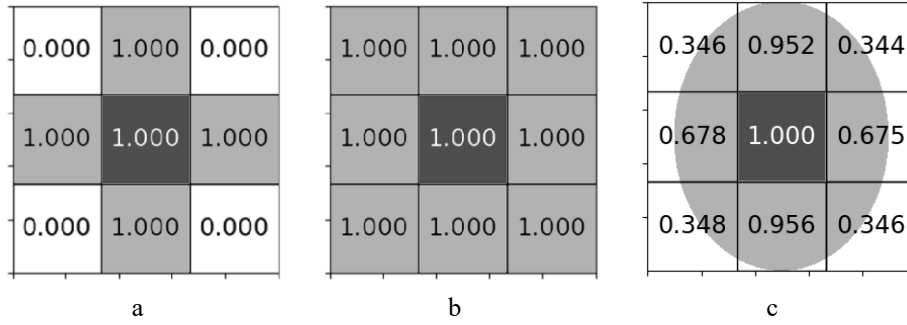


Fig. 1. Cell neighbours: a – von Neumann; b – Moore; c – probabilistic neighbour

4 Nucleation rate models

One of the characteristics affecting the microstructure of the material is the nucleation rate ($J(n)$). Researchers in their papers [35–38] note that the nucleation rate depends on the Gibbs free energy. But it is difficult to introduce these models in discrete space and time of cellular automata. Therefore, to evaluate the influence of crystallization parameters on the geometric characteristics of the formed microstructures, it is proposed to use the generalized models given [35–38] as the following models (3) – (7). These models are based on the assumption that the number of crystallization centres N_{gr} for all models has the same value at N_p iteration of the algorithm. The nucleation rate is defined as a derivative of the number of grains $N(n)$ function, n is the iteration number of the algorithm. The visualization of the proposed models is shown in Fig. 2.

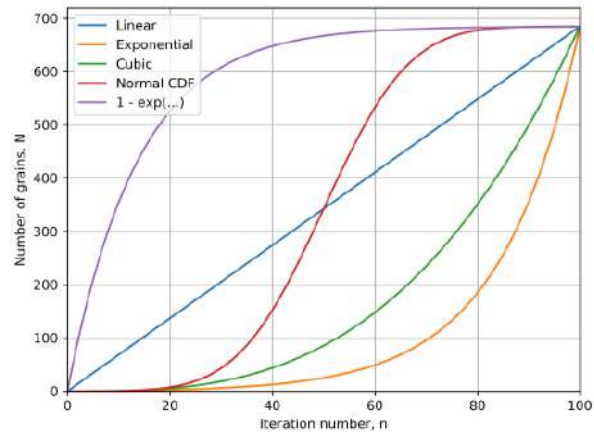
$$N(n) = \frac{N_{gr}}{N_p^3} n^3, \quad J(n) = \frac{3N_{gr}}{N_p^3} n^2, \quad (3)$$

$$N(n) = \frac{N_{gr}}{N_p} n, \quad J(n) = \frac{N_{gr}}{N_p}, \quad (4)$$

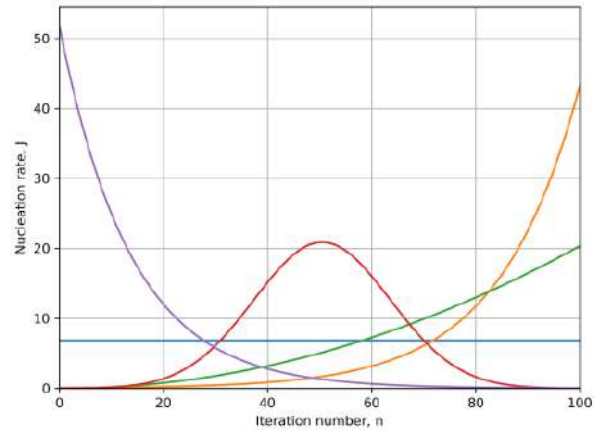
$$N(n) = \exp\left(\frac{\ln(N_{gr})}{N_p} n\right), \quad J(n) = \frac{\ln(N_{gr})}{N_p} \exp\left(\frac{\ln(N_{gr})}{N_p} n\right), \quad (5)$$

$$N(n) = \frac{1}{2} \left[1 + \operatorname{erf}\left(\frac{n - N_p/2}{\sqrt{N_g}}\right) \right], \quad J(n) = \frac{1}{\sqrt{N_{gr}\pi}} \exp\left(\frac{N_p/2 - n}{\sqrt{N_{gr}}}\right), \quad (6)$$

$$N(n) = N_{gr} \left[1 - \exp\left(-\frac{N_p}{2N_{gr}} n\right) \right], \quad J(n) = \frac{N_p}{2} \exp\left(-\frac{N_p}{2N_{gr}} n\right). \quad (7)$$



a



b

Fig. 2. Number of grains (a) and nucleation rate (b)

5 Simulation results

The following parameters for the test calculations have been chosen. The image size is 513×565 pixels and contains from $N_{gr} = 693$ grains, $N_p = 100$. These parameters are used for modelling by the method of cellular automata. Figure 4 shows the results of microstructure generation with different nucleation rates using the probabilistic neighbour from fig. 1c. To compare results with real microstructures two images have been selected. The first one –is pure iron microstructure, the second one is H62 copper alloy [39] (Fig. 3). For these images, geometric parameters have also been calculated. For assessing the uniformity of nuclei centre occurrence side histogram has been plotted. It allows estimating the probability density function of nucleation centres.

All simulations have been performed using the original program. It is written using python language with numpy library. This program allows modelling cellular microstructures according to specified parameters.

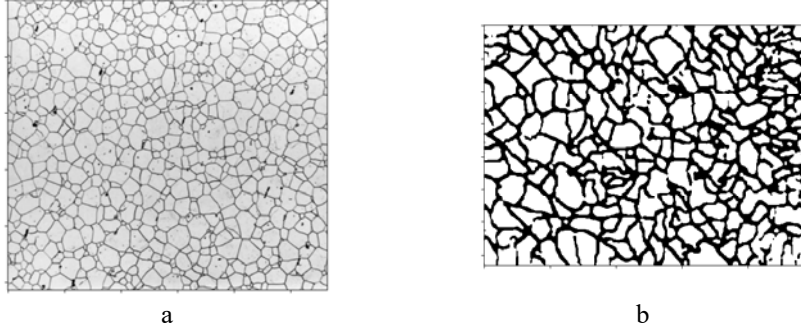


Fig. 3. Microstructures: a – pure iron; b – copper alloy H62 copper alloy [39]

For the quantitative analysis of the similarity between microstructures, the following features have been chosen. The form factor C_s . This coefficient is defined as the normalized ratio of the grain area A_{gr} to the square of the grain perimeter P_{gr} . If the grain has a perfect circle shape, then $C_s = 1$. For all other cases $0 < C_s < 1$. For example, for the square $C_s = \pi / 4 = 0.7853$.

$$C_s = 4\pi \frac{A_{gr}}{P_{gr}^2} . \quad (8)$$

Also the normalized grain area A_n is used as the ratio of the A_{gr} grain area to the total area of the image A .

$$A_n = \frac{A_{gr}}{A} \quad (9)$$

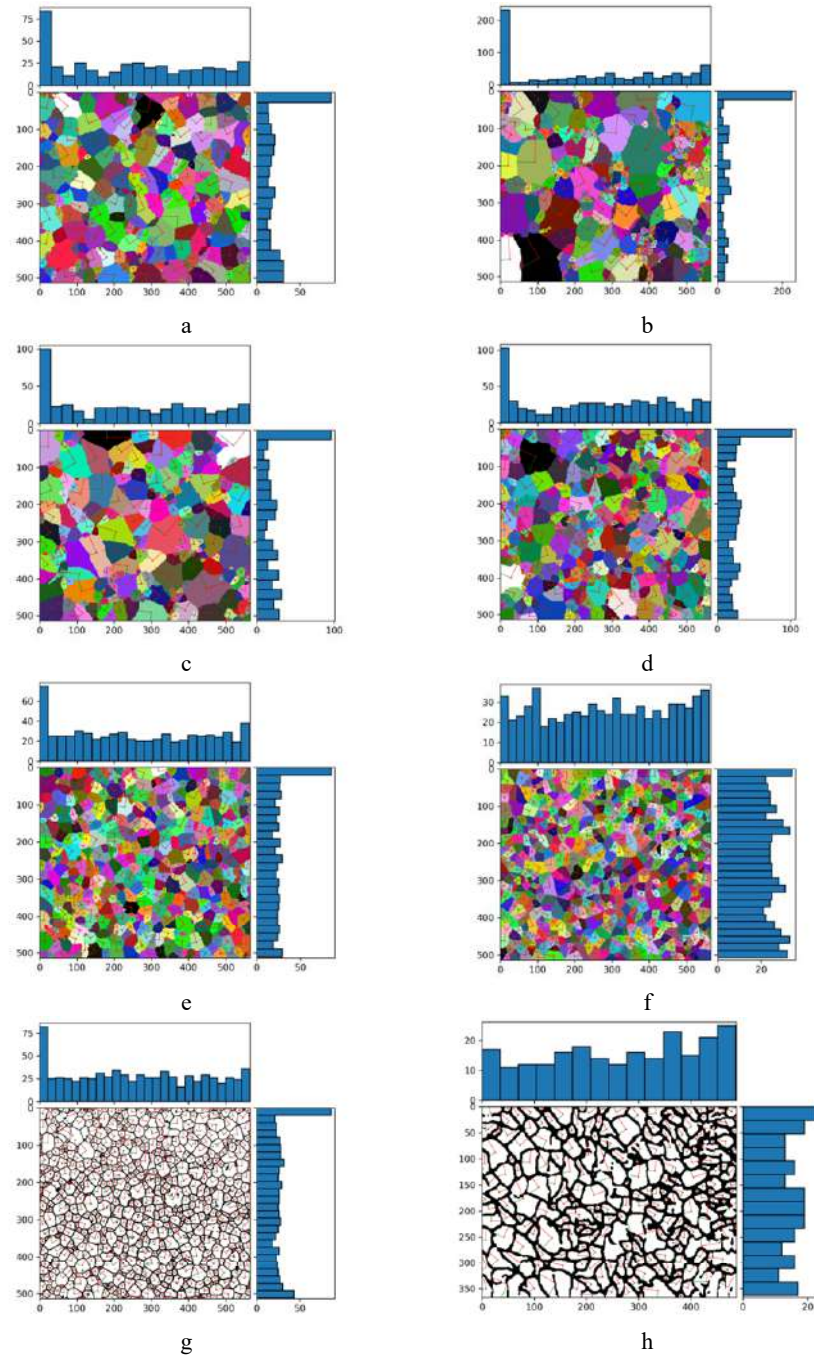


Fig. 4. Microstructures generated by cellular automata method (a-f), pure iron (g), H62 copper alloy (h)

Often, a scale factor S_c is used to compare grain parameters. It is introduced as a ratio of a larger characteristic (S_x) grain size to lesser (S_y) (fig. 5).

$$S_c = \frac{S_x}{S_y} \quad (10)$$

To determine the angle of grain orientation, the angle ψ is introduced, as is the angle between the horizontal and the largest characteristic diameter of the grain (fig. 5).

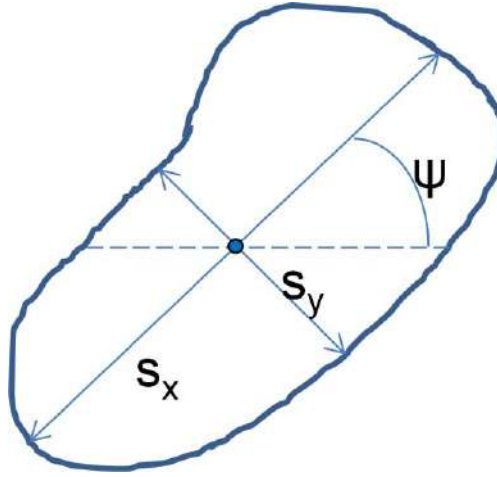


Fig. 5. Sketch of one cell parameters

For the comparison of microstructures for each grain of the generated microstructure, the values of four coefficients have been calculated. These values have been statistically processed and histograms have been constructed. These histograms have been approximated by the kernel density estimation method [40]. The approximated probability densities for the coefficients are shown in Figure 5. Descriptive statistic parameters are collected in table 1.

Table 1. The mean and standard deviation of microstructure parameters

Nucleation rate model	N_{gr}	$A_m \times 10^{-3}$		$C_{ss} \times 10^{-1}$		S_c		ψ , rad	
		mean	std. dev.	mean	std. dev.	mean	std. dev.	mean	std. dev.
CDF (6)	580	1.72	2.52	5.87	0.87	1.65	0.51	1.61	0.87
Constant	679	1.47	0.77	6.05	0.80	1.71	0.52	1.55	0.86
Cubic (3)	364	2.74	4.46	6.04	0.78	1.60	0.44	1.50	0.81
Exponential (5)	513	1.94	4.64	5.93	0.79	1.60	0.48	1.56	0.85
H62 Copper Al.	223	2.82	3.10	5.70	1.48	1.94	0.72	1.73	0.90
Linear (4)	346	2.89	3.40	6.02	0.90	1.64	0.56	1.58	0.85
Pure Iron	693	1.10	1.06	5.84	1.31	1.73	0.94	1.57	0.85
Reverse Exp (7)	590	1.69	1.38	6.08	0.79	1.63	0.47	1.64	0.83

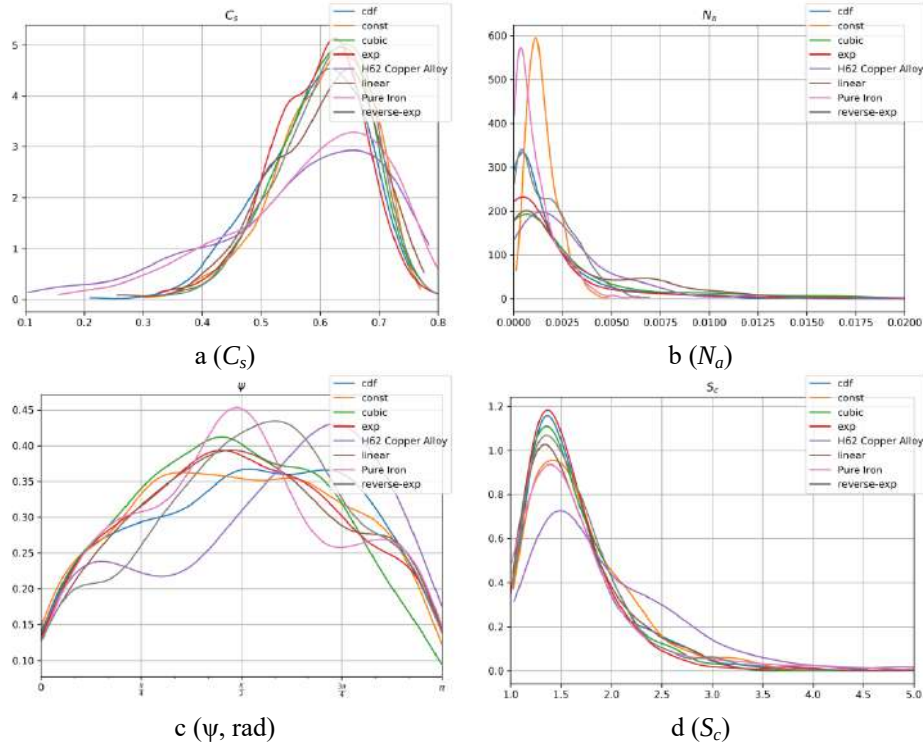


Fig. 6. The probability density function of microstructure parameters: a - C_s ; b - N_a , c - ψ , d - S_c

As can be seen from the figures, the distribution of parameters is very similar. However, there are some differences. The probability density distribution of the parameter C_s for all microstructures which is created using the cellular automata method has a similar structure. Microstructures of pure iron and copper H62 alloy stand out from the total number. They have a heavier left edge of the distribution. This suggests that more elongated grains are more common in these microstructures than is obtained as a result of modelling.

The distribution of the parameter A_n also has significant similarities. However, for the case of pure iron, the best match is shown by the constant model, and in the case of copper alloy – exponential or cubic.

The probabilistic neighbourhood was initially tuned to obtain the expected orientation $\psi = \pi / 2$. This is confirmed by the obtained probability densities. The copper alloy has another preferred grain orientation $\psi = 3\pi / 4$, which is confirmed by the obtained values. According to S_c scales, there is the best coincidence of results.

Acknowledgement

This work has been supported by the Ministry of Education and Science of Ukraine in the framework of the realization of the research project "Development of methods

for mathematical modelling of the behavior of new and composite materials aims to structural elements lifetime estimation and prediction of engineering designs reliability" (State Reg. Num. 0117U004969).

Conclusions

In this work, the implementation of the method of cellular automata probabilistic neighbours is developed. The developed method shows the results of the generation of the microstructure of the material.

A comparison of the obtained microstructure of pure iron and H62 copper alloy with synthetic microstructures has been made on the coefficient of the form, the normalized grain area, the scale factor and the angle of the grain orientation.

The results of the comparison showed that the probability densities of the corresponding parameters have a qualitative similarity. However, by the parameter of the normalized grain area, the distribution for pure microstructure has differences.

However, identifying the best way to reproduce microstructures is problematic. For some situations, better results can be achieved by one criterion, but this results in worse performance by another.

References

1. Wolfram, S.: Theory and applications of cellular automata : including selected papers, 1983-1986. World Scientific (1986)
2. Schwartz, J.T., von Neumann, J., Burks, A.W.: Theory of Self-Reproducing Automata. Math. Comput. (1967). <https://doi.org/10.2307/2005041>
3. Wolfram, S.: Cellular Automata and Complexity. CRC Press (2018)
4. Nandi, S., Kar, B.K., Pal Chaudhuri, P.: Theory and applications of cellular automata in cryptography. IEEE Trans. Comput. 43, 1346–1357 (1994). <https://doi.org/10.1109/12.338094>
5. Seredynski, F., Bouvry, P., Zomaya, A.Y.: Cellular automata computations and secret key cryptography. Parallel Comput. 30, 753–766 (2004). <https://doi.org/10.1016/J.PARCO.2003.12.014>
6. Hernández, G., Herrmann, H.J.: Cellular Automata for Elementary Image Enhancement. Graph. Model. Image Process. (1996). <https://doi.org/10.1006/gmip.1996.0006>
7. Chen, R.-J., Lai, J.-L.: Image security system using recursive cellular automata substitution. Pattern Recognit. 40, 1621–1631 (2007). <https://doi.org/10.1016/J.PATCOG.2006.11.011>
8. Hyon, S.H., Torres, D., Groppa, R., Pekolj, J., Giudice, C., Litwak, L., Argibay, P.: Transplante simultáneo de páncreas y riñón. Medicina (B. Aires). 59, 93 (1999)
9. Rosin, P.L.: Training cellular automata for image processing. IEEE Trans. Image Process. 15, 2076–2087 (2006). <https://doi.org/10.1109/TIP.2006.877040>
10. Ermentrout, G.B., Edelstein-Keshet, L.: Cellular automata approaches to biological modeling. J. Theor. Biol. (1993). <https://doi.org/10.1006/jtbi.1993.1007>
11. Xiao, X., Shao, S., Ding, Y., Huang, Z., Chen, X., Chou, K.-C.: Using cellular automata to generate image representation for biological sequences. Amino Acids. 28, 29–

- 35 (2005). <https://doi.org/10.1007/s00726-004-0154-9>
12. Alber, M.S., Kiskowski, M.A., Glazier, J.A., Jiang, Y.: On Cellular Automaton Approaches to Modeling Biological Cells. Presented at the (2003)
13. Vichniac, G.Y.: Simulating physics with cellular automata. *Phys. D Nonlinear Phenom.* (1984). [https://doi.org/10.1016/0167-2789\(84\)90253-7](https://doi.org/10.1016/0167-2789(84)90253-7)
14. Chopard, B.: Cellular automata modeling of physical systems. In: *Computational Complexity: Theory, Techniques, and Applications* (2012)
15. Madej, L.: Digital/virtual microstructures in application to metals engineering – A review. *Arch. Civ. Mech. Eng.* 17, 839–854 (2017). <https://doi.org/10.1016/J.ACME.2017.03.002>
16. Raabe, D.: Cellular Automata in Materials Science with Particular Reference to Recrystallization Simulation. *Annu. Rev. Mater. Res.* 32, 53–76 (2002). <https://doi.org/10.1146/annurev.matsci.32.090601.152855>
17. Bakhtiari, M., Seyed Salehi, M.: Reconstruction of deformed microstructure using cellular automata method. *Comput. Mater. Sci.* 149, 1–13 (2018). <https://doi.org/10.1016/J.COMMATSCI.2018.02.053>
18. Raabe, D.: Mesoscale simulation of spherulite growth during polymer crystallization by use of a cellular automaton. *Acta Mater.* 52, 2653–2664 (2004). <https://doi.org/10.1016/J.ACTAMAT.2004.02.013>
19. Popova, E., Staraselski, Y., Brahme, A., Mishra, R.K., Inal, K.: Coupled crystal plasticity – Probabilistic cellular automata approach to model dynamic recrystallization in magnesium alloys. *Int. J. Plast.* 66, 85–102 (2015). <https://doi.org/10.1016/J.IJPLAS.2014.04.008>
20. Yazdipour, N., Davies, C.H.J., Hodgson, P.D.: Microstructural modeling of dynamic recrystallization using irregular cellular automata. *Comput. Mater. Sci.* 44, 566–576 (2008). <https://doi.org/10.1016/J.COMMATSCI.2008.04.027>
21. Qian, M., Guo, Z.: Cellular automata simulation of microstructural evolution during dynamic recrystallization of an HY-100 steel. *Mater. Sci. Eng. A.* 365, 180–185 (2004). <https://doi.org/10.1016/J.MSEA.2003.09.025>
22. Zinoviev, A., Zinovieva, O., Ploshikhin, V., Romanova, V., Balokhonov, R.: Evolution of grain structure during laser additive manufacturing. Simulation by a cellular automata method. *Mater. Des.* 106, 321–329 (2016). <https://doi.org/10.1016/J.MATDES.2016.05.125>
23. Kühbach, M., Gottstein, G., Barrales-Mora, L.A.: A statistical ensemble cellular automaton microstructure model for primary recrystallization. *Acta Mater.* 107, 366–376 (2016). <https://doi.org/10.1016/J.ACTAMAT.2016.01.068>
24. Reyes, L.A., Páramo, P., Salas Zamarripa, A., de la Garza, M., Guerrero Mata, M.P.: Grain size modeling of a Ni-base superalloy using cellular automata algorithm. *Mater. Des.* 83, 301–307 (2015). <https://doi.org/10.1016/J.MATDES.2015.06.068>
25. Akbari, M., Asadi, P., Givi, M.B., Zolghadr, P.: A cellular automaton model for microstructural simulation of friction stir welded AZ91 magnesium alloy. *Model. Simul. Mater. Sci. Eng.* 24, 035012 (2016). <https://doi.org/10.1088/0965-0393/24/3/035012>
26. Zaitsev, R. V., Kirichenko, M. V., Khrypunov, G.S., Radoguz, S.A., Khrypunov, M.G., Prokopenko, D.S., Zaitseva, L. V.: Operating temperature effect on the thin film solar cell efficiency. *J. Nano- Electron. Phys.* 11, 04029-1-04029–5 (2019). [https://doi.org/10.21272/jnep.11\(4\).04029](https://doi.org/10.21272/jnep.11(4).04029)
27. Kudii, D.A., Khrypunov, M.G., Zaitsev, R. V., Khrypunova, A.L.: Physical and technological foundations of the «Chloride» treatment of cadmium telluride layers for thin-film photoelectric converters. *J. Nano- Electron. Phys.* 10, (2018).

[https://doi.org/10.21272/jnep.10\(3\).03007](https://doi.org/10.21272/jnep.10(3).03007)

28. Zaitsev, R. V., Kirichenko, M. V., Khrypunov, G.S., Prokopenko, D.S., Zaitseva, L. V.: Development of hybrid solar generating module for high-efficiency solar energy station. In: 2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON). pp. 360–364. IEEE (2017)
29. Avdieieva, O., Lytvynenko, O., Mykhailova, I., Tarasov, O.: Method for Determination of Flow Characteristic in the Gas Turbine System. In: Lecture Notes in Mechanical Engineering. pp. 499–509. Pleiades Publishing (2020)
30. Usatyi, O., Avdieieva, O., Maksuta, D., Tuan, P.: Experience in applying DOE methods to create formal macromodels of characteristics of elements of the flowing part of steam turbines. In: AIP Conference Proceedings. p. 020025. American Institute of Physics Inc. (2018)
31. Zaharchuk, I.I.: On the complexity of one-dimensional universal cellular automata. *Discret. Anal. Oper. Res.* 9, 50–56 (2002)
32. Zaharchuk, I.I., Zaharchuk, I.I., Veselov, Y.G., Ostrovskiy, A.S.: Providing information protection for wireless sensor networks based on cellular automata. *Eng. J. Sci. Innov.* 45 (2013)
33. Naumov, L., Shalyto, A.: SoftCraft: Cellular automata. Implementation and experiments, <http://www.softcraft.ru/auto/switch/kla/>
34. Vodka, O.: Analysis of quantitative characteristics of microstructures that are generated by the probabilistic cellular automata method. In: 2019 IEEE 2nd Ukraine Conference on Electrical and Computer Engineering, UKRCON 2019 - Proceedings. pp. 990–994. Institute of Electrical and Electronics Engineers Inc. (2019)
35. Sangwal, K.: Additives and crystallization processes: from fundamentals to applications. Wiley (2007)
36. McGinty, J., Yazdanpanah, N., Price, C., ter Horst, J.H., Sefcik, J.: CHAPTER 1. Nucleation and Crystal Growth in Continuous Crystallization. In: *The Handbook of Continuous Crystallization*. pp. 1–50. Royal Society of Chemistry, Cambridge (2020)
37. Fan, H., Chen, D., Liu, T., Duan, H., Huang, Y., Long, M., He, W.: Crystallization Behaviors of Anosovite and Silicate Crystals in High CaO and MgO Titanium Slag. *Metals (Basel)*. 8, 754 (2018). <https://doi.org/10.3390/met8100754>
38. Sangwal, K.: Nucleation and crystal growth: metastability of solutions and melts.
39. Lei, L., Zhiming, S., Li, Z.: Microstructure Evolution of H62 Copper Alloy and T2 Pure Copper in Different Plastic Deformation under Tensile Loading. *Rare Met. Mater. Eng.* 46, 3589–3594 (2017). [https://doi.org/10.1016/S1875-5372\(18\)30042-0](https://doi.org/10.1016/S1875-5372(18)30042-0)
40. Silverman, B.W.: *Density Estimation for Statistics and Data Analysis*. (1986)

Assessing the Possibility of a Country's Economic Growth Using Static Bayesian Network Models

Mariia Voronenko¹[0000-0002-5392-5125], Dmytro Nikytenko²[0000-0003-4989-0879],
Jan Krejci³[0000-0003-4365-5413], Nataliia Savina²[0000-0001-8339-1219],
Volodymyr Lytvynenko¹[0000-0002-1536-5542]

¹Kherson National Technical University, Kherson, Ukraine

²National University of Water Management and Environmental Engineering, Rivne, Ukraine

³Jan Evangelista Purkyně University in Ústí nad Labem, Ústí nad Labem, Czech Republic

mary_voronenko@i.ua, d.v.nikytenko@nuwm.edu.ua,
jan.krejci@ujep.cz, n.b.savina@nuwm.edu.ua
immun56@gmail.com,

Abstract: This article is devoted to the use of Bayesian networks to analyze the possibility of economic growth in Ukraine. It was found that at the maximum level of external investment, direct investment in Ukraine will increase and this creates the conditions for increasing the country's economic growth. It has also been shown that Noisy-max nodes, compared to General nodes, provide a relatively high initial accuracy. General nodes require retesting. However, Noisy-max nodes entail an increase in time and computational cost.

Keywords: Economic growth; Innovative development; General nodes; Noisy-max nodes; Bayesian networks; Structural learning; Sensitivity analysis; Validation

1 Introduction

Economic growth can be considered a major factor in the well-being and prosperity of the country. Industrialization, technology development, and innovation activity are widening the gap between developed countries and developing countries. The innovative development of enterprises is one of the basic needs of the national economy.

The activities of the enterprise reveal innovations by transforming and reforming production through the use of inventions or various opportunities for the release of new goods, the opening of new sources of raw materials, markets, modernization of production, etc., i.e. the implementation of new combinations of factors of production. Innovative activity is a factor that gives a dynamic character to the economy and has a two-sided influence: on the one hand, it opens new opportunities for economic expansion, on the other hand, it requires a change of traditional directions for further development [1].

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

Today, there are 42,564 industrial enterprises in Ukraine, representing 12.4% of their total. In the countries of the European Union (EU), the share of enterprises engaged in innovation activity is about 53%. The largest number of innovative enterprises among EU countries is in Germany (79.3% of the total number of enterprises), the smallest in Bulgaria (27.1% of the total number of enterprises) [2]. Theories and models of economic growth highlight the ways in which current economic activity can influence future economic events. Therefore, it will be advisable to determine the informative economic indicators that have the greatest impact on the dynamics of the economic growth of Ukraine.

This will create the necessary prerequisites for the growth of production and expanded reproduction of GDP in order to increase the welfare of the country's population. This paper presents the results of studies on the development of probability-determined models based on Bayesian networks to assess the degree of economic development of Ukraine. Analysis of the country's economic growth is associated with the level of external investment resource, the level of internal investment potential, financial development and the level of manufacturability (innovation) of industrial enterprises.

The aim of the work is to develop static Bayesian network models based on noisy-MAX nodes to analyze the country's economic growth trends.

2 Problem Statement

A mathematical model that, when analyzing the financial, investment, and economic indicators of an enterprise would help assess the level of economic growth of a country, requires the availability of input data. Bayesian network methods, with a certain degree of probability, make it possible to achieve the goal.

Having input indicators, such as level of manufacturability (innovativeness), level of financial security (financial development), UAH, an indicator of external investment, UAH, internal investment potential, UAH, which interact with each other as shown in Figure 1, it is necessary to design a static Bayesian network for assessing the country's economic growth opportunities.

Considering that one of the problems in the development of Bayesian networks is the exponential increase in the number of parameters in conditional probability tables (CPT), this study proposes a technique for using noisy-MAX nodes to model economic processes.

The Noisy-MAX node, which in the case of noisy variable reduces to the noisy-OR, consists of a child node, Y , taking on n_Y possible values that can be labeled from 0 to $n_Y - 1$, and N parents, $Pa(Y) = \{X_1, \dots, X_N\}$, which usually represent the causes of Y . Each X_i has a certain zero value, so that $X_i = 0$ represents the absence of X_i . Two basic axioms define the Noisy-MAX [3].

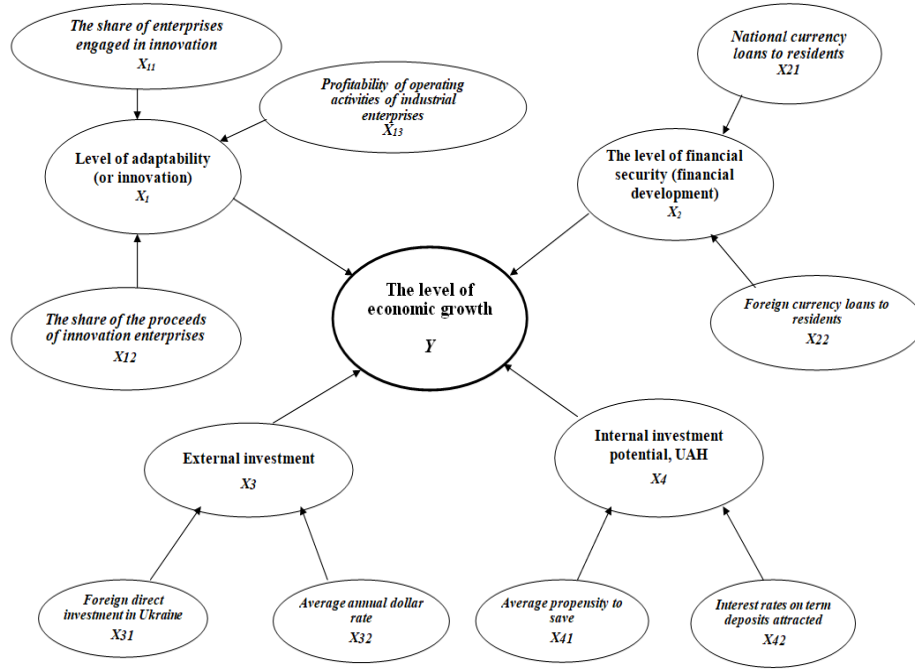


Fig. 1. Conceptual model of a static BN for assessing a country's economic growth

When all the causes are absent, the effect is absent:

$$P(Y = 0 | X_i = 0_{[\forall_i]}) = 1, \quad (1)$$

The degree reached by Y , is the maximum of the degrees produced by the X , if they were acting independently:

$$P(Y \leq y | x) = \prod P(Y \leq y | X_i = x_i, X_j = 0_{[\forall_j, j \neq i]}), \quad (2)$$

where x represents a certain configuration of the parents of Y , $x = (x_1, \dots, x_N)$. The parameters for link $X_i \rightarrow Y$ – are the probabilities that the effect assumes a certain value y , when X_i takes on the value x_i , and all the other causes of Y are absent:

$$c_y^{x_i} = P(Y = y | X_i = x_i, \dots, X_j = 0_{[\forall_j, j \neq i]}), \quad (3)$$

If X_i has n_{x_i} values, the number of parameters required for the link $X_i \rightarrow Y$ is $(n_{x_i} - 1) \times (n_y - 1)$ – because of Equation 1. Since all the variables involved in a noisy OR are binary, this model only requires one parameter per link. Alternatively, it is possible to define new parameters:

$$C_y^{x_i} = P(Y \leq y | X_i = x_i, X_j = 0_{[\forall j, j \neq i]}) = \sum_{y'}^y c_y^{x_i}, \quad (4)$$

so that Equation 2 can be rewritten as:

$$P(Y \leq y | x_1, \dots, x_n) = \prod_i C_y^{x_i}, \quad (5)$$

The CPT is obtained by taking into account that:

$$P(y|x) = \begin{cases} P(Y \leq 0|x) & \text{if } y = 0 \\ P(Y \leq y|x) - P(Y \leq y-1|x) & \text{if } y > 0 \end{cases}, \quad (6)$$

3 Review of the Literature

An analysis of the current state, tendency of innovation activity of industrial enterprises of Ukraine and generalization of theoretical approaches, directions, and measures of increasing the innovative activity of the country is covered in [1].

Authors [4,5] concluded that the actions of large financial institutions have significant implications for the stability of the entire economic system and may be a threat to economic risk.

Network models have become attractive for modeling dependencies in real-world phenomena because of their ease of representation and the ability to provide an intuitive way visualization and interpretation of complex economic relations [6].

To understand vulnerabilities in the financial system, the idea of what-if network analysis has proven to be a promising tool that can help monitor the interconnectedness of financial institutions and markets. This has led to a significant increase in research on the statistical properties of network indicators to analyze systemic risk and the possibility of economic development of the country [7].

Directional acyclic graphical models and graphical Gaussian models were dealt with by the authors [8-10] and achieved results in their research.

However, when analyzing and modeling economic processes, risks, and to predict future results, Bayesian network models have proven themselves best [11, 12]. They are currently widely used for working with discrete data [13].

4 Materials and Methods

4.1 Data

As experimental data for assessing the economic growth of Ukraine, that take into account the statistical capabilities and the modification of existing methods for studying the financial activity of an enterprise (Table 1) macroeconomic indicators were used.

Table 1. Matrix of economic indicators

Indicators	Appointment
X1	Level of adaptability (or innovation)
X11	The share of enterprises engaged in innovation
X12	The share of the proceeds of innovation enterprises
X13	Profitability of operating activities of industrial enterprises, %
X2	The level of financial security (financial development), UAH
X21	National currency loans for a term of 5 years to residents (excluding deposit-taking corporations), average value, UAH million
X22	Foreign currency loans to residents (excluding deposit-taking corporations) for a term of 5 years, average value, UAH million
X3	External investment potential, UAH
X31	Foreign direct investment in Ukraine
X32	Interest rates on term deposits attracted, %
X4	Internal investment potential, UAH
X41	Average propensity to save
X42	Average annual dollar rate, UAH
Y	The level of economic growth (nominal GDP at actual prices)

In our study, we are dealing with a set of statistical data that are interconnected (Fig. 1) consisting of 14 indicators for the period 2005-2018. The matrix of indicators is divided into four blocks that most fully characterize the financial-economic and business activity of enterprises, as well as the course of economic processes in the country (Table 1). The resulting indicator Y is an integral indicator of the level of economic growth of Ukraine.

4.2 Materials and Methods

A **Bayesian network** (BN) is a pair $\langle G, B \rangle$, in which the first component G is a directed acyclic graph corresponding to random variables [14, 15]. Each variable is independent of its parents in G. So, the graph is written as a set of independence conditions. The set of parameters defining the network is the second component B. It

contains parameters $Q_{x^i|pa(X^i)} = P(x^i | pa(X^i))$ for each possible x_i value from X_i and $pa(X^i)$ from $pa(X^i)$, where $pa(X^i)$ denotes the set of parents of the variable X_i in G . Each variable X_i is represented as a vertex. We use the notation to identify the parents $pa^G(X^i)$ if we consider more than one graph. The total joint probability of BN is calculated by the formula $P_B(X^1, \dots, X^N) = \prod_{i=1}^N P_B(X^i | pa(X^i))$.

BN is a probabilistic model for representing probabilistic dependencies, as well as the absence of these dependencies. At the same time, the $A \rightarrow B$ relationship is causal, when event A causes B to occur, that is, when there is a mechanism whereby the value accepted by A affects the value adopted by B .

Validation was proposed for the first time in 1977 in [16]. Validation of the network that we design was carried out according to the algorithm for maximizing expectations. The algorithm finds local optimal estimates of the maximum likelihood of arguments. The concept of the algorithm is that if we knew the values of all nodes, then training would be simple at some step M . Therefore, at stage E , estimations of the expected likelihood value are made, including latent variables, as if we were able to observe them. In step M , the maximum likelihood values of the parameters are estimated using the maximization of the expected likelihood values obtained in step E . Then, the algorithm performs step E using the parameters obtained in step M again and so on.

The goal of parametric learning is to find the most likely θ variables that explain the data. Let $D = \{D_1, D_2, \dots, D_N\}$ be a composition of the learning data, where $D_i = \{x_1[l], x_2[l], \dots, x_n[l]\}$ consists of instances of Bayesian network nodes. So the learning parameter is quantified by a log-likelihood function, denoted as $L_D(\theta)$ [17].

5 Experiments

When developing the BN, the GeNIe 2.4 Academic software environment was used. The original Bayesian network model is built on General nodes. The block diagram of the BN of the country's economic growth is presented in Fig. 2.

When developing the BN, the GeNIe 2.4 Academic software environment was used. As can be seen from Fig. 1, the network contains 5 key nodes:

- X1 - the level of manufacturability (innovation),
- X2 - the level of financial security (financial development), UAH
- X3 - external investment, UAH
- X4 - domestic investment potential, UAH,
- Y – the level of economic growth.

It should be noted that due to the specifics of the Bayesian networks, all the conclusions of this model regarding the information sought are probabilistic in nature and are presented in the form of a ranked list (according to the values of the probability of fidelity of a particular conclusion).

Data were taken from 2005 to 2018. The dynamics of changes in the initial indicators for the observed period are presented in Figure 3. All nodes have five states: s1, s2, s3, s4, s5.

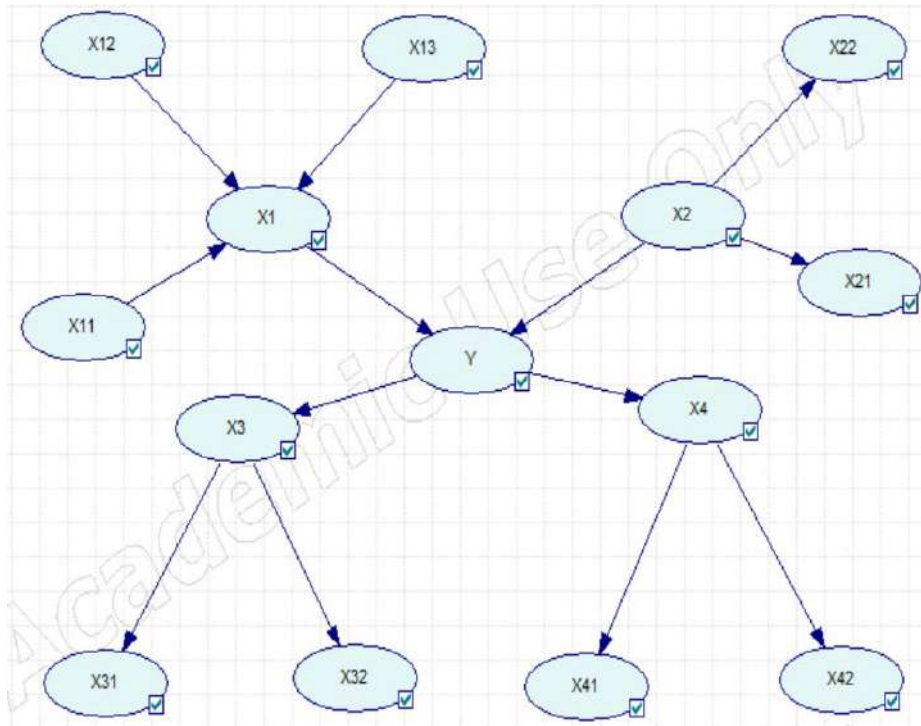


Fig. 2. Structural model of a static BN of the country's economic growth

For example, for the node X1, the intervals of state discretization will be as follows:

```
s1_below_85879;
s2_85879_114478;
s3_114478_150998;
s4_150998_218982;
s5_218982_up
```

We carry out, parameterization, and network validation on the nodes of General. The initial overall accuracy of the network was 48.8%, the accuracy of the result was 42%. At the next stage of the study, we changed the type of all nodes to Noisy-max with five states s1-s5, and the resulting node Y.

The network remains the same, the data file also does not change. We carry out training in parameters, primary validation. The overall accuracy of the network increased by 6.6% (and amounted to 55.4%), the accuracy of the result increased by 4% (and amounted to 46%).

	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
Y	441452	544153	720731	948056	913345	1079346	1299991	1404669	1465198	1586915	1988544	2383182	2982920	2502130
X1	0.253	0.355	0.44	0.39	0.26	0.308	0.352	0.307	0.295	0.21	0.154	0.188	0.218	0.399
X11	0.082	0.1	0.115	0.108	0.107	0.115	0.128	0.136	0.136	0.121	0.152	0.166	0.143	0.166
X12	0.065	0.067	0.067	0.059	0.048	0.038	0.038	0.033	0.033	0.025	0.014	0.005	0.007	0.059
X13	0.016	0.035	0.058	0.049	0.018	0.035	0.047	0.034	0.03	0.016	0.009	0.042	0.066	0.034
X2	0.335	0.365	0.18	0.28	0.414	0.338	0.273	0.222	0.197	0.215	0.196	0.163	0.143	0.168
X21	41802	51979	13590	34691	41311	47002	53498	62168	68997	72530	72763	90300	116728	146944
X22	46798	67256	64037	124453	185337	171670	159721	125079	104503	132053	161182	142275	139388	252627
X3	0.995	1	1	0.998	0.76	0.978	0.94	0.978	0.592	0.175	0.513	0.669	0.529	0.487
X31	7808	5604	9891	10913	4816	6495	7207	8401	4499	410	2961	3284	2202	1526
X32	0.147	0.197	0.1	0.116	0.159	0.158	0.147	0.184	0.18	0.197	0.215	0.182	0.149	0.18
X4	0.512	0.445	0.567	0.587	0.586	0.578	0.616	0.543	0.516	0.305	0.279	0.33	0.335	0.305
X41	0.2028	0.2003	0.243	0.1951	0.1184	0.1479	0.1209	0.1177	0.1039	-0.0749	-0.0801	-0.0194	0.0362	0.0362
X42	5.12	5.05	5.05	5.27	7.79	7.94	7.97	7.99	7.99	11.89	21.84	25.55	26.6	27.2

Fig. 3. The dynamics of changes in the initial indicators for the observed period

Next, we analyze the sensitivity [18] of the network using influence charts. Repeated training in parameters and repeated validation led to an increase in overall accuracy by 2% (from 55.4% to 57.4%), and the accuracy of the result increased by 14% (from 46% to 60%). The results are shown in Table 2:

Table 2. Comparison of the initial accuracy of the model with accuracy after sensitivity analysis on nodes Noisy - MAX

	Initial accuracy		Accuracy after a sensitivity analysis	
	Overall network accuracy,%	Accuracy of the result ,%	Overall network accuracy,%	Accuracy of the result ,%
Noisy-MAX nodes	55,4	46,0	57,4	60,0

6 Results and Discussion

If the level of innovation is increased to the maximum, the share of enterprises' bargaining will increase by 7% (from 34% to 41%), and this, in turn, will lead to an increase in the country's economic growth by 12% (from 24% to 36%), as shown in Figure 4.

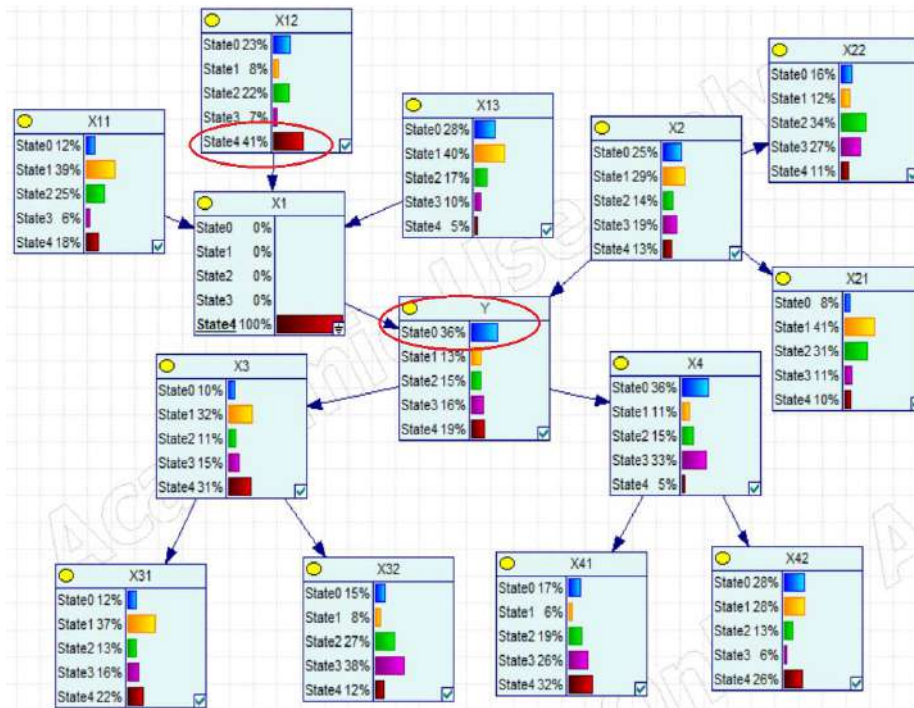
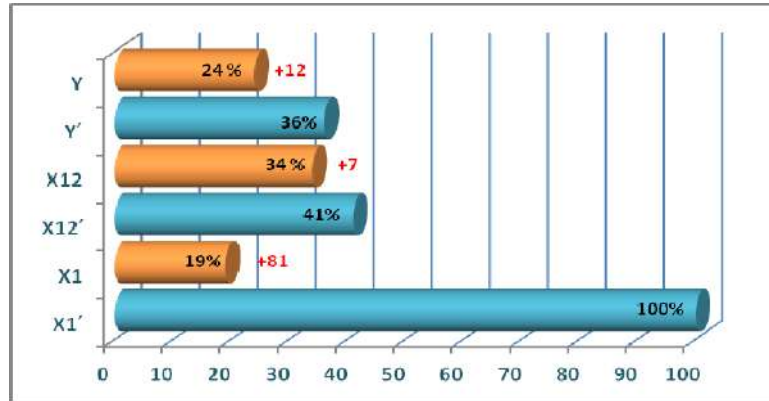


Fig. 4. Experiment Results

At the maximum level of external investment, the level of innovation will increase by 4% (from 14% to 23%), direct investment in Ukraine will increase by 28% (from 19% to 47%), the tendency of the population to save will increase by 14% (from 28% to 42%). All this together will create conditions for increasing the country's economic growth, which will increase by 42% (from 24% to 66%), as shown in Figure 5.

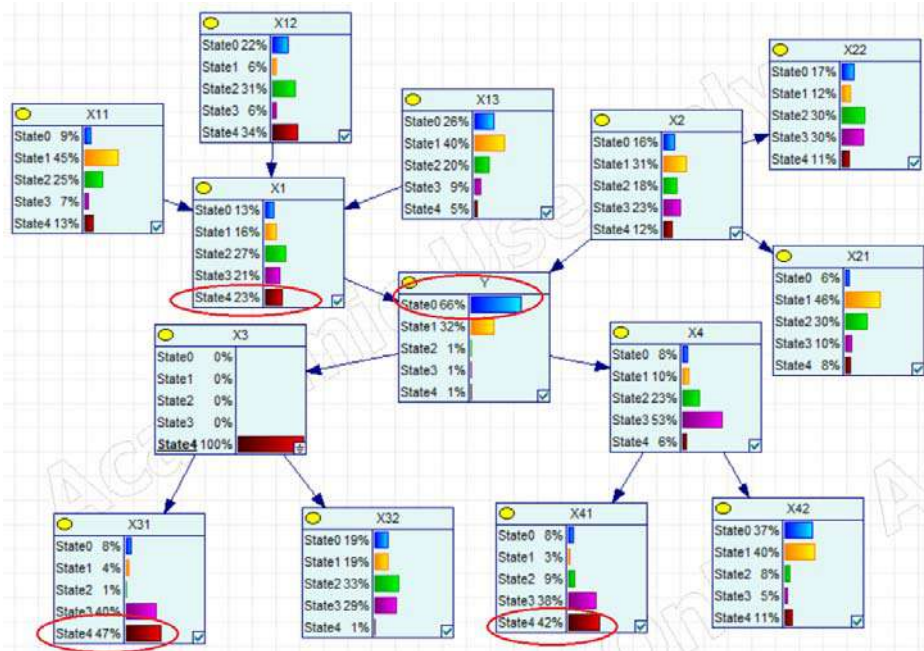
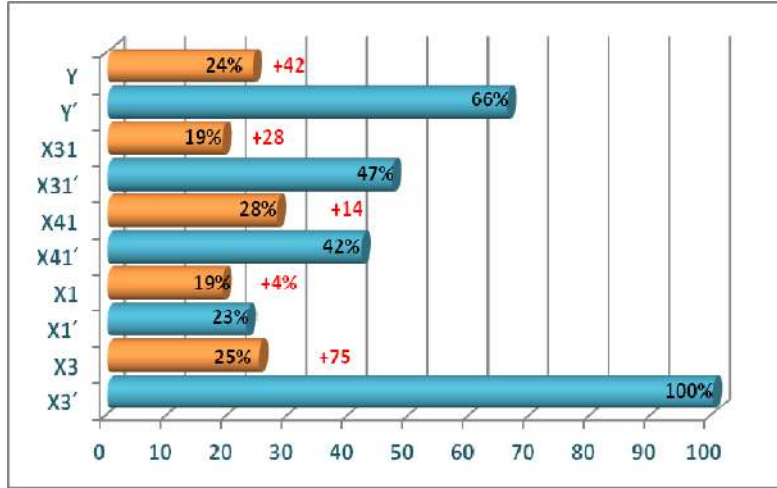


Fig. 5. Conditions for ensuring economic growth

7 Conclusion

This article has conducted a comparative study of the behavior of Noise Max nodes and common nodes when designing a Bayesian network. Noisy-max nodes have been shown to provide relatively high initial accuracy compared to conventional nodes.

Shared nodes require retesting. However, Noisy-max nodes entail an increase in time and computational cost.

Along with this, experiments were conducted to assess the general trends of the country's economic growth potential. It was found that at the maximum level of external investment, direct investment in Ukraine will increase by 28% (from 19% to 47%) and this creates the conditions for increasing the country's economic growth, which will increase by 42% (from 24% to 66%).

In our future research, we will try to trace the country's economic growth over time using the dynamic Bayesian network tool.

References

1. Shumpeter, Y.: *Teoriia ekonomichnoho rozvytku*. Prohress, 455p. (in Ukrainian) (1982).
2. *Naukova ta innovatsiina diialnist v Ukraini u 2012–2017 rokakh*. Statystychnyi zbirnyk. KDP, 289p. (in Ukrainian) (2017).
3. Díez, F., Galán, S.: Efficient computation for the Noisy-MAX, *Int. J. Int. Syst.*, vol. 18/2, 165–177. (2004)
4. Billio, M., Getmansky, M., Lo, A., Pelizzon, L.: Econometric Measures of Connectedness and Systemic Risk in the Finance and Insurance Sectors. *Journal of Financial Economics*, vol. 104 (3), 535 – 559. (2012)
5. Battiston, S., Delli-Gatti, D., Gallegati, M., Greenwald, B., Stiglitz, J.: Liaisons Dangereuses: Increasing Connectivity, Risk Sharing, and Systemic Risk. *Journal of Economic Dynamics and Control*, vol. 36 (8), 1121–1141. (2012)
6. Brunnermeier, M., Pedersen, H.: Market Liquidity and Funding Liquidity/ Review of Financial Studie, vol. 22 (6), 2201–2238. (2009)
7. Granger, C. W.: Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometric*, vol. 37 (3), 424–438. (1969)
8. Geiger, D., Heckerman, D.: Parameter Priors for Directed Acyclic Graphical Models and the Characterization of Several Probability Distributions. *Annals of Statistics*, vol. 30 (5), 1412– 1440. (2002)
9. Giudici, P., Green, P.: Decomposable Graphical Gaussian Model Determination. *Biometrika*, vol. 86 (4), 785–801. (1999)
10. Miyamura, M., Kano, Y.: Robust Gaussian Graphical Modelin. *Journal of Multi-variate Analysis*, vol. 97 (7), 1525–1550. (2006)
11. Heckerman, D., Chickering, D.: Learning Bayesian Networks: The Combination of Knowledge and Statistical Data. *Machine Learning*, 20–197. (1995)
12. Heckerman, D., Geiger, D.: Learning Gaussian Networks. *Uncertainty in Artificial Intelligence*, 235–243. (1994)
13. Madigan, D., York, J.: Bayesian Graphical Models for Discrete Data. *International Statistical Review*, vol. 63 (2), 215–232. (1995)
14. Cheeseman, P.S., Kelly, M., Taylor, W., Freeman, D., Stutz, J.: Bayesian classification, In: *Proceedings of AAAI*, St. Paul, MN, 607-611. (1988)
15. McCann, R.K., Marcot, B.G, Ellis, R.: Bayesian belief networks: applications in ecology and natural resource management. *Canadian Journal of Forest Research*, Vol. 36/12, 3053-3062. (2006)

16. Friedman, N.: The Bayesian structural EM algorithm. Fourteenth conference on Uncertainty in Artificial Intelligence (UAI'98), Madison, Wisconsin, USA, 24–26 July, SF.: Morgan Kaufmann, 129-138. (1998)
17. Pollino, C. A., Woodberry, O., Nicholson, A., Korb, K., Hart, B.: Parameterisation and Evaluation of a Bayesian Network for Use in an Ecological Risk Assessment. Environmental Modelling and Software, February 5: In proceedings, vol. 22(8), 1-13. (2007)
18. Darwiche, N.: A differential approach to inference in Bayesian networks. In Uncertainty in Artificial Intelligence: Proceedings of the Sixteenth Conference (UAI 2000), San Francisco, CA: Morgan Kaufmann Publishers, 123–132. (2000)

Computer Modeling of Multi-Mass Electromechanical Systems

Olena Nazarova [0000-0002-0784-7621]

National University “Zaporizhzhia Politecnic”, Zhukovsky str., 64, Zaporizhzhia, 69063,
Ukraine
nazarova@zntu.edu.ua

Abstract. Industrial process plants and units are mostly multi-mass electromechanical systems with a number of features such as damping friction, elastic coupling, clearance and others. The article is devoted to computer modeling of multi-mass electromechanical systems taking into account elastic bonds of the first and second kind on the example of electromechanical systems of cold rolling mills. A comparison of the electromechanical processes obtained during the simulation with the monitoring data of real equipment confirmed their adequacy (the relative modeling error does not exceed 9%), that allows to use these models at modernization of existing and designing of new automatic control systems of electric drives of multi-mass electromechanical systems.

Keywords: modeling, multi-mass system, electromechanical system, electric drive, cold rolling mill, the first and second kind elastic bonds

1 Introduction

Currently, automation of technological and production processes using mathematical and computer models is the most popular and progressive approach to the modernization of existing and the development of new automatic control systems. The electric drive as an actuator of many units and complexes used in various industries, science and technology, largely determines their operational characteristics. In transient electromechanical processes, there is a continuous conversion of energy from one type to another due to the presence of various energy storage devices (electrical, electromagnetic, thermal, mechanical), as well as various energy converters interconnected by mechanical (elastic), electrical or magnetic bonds. The mechanical part of the electric drive may have a branched multi-mass system with gaps (backlashes) and viscoelastic mechanical elements (long shaft, cable, rolled metal strip), interconnected by a constant or variable gear ratio. During operation of an industrial unit, the magnitude of the moving mass and the moment of inertia (casting machines, winding and unwinding mechanisms of the rolled strip) can change, which affects the dynamics and statics of the entire system. All this leads to the complexity of the analysis and synthesis of automatic control systems, so the development of mathematical and computer models of multi-mass electromechanical systems is a relevant task.

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

2 Formal problem statement

The purpose of the work is the development and study of computer models of multi-mass electromechanical systems taking into account the elastic bonds of the first and second kind on the example of electromechanical systems of cold rolling mills (CRM).

3 Literature review

A fairly large number of works are known related to the study of the effect of elastic transmissions on the quality of transients of one-, two- [1-3], five- [4] and multi-mass systems [5]. Most of these systems are unbranched and take into account the influence of elastic bonds described by Hooke's law. Some of the researchers consider such systems taking into account dry friction [6, 7], as well as other nonlinearities [8, 9] to obtain properties that are closer to a real object. Some researchers use the procedure for identifying multi-mass systems, which is promising for obtaining mechanical parameters, such as inertia or shaft stiffness [10-12]. Multimass systems with elastic coupling are found in many industrial applications, for example, in wind power plants [13, 14], in public utilities, for example, when considering a passenger or freight elevator as a two- or three-mass system (engine, elevator car and counterweight) [15, 16], as well as in the automotive industry, for example, when considering the drive of electric vehicles taking into account the number of drive wheels, the distribution of torque between them [17, 18]. A number of works are devoted to the study of the dynamic properties of adaptive neuroregulators, optimal, relay, modal, and robust regulators based on a two-mass system with a complex mechanical structure [19-21] and a variable moment of inertia [20, 23], which confirms the interest of scientists in such objects. Models of two-mass and multi-mass systems are used in the educational process [24] and in developing educational and research laboratory stands based on them [25-27]. However, not enough attention is paid to the consideration of multi-mass systems with elastic gears of the first and second kind.

4 Mathematical description of multi-mass electromechanical systems

When creating automation systems using computer and microprocessor technology for predicting, designing and optimizing equipment parameters and cold rolling technology, the role of mathematical modeling increases [3]. The development of methods and universal mathematical models for studying the dynamic modes of operation of electromechanical systems is a very urgent task, which is caused by the desire of researchers to most fully reflect the features of the object in the model (complicate the model), on one hand, and present it as an element of the system (to simplify the model if possible), on the other [2]. One of the most common purposes of models is their application in the study and prediction of the behavior of complex processes and phe-

nomena. Another, no less important, purpose of the models is to identify the most significant factors that form certain properties of the object, the consideration of which is necessary in the study of various processes or phenomena [4].

Since mathematical modeling is the most perfect and effective modeling method, which, when researching and optimizing, is based on modern methods of mathematical analysis, computational mathematics, and programming, it is advisable to study electromechanical complexes on mathematical models. However, the models of most EMSs are not always adequate to their originals because of the complexity of taking into account all the circumstances and features of real processes in the mathematical description.

Due to the high level of modern programming, it is possible to take into account a set of EMS identification factors when choosing a software product, such as the type of mathematical description of the EMS under study; features of data presentation; type of presentation of calculation results; the possibility of flexible changes or editing of the mathematical model, algorithms, accuracy, selection and automation of the use of numerical methods in the calculation of systems of differential equations, the ability to automate the calculation process [5]. To solve the problems of researching automated EP systems, the following software tools are used: MATLAB, MATHCAD, LABVIEW, MATHEMATICA and others.

It is convenient enough to consider EMS as a set of structural schemes, especially when it is necessary to synthesize the structure and parameters of control systems. Conventionally, it is possible to single out CRM aggregates that are universal for most mills: winding and unwinding devices, rolling stand, leveling machine, scissors cutting rolled metal. In addition, all these units are interconnected: the unwinding device feeds the rolled strip to the work rolls of the rolling stand, from the stand the strip enters either the subsequent stands (if it is the multi-train mill), or to the leveling machine, from which the strip comes out, which is fed to the winder, or scissors for transverse or longitudinal cutting for the subsequent formation of sheets of the required format. Each of these units is driven by an electric motor, which is elastically connected to the actuator through a long shaft. Therefore, in the above set of basic elements of CRM, it is necessary to add such structural units as the “long shaft” and the “rolled strip”, which are elastic bonds of the first and second genera. The selection of the most suitable method for solving practical rolling problems is often a rather difficult task, the solution of which when creating a model of a specific production equipment, namely, CRM, will require significant time costs.

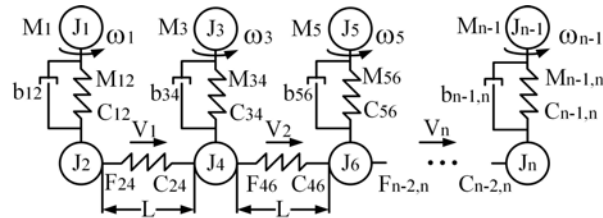


Fig. 1. Kinematic diagram of a multi-mass interconnected system taking into account elastic bonds of the first and second kind

Therefore, the idea of creating a library of blocks for constructing the necessary CRM is relevant.

As a multi-mass interconnected system, a mechanical system with rotating masses connected in series by elastic elements taking into account viscous friction is considered. In addition, in the general case, it is possible to take into account dry friction, changes in the moments of inertia of rotating masses and gear ratio in gears.

Figure 1 shows the system in question, where $J_1, J_2, J_3 \dots J_n$ - rotating mass of system, $\text{N}\cdot\text{m}\cdot\text{sec}$, $\omega_1, \omega_2, \omega_3 \dots \omega_n$ - angular velocities of specified masses, sec^{-1} , $M_1, M_2, M_3 \dots M_n$ - moving mass moments, $\text{N}\cdot\text{m}$, $M_{12}, M_{34}, M_{56} \dots M_{n-1,n}$ - moments created in the elastic coupling of the first kind of "long shaft", $\text{N}\cdot\text{m}$, $C_{12}, C_{34}, C_{56} \dots C_{n-1,n}$ - elasticity of the long shaft connection, $\text{N}\cdot\text{m}/\text{rad}$, $b_{12}, b_{34}, b_{56} \dots b_{n-1,n}$ - intrinsic viscosity coefficient in gears between J_{n-1} and J_n , $\text{N}\cdot\text{m}\cdot\text{sec}$, $F_{24}, F_{46} \dots F_{n-2,n}$ - tensile forces occurring in a second kind of elastic coupling (e.g. a metal strip being rolled), N , L - the distance between adjacent masses connected by a second kind of elastic coupling, m .

As is known, the total moment acting on n rotating mass is equal to the algebraic sum of moments which are attached to this mass. In general, these are: motor torque, load resistance torque, deformation torque of the elastic element (elastic torque), dry and viscous (internal and external) friction moments, as well as variable elastic torque due to non-uniform machine running [28].

Mass with inertia moment J_n is an electric motor that drives a rotary actuator J_{n-1} . They are connected by the elastic coupling of the first kind of "long shaft", the deformation of which is linear and subject to Hook's law.

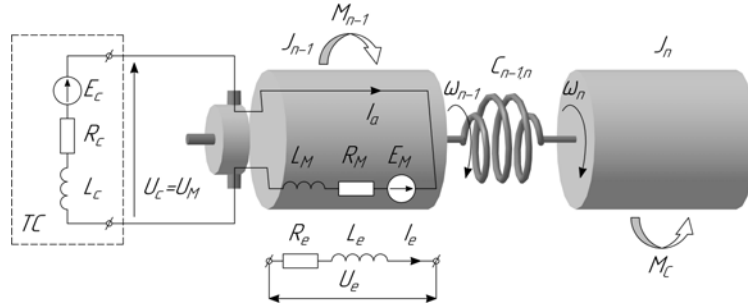


Fig. 2. Two-mass DC motor electric drive model with independent excitation

At the Figure 2 TC - thyristor converter equivalent circuit; E_c , R_c , L_c - EMF, internal impedance and inductance of the thyristor converter that feeds the motor anchor; U_c , U_M , U_e - voltage of converter, motor and excitation winding; I_a , I_e - anchor circuit current and excitation winding; R_e , L_e - internal impedance and inductance of the excitation winding; M_c - resistance torque.

$$\begin{cases} E_c = E_M + I_a R_\Sigma + L_\Sigma \frac{dI_a}{dt}, \\ R_\Sigma = R_c + R_M, \\ L_\Sigma = L_c + L_M, \\ E_M = C_c \cdot \omega_{n-l}, \\ M_{n-l} = C_m \cdot I_a, \end{cases} \quad (1)$$

where $C_c = \frac{U_{nom} - I_{nom} \cdot R_M}{\omega_{nom}}$ - constructive engine coefficient, V·sec; $C_m = \frac{M_{nom}}{I_{nom}}$ - mechanical engine steel, N·m/A, $U_{nom}, I_{nom}, \omega_{nom}, M_{nom}$ - nominal voltage, current, angular speed and torque, which are calculated according to the passport data of the engine.

The elastic torque is proportional to the angle of rotation of the shaft or the difference in rotation angles of the individual masses. $M_{n-l,n}$, viscous friction torque M_{vf} in the elastic element is proportional to the difference in velocity and very small, the torque of external viscous friction. M_{evf} (damping torque) is proportional to speed and affects the dynamic torque, but its value is negligibly small and often not taken into account in the calculations. The elastic coupling of the first kind is described by the system (2), where p - Laplace operator, β_n - mass damping coefficient J_n N·m·sec.

$$\begin{cases} M_{n-l,n}(p) = C_{n-l,n}(\varphi_{n-l}(p) - \varphi_n(p)) = \frac{C_{n-l,n}}{p}(\omega_{n-l}(p) - \omega_n(p)), \\ M_{vf}(p) = b_{n-l,n}(\omega_{n-l}(p) - \omega_n(p)), \\ M_{evf}(p) = \beta_n \omega_n(p), \\ M_{n-l,n}(p) - M_c = J_n p \omega_n(p), \\ M_{n-l}(p) - M_{n-l,n}(p) = J_{n-l} p \omega_{n-l}(p). \end{cases} \quad (2)$$

Positioned actuators one after the other J_n, J_{n-2} (for cold rolling mills, it is a unwinding mechanism, stand or stands, a winding mechanism) are connected by elastic bonds of a second kind, such as the transported material or the rolled metal strip, which are described by the system (3).

$$\begin{cases} F_{n-2,n}(p) = \frac{C_{n-2,n}}{p} \{V_{n-l}(p) - [l + \varepsilon_{n-2,n}(p)]V_n(p)\}; \\ \varepsilon_{n-2,n}(p) = \frac{l}{C_{n-2,n}L} F_{n-2,n}(p); \end{cases} \quad (3)$$

where $\varepsilon_{n-2,n}$ - percentage elongation.

The mathematical description of the state of the metal between the cage and the winding-unwinding mechanism was supplemented by the ability of reversing

$$F'_{n-2,n}(p) = \begin{cases} \frac{C_{n-2,n}}{P} \{V_{n-1}(p) - [I + \varepsilon_{n-2,n}(p)]V_n(p)\}, & \text{if } V_{n-1} - V_n > 0 - \text{winding}; \\ \frac{C_{n-2,n}}{P} \{V_n(p) - [I + \varepsilon_{n-2,n}(p)]V_{n-1}(p)\}, & \text{if } V_{n-1} - V_n < 0 - \text{unwinding}; \\ 0, & \text{if } V_{n-1} - V_n = 0 - \text{lack of tension during a technological pause.} \end{cases} \quad (4)$$

During the rolling process, the strip along the entire length has a different thickness, which, however, falls within the tolerance limits for deviation. Changing the value of h affects the tension of the strip F_c , as well as its elastic properties. The thickness variation should also be taken into account when calculating the current value of the radius of the roll, mass, and the moment of inertia of the roll.

Let us supplement the well-known mathematical description of elasticities of the second kind [28] by taking into account the variable thickness of the strip, as well as the following condition under which the effect of breaking the strip is simulated, that is, when the current value of the tension force in the strip $F_{n-2,n}$ achieves the critical $F'_{(n-2,n)max}$, the metal strip breaks, and in the mathematical description the output value of the tension force $F'_{n-2,n}$ is nullified:

$$F'_{n-2,n} = \begin{cases} F_{n-2,n}, & \text{if } F_{n-2,n} < F'_{(n-2,n)max}, \\ 0, & \text{if } F_{n-2,n} \geq F'_{(n-2,n)max}, \end{cases} \quad (5)$$

$$F'_{(n-2,n)max} = K_r \cdot F_{(n-2,n)rab}, \quad (6)$$

where K_r – coefficient of tensile strength of the material (in the model it is accepted as $K_r=1,5$); $F_{(n-2,n)rab}$ – metal tension force during operating modes of rolling.

Quantities K_r , $F_{(n-2,n)rab}$ are determined taking into account the physical and mechanical properties of the material and the geometric dimensions of the rolled strip.

When describing the elastic coupling of the second kind, the following assumption is accepted [28]: the material is homogeneous and has the same thickness and length along the whole length; the influence of the mass of the material on its deformation can be neglected; the deformation has a purely elastic character and is uniformly distributed along the cross-section; wave processes related to the distribution of deformation along the length can be neglected.

$$\begin{cases} J'_{n-1}(p) = m_{n-1}(p)(R_{n-1}(p) - R_0)^2 / i^2; \\ m_{n-1}(p) = \rho \cdot \pi(R_{n-1}^2(p) - R_0^2); \\ R_{n-1}(p) = \sqrt{R_0^2 + \frac{h}{\pi} V_{n-1}(p)}; \\ V_{n-1} = R_{n-1} \omega_{n-1}(p); \\ V_n = R_n \omega_n(p). \end{cases} \quad (7)$$

The mathematical description of the winding (unwinding) device takes into account J'_{n-l} - variable component of the inertia moment of the winding (unwinding) device of CRM, m_{n-l} - metal coil mass; i - gear ratio; ρ - specific gravity of rolling material; R_0 - winder drum radius, h - the thickness of the metal strip rolled, as shown in the system (7).

5 Computer models of multi-mass electromechanical systems

When modeling using the Simulink libraries of the MATLAB, the principle of visual programming according to structural schemes is implemented, according to which the user draws a model of the object on the screen from the library of standard blocks and performs calculations in automatic mode, with the ability to control the calculation time and establish initial conditions. Also, user has the opportunity to choose a method for solving differential equations (Runge-Kutt, Rosenbrock, Dormand-Prince, Adams, etc.), choose the type of solution (with variable or constant step), upgrade library blocks, create their own, as well as creating new block libraries, which is especially important when conducting research in a fairly narrow industry, for example, the production of cold rolled sheets.

Based on the mathematical description of the multi-mass electromechanical systems, computer models have been created that simulate the operation of winding-unwinding mechanisms, a working stand, as well as a model of a rolled strip of metal, taking into account changes in strip thickness and an emergency caused by strip breakage. These models were controlled using models of DC motors of independent excitation and automatic control systems. Each model is an independent subsystem with ports of input and output coordinates. For convenience, a separate logo was created, which is depicted on the model above, an interface for introducing and changing the parameters of each of the subsystems. This approach has advantages, since any unit has the ability to connect to another through mechanical, electrical and information communication channels. It is possible to observe any electromechanical coordinates and set various rolling modes depending on the process. One of the main units of the cold rolling mill is a winding (unwinding) mechanism *Motalka*. Based on system (7), a of the winding mechanism of the mill, shown in Figure 3, and the similar models structure shown in Figure 4, created in Matlab / Simulink.

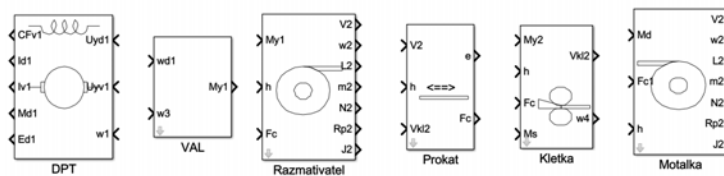


Fig. 3. Computer models of multi-mass electromechanical systems

The difference between the computer model of the unwinder from the one of the winder is that in the model of unwinder at the initial instant of time, the moment of inertia, diameter, and mass of the roll are maximum, and in the model of the winder are minimum, they take into account the direction of movement of the strip and the tasks of initial conditions.

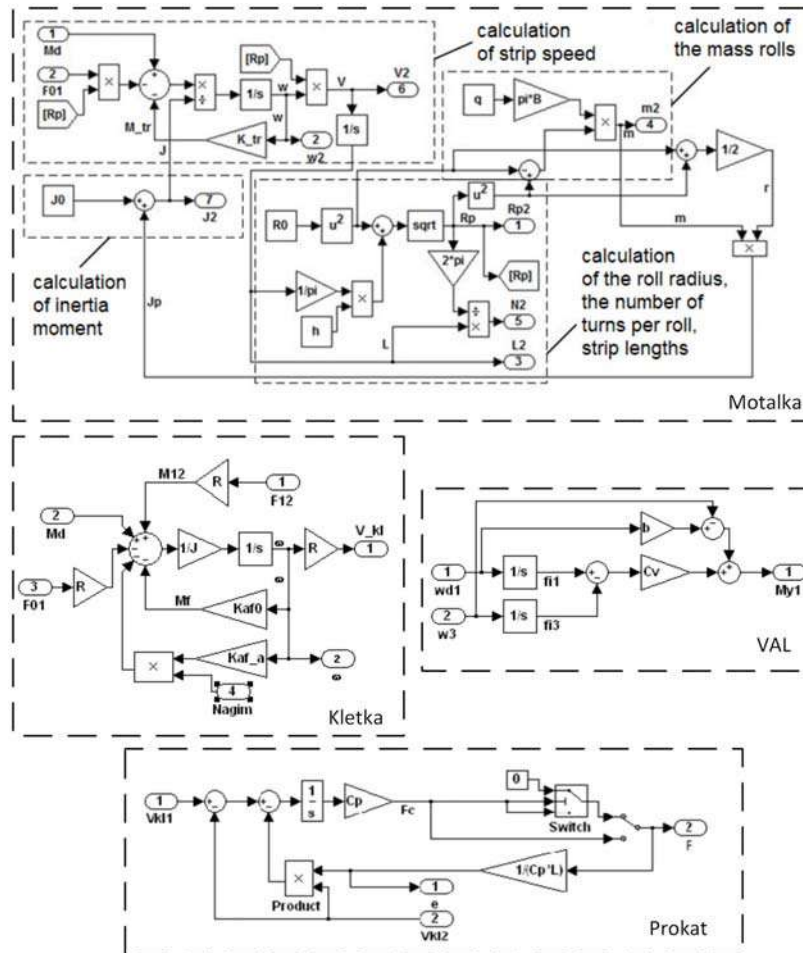


Fig. 4. - Model structures of the winding mechanism (*Motalka*), working stand (*Kletka*), "long shaft" (*VAL*) and rolled metal (*Prokat*)

The model of the working stand (*Kletka*) takes into account not only the rotating masses and the balance of moments, but also provides the possibility of loading the stand from the pressure channel (pressure screws) through the *Nagim* input port. The coefficients of friction and rolling designated in the model as K_{af0} , K_{af_a} , the pressure force F_N from the side of the pressure screws is indicated by *Nagim*, since there is no possibility to use indexes in the application package in which the simulation was

performed. Therefore, the linear speed of the strip at the exit of the stand is indicated by V_{kl} , the forces forming the front and rear tension in the stand are indicated by F_{12} and F_{01} , respectively.

According to (1), a model *VAL* was created that simulates the elastic “long shaft” connection between the engine and the drive (Figure 4). In the *VAL* model, the internal damping coefficient is denoted by b , the angle of rotation of the motor shaft is $fi1$, the angle of rotation of the shaft of the drive mechanism (of winder, unwinder, stand and others) is $fi3$, ω_{d1}, ω_3 – are the angular speeds of rotation of the shaft ends. This model does not take into account gaps in the mechanical part and, if necessary, can be supplemented by dead zones or non-linear blocks. But this will complicate the whole model, which is not always advisable.

Also, Figure 4 shows a metal model connecting two neighboring units. Note that by its structure it is universal, suitable for all sections of the stand. This model is a second kind elastic link, since the force (moment) between adjacent drives is transmitted through the transported material. To illustrate the consideration of material's properties, namely the residual elongation when the strip is broken, the dependence $F = f(\varepsilon)$ is simulated, which displays the tensile strength of the material (Figure 5). Comparing the obtained dependence with the stretching diagram of the rolled strip, it can be seen that it corresponds to theoretical [29]

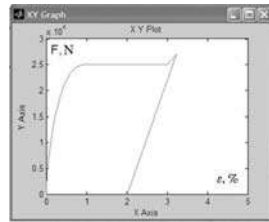


Fig. 5. Dependence $F=f(\varepsilon)$ representing the tension of the material

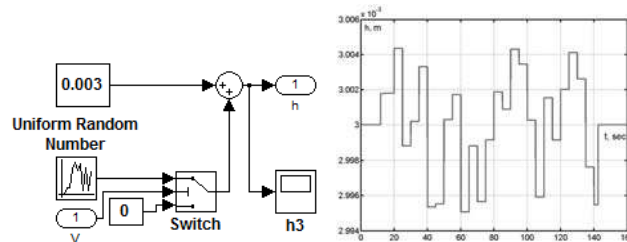


Fig. 6. Model of the strip thickness former and signal shaper thickness of a random nature

To simulate random influences from the rolled metal's side, *Motalka*, *Prokat*, *Kletka* blocks can be supplemented with input ports h , which allows us to study the operation of these mechanisms with a random change in the thickness of the rolled material. The inconsistency of the value of h leads to a change in the stiffness of the strip C_p and affects its elastic properties. Also, the influence of the thickness of the strip is taken into account in the model of the winder (when calculating the current value of

the radius, mass, moment of inertia and mass of the roll) and the stand, where the resistance moment changes with h due to an increase in the rolling force.

For example, Figure 6 shows a model of a shaper of the thickness of a strip of random nature and the generated signal. To simulate changes in the strip thickness, a source of a random signal with a uniform distribution (uniform random number) was used. The signal level is limited by the values of Maximum and Minimum above and below, the frequency of signal changes is set by the Sample time parameter, the range of variation of the value of h is 0,4%.

6 Experiments and results

To check the adequacy of the developed computer models, the multi-mass electro-mechanical system of the single-stand cold rolling mill of cold rolling shop No. 1 of Zaporizhstal JSC was simulated (Figure 7) and the simulation results under normal rolling conditions and in case of a strip break were compared with the results of monitoring of this mill [30] in an industrial environment (Figure 8, Figure 9). The data for modeling coincide with the data of real equipment with existing automatic control systems (SAU), rolling of one roll is considered.

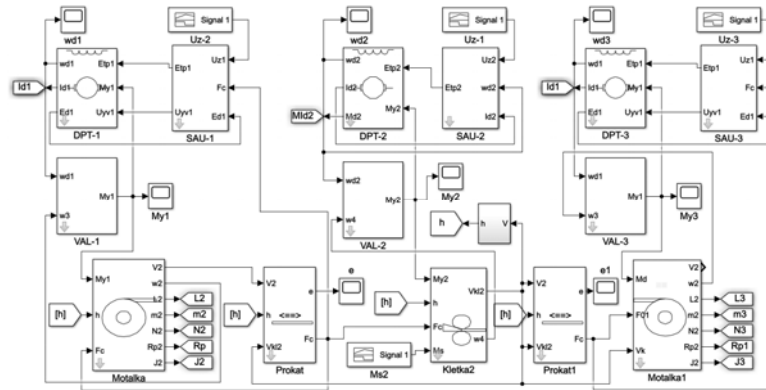


Fig. 7. Multi-mass electromechanical system of the single-stand cold rolling mill model

Comparison of the model obtained and experimental values shows that the relative error does not exceed 8.12%. The average statistical relative error of the entire model was calculated $\bar{a}_m = 3,79\%$.

Using the developed computer models, the work of a four-stand CRM was simulated (Figure 10). The ZADANIE subsystem is created on the basis of the “speed wedge” principle, that is, the angular speed of the engines of each next stand has a steeper section of acceleration to the operating speed compared to the previous one and, accordingly, a higher working speed. This is due to the increase in the length of the strip after processing it with each stand. For each subsequent stand, the corresponding value of the pressing force is set, that is, the force with which the pressing

screws act on the rolled strip is regulated. In the SAU units, a linear change in the voltage across the windings of the armature of the motors was set, and the voltage in the field winding was varied based on maintaining a constant linear velocity and tension of the rolled metal during winding and unwinding (taking into account the variation of the current radial size of the roll).

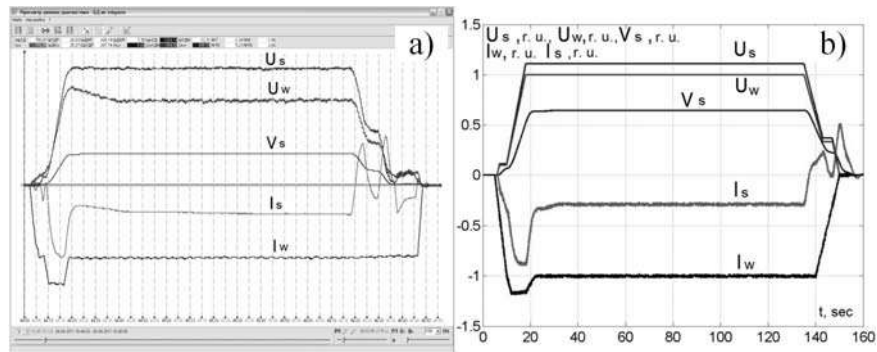


Fig. 8. Average CRM electromechanical processes (a), simulation results (b) under normal rolling conditions

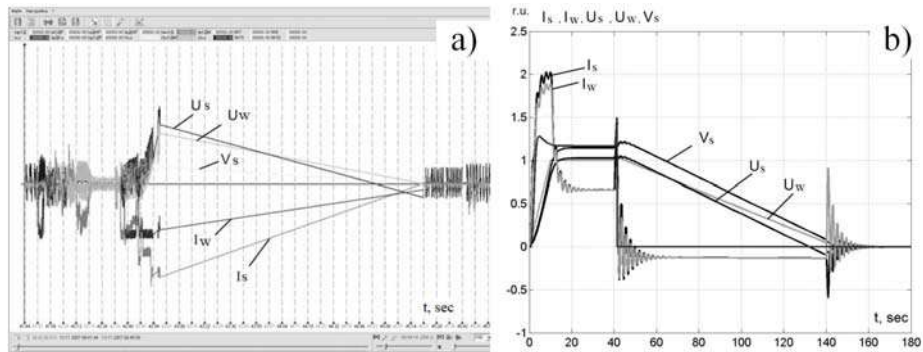


Fig. 9. Average CRM electromechanical processes (a), simulation results (b) in case of a strip break

At the Figure 11: currents of armature windings of engines (a); angular velocities of the motors of the stands (b); linear speeds of the strip movement (c); changing the radius of the roll on the winder (d); a change in the mass and moment of inertia of each of the stands (e), (f). The numbers 1, 2, 3, 4 indicate the working stands of the mill: the first, second, third and fourth, respectively. The speed task for these stands is set based on the principle of “speed wedge”, that is, the speed of metal passage through the stands increases from the first stand to the fourth, as it is necessary to maintain a constant volume of the metal. Each subsequent stand increases the length of the rolled strip and reduces the thickness, so the linear speed of the strip increases. When winding a coil, the mass, moment of inertia and radius of the roll increase, as shown in Figure 11 (d), (e), (f).

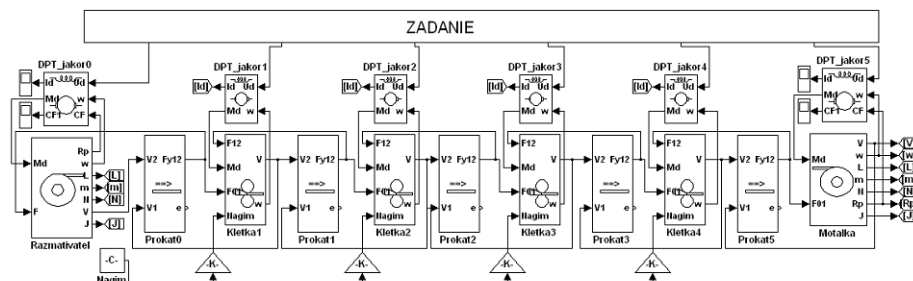


Fig. 10. Multi-mass electromechanical system of the four-stand cold rolling mill model

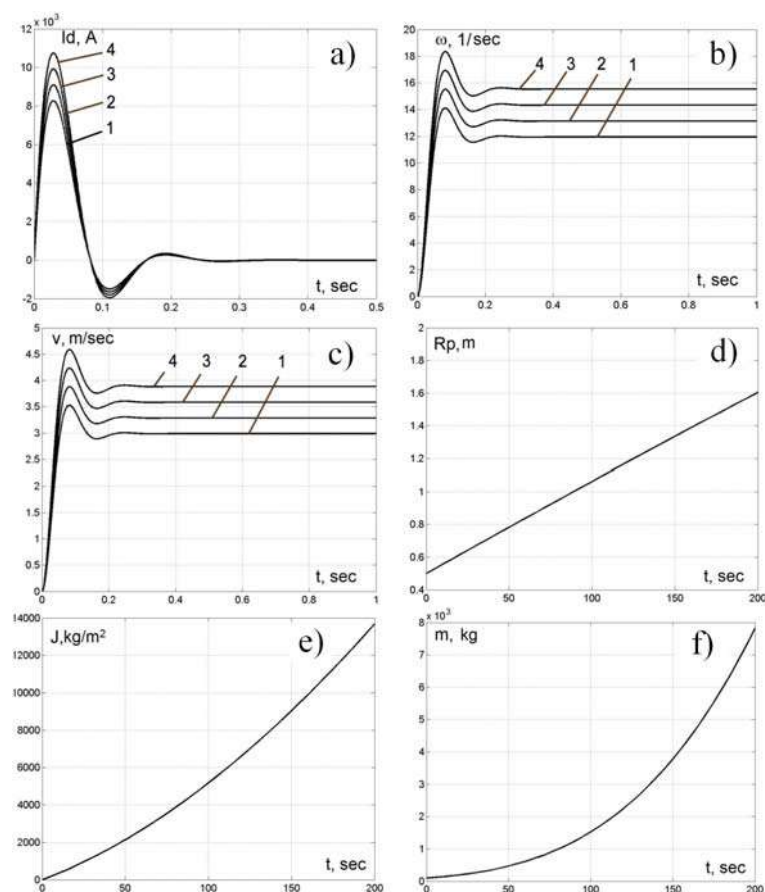


Fig. 11. Graphs of transients on four-stands cold rolling mill

Comparison of the transient graphs obtained during the simulation with the data of the operating equipment of Zaporizhsal JSC confirmed the adequacy of the developed model of the four-stand mill. Thus, we can conclude that the developed mathematical

and computer models of the main drives, automatic control systems, actuators and transients are adequate for the real equipment and transient schedules of the mills cold rolling shop No. 1 of Zaporizhstal JSC.

7 Conclusion

Complex of interconnected computer models of the basic elements of rolling production has been developed, which allows you to take into account the elastic bonds of the first and second kinds between the electric drives of the rolling stand and the winding-unwinding mechanism.

Method for studying multi-mass electromechanical systems on example of the main mechanisms of cold rolling mills with variable moments of inertia, static resistance, and metal strip thickness by using the developed complex of computer models of the basic elements of rolling production is proposed.

The mathematical description of the rolling processes of the strip has been improved by taking into account its variable thickness, the effect of breakage and the dependence of the tension force on the relative elongation, which is the basis for the development of mathematical models that allow you to simulate the work and study the quality of control processes for the interconnected electric drives of the main mechanisms of cold rolling mills. A parameter is introduced that takes into account the variable thickness of the metal strip and allows to increase the adequacy and accuracy of mathematical models of the winding-unwinding mechanism and rolling stand.

A comparison of the electromechanical processes obtained during the simulation with the monitoring data of real equipment confirmed their adequacy (the relative modeling error does not exceed 9%) and the possibility of using the developed set of computer models for the synthesis and study of cold rolling mills control systems.

Thus, the consistent use of well-known packages in solving problems of a particular industry allows us to obtain the desired result in the study and modernization of existing control systems, as well as in the development of new control systems for multi-mass electromechanical systems.

References

1. Seppo, E. Saarakkala and Hinkkanen, Marko: Identification of Two-Mass Mechanical Systems Using Torque Excitation: Design and Experimental Evaluation. *IEEE Transactions On Industry Applications*, vol. 51, no. 5, pp. 4180-4189 (2015) doi: 10.1109/TIA.2015.2416128
2. Kaniuk, G., Vasylets, T., Varfolomiyev, O., Mezeria, A., Antonenko, N.: Development of neural-network and fuzzy models of multimass electromechanical systems. *Eastern-European Journal of Enterprise Technologies*, no. 3/2 (99), pp. 51-63 (2019) doi: 10.15587/1729-4061.2019.169080
3. Szabat, K. and Orłowska-Kowalska, T.: Vibration Suppression in a Two-Mass Drive System Using PI Speed Controller and Additional Feedbacks - Comparative Study. *IEEE Transactions on Industrial Electronics*, vol. 54, no. 2, pp. 1193-1206, (2007) doi: 10.1109/TIE.2007.892608

4. Aboah, Boateng Emmanuel, Normanyo, Erwin: Investigating Efficiency of a Five-Mass Electromechanical System having Damping Friction, Elastic Coupling and Clearance. In: International Journal of Engineering Research & Technology (IJERT), vol. 6, pp. 94-98 (2017)
5. Łuczak, Dominik: Mathematical model of multi-mass electric drive system with flexible connection. In: 19th International Conference on Methods and Models in Automation and Robotics, (MMAR), (2014) doi: 10.1109/MMAR.2014.6957420
6. Lukichev, D. V., Demidova, G. L. and Brock, S.: Application of fuzzy cooperative coevolution adaptive control for damping of torsional vibrations in multi-mass drive system. In: 19th European Conference on Power Electronics and Applications (EPE'17 ECCE Europe), Warsaw, 2017, pp. P.1-P.10. (2017) doi: 10.23919/EPE17ECCEEurope.2017.8099208
7. Li, K., Wang, J., Wang J., Wei, D., Chen, A. and Zhang, C.: Control of multi-mass flexible transmission system based on the notch filter. In: Proceedings of the 32nd Chinese Control Conference, Xi'an, pp. 4282-4285 (2013)
8. Hackl, C. M. and Schroder, D.: Funnel-Control for Nonlinear Multi-Mass Flexible Systems. In: IECON 32nd Annual Conference on IEEE Industrial Electronics, Paris, pp. 4707-4712. (2006) doi: 10.1109/IECON.2006.347369
9. Voliansky, R., Sadovoi, O. and Volianska, N.: Defining of Lyapunov Functions for the Generalized Nonlinear Object. In: 2018 IEEE 5th International Conference on Methods and Systems of Navigation and Motion Control (MSNMC), Kiev, pp. 222-228 (2018) doi: 10.1109/MSNMC.2018.8576315
10. Łuczak, D. and Zawirski, K.: Parametric identification of multi-mass mechanical systems in electrical drives using nonlinear least squares method. IECON 2015 - 41st Annual Conference of the IEEE Industrial Electronics Society, Yokohama, pp. 004046-004051 (2015) doi: 10.1109/IECON.2015.7392730
11. Łuczak, D. and Nowopolski, K.: Identification of multi-mass mechanical systems in electrical drives. In: Proceedings of the 16th International Conference on Mechatronics - Mechatronika 2014, Brno, pp. 275-282 (2014) doi: 10.1109/MECHATRONIKA.2014.7018271
12. Zoubek, H. and Pacas, M.: An identification method for multi-mass-systems in speed-sensorless operation. In: 2011 IEEE International Symposium on Industrial Electronics, Gdansk, pp. 1895-1900 (2011) doi: 10.1109/ISIE.2011.5984447
13. Muyeen, S. M. et al.: Transient Stability Analysis of Wind Generator System with the Consideration of Multi-Mass Shaft Model. International Conference on Power Electronics and Drives Systems, Kuala Lumpur, pp. 511-516 (2005) doi: 10.1109/PEDS.2005.1619740
14. Joost, C. Mehler and Orlik B.: Dynamic load change stress minimizing control of elastically coupled multi-mass systems. In: 12th International Conference on Optimization of Electrical and Electronic Equipment, Basov, pp. 754-759 (2010) doi: 10.1109/OPTIM.2010.5510374
15. Nazarova, O., Osadchyy, V., Shulzhenko, S.: Accuracy improving of the two-speed elevator positioning by the identification of loading degree. In: IEEE International Conference on Modern Electrical and Energy Systems (MEES), Kremenchuk, Ukraine, pp. 50 - 53 (2019) doi: 10.1109/MEES.2019.8896414
16. Osadchyy, V.V., Nazarova, O.S., Shulzhenko, S.S.: Identification of the level of completion of a two-part elevator. Electrotechnical and computer systems, No. 27 (103), pp 103-111 (2018) doi: <https://doi.org/10.15276/eltecs.27.103.2018.11>

17. Osadchy, V.V., Nazarova, E.S., Brylystiy, V.V.: The structure of the control system of a 4-drive power plant for electric vehicles. Problems of Regional Energy (special issue), No. 1-2 (41), pp. 65-73 (2019) doi: 10.5281/zenodo.3239150
18. Voliansky, R. Sadovoi, O. Sokhina, Y. and Volianska, N.: Active Suspension Control System. In: IEEE International Conference on Modern Electrical and Energy Systems (MEES), Kremenchuk, Ukraine, pp. 10-13 (2019) doi: 10.1109/MEES.2019.8896419
19. Ha, Q. P. and Negnevitsky, M.: A robust modal controller with fuzzy tuning for multi-mass electromechanical systems. In: Proceedings of Third Australian and New Zealand Conference on Intelligent Information Systems. ANZIIS-95, Perth, WA, Australia, pp. 214-219 (1995) doi: 10.1109/ANZIIS.1995.705743
20. Luczak, D. and Pajchrowski, T.: Application of Adaptive Neural Controller and Filter Tune for Multi-Mass Drive System. In: 20th European Conference on Power Electronics and Applications (EPE'18 ECCE Europe), Riga, pp. P.1-P.9 (2018)
21. Lukichev, D. V., Demidova, G. L., Kuzin A. Y. and Saushev, A. V.: Application of adaptive Neuro Fuzzy Inference System (ANFIS) controller in servodrive with multi-mass object. In: 25th International Workshop on Electric Drives: Optimization in Control of Electric Drives (IWED), Moscow, pp. 1-6 (2018) doi: 10.1109/IWED.2018.8321388
22. Burian, S.O., Kiselychynk, O.I., Pushkar, M.V., Reshetnik V.S., Zemlianukhina, H.Y.: Energy-Efficient Control of Pump Units Based On Neural-Network Parameter Observer. Technical Electrodynamics, no.1, pp. 71-77(2020)
23. Nazarova, O.S., Yefimenko, R.A.: Features of modeling electromechanical systems with a variable moment of inertia. Visnik NTU «KhPI», no. 27(1249), pp. 71-74 (2017)
24. Zbaravska, L., Chaikovska, O., Hutsol T., Slobodian, S., Dumanskyi, O.: Professional competence as a key factor in improving the quality of engineering education. Environment. Technology. Resources. Rezekne, Latvia, Proceedings of the 12th International Scientific and Practical Conference, Vol. III, pp. 253-256. (2019) doi: <http://dx.doi.org/10.17770/etr2019vol3.4118>.
25. Osadchiy, V.V., Nazarova, E.S., Brylisty, V.V., Savilov, R.I.: Laboratory stand for research of microprocessor control systems for a two-mass electric drive. Electrotechnical and computer systems, No. 22 (98), pp. 33-38 (2016) doi: <https://doi.org/10.15276/eltecs.22.98.2016.05>
26. Osadchiy, V.V., Nazarova, E.S., Tobolkin S.Yu.: Laboratory bench for researching algorithms of microprocessor control systems for stepper motor. Elektromechanichni i energosberigauchi systemy, vol. 2/(26), pp.102-108 (2014)
27. Chornyi, O.P., Serhiienko, S.A.: A virtual complex with the parametric adjustment to electromechanical system parameters. Tekhnichna elektrodynamika, No 1, pp. 38-41 (2019) doi: <https://doi.org/10.15407/techned2019.01.038>
28. Nazarova, E.S.: Mathematical modeling of electromechanical systems of cold rolling mills. Tekhnichna elektrodynamika, No. 5, pp. 82-89 (2015)
29. Nazarova, E.S., Pirozhok, A.V., Nechpay, A.S., Podpruzhnikov, P.A.: Taking into account the effect of breaking the rolled strip when modeling electric drives of two adjacent stands of a cold rolling mill. Electrical engineering and electricity, No. 2, pp. 37-41 (2011) doi: <https://doi.org/10.15588/1607-6761-2011-2-6>
30. Nazarova, E.S.: On the issue of developing diagnostic systems for electromechanical systems of cold rolling mills. Electrical engineering and power engineering, No. 1, pp. 36-41 (2013) doi: <https://doi.org/10.15588/1607-6761-2013-1-6>

Decision Support System for Microclimate Control at Large Industrial Enterprises

Nataliya Pankratova¹[0000-0002-6372-5813], Petro Bidyuk²[0000-0002-7421-3565],

Igor Golinko³[0000-0002-7640-4760]

^{1, 2, 3} Igor Sikorsky Kyiv Polytechnic Institute, Peremogy str. 37, 03056 Kyiv, Ukraine

¹natalidmp@gmail.com,

²pbidyuk_00@ukr.net,

³conis@ukr.net

Abstract. The structure and mathematical support for the climate control system for large enterprises is proposed. The developed air conditioning control algorithms and multifunctional software are used at the middle level of an enterprise management using decision-making system. Criteria are defined and algorithms for optimization and adaptation of control systems for industrial air conditioners are proposed. The use of the proposed decision-making system can improve efficiency of the functioning of the microclimate systems of industrial premises when changing their operating conditions. The developed models, methods and control algorithms are recommended for use at the stage of design, commissioning and operation of industrial air conditioners at the middle level of operational production management.

Keywords: integrated enterprise management system, manufacturing execution system, automatic control system, industrial air conditioning, decision-making system, mathematical model.

1 Introduction

Modern production processes place high demands on industrial air conditioning systems. The quality of the product and the reduction in its cost [1] depend to a large extent on the microclimate of industrial premises. The maximum effect of the industrial air conditioners automation is achieved by integrating the air conditioning control system into the enterprise management system.

The world practice of introducing integrated enterprise management systems shows a significant increase in the efficiency of their work by reducing: energy costs, production downtime, optimal distribution of material and energy flows, the use of hidden reserves. Modern automation systems are integrated and include control subsystems interconnected by functions and levels. The architecture of the modern integrated enterprise management system is shown in Fig. 1. The functions of the first three levels are implemented by the automatic control system (ACS) of technological processes, which ensures optimal flow of the technological process in the workshop

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

or on a separate production area. At the fourth level of enterprise management, a manufacturing execution system (MES) is used, which plays the role of an information bridge between the supervisory control and data acquisition (SCADA) and the enterprise resource planning (ERP) system in a single information space of the enterprise.

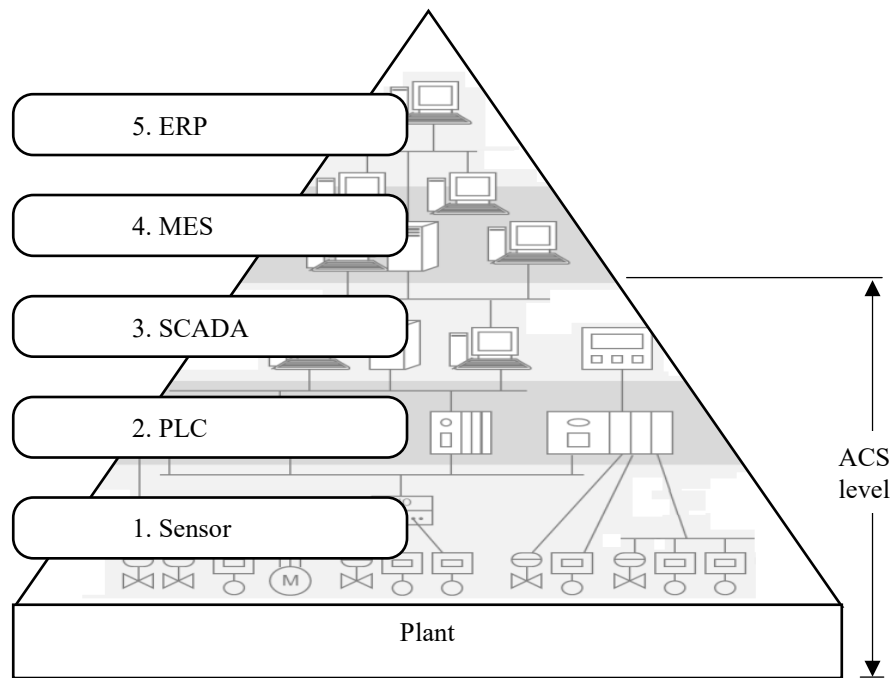


Fig. 1. Integrated enterprise management system hierarchical architecture

At most enterprises, SCADA and ERP are implemented and operated. The next step in improving the quality of products while reducing their cost is the introduction of the MES for operational management of production. MES were first proposed by the Manufacturing Enterprise Solutions Association (MESA) International in 1994. MESA International has established a basic set of functions for MES, which has subsequently been repeatedly adjusted. These results are published in a large number of articles and manuals [2–4]. The most advanced ideas of integrated production were approved by International Society of Automation (ISA) in the international standards ISA – 88, ISA – 95, ISA – 106 [5–7]. The listed standards are the most significant for the development of integrated enterprise management systems.

The developed international standards describe the basic rules for the operation of MES, but do not disclose the mathematical support for the operation of specific systems, in particular for the climate control systems of industrial premises, which are discussed here. In publication [8], a new concept for the automation of industrial air conditioners was proposed. In [9], the principles of automation of industrial air conditioners are considered, which provide an increase in the efficiency of the operation of

HVAC equipment through the use of control algorithms for air conditioners and multifunctional mathematical models of climatic equipment [8–12]. A literature review of industrial air conditioning automation systems confirms that all the prerequisites for integrating industrial air conditioning control systems into the enterprise management system have been created.

2 Problem Statement

At all levels of integrated enterprise management, operator jobs are used in which dispatching decision support system (DSS) are used [13]. The purpose of this research is to develop the structure, software and functioning algorithms of the DSS to support the microclimate in large industrial enterprises. The developed system will improve the overall production efficiency when changing the operating conditions of industrial rooms.

3 Decision Support System for Control the Microclimate

According to Fig. 1, DSS for the industrial air conditioners operational management is realized at the MES level of enterprise management. The block diagram of the developed DSS is shown in Fig. 2.

The proposed DSS is designed and implemented using the following system analysis principles [14]: the hierarchy principle; optimal solutions principles; mathematical models' adaptability; identification and minimization of the uncertainties influence that appear in the control system; the functional orthogonality principle and others. So, the integrated enterprise management system architecture is built according to the hierarchical principle, which assumes the management of several levels plant, united by information links. The operational control system assumes the optimization control actions, which ensures the air conditioning system optimization necessary to determine the acceptable temperature and humidity limits in an industrial room and of the recirculation coefficient, which affects the economical use of energy resources by the microclimate system. DSS provides the functions implementation providing for the structural and parametric adaptation of mathematical models, thereby achieving their high degree of adequacy to the object. Stochastic uncertainties in the form of random external disturbances and measurement errors are minimized by using optimal filtering methods for measurements at the stage of preparing data for modeling. The existing parametric uncertainty of the models is minimized by using alternative methods for estimating the mathematical models' parameters of the studied objects. The functional orthogonality principle ensures the rational implementation of all the control system basic functions and the elimination of their duplication. Thus, the use of the system analysis principles in the design of DSS makes it possible to improve quality of the intermediate computation results, and the management quality in general.

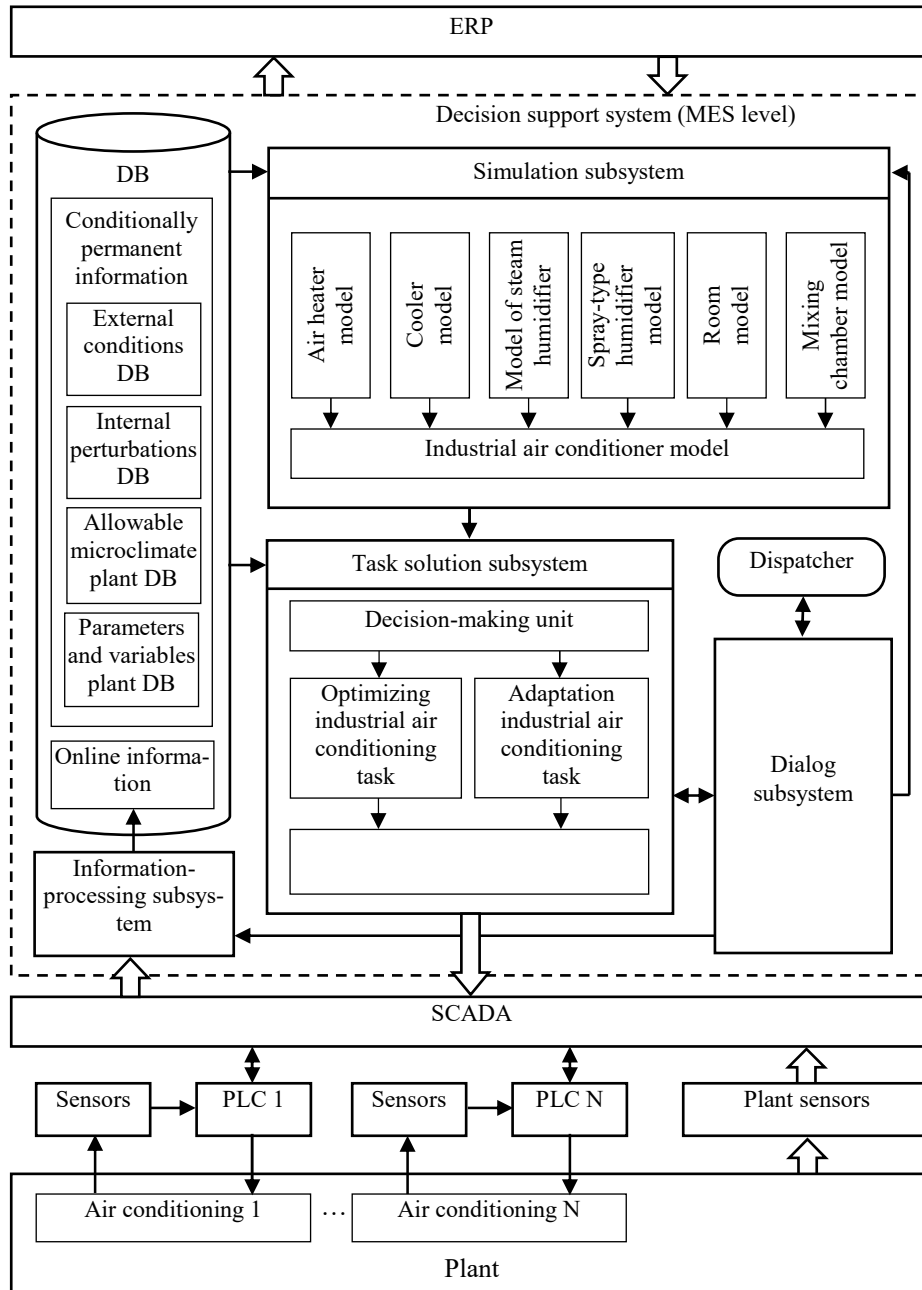


Fig. 2. DSS block diagram for microclimate management

At the lower level, local air conditioning control systems are implemented on program logic controllers (PLC). The SCADA performs cyclical measurements of all

current process variables, including the microclimate of industrial premises. The information-processing subsystem DSS performs analysis of input information, replenishes the database (DB) with new information and provides operational information about the condition of industrial air conditioners: the disturbing effects magnitude; operating modes of mixing chambers, water heaters, coolers, steam humidifiers, and industrial rooms.

Using models of climatic equipment, the simulation subsystem implements a model of an industrial air conditioner. The task of control solving subsystem offers the operator a set of possible scenarios for optimization and adaptive control of industrial air conditioners that are used in production. The effectiveness of the proposed solutions is evaluated according to the criteria chosen by the operator. The decisions made by the operator in the form of PLC settings are transferred to the lower hierarchical system control level for execution.

3.1 Mathematical model of industrial air conditioning

The methodology for implementing the air conditioner model was considered in [10, 11]. In general, it is proposed to implement a model of an industrial air conditioner in a state space

$$\begin{cases} \mathbf{X}' = \mathbf{A}\mathbf{X} + \mathbf{B}\mathbf{U} + \mathbf{W}, \\ \mathbf{Y} = \mathbf{C}\mathbf{X} + \mathbf{V}. \end{cases} \quad (1)$$

The vectors of the variable statistical perturbations model $\mathbf{W}$ and $\mathbf{V}$ are determined in the DSS using the Kalman filtering methods considered in [14, 15]. The matrices content of deterministic component model ($\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$) depends on the technological scheme for processing air by an industrial air conditioner. For example, the state matrix $\mathbf{A}$ is determined by for:

- a direct-flow conditioning system

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}^{P1} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{C}_{P1} & \vdots & \vdots & \mathbf{0} & \vdots & \vdots \\ \mathbf{0} & \dots & \mathbf{A}^{Pi} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \vdots & \mathbf{C}_{Pi} & \dots & \mathbf{A}^{PN-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{C}_{PN-1} & \mathbf{A}^{PN} \end{bmatrix},$$

where $\mathbf{A}^{Pi}$ is the state matrix of the i -th equipment, N is the number of devices in the technological chain;

- a system with recirculation or heat recovery

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}^{P1} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{C}_{PN} \\ \mathbf{C}_{P1} & \mathbf{A}^{P2} & \vdots & \vdots & \vdots & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_{P2} & \mathbf{A}^{Pi} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{C}_{Pi} & \mathbf{A}^{Pi+1} & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots & \mathbf{C}_{Pi+1} & \mathbf{A}^{PN-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{C}_{PN-1} & \mathbf{A}^{PN} \end{bmatrix};$$

- multi-zone conditioning systems, the filling of matrix **A** must be analyzed additionally, since it depends on the number of rooms with a microclimate system.

The control matrix **B** is formed on the basis of the existing climate equipment that is involved in controlling the air conditioning. The observation matrix **C** is formed based on the number of state variables that take part in controlling the air conditioning. Among the state variables that are necessarily involved in the management of mandatory are temperature and humidity at the outlet of the air conditioner (or indoors).

A more detailed acquaintance with the methodology for the synthesis of industrial air conditioners complex dynamic models can be found in [10, 11]. For control systems for industrial air conditioners, a synthesis procedure has been developed for a multidimensional linear-quadratic digital controller (LQDC) [10], which takes into account the logical switching of the climate equipment and allows the ACS of industrial air conditioner to adapt to changing of dynamic properties plant. At the ASC-level, direct digital control algorithms are used [9–11]. On a PLC, the control system can be implemented using one-dimensional digital controllers (DC), or using multidimensional LQDC. The use of ACS with LQDC improves the integral quality indicators by 1.2 – 2.3 times in comparison with ACS where one-dimensional DC are used [12].

Consider the tasks of operational management to optimize and adapt the ACS of industrial air conditioners, which are implemented by the task solution sub-system.

3.2 Operational management for the optimization ACS air conditioning

The task of operational management to optimize the air conditioning system is necessary to determine the acceptable boundaries of temperature and humidity in an industrial room, as well as the recirculation coefficient (for air conditioners with recirculation), which affects the economical use of microclimate system energy.

As an example, consider an air conditioner with a steam humidifier and recirculation. For this air conditioner, see Fig. 3 shows a $h - d$ diagram of the preparation of outdoor air with an inlet temperature $\theta_{A0}^{\min} \dots \theta_{A0}^{\max}$, and moisture content $d_{A0}^{\min} \dots d_{A0}^{\max}$. Region V characterizes the permissible microclimate in an industrial room with a range of a given temperature $\theta_{ref}^{\min} \dots \theta_{ref}^{\max}$ and moisture content, $d_{ref}^{\min} \dots d_{ref}^{\max}$. The smaller is the area V, the higher are the production requirements for the indoor microclimate.

When processing outdoor air (point A or A*) to a predetermined value (point D or D*), the minimum energy consumption of the air conditioner will be at the edges of the area V boundary. Therefore, the larger is the area V, the less energy is needed to prepare the outdoor air. Region V, on the other hand, is limited by production requirements for indoor microclimate. To optimize the area of permissible microclimate, the DB of the DSS should contain information on the permissible microclimate parameters for all plants of production.

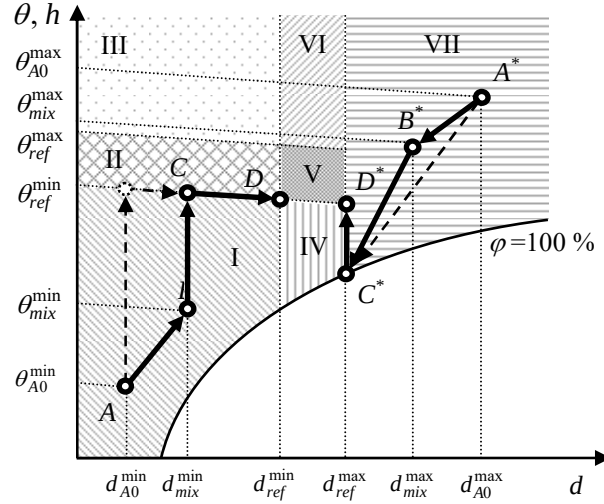


Fig. 3. $H-d$ preparation diagram air

Optimization of the permissible microclimate in industrial premises of production is carried out during the transition of the technological line to the manufacture of new products and consists of the following steps:

- 1) the decision-making unit initiates a dialogue with the operator and offers acceptable microclimate limits in the room for manufacturing of new products;
- 2) the simulation model of an industrial air conditioner is formed according to the methodology given in [10];
- 3) the simulation subsystem performs the ambient air preparation quantitative modeling: the transition from point A (A^*) to point D (D^*);
- 4) then, the assessment is made of the industrial air conditioner the efficiency of which is based on the criterion of minimizing energy consumption:

$$\sum_{i=1}^N Q_i \rightarrow \min \quad (2)$$

where, Q_i is the i -th apparatus power consumption; for the water heat-transfer agent $Q_i = G_i c_i (\theta_{i0} - \theta_i)$, where G_i, c_i are respectively the flow rate and the coolant heat capacity; θ_{i0}, θ_i are the heat-transfer agent temperatures at the inlet and the apparatus outlet;

- 5) the operator analyzes the accumulated statistical information and decides to change the specified reference microclimate zone;

- 6) the laboratory section analyzes influence of the microclimate of the industrial premise on the product quality; if necessary, the range of the boundaries permissible region V is modify by the operator according to steps 1 – 6; thus, statistical information is collected that allows one optimize the energy consumption of the air conditioner due to the permissible range parameters of microclimate in the room.

A significant reduction in energy consumption of the air conditioner is possible through the use of up-take air recirculation. Utilization of the exhaust air heat makes it possible to bring the parameters of the microclimate ambient air from point A to point B for the winter season, and from point A^* to point B^* for the summer season (see Fig. 3). At the same time, the energy consumption for air processing is significantly reduced (transition from point B / B^* to point D / D^*). At 100% recirculation, the energy efficiency air conditioner is maximum. However, according to sanitary norms and rules, there are standards for air exchange with the external environment, which must be maintained in the production premises if there are workers in the room. To optimize the recirculation coefficient of an industrial air conditioner, an air conditioning system database should contain regulatory information on air exchange for industrial premises with employees.

The air conditioner recirculation coefficient is optimized during the transition of the artificial microclimate system from winter to summer mode and vice versa, as well as when the number of workers in the industrial premises changes and consists of the following steps:

- 1) decision making unit initiates a dialogue with the operator and offers permissible recirculation value for the number of employees in the room;
- 2) the simulation model of an industrial air conditioner is formed according to the methodology given in the studies [10];
- 3) the simulation subsystem performs an ambient air preparation quantitative modeling: this is the transition from point A (A^*) to point D (D^*);
- 4) the assessment of the effectiveness of the operation air conditioner based on minimizing the criterion of energy consumption (1);
- 5) the operator changes the recirculation coefficient for the air conditioner model based on the experience and DSS statistics;
- 6) an analysis is performed of the energy efficiency air conditioner; when changing the number of employees or the operation mode air conditioner, steps 1 – 6 are repeated to optimize the recirculation coefficient.

Thus, the DSS helps the operator to optimize the energy efficiency of industrial air conditioning equipment without loss of product quality in production.

3.3 Operational management for the adaptation ACS air conditioning

The operational management task under adaptation is necessary for the correction of ACS settings when changing the dynamic properties of an industrial air conditioner. Any adaptive control system has two loops: internal – a direct-digital control circuit; external – the circuit for evaluating the parameters of plant and calculating the parameters of the controller. The internal adaptive control loop is implemented on the PLC, and the external one in the frames of the constructed DSS (see Fig. 2).

The ACS adaptation of an air conditioner is carried out during the transition of the artificial microclimate system from winter to the summer mode and vice versa. The operational control algorithm for adaptation includes of the following steps:

- 1) the decision-making unit initiates a dialogue with the operator and offers to evaluate the parameters of mathematical models of HVAC equipment; for models, heat- and mass- transfer coefficients are estimated, which depend on many factors; to eval-

uate the parameters, one can use the scanning or gradient search methods; as a reference model, the accumulated information of measured data of the operated air conditioning systems, which are contained in the DB of the DSS, is used;

2) an industrial air conditioner simulation model is formed according to the method given in the study [10];

3) the LQDC feedback matrix is computed for the diagram regions $h - d$ [10];

4) the assessment is made of the simulation ACS effectiveness functioning with LQDC using a quadratic criterion for all areas of the $h - d$ diagram:

$$I = \int [\mathbf{X}^T \mathbf{Q} \mathbf{X} + \mathbf{U}^T \mathbf{R} \mathbf{U}] dt, \quad (3)$$

where $\mathbf{Q}$, $\mathbf{R}$, are weighting coefficient matrices for elements of state and control vectors; $\mathbf{X}$ is state vector, which includes variable temperatures θ and moisture content d ; $\mathbf{U}$ is the control vector;

5) the operator changes the LQDC parameters if the air conditioner ACS efficiency with the calculated parameters is higher than the existing one;

6) the steps 1 – 5, aiming to adapt LQDC are repeated when the industrial air conditioner changes its operating mode.

4 Conclusions

The DSS structure for the integrating industrial air conditioners ACS into the integrated enterprise management system is proposed. It was proposed to use mathematical models and control algorithms developed in the previous developments [8–12] at the enterprise management (MES) level in the decision support system proposed, which is implemented to improve the industrial air conditioners operation efficiency under changing operating conditions. As a result of applying the system analysis methodology, the multifunctional mathematical and algorithmic software was developed for integrating an industrial air conditioning control system into computer-integrated production management system. The developed models, methods and algorithms are recommended to be used at the stage of design, commissioning and operation at middle level of operational production management. The proposed mathematical models, methods and algorithms for controlling air conditioners are brought to practical implementation and theirs are proposed to be implemented in DSS. The industrial implementations results have confirmed the high efficiency of automatic control systems developed on the basis of the systemic approach to building conceptually integrated industrial control systems discussed above.

References

1. Belova, E.: Central air conditioning systems in buildings. Euroclimate, Moscow (2006).
2. Kletti, J., Berres, B., Brauckmann O.: Manufacturing Execution Systems – MES. Springer, Heidelberg (2007).
3. Meyer, H., Fuchs, F., Thiel, K.: Manufacturing Execution Systems. Optimal Design, Planning, and Deployment. McGraw–Hill, USA (2009).

4. Itskovich, E.: Production enterprises automation complex methods in technological industries. KRASAND, Moscow (2013).
5. Enterprise/Control System Integration. Part 1: Models and Terminology: ANSI/ISA-95.00.01-2010. International Society of Automation, USA (2010).
6. Enterprise/Control System Integration. Part 2: Object Model Attributes: ANSI/ISA - 95.00.02-2010. International Society of Automation, USA (2010).
7. Enterprise/Control System Integration. Part 3: Activity Models of Manufacturing Operations Management: ANSI/ISA-95.00.03-2013. International Society of Automation, USA (2013).
8. Golinko, I.: Industrial air conditioners automatic control systems synthesis principles. Automation technological and business-processes, 8, 33–42 (2016).
9. Golinko, I., Galytska I.: Microclimate control system synthesis using relative humidity and humidity content as a stabilization parameter. Information systems, mechanics and control, 13, 88–98 (2015).
10. Golinko, I., Galytska, I.: Optimization of multidimensional control systems for complexes of artificial microclimate. Precarpathian bulletin of the Shevchenko Scientific Society. Series Number, 1(33), 61–73 (2016).
11. Golinko, I., Galytska, I.: Mathematical model of heat exchange for non-stationary mode of water heater. ICCSEEA 2019: Advances in Computer Science for Engineering and Education II, pp. 58–67. Springer, Switzerland (2020).
12. Golinko, I., Galytska, I.: An artificial microclimate control system synthesis using relative humidity and moisture content as a parameter of stabilization. Information systems, mechanics and control, 13, 88–98 (2015).
13. Zgurovsky, M., Pankratova, N.: Basics of system analysis. BHV Publishing Group, Kiev (2007)
14. Zgurovsky, M., Podladchikov, V.: Analytical methods of Kalman filtration for systems with a priori uncertainty. Naukova Dumka, Kiev (1995).
15. Ponyatsky, V.: Study of ways to implement an adaptive control system with a Kalman filter. Stochastic optimization in computer science, 1, 196–210, (2008).

Enterprise Architecture Modeling Support based on Data Extraction from Business Process Models

Dmytro Orlovskyi^[0000-0002-8261-2988] and Andrii Kopp^[0000-0002-3189-5623]

National Technical University “KhPI”, Kyrpychova str. 2, 61002 Kharkiv, Ukraine
orlovskyi.dm@gmail.com, kopp93@gmail.com

Abstract. This paper presents a problem of enterprise architecture artifacts extraction from business process model collections, which organizations of higher maturity levels tend to manage, in order to build the architecture landscape and apply enterprise architecture management practices. Existing enterprise architecture frameworks, languages, and methodologies are discussed. Proposed approach is based on mapping between business process and enterprise architecture elements in order to present all business architecture artifacts in a single coarse-grained model. The software implementation allows generating business architecture landscapes that could be used for architecture evolution purposes, such as transformation planning or maintenance efforts evaluation.

Keywords: Enterprise Architecture, Business Process Model, Business Architecture, Architecture Landscape, ArchiMate.

1 Introduction

These days Enterprise Architecture (EA) has two definitions upon the context. It can be considered as a formal description of a system or a detailed plan of the system at component level to guide its implementation. On the other hand, EA may be defined as the structure of components, their interrelations, principles and guidelines governing their design and evolution over time [1]. The term “enterprise” denotes any organization linked by a common set of goals. Despite existing misconceptions, EA is covering not only operational software systems, but strategic, business, and organizational aspects as well. Therefore, EA deals with requirements and strategies, as well as business processes, technical applications, and infrastructures. EA strives for optimal articulation between these different facets [2].

The EA is expressed using models, which are based on a metamodel that defines model types and their relationships. Various EA frameworks have their own particular metamodel and taxonomy, but in general, all the EA frameworks cover the following domains [1, 3, 4]:

- Business Architecture: business goals, business functions or capabilities, and business processes. It defines the business strategy, governance, organization, and key business processes.

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

- **Data Architecture:** data stores and data objects mapped to the business functions and business processes that need specific data. It describes the structure of an organization's logical and physical data assets and data management resources.
- **Application Architecture:** the structure and behavior of software applications that produce and consume data used by business functions and business processes. It provides a blueprint for the individual application systems to be deployed, their interactions, and their relationships to the core business processes of the organization.
- **Technology Architecture:** the structure and behavior of the IT (Information Technology) infrastructure (client/server nodes, system software, protocols and networks). It describes the logical software and hardware capabilities that are required to support the deployment of business, data, and application services.

As it was already mentioned, EA is often positioned only within the context of IT governance. However, EA is actually related with a number of well-known best practices and standards in IT and general management [4]. The set of the best practices, frameworks, guidelines, and standards outlined below serves to support each of the considered EA viewpoints (business, data, applications, and technical architectures):

- **Strategic Management: Balanced Scorecard (BSC).** The scorecard measures organizational performance across four perspectives: financial, customers, business processes, learning and growth. The BSC enables companies to track financial results while monitoring progress in building the capabilities and acquiring intangible assets they need for future growth [5].
- **Business model development: Business Model Canvas.** A Business Model Canvas (BMC) helps users visually represent the elements of a business model and the potential interconnections and impacts on value creation. As a visual tool, the BMC can facilitate discussion, debate, and exploration of potential innovations to the underlying business model itself; with users developing a more systemic perspective of an organization and highlighting its value creating impacts [6].
- **Business architecture: BIZBOK (Business Architecture Body of Knowledge) and O-BA (Open Business Architecture).** BIZBOK defines business architecture as a blueprint of the enterprise that provides a common understanding of the organization and is used to align strategic objectives and tactical demands [7].
- **Quality Management: EFQM (European Foundation for Quality Management) and ISO (International Organization for Standardization) 9001.** The EFQM model has a much broader scope than ISO 9001, since it not only focuses on quality management, but provides an overall management framework for performance excellence of the entire organization [4].
- **IT governance: COBIT (Control Objectives for Information and Related Technologies) framework** that assists enterprises in achieving their objectives for the governance and management of enterprise IT. It helps enterprises create optimal value from IT by managing a balance between realization benefits and optimization risk levels and resource use [8].
- **IT delivery and support: ITIL (IT Infrastructure Library).** The ITIL is a technique to manage the technology and communications in an optimal way. The primary ob-

jective of the ITIL is to establish the best practices and improving the standard of IT service quality that customers should demand and providers should supply [9].

- IT implementation: CMM (Capability Maturity Model) and CMMI (Capability Maturity Model Integration). CMMI contains practices that cover project management, process management, systems engineering, hardware engineering, software engineering, and other supporting processes used in development and maintenance of both products and services [10].

In order to organize these diverse EA perspectives into a holistic and unified view, it is required to use an Enterprise Architecture Framework (EAF) [11]. According to reference [1], EAF should describe a method for designing an information system in terms of a set of building blocks, and for showing how the building blocks fit together. However, EA is not only about information systems. Thus, the goal of EAF is to provide a language, an approach, and a set of recommendations covering all facets of the EA, from organization and strategy, to business and technology, to planning and change management [2].

Modern EA frameworks define the Architecture Landscape (AL) as the representation of EA assets that are planned or already in use by the enterprise. The AL contains models of the existing architecture across the entire enterprise. Such models deal with business processes, applications, and data. Naturally, AL content is constantly evolving as architectural transformation take place. All the EA content, such as models and other architectural descriptions, is contained in the EA repository that serves as the source of EA models for the AL [2]. It is necessary to understand, that the EA repository is not concrete software or data storage but just an abstract concept of a single place where the enterprise descriptions should be stored for their further reuse for EA evolution purposes. Also, gathering the information about all or the most valuable EA assets and their preparing according to the practiced EAF might be a long-term and quite expensive project. However, many organizations maintain repositories of business process models that serve as a knowledge base for their ongoing business process management efforts [12].

Hence, in this paper we present the idea on how the EA artifacts might be extracted from the collection of business process models in order to build the AL and apply EA management practices. Processing of a large business process model collection might help to retrieve information about business processes and business functions, corresponding inputs and outputs, participants of business processes, and triggering events. This approach is supposed to shorten time, save costs and efforts for gathering information in order to design the business architecture view of the whole AL, which then might be complemented with required data, applications, and technical architectures artifacts that support business process execution.

2 Literature Review

Many EA methodologies have come and gone in the last decades. According to [13], most enterprises use one of these four EAFs: The Zachman Framework for Enterprise Architecture, The Open Group Architecture Framework (TOGAF), Federal Enterprise

Architecture Framework (FEAF), and the Gartner Enterprise Architecture Framework (GEAF). Additionally to these enterprise frameworks, the latest research [11] considers the United States Department of Defense Architecture Framework (DoDAF) and derived from it NATO (North Atlantic Treaty Organization) Architecture Framework (NAF), British Ministry of Defense Architecture Framework (MODAF), and Unified Architecture Framework (UAF) responding to the needs of military communities to create a standardized and consistent EA based on DoDAF and MODAF frameworks.

Originally, the first EAF was designed by John Zachman. He described the EAF as a logical structure for classifying and organizing the descriptive representations of an enterprise that are significant to the management of the enterprise, as well as to the development of the enterprise's systems [14]. The Zachman Framework is actually the taxonomy for organizing architectural artifacts (design documents, specifications, EA models etc.), which takes into account both who the artifact targets (e.g., business owner or another stakeholder) and what particular issue (e.g., data and functionality) is being addressed [13]. Thus, authors of [15] have formalized the Zachman Framework using the ontology model. Detailed overview of the Zachman Framework for Enterprise Architecture (ZFEA) is described in [16]. One of the authors of this paper is John Zachman, the originator of the framework. They have described the ZEFA as a bounded matrix with six rows and six columns. The columns are the interrogatives, answers to which allow describing all aspects of any enterprise, while the rows represent different perspectives on the enterprise from the viewpoint of different stakeholders [16]. Besides the Zachman Framework, military frameworks (DoDAF, NAF, MODAF, and UAF), and the United States federal methodology FEAF, TOGAF was identified in [11] as the best EAF, mostly according to the criteria: information availability, tool support, and prevalence by researchers. In the overview of architectural frameworks authors of [17], together with DoDAF and FEAF, mentioned another U.S. federal EA methodology, such as Treasury Enterprise Architecture Framework (TEAF). However, they declared only TOGAF and the Zachman Framework as the two popular EAFs that are used nowadays [17].

TOGAF is the one of the leading EA frameworks worldwide. It is developed and is currently maintained as a standard by The Open Group (TOG). The TOGAF documents focus on EA key concepts and Architecture Development Method (ADM), an iterative approach to developing the EA [18]. The ArchiMate EA modeling language is a TOG standard as well; it provides the architect with instruments that support and improve the architecture process [18]. The generic ArchiMate metamodel consists of two main types of elements: structure and behavior elements. Structure elements can be subdivided into active structure elements and passive structure elements. Active structure elements can be further subdivided into external active structure elements (also called interfaces) and internal active structure elements. Behavioral elements can be subdivided into internal behavior elements, external behavior elements (also called services), and events [19].

The ArchiMate specification also declares its relationship to other standards, specifications, and guidance documents (see Fig. 1). Among these standards are TOGAF, BIZBOK, BPMN, and UML (Unified Modeling Language) [19]. BPMN stands for Business Process Model and Notation. It is the leading standard for modeling business

processes introduced by the Object Management Group (OMG). Due to its popularity, BPMN is considered to be the true lingua franca of Business Process Management (BPM). By allowing BPM practitioners to create business process models using a common graphical notation, BPMN makes it easier to communicate processes in a compact way across companies and continents [20]. Latest BPTrends survey [21] declares that BPMN process models are used by 64% of questioned organizations.

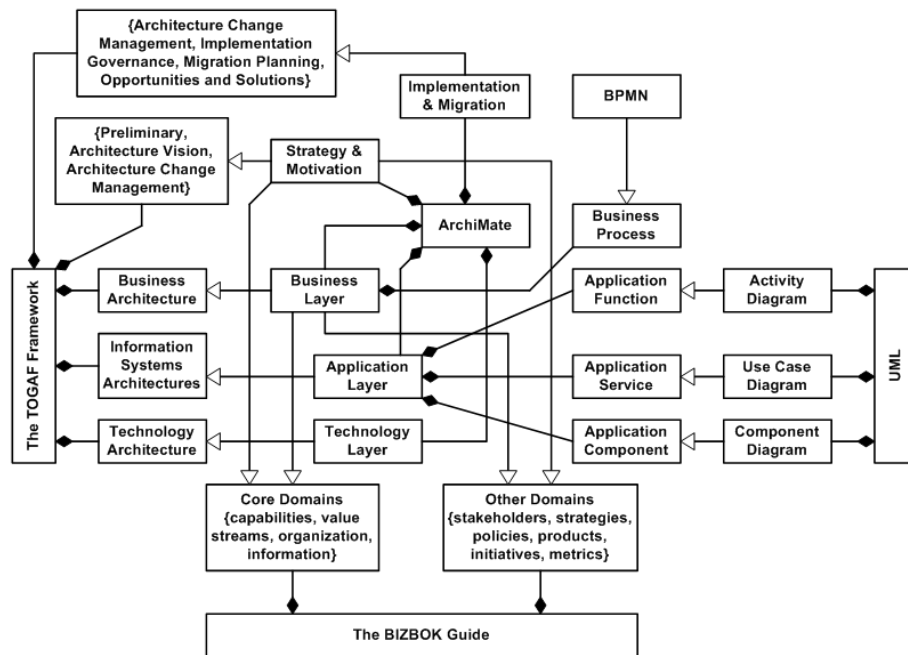


Fig. 1. ArchiMate relationship to other standards

Both ArchiMate and BPMN can be used for modeling business processes, but their aims are different. ArchiMate is used for high-level processes and their relations to the enterprise context, whereas BPMN offers detailed workflow modeling, but lacks the application services that support a process or goals it has to fulfill [19]. Therefore, BPMN business process models might be used as the source of artifacts only for the BA perspective of the EA landscape according to a bottom-up approach.

Authors of the paper [22] have proposed the way, in which the ArchiMate Business Layer and BPMN meta-models can be linked. However, this paper does not provide any direct mapping between corresponding ArchiMate and BPMN elements. Author of [23] proposes mapping of ArchiMate elements to corresponding BPMN elements. However, the linkage between BPMN and ArchiMate defined at the metamodel level is not outlined in details and, therefore, cannot be checked or proven. More detail and cogent mapping is proposed in [24]. Author of this paper have performed the syntactic, semantic, and structural analysis of ArchiMate and BPMN meta-models. However, mapping results presented in paper [24] contain many-to-many relations between ArchiMate and BPMN elements, which cause uncertainty. Thus, it is

ArchiMate and BPMN elements, which cause uncertainty. Thus, it is required to define one-to-one linkage between ArchiMate and BPMN elements by matching meta-models of the considered standards.

3 Formal Problem Statement

ArchiMate was chosen for description of the EA, since it is a contemporary, open and independent language. It comprises three main modeling layers: business, application, and technology. ArchiMate allows presenting a whole EA in the form of views which, depending on the needs, can include only items in one layer or can show vertical relations between layers. The internal structure of an ArchiMate model constructs a graph of nodes linked by directed edges. Both nodes and edges are attributed with information indicating a type of element or relation [25].

Formally ArchiMate EA model can be represented using the tuple (1) [25]:

$$AM = \langle V, E, C, R, vt, et \rangle. \quad (1)$$

Here V – the set of vertices; $E \subset V \times V$ – the set of edges; C – the set of element types; R – the set of relations; $vt: V \rightarrow C$ – the function that assigns element types to graph vertices; $et: E \rightarrow R$ – the function that assigns relation types to edges.

In order to provide the AL template in ArchiMate language, it is required to define a set of vertices, where each vertex $v \in V$ describes the EA asset of a certain type. To do this, it is necessary to define the mapping between BPMN and ArchiMate artifacts. The software implementation of the proposed approach should take the collection of BPMN models as input and produce the ArchiMate model with pre-defined EA elements of the business layer ready to be used for AL design purposes as output. The conceptual scheme of a considered problem is demonstrated in Fig. 2.

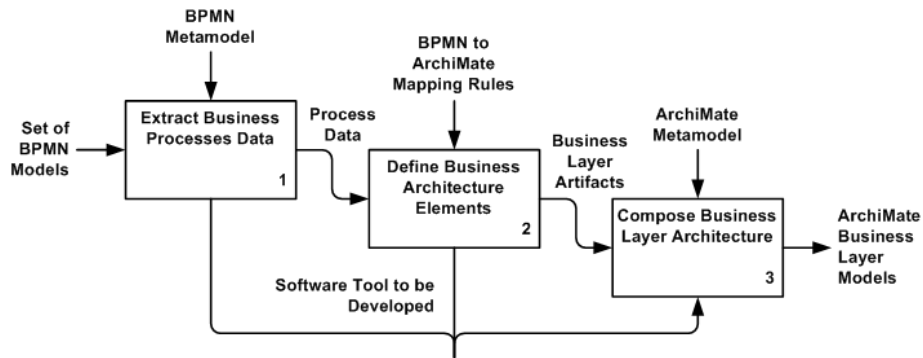


Fig. 2. Conceptual scheme of the considered problem

Such bottom-up approach might save time and resources required to design the EA by providing a ready-to-use ArchiMate document with pre-defined business architecture elements (Fig. 2).

4 Proposed Approach

In order to define mapping of BPMN elements to ArchiMate elements, we have come up with the idea to use the corresponding meta-models of these standards. Such meta-models are provided in paper [22] and their fragments are depicted in Fig. 3.

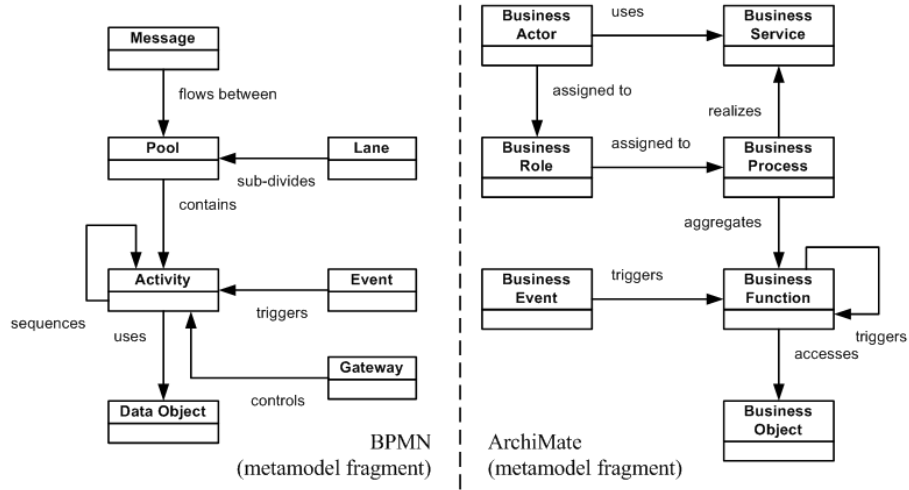


Fig. 3. Fragments of BPMN and ArchiMate Business Layer meta-models

We have translated such meta-models into a collection of RDF (Resource Description Framework) statements (provided in a form of “subject-predicate-object” triples), originally designed to represent a metadata better than other relational or ontological models [26]. As the result, we obtained RDF-graphs that have been queried in order to define similarities and then to conclude mapping rules between their nodes. For this purpose we have used Apache Jena, the open-source Java-based framework used in Linked Data and Semantic Web applications [27]. It provides API (Application Programming Interface) that allows querying RDF models directly or with the help of the query language SPARQL.

The starting point is our awareness that a business event in ArchiMate is similar to BPMN event elements, which is outlined in the ArchiMate specification [19]. Since in BPMN an event triggers an activity object, we can conclude that in ArchiMate a business event also triggers some sort of a behavior element. It is shown that SPARQL is an SQL-like language, so queries used to define relations between the event and other elements are the following:

```
SELECT * WHERE { <Event> ?a ?x . }
SELECT * WHERE { <Business Event> ?b ?y . }
```

Results of these SPARQL queries are shown in Fig. 4. Therefore, it is seen, that the BPMN activity might be represented as the business function in ArchiMate.

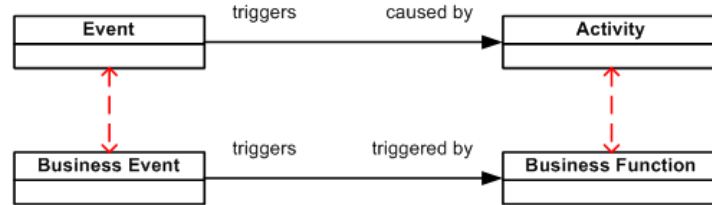


Fig. 4. Matching BPMN Activity to ArchiMate Business Function

Then, we can check the “Activity” and “Business Function” nodes of BPMN and ArchiMate meta-models respectively:

```
SELECT * WHERE { <Activity> ?a ?x . }
SELECT * WHERE { <Business Function> ?b ?y . }
```

Query results of outgoing arcs in RDF-graph have shown that both activity and business function have cycles (“sequences” and “triggers” relations) used to model a flow of activities. Also both activity and business function are related to data objects in BPMN and business objects in ArchiMate. These relations have labels with similar terms “uses” and “accesses”, so we can assume that BPMN data objects could be mapped to ArchiMate business objects (Fig. 5).

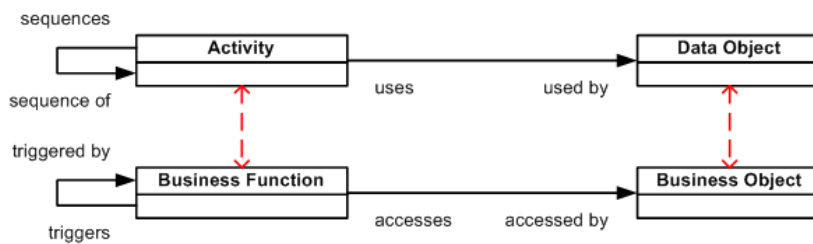


Fig. 5. Matching BPMN Data Object to ArchiMate Business Object

In order to analyze incoming arcs we have used the following SPARQL queries:

```
SELECT * WHERE { ?x ?a <Activity> . }
SELECT * WHERE { ?y ?b <Business Function> . }
```

Query results of incoming arcs in RDF-graph have shown that both activity and function are parts of (by analyzing relationships “contains” and “aggregates”, which meaning is very close) other elements “Pool” and “Business Process” of BPMN and ArchiMate meta-models respectively. Also it is known that a pool in BPMN sets the boundaries of a business process. Hence, we can assume that BPMN pools could be mapped to ArchiMate business processes (Fig. 6).

Incoming arcs of “Pool” in the BPMN metamodel and “Business Process” in the ArchiMate metamodel were defined using the following queries:


```

    if !visited.contains(?y):
        queue.add(?y)
        visited.add(?y)
    subjects = execute(SELECT * WHERE { ?x ?y node . })
    while statement = subjects.next():
        if !queried.contains(statement):
            queried.add(statement)    # mark queried incoming
                                     # relation as traversed
        if !visited.contains(?x):
            queue.add(?x)
            visited.add(?x)

```

Unlike the classic BFS algorithm, the proposed modification uses the additional list “queried” in order to memorize already queried RDF-triples. It is also harder than in ordinary graphs to get the adjacent edges of a certain vertex. For this purpose we need to execute two SPARQL queries in order to retrieve outgoing and incoming relations separately (see the pseudo code above). Using the RDF-BFS algorithm it is possible to traverse more complicated BPMN and ArchiMate ontologies in order to support the BPM-EA interoperability. Results of BPMN and ArchiMate meta-models traversing are shown in Table 1.

Table 1. Results of BPMN and ArchiMate meta-models traversing

Node	Statements (BPMN)	Node	Statements (ArchiMate)
Event	Event triggers Activity	Event	Event triggers Function
Activity	Activity uses Data Object, Activity sequences Activity, Gateway controls Activity, Pool contains Activity	Function	Function accesses Object, Function triggers Function, Process aggregates Function
Pool	Lane sub-divides Pool, Message flows between Pool	Process	Process realizes Service, Role assigned to Process

Obtained results (Table 1) correspond to the results of manual RDF-graphs querying, which proves concluded mapping rules (Fig. 4-7). By analyzing results of BPMN and ArchiMate meta-models traversing (i.e. correspondence between nodes and relations) we have defined one-to-one linkage between ArchiMate and BPMN elements.

5 Experiments and Results

Developed software solution is based on the provided mapping rules between BPMN and ArchiMate elements. Its workflow is described using the UML activity diagram that is shown in Fig. 8. Business process models described using the BPMN notation are usually stored in a form of the XML-based (eXtensible Markup Language) language called BPMN 2.0. Its structure like of any other XML document consists of nodes and attributes, which represent elements and sequence flows within a business

process. With the help of a BPMN model API [29] we can easily extract information from an existing business process definition, edit an existing business process definition or create a complete new one without manual XML parsing.

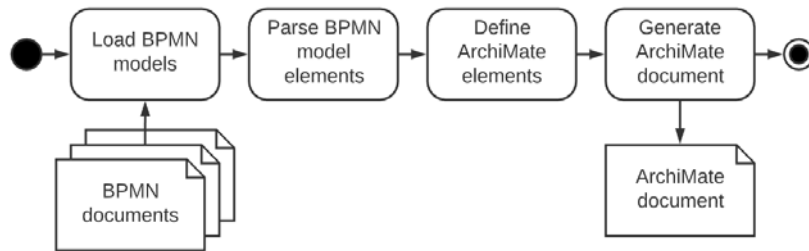


Fig. 8. Activity diagram that represents a workflow of the software solution

As for ArchiMate, its exchange file format is defined by The Open Group Standard and can be used to exchange data between tools and/or systems that wish to import, and export ArchiMate models [30]. The ArchiMate exchange file format is also based on XML language and supported by multiple EA modeling tools.

The UML deployment diagram in Fig. 9 demonstrates the architecture of the software solution. The Java-based web application, created using the Spring Boot framework, loads BPMN 2.0 files from a catalog on the FTP (File Transfer Protocol) server and parses business process definitions using the BPMN model API [29]. Then, for each element of a business process extracted from the BPMN 2.0 file, a corresponding BA artifact (business process, function, event, etc.) is defined according to the mapping rules provided in previous section (Fig. 4-7).

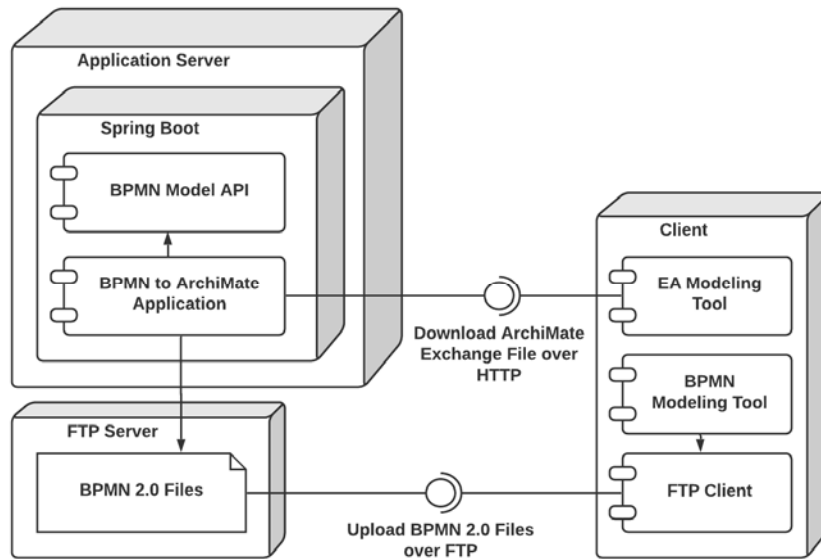


Fig. 9. Architecture of the software solution

On output the application generates an XML document formed according to the ArchiMate exchange file format. It could be downloaded over the HTTP (HyperText Transfer Protocol) manually or directly accessed by any EA modeling tool that supports the ArchiMate exchange file format.

Sample set of BPMN models that describe business processes in the field of products supply was used to validate proposed solution (Fig. 10).

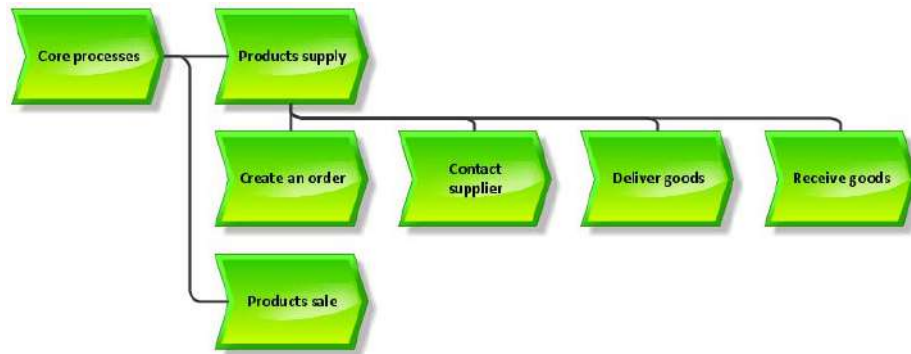


Fig. 10. Generic business process landscape

Since we are limited in space, let us demonstrate a single BPMN model of a whole set, which describes “Create an order” business process (Fig. 11). However, the set of considered business process models in BPMN 2.0 format could be found at [31].

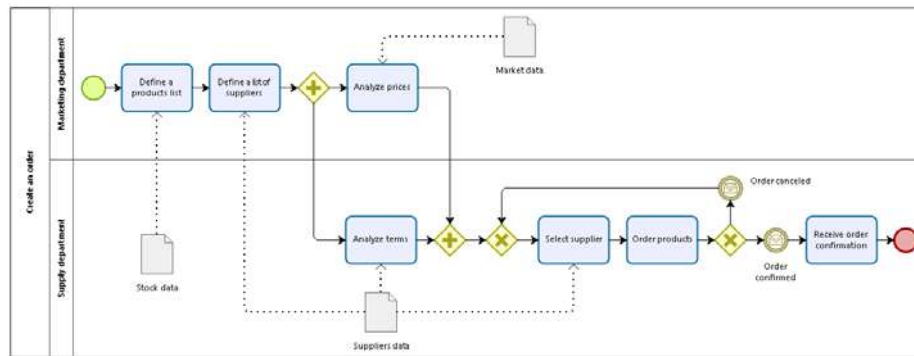


Fig. 11. “Create an order” business process model

Complete BA landscape for the given business processes (Fig. 10) is demonstrated in Fig. 12. As the EA modeling tool we have used Archi, which is open source, cross-platform solution that also supports the ArchiMate exchange file format. As the result, we have obtained business architecture elements that form the set of vertices V that could be augmented by relations (composition, assignment, triggering etc.) that form the set R according to the formal definition AM of the EA model. Besides relations,

the set of BA elements could be extended by adding data, applications, and technical architecture elements, and relations.

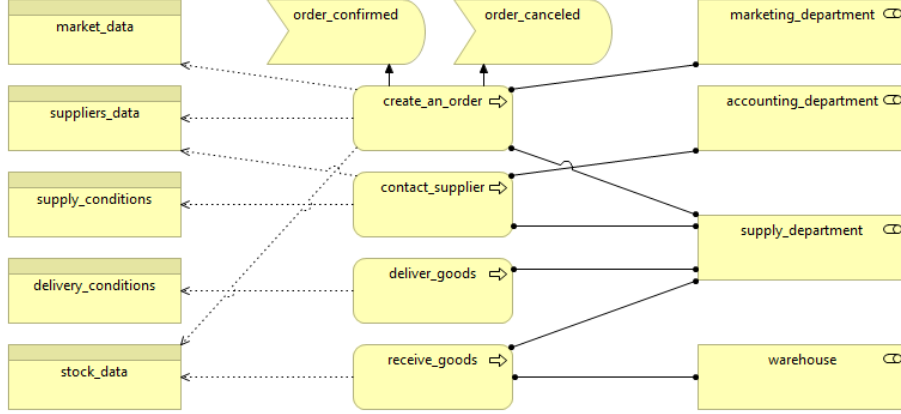


Fig. 12. Business architecture landscape generated from the given process models

Business functions are avoided on the presented EA model (Fig. 12) in order to achieve clearer and less congested diagram. It is shown that all labels are presented in lower case with space characters replaced with underscores in order to avoid duplicated strings (e.g., the “Create an order” label of the pool in the BPMN diagram is transformed into the “create_an_order” label of the business process in the EA model). In order to evaluate the obtained business architecture landscape, we used the Propagation Cost (2), which equals to the fraction of the architecture affected when a change is made to a randomly selected element [32]:

$$\text{PropagationCost} = \frac{1}{|V|^2} \cdot \sum_{v \in V} \deg^+(v) = \frac{1}{|V|^2} \cdot \sum_{v \in V} \deg^-(v). \quad (2)$$

It is computed from $\deg^+(v)$ – the fan-in (the number of elements that depend on a certain element) or $\deg^-(v)$ – the fan-out (the number of elements that a certain element depends on) [32]. Obtained propagation cost (0.10) means that only 10% of the business architecture may be affected in case of required transformation.

6 Conclusion and Future Work

In this study we have discussed the problem of EA artifacts extraction from the collection of business process models, which may exists in BPM-mature organizations, in order to build the architecture landscape and apply EA management practices. We have proposed the approach based on ArchiMate and BPMN metamodels comparison used to define direct mapping between business process and EA modeling elements of

business architecture layer. In contrast to existing studies in this field, such as [23, 24], we have provided one-to-one mapping between BPMN and ArchiMate modeling elements (Fig. 4-7), which is based on the meaning of nodes and relations of the corresponding meta-models.

Proposed approach was formalized using the BFS algorithm extended by the RDF-querying features. Developed software that implements proposed approach is web-based and interoperable, since it supports BPMN 2.0 and ArchiMate exchange formats. Provided example demonstrates the business architecture landscape extracted from the set of BPMN models. In order to demonstrate utility of the EA landscape, we evaluated EA maintenance efforts and evolvability with the help of the propagation cost measure. However, instead of or together with the propagation cost there might be used another link analysis metrics in order to evaluate obtained business architecture landscape.

Future work includes elaboration in the field of automatic EA modeling, which can be referred as “EA-mining” similarly to the process mining in BPM field.

References

1. Pieterse, R.: Enterprise Architecture Frameworks, Methods and Tools. Lulu.com (2015)
2. Desfray, P., Raymond, G.: Modeling Enterprise Architecture with TOGAF: A Practical Guide Using UML and BPMN. Morgan Kaufmann (2014)
3. Jensen, C. T., Cline O., Owen M.: Combining Business Process Management and Enterprise Architecture for Better Business Outcomes. IBM Redbooks (2011)
4. Lankhorst, M.: Enterprise Architecture at Work: Modelling, Communication and Analysis. Springer (2017)
5. Pramudita, C. D.: The Balanced Scorecard as Strategic Controlling Instrument. Introducing the Indicators-based BSC for Implementation of a Corporate Strategy from Four Different Perspectives. Anchor Academic Publishing (2016)
6. Joyce, A., Paquin, R. L.: The triple layered business model canvas: A tool to design more sustainable business models. *Journal of cleaner production*, vol. 135, p. 1474-1486 (2016)
7. Swenson, K. et. al.: Business and Dynamic Change: The Arrival of Business Architecture. Future Strategies Inc. (2019)
8. COBIT 5: A Business Framework for the Governance and Management of Enterprise IT. ISACA (2012)
9. Mora, M., Gómez, J. M., Garrido, L., Pérez, F. C.: Engineering and Management of IT-based Service Systems: An Intelligent Decision-Making Support Systems Approach. Springer Science & Business Media (2013)
10. Chrissis, M. B., Konrad, M., Shrum, S.: CMMI for Development: Guidelines for Process Integration and Product Improvement. Pearson Education (2011)
11. Bankauskaite, J.: Comparative analysis of enterprise architecture frameworks. *CEUR Workshop Proceedings*, vol. 2470, p. 61-64 (2019)
12. Dijkman, R., Dumas, M., Van Dongen, B., Käärik, R., Mendling, J.: Similarity of business process models: Metrics and evaluation. *Information Systems*, vol. 36, no. 2, p. 498-516 (2011)
13. Sessions, R.: A comparison of the top four enterprise-architecture methodologies. Microsoft Developer Network Architecture Center, p. 1-31 (2007)

14. Zachman, J. A.: The framework for enterprise architecture: background, description and utility. Zachman International, p. 1-5 (1996)
15. Lee, J. Y., Ahn, K., Lee, S.: Applying OWL Ontology to Zachman Framework for Requirement Analysis. *Advanced Science and Technology Letters*, vol. 46, p. 34-37 (2014)
16. Lapalme, J., Gerber, A., Van der Merwe, A., Zachman, J., De Vries, M., Hinkelmann, K. Exploring the future of enterprise architecture: A Zachman perspective. *Computers in Industry*, vol. 79, p. 103-113 (2016)
17. Raj, P., Raman, A., Subramanian, H.: *Architectural Patterns: Uncover essential patterns in the most indispensable realm of enterprise architecture*. Packt Publishing (2017)
18. Vicente, M., Gama, N., da Silva, M. M.: Using ArchiMate and TOGAF to understand the enterprise architecture and ITIL relationship. *International Conference on Advanced Information Systems Engineering*. Springer, Berlin, Heidelberg, p. 134-145 (2013)
19. ArchiMate 3.1 Specification. Generic Metamodel. <https://pubs.opengroup.org/architecture/archimate3-doc/chap04.html>
20. Bonnet, F., Decker, G., Dugan, L., Kurz, M., Misiak, Z., Ringuette, S.: *Making BPMN a true lingua franca*. BPM Trends (2014)
21. Harmon, P.: *The State of Business Process Management 2016*. BPTrends (2016)
22. van den Berg, M., Solutions, R. I.: *ArchiMate, BPMN and UML: An approach to harmonizing the notations*. Orbus, software, white paper (2012)
23. Penicina, L.: Linking BPMN, ArchiMate, and BWV: Perfect match for complete and lawful business process models?. *CEUR Workshop Proceedings*, vol. 1023, p. 156-165 (2013)
24. Gill, A. Q., Qureshi, M. A.: Adaptive Enterprise Architecture Modelling. *JSW*, vol. 10, no. 5, p. 628-638 (2015)
25. Klimek, R., Szwed, P.: Verification of ArchiMate process specifications based on deductive temporal reasoning. *IEEE 2013 Federated Conference on Computer Science and Information Systems*, pp. 1109-1116 (2013)
26. Resource Description Framework (RDF). <https://www.w3.org/RDF/>
27. Apache Jena. https://jena.apache.org/getting_started/
28. Even, S.: *Graph Algorithms*. Cambridge University Press (2011)
29. BPMN Model API. <https://docs.camunda.org/manual/7.7/user-guide/model-api/bpmn-model-api/>
30. ArchiMate Model Exchange File Format. <http://www.opengroup.org/xsd/archimate/>
31. BPMN models. https://github.com/andriikopp/research-calculations/tree/master/research-calculations-java/processModelsStorage/Supply_process_BPMN
32. Lagerström, R., Baldwin, C., MacCormack, A., Dreyfus, D.: Visualizing and Measuring Enterprise Architecture: An Exploratory BioPharma Case. *IFIP Working Conference on The Practice of Enterprise Modeling*. Springer, Berlin, Heidelberg, p. 9-23 (2013)

Software for Simulation of Bio-Like Systems and Processes Using Movable Cellular Automata

Volodymyr Zhikharevich¹ [0000-0003-4882-2954], Kateryna Hazdiuk² [0000-0002-7568-4422] and Serhiy Ostapov³ [0000-0002-4139-4152]

¹⁻³Yuriy Fedkovych Chernivtsi National University, Kotsyubynsky str., 2,
Chernivtsi, 58012 Ukraine

¹vzhikhar81@gmail.com
²kateryna.gazdyik@gmail.com
³sergey.ostapov@gmail.com

Abstract. This paper deals with the development of a universal computer system for simulation biosimilar processes and structures. The main purpose of the system is to provide a wide range of users (scientists, engineers or just interested people) a convenient tool for their research. The simulating method of the proposed computer system is the method of movable cellular automata. An object of simulation, according to the method of movable cellular automata, is divided into small parts (automata), which interact with each other through the interaction's functions. The proposed system contains a library of elementary functions, the parameters of which can be customized according to the requirements of the model. The formation the combinations of elementary functions will allow researchers to visualize the model not only through a set of built-in interaction functions, but also to create their own interactions functions. The basic principles of these interactions are the same, although the examples can be diverse. It should be noted that the models are qualitative, not quantitative, and allow you to demonstrate the fundamental possibility of movable cellular automata for such simulation.

Keywords: Software, Movable cellular automata, Bio-like systems, Computer simulation.

1 Introduction

Simulation of biological processes and systems is one of the most relevant and promising areas of scientific research. The mechanisms of many bioprocesses are currently unknown. Therefore, one of the tasks, that is solved by simulation, is finding answers to questions that are open remain. Those include, in particular, the search for ways of self-organization and the evolution of living matter (Origin of Life).

Nowadays, simulation problems for biosystems are widely developed and require more complex calculations, so traditional approaches become less effective. New ways of improving classical direct imitation are also being explored in this field by

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

different approaches to the evolution and development of artificial organisms (indirect coding).

Many biological phenomena, such as intracellular localization of molecules, morphogenesis, growth, and dynamics of forest distribution, etc., are related to a spatial component that requires an understanding of spatiotemporal systems. Investigation of the localization and distribution of system components leads to the need to reproduce the shapes and geometry of the studied objects, as well as their possible deformation over time. In image systems, nature, forms, spatial distributions and their temporal dynamics are derived from images. This interprets the image not only as a visualization but also as a quantitative measurement and makes it a major source of data.

However, the use of images as a quantitative dimension raises several key problems. First, we need to investigate and take into account all prior knowledge and hypotheses about the system we are studying. Secondly, we need universal methods that can be applied to more than one specific situation. Finally, we need user-friendly and efficient software that facilitates the transfer of new methods to everyday scientific practice.

2 Previous investigation

A review of existing sources has identified a number of software systems that allow simulating of bio-like processes and structures.

One of these programs is COMPUCELL 3D [1, 2], the creation of which was started in 2004. This development allows us to simulate morphogenetic processes in different organisms, as well as to visualize the phenomena of cell growth, changes in the direction of distribution of chemicals, etc. CompuCell 3D is a three-dimensional modelling environment for complex biological processes that integrates several mathematical models. In CompuCell 3D, the space is discrete and is divided into identical cells that can be square or hexagonal in shape. Thus, the shape of the simulated objects (cells, tissues, or organs in general) is given by a set of such cells, and the processes occurring in these objects are modelled by differential equations. CompuCell3D has modelled such processes as tumour cell growth [3, 4, 5], bacterial colonies growth [6], development of blood vessels [7], and others. However, the COMPUCELL 3D morphogenesis model is limited by the fact that the shape of the modelled organism is a priori a continuous function over time.

Although CompuCell 3D provides the ability to test and experiment with biological models by providing a flexible package with many different levels of control, the user must have a broad range of bioinformatics knowledge to use this system. The disadvantages can be attributed to the poor visualization of the simulated systems. Additions and improvements to the system are still ongoing.

Other interesting computer simulation system (CSS) is Morpheus [8]. It is a simulation environment for studying high-scaled systems. Morpheus, developed for the study of cellular models under the influence of both internal and extracellular dynamics, which is described in a special declarative language, using the means of numerical

solution of ordinary differential equations, partial differential equations and cellular models of Potts [9], as in the previous CompuCell 3D system.

However, despite the wide range of settings and possible functionality, the construction and study of the dynamics of a multicellular organism is not foreseen.

Ascalaph is the next CSS that was considered [10]. It is a versatile program for molecular dynamics simulating. Under a single graphical environment are presented as authors' own implementations of molecular dynamics, and methods of classical and quantum mechanics of popular programs.

All user manipulations apply to molecular models. In this program, each model has a separate graphic application and file. A model is a set of atoms and their associations, such as bonds, molecules, and residues. Models may also include operating methods, such as simulating parameters, periodic boundary conditions, etc. Other methods may be common to all models, such as geometric measurements.

The disadvantages of this software are that its application to reflect the dynamics of biosimilar processes is quite limited, since the main area of modelling and related relationships are focused on chemical-physical processes.

The concept of a two-dimensional dynamic cellular automaton with a homogeneous structure is proposed in next reviewed study [11]. It extends the capabilities of two-dimensional cellular automata (CA) and allows to apply a CA approach to problems related to the dynamics of the structure of a simulated space in a two-dimensional case. In the process of the research the author has developed a software package that allows the user to program two-dimensional static CA models, two-dimensional dynamic CA models with a homogeneous structure and three-dimensional CA models with a dynamic inhomogeneous cell structure. The visualizer included with the OpenGL library allows you to visualize the simulation process in real time using a graphical accelerator, as well as to build two-dimensional graphs and histograms from the statistics obtained during the simulation process.

The user interface allows you to control the simulating process and set model parameters, most of which can be changed directly during the simulating process. The core of the software package is designed as an open source algorithms and class library that allows any user to add and modify the library code at their own discretion.

All models mentioned above have both advantages and disadvantages. Also, it should be noted that there are other software complexes for simulating of bio-like processes that achieve more precise characteristics, but are narrowly focused. The basis for their creation was a specific task, and therefore the result is aimed at specialists in a particular field (medicine, genetics, etc.). For example, in [12, 13], a 2-dimensional programmatic implementation of a tumour is used. Also, the development of Java 3D imaging of cancer is described in [14]. Similar software complexes also exist to simulate morphogenesis, embryogenesis, etc.

In contrast to the simulation systems mentioned above, this paper proposes the development of a universal computer system that would allow researchers to visualize the model not only through a set of built-in interaction functions, but also to create their own. The main purpose of the system is to provide a wide range of users (scientists, engineers or just interested people) a convenient tool for simulating biosimilar processes and structures. This tool should be simple and flexible when used. In this

regard, the proposed CAD system is based on the method of movable cellular automata (MCA). The core of the system implements an iterative algorithm for selecting automata, searching for the nearest neighbours and their interaction with each other, and the interface part provides visualization of the model and setting of input parameters (set of rules of interactions, properties of automata, etc.). Below we provide a description of the structure and features of the functioning of the CAD-system BLSD (Bio-Like Structure Design, Version 1.0).

3 BLSD CAD system architecture

The system has a modular architecture and is created by means of structured programming. The central part of the system is a procedure that provides the implementation of the algorithm of the method of movable cellular automata (Fig. 1). The method involves the operation of an iterative loop. It means the asynchronous choosing of the automaton from the set of automata, the search for the nearest neighbor and interaction with it. This may be the removal of the automaton or the addition of many new automata. In parallel, there is a graphical representation of a set of movable cellular automata. These two cycles are implemented as separate threads, i.e. the system in this respect has a multi-threaded architecture, which can significantly increase system performance in terms of multiprocessor hardware.

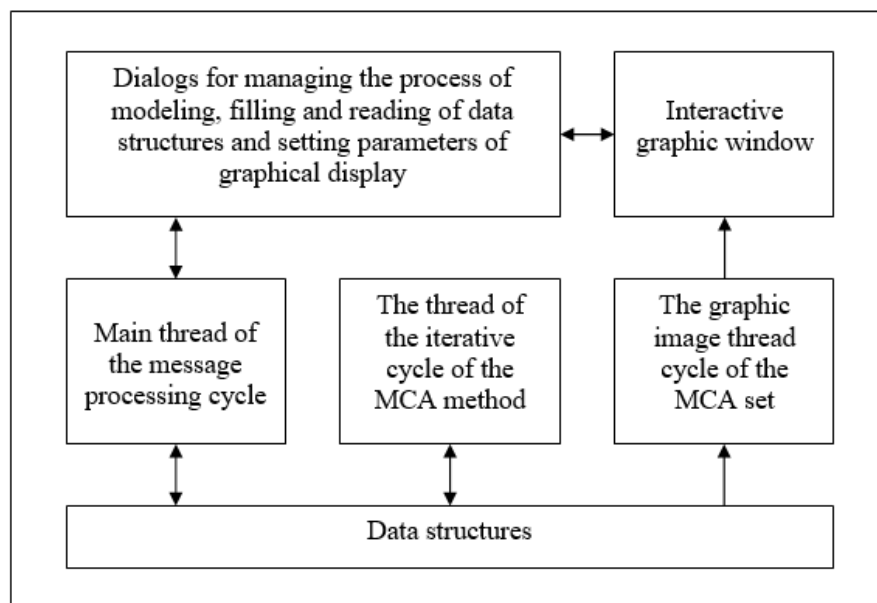


Fig. 1. BLSD CAD-system architecture

The main purpose of the interface part of the CAD system is to control the process of simulation, filling and reading of data structures of models and setting the param-

ters of graphical display. Graphical rendering is performed using a separate interactive graphical window using the OpenGL graphical pipeline functionality. Interactivity not only displays 3D models on the screen, but also manipulates them with a computer mouse, including: rotate, scale, add or remove individual elements, shift and link them. All system actions, like any standard Windows application, occur when processing the corresponding messages (Windows Messages), which is implemented in the main thread of the message processing cycle.

4 BLSD CAD system data structure

Data structures are implemented in the form of arrays, which store all the necessary information about the models and their parameters in the graphical window. These arrays are read or written to files with the extension "*.blsd".

All information about the set of movable cellular automata is stored in a two-dimensional array E (Elements), the structure of which is shown in Tab.1. The line numbers of the array correspond to the numbers (indices) of movable cellular automata. The number of columns is

$$1+3+N_{\max}+P_{\max}$$

The first column corresponds to the type of automaton, then the three columns contain the x -, y - and z -coordinates of the automaton in space. The following $N_{\max}$ columns contain indexes of adjacent automata. To determine the direct index of the neighbor, it is necessary to ignore (mask) the higher bit of the number obtained from the corresponding cell, and to determine the type of communication with the neighbor one must analyze this bit, since there are two types of neighborhood (unrelated or linked), and this sign is indicated in the higher bit of the neighbor index (0 or 1). Finally, the last columns of $P_{\max}$ contain the possible values of continuous parameters of cellular automata, for example: the sign of the active or passive state of the automaton, the value of some internal energy, the components of the momentum vector, etc.

Table 1. The structure of E array

Index of an element in an array	Type of cells	Coordinates			Neighbours			Parameters		
		x	y	z	N1	...	$N_{\max}$	P1	...	$P_{\max}$
1	S1	x_1	y_1	z_1	N11	...	$N_{1\max}$	P11	...	$P_{1\max}$
2	S2	x_2	y_2	z_2	N21	...	$N_{2\max}$	P21	...	$P_{2\max}$
...	...	...	...	...	...	...	...	...	...	...
m	S _m	x_m	y_m	z_m	N _{m1}	...	$N_{m\max}$	P _{m1}	...	$P_{m\max}$

In addition to the described data structure, which contains information about the set of movable cellular automata, the system uses structures to store information about the properties of the types of automata, as well as the functions of their interactions. All these data structures are implemented as corresponding arrays. And all arrays of cells of real type, except for the matrix of interactions containing indices of complex functions, for which there are enough cells of integer type. In addition, the names of automata types, their parameters, the names of elementary and complex functions, as well as other text information are contained in the corresponding arrays of cells of the string type.

The parameters of elementary functions that are part of complex functions are contained in the three-dimensional array FP (Function Parameters). Cells with zero addresses indicate the number of elementary functions that are included in the complex, as well as the indices of these elementary functions. A separate one-dimensional FS (Function String) array contains the names of complex functions. The two-dimensional FES (Function Elementary String) array contains the names of elementary functions (in null columns) and the names of their parameters. The interaction matrix is contained in a three-dimensional FM (Function Matrix) array, which has two layers – one (0) for the two-dimensional matrix of rules for interactions between unrelated neighbours and the other (1) for connected neighbours.

Properties of automata types are contained in the corresponding arrays, the cells of which are filled when working with the Properties of Elements dialog box. In particular, the upper part of the dialog handles elements of the two-dimensional array A (Attributes of Elements), and the lower part - elements of the three-dimensional array C (Characteristic of Continuous Parameters). The names of automata types and their parameters also include two separate arrays: one-dimensional AS (Attributes String) and two-dimensional CS (Characteristic String).

5 BLS D interface

The system has a multiple window interface. The main window (Fig. 2) is intended for manipulation of the model and consists of two parts: the field of control of the modes of operation of the program and the field of output of three-dimensional graphic image of the model.

The main menu contains options for invoking standard File Save or Open dialogs, Editing Rules for Interaction Rules and movable automata Properties (Edit), and a System Information (Info) dialog. When you save or open an arbitrary model in a file with the extension "*.blsd", it writes (or reads) all data structures containing information about the state of the set of movable cellular automata, as well as the rules of their interactions. Selecting the Edit menu shows the following dialogs:

- a window for editing properties of automata (Properties of Elements);
- a window for editing the DataBase of interaction rules (Database of Functions);
- window for editing Matrices of Functions.

In general, the system can operate in two modes: construction and simulation, which is determined by the state of the respective controls of the GUI system. The upper left part of the main window contains buttons for controlling the simulation process (start, pause, step, stop). A field of choice of the mode of construction of model elements (Elements), containing three items exist below:

- Adding / Deleting the automaton (Add / Del);
- Moving the automaton (Move);
- Tying up automata (Tie).

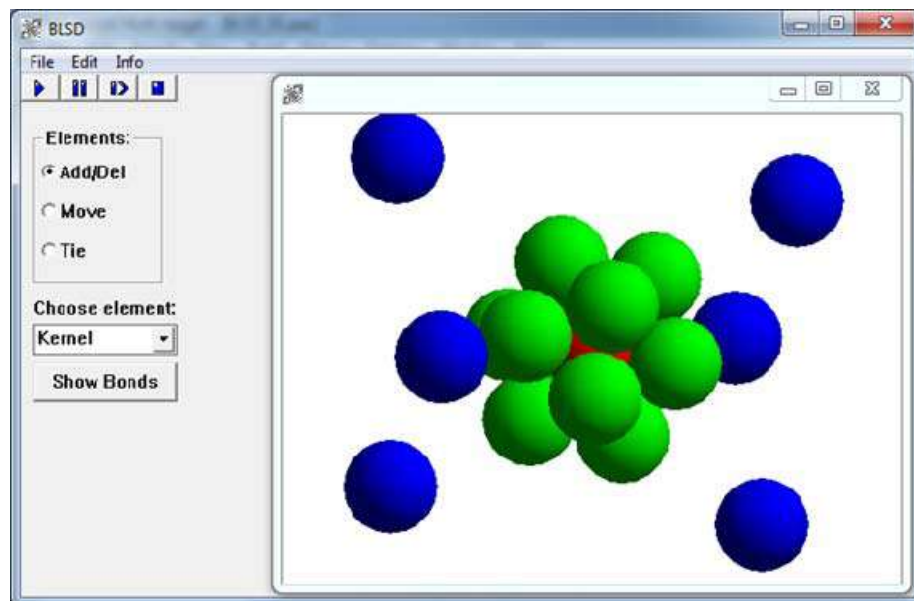


Fig. 2. BLSD main window

Selecting the first item in this field gives you the option to add or delete a movable automaton. Adding happens when you left-click. At the point corresponding to the position of the mouse cursor, a new movable automaton (with a zero-coordinate $z = 0$) will appear in the 3D model display window. The automaton type is being select in the appropriate Choose Element dropdown lists. If you select the second item of the field (Move), you can move an arbitrary automaton by left-clicking on it and moving the cursor without releasing it. Moving occurs only in the plane of the X and Y axes. If there is a need to move along the Z axis, then you should rotate the model 90 degrees relative to the centre, move the required automata and rotate the model back. The model rotates with the mouse so that it is in a transparent sphere that can be “touched” by the surface. In addition, you can scale the model image by rotating the Scroll Wheel and expand the model display window to full screen. Selecting the third item of the field (Tie) gives you the opportunity to connect automata which property assumes the presence of fixed neighbours. First of all, you need to set the connection

mode by clicking the appropriate Show Bonds button. In this mode, the lines will show all the connections between the movable automata that have fixed neighbours.

The properties of movable automata and the rules for their interaction can be edited in the dialogs that appear when you select the appropriate menu item Edit. Consider the window of editing properties of automata (Properties of Elements), shown in Fig.3.

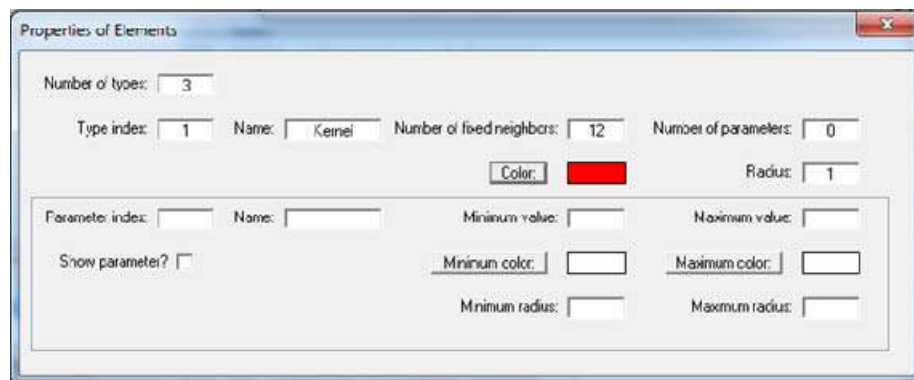


Fig. 3. BLSD Properties of Elements window

This window allows ones to specify the full number of types of cellular automata that can be used when the model is being implement. This number is entered in the "Number of types" field and determines the dimensions of the Matrices of Functions shown in Fig. 5. Selecting a specific type of automaton for editing its properties is made when its number is entered in the field "Type index".

When editing the properties of the automaton, it should be given a name in the "Name" field, specify the number of fixed neighbours (if any) in the field "Number of fixed neighbours", the number of continuous parameters of the automata (if any) in the field "Number of parameters", the radius of the displayed sphere in the "Radius" field in relative units. When you click the Colour: button, a standard colour selection dialog box appears, which allows you to select the colour of the displayed sphere that matches that type of machine.

Consider editing the DataBase of Functions shown in fig.4. This window allows you to form a complete set of complex interaction functions, the indices of which can be entered into the corresponding cells of the interaction matrices (Fig. 5).

This allows you to define the full range of rules for automated interactions. Complex interaction functions are a set (composition) of elementary rules. For a complete list of all elementary interaction rules, see the right in the Set of elementary functions window. These elementary rules implement the corresponding elementary iterative steps to modify the state of the automatons. For example, the Fluctuation rule realizes an elementary probabilistic shift of a movable automaton in an arbitrary direction, and the Attraction rule is an elementary step of displacement toward a selected neighbouring automaton.

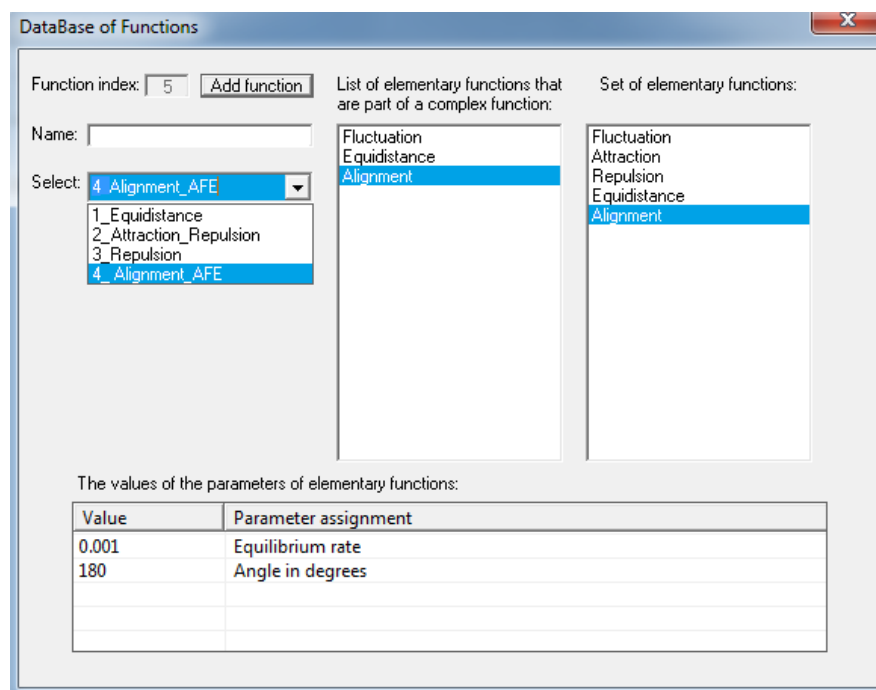


Fig. 4. Editing window of interaction functions DataBase

When you right-click on an elementary rule from the list of “Set of elementary functions”, an information box will appear with a description of this rule and its parameters (Fig. 5). The window contains a simple illustrative image, which may include an animated image obtained from the corresponding gif file.

The edit rule base editing window allows you to create complex rules that consist of a set of elementary rules. To do this, ones should enter the name of the complex rule in the "Name:" field and create a list of elementary functions that will be included in it. An elementary function is added to the List of elementary functions that are part of a complex function by double-clicking the left mouse button or pressing Enter on the selected function in the Set of elementary functions field. Deleting a function by right-clicking or deleting a selected function in the "List of elementary functions that are part of a complex function" field.

When you select an elemental function that is part of a complex function, a table is displayed at the bottom of the dialog box in the field "The values of the parameters of elementary functions:" and a description of the parameters included in this function. It is thus possible to change the values of these parameters by editing the corresponding cells in the table.

Then, when you click the "Add function" button, a newly created function is added to the DataBase of complex functions, which is automatically assigned a new index (Function index) and is assigned at the beginning of the function name. Further, if necessary, it is possible to edit existing rules in the DataBase. To do this, select the

appropriate function from the "Select" drop-down list, and then perform the steps described above. In particular, add or remove elementary functions and change the values of their parameters.

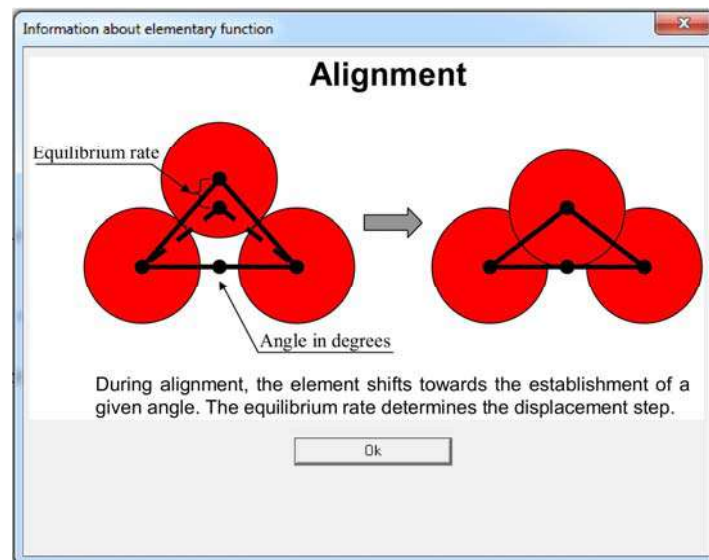


Fig. 5. Tooltip with information about elementary interaction function

Let us consider editing the Matrices of Functions shown in Fig. 6.

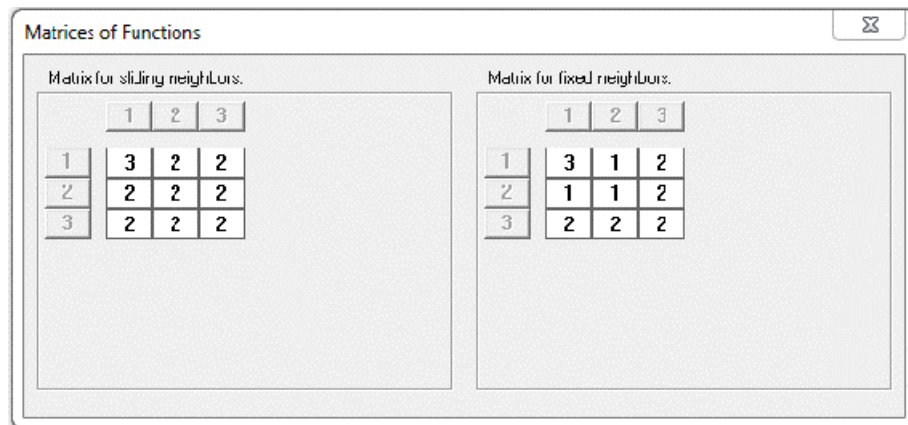


Fig. 6. Editing window of interaction rules matrices

This window allows you to define the rules of interaction between all possible combinations of types of two interacting automata. As mentioned above, the dimen-

sions of the matrices correspond to the number of automaton types in the simulated system. If the dimensions of the matrix are larger than 7×7 , which exceeds the size of the fixed field of the dialog window, the Scroll Bar buttons appear, which allow to move the visible fragment of the matrix. In the cells of the matrices, a Function index should be specified, which will be used for automata interactions of the corresponding types. Moreover, since there may be different rules of behaviour between unbound and connected automata of the same type, two matrices are also required: Matrix for sliding neighbours and Matrix for fixed neighbours.

It should be noted, that examples which demonstrate the possibilities of the system and proposed method was described in our previous investigations [15-17].

6 Conclusion and future research

In this paper the problem of the construction of a CAD-system for a simulation bio-like structures and processes using movable cellular automata was considered. The paper gave an overview of available computer systems deals with such problem, outlined the advantages and disadvantages each of them. The detailed description of main components of the BLSD system is provided.

The evolution of the proposed system is expected in two main areas:

- Development of a base of elementary rules and means of their editing;
- Enhancing interactive graphics and improving quality.

The current version of the system includes a primary rule base, which does not support flexible modification, like adding new features without having to recompile the system and other related inconveniences. In the future, it is planned to create the ability to download a library of elementary rules from the appropriate DLL (Dynamic Link Libraries), which will allow you to develop a DataBase of rules in any programming language.

The graphics part of the system will also be improved. In particular, it will be possible to choose the type of to display models, in addition to Orthographic projection also Perspective projection. You will also be able to choose from a variety of graphic effects, such as translucency, fog, blur (photographic depth effect) and more.

The interactive part, which allows you to manipulate the elements of the model, will be complemented by the possibility of group selection of elements and their simultaneous copying. This will greatly speed up the process of building structures that contain the same fragments.

References

1. Kumar, S., Das, A., & Sen, S. Multicompartment cell-based modeling of confined migration: regulation by cell intrinsic and extrinsic factors. *Molecular biology of the cell*, 29(13), pp. 1599–1610., (2018). DOI:10.1091/mbc. E17-05-0313
2. Swat, M. H., Belmonte, J. B., Heiland, R. W., Zaitlen, B. L., Glazier, J. A., Shirinifard A., CompuCell3D Reference Manual. Source page: www.compuCell3d.org/Manual/

3. Scianna, M., Preziosi, L., Preziosi L.; Ed. by M. Chaplain, J. Batzel, M. Bachar: New Challenges for Cancer Systems Biomedicine. Hybrid cellular Potts model for solid tumor growth. Springer, pp. 205–224, (2012).
4. Andasari, V., Roper, R. T., Swat, M. H., Chaplain, M. A.: Integrating Intracellular Dynamics Using CompuCell3D and Bionetsolver: Applications to Multiscale Modelling of Cancer Cell Growth and Invasion. PLoS ONE, Vol. 7, no. 3 (2012).
5. Gao, X., Tangney, M., Tabircam, S. A. Multiscale Model for Hypoxia-induced Avascular Tumor Growth. Proceedings of 2011 International Conference on Bioscience, Biochemistry and Bioinformatics Vol. 5, pp. 53–58 (2011).
6. Poplawski, N. J., Shirinifard, A., Swat M., Glazier, J. A.: Simulation of single-species bacterial-biofilm growth using the Glazier-Graner- Hogeweg model and the CompuCell3D modeling environment. Math Biosci Eng. Vol. 5, no. 2, pp. 355–388 (2008).
7. Kleinstreuer, N. A, Dix, D., Rountree, M., Baker, N., Sipes, N., Reif, D., Spencer, R., Knudsen, T.: Computational Model Predicting Disruption of Blood Vessel Development PLoS Comput Biol. Vol. 9, no. 4. 20 pp. (2013).
8. Source page: <http://morpheus.gitlab.io>
9. Wu, F.Y.: The Potts model. Reviews of Modern Physics. Vol. 54, Issue 1, p. 235. (1982).
10. Source page: <http://www.biomolecular-modeling.com/Ascalaph/Documents/TOC.html>
11. Vitvitskiy A.A.: Razrobotka i issledovanie metodov kletочно-avtomatnogo modelirovaniya rosta i deleniya kletok jiviyh organizmov [Development and research of cellular-automata modelling methods of growth and division of living organisms' cells]: diss. for the phd degrees of Phys.-Math. sciences: Speciality. 05.13.18, "Mathematical modelling, numerical methods and program complexes". Novosibirsk, (2015).
12. Sottoriva, A., Verhoeff, J.J., Borovski, T., McWeeney, S.K., Naumov, L., et al.: Cancer stem cell tumor model reveals invasive morphology and increased phenotypical heterogeneity. Cancer Res 70: pp. 46–56 (2010).
13. Anderson, A.R., Weaver, A.M., Cummings, P.T., Quaranta, V. Tumor morphology and phenotypic evolution driven by selective pressure from the microenvironment. Cell 127: pp. 905–915 (2006).
14. Chen, Y.J., Tabirca, S., Tangney M. Interactive 3D Graphics for Cancer Experiment Data Visualisation. In: IEEE eds. The 2010 International Conference on Bioinformatics and Biomedical Technology (ICBBT 2010): pp. 42–46 (2010).
15. Hazdyuk, K.P., Zhikharevich, V.V., Nikitina, A.M., Ostapov, S.E.: Application of movable cellular automata to modeling the locomotion of worm-like organisms: Mathematical and computer simulation. Series: Technical Sciences: Coll. of sciences. Kamianets-Podilskyi: Kamianets-Podilsk. nats. un-t: pp. 16–25 (2018) (In Ukrainian).
16. Hazdyuk, E.P., Zhikharevich, V.V., Nikitina, A.M., Ostapov, S.E.: The unicellular micro-organisms "Amoeba Proteus" locomotion simulation with the use of movable cellular automata method: Applied Discrete Mathematics, № 42: pp. 104–119 (2018). DOI: 10.17223/20710410/42/8
17. Zhikharevich, V., Hazdiuk, K. and Ostapov, S.: Simulation of bio-like systems and processes using movable cellular automata: [Electronic resource]. CEUR Workshop Proceedings, Vol. 2353: pp. 664-673 (2019). Access mode: <http://ceur-ws.org/Vol-2353/paper63.pdf>

An Integrated Approach to Improve Effectiveness of Industrial Multi-factor Statistical Investigations

Victoria Miroshnichenko¹[0000-0002-5956-7867], Alexander Simkin²[0000-0002-9939-7866]

¹Priazovsky State Technical University, University str., 7, 87555, Mariupol, Ukraine
miroviktoria@gmail.com

²Priazovsky State Technical University, University str., 7, 87555, Mariupol, Ukraine
simkin@ukr.net

Abstract. An approach was developed for computer statistical analysis of big, multi-dimensional arrays of technology parameters and industrial product quality indexes. It provides fully objective, mathematically comprehensive, scientifically grounded and physically interpretable description of the manufacturing factor effects on the performance of an industrial product. The approach integrates a basic Data Mining exploratory technique, multiple regression models construction and Monte-Carlo simulations. The approach was applied to industrial statistical arrays investigations for the ASTM A514 steel. The results obtained are in a good accordance with the known Material Science data and were confirmed in industry

Keywords: multi-dimensional data; exploratory technique; multiple regression models; Monte-Carlo simulations.

1 Introduction

One of the current trends of the modern stage of the Industry 4.0 development is to improve the big manufacturing technology data analysis techniques for increasing effectiveness of the Industry 4.0 platform components [1, 2]. The basic finishing goals of the components are, as it known, to increase manufacturing productivity, improve quality and reliability of a product by eliminating employed technologies lacks with minimal expenses. Cardinal role in the situation is played by the industrial computer statistical investigations because of: their high potentials in treating multifactor industrial phenomena; principal low effectiveness of laboratory researches, not enabling to simulate exactly the real industrial manufacturing environment; practical impossibility to conduct the real, in depth industrial experiments in an operating plant conditions. Nevertheless, the statistical techniques currently applied in the industry are not enough effective in meeting actual practical and theoretical challenges. Typically as-recorded, raw industrial data require the preliminary treatment. The most widely currently used relevant tool in the case is the Data Mining technology, which is the collection of several computer aided statistical techniques [3-5]. Among the techniques

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

the most effective ones today are the artificial neural networks (ANN) [6, 7] and classification and regression trees (C&RT) [8, 9].

2 Brief literature overview

Both ANN and C&RT techniques have been effectively applied to solving a number of industrial technology improvement and product quality problems [10-13], particularly in the fields of Material Engineering. The obtained results are in contrast [14, 16] to those ones provided by direct use of the traditional statistical analysis techniques: multiple regression models, MANOVA, ANOVA etc., under the same conditions. Nevertheless, both ANN and C&RT procedures as such have some lacks, considerably decreasing their application effectiveness. Particularly, ANN is not capable of to express a regression dependence revealed in the conventional visual, mathematic and physically interpretable forms, while C&RT does not provide the discovered visual relations in a quantitative form of a regression equation. Such features are typical for the most of the statistical techniques that restricts the ability of the modern computer modeling and simulation technologies to be effectively applied in industrial practice and fundamental multifactor phenomena researches. The features may also be considered as probable reasons of low effectiveness of the modern statistical technologies in revealing the factors which determine spreading the last decade's epidemic deceases.

Besides, such widely used statistical analysis tool as the multiple regression modeling under the sole, direct application for treating multi-dimensional data arrays is also extremely ineffective due to typical simultaneous changes of numerous input variables that makes it impossible to reveal any regression dependence [13-15] within the arrays. In addition, a powerful tool of modern computer technologies known as Monte-Carlo simulations or computer experiments is not practically involved now in statistical investigations contrary to other areas of scientific researches [16].

Aim of the paper is to outline the main features of a developed integrated approach to industrial statistical investigations together with basic results of its application for the case of ASTM A514 steel.

The steel is one of a modern mass produced multi-alloy steels, characterized by extreme performance instabilities due to its complex alloying and heat treating technologies interactions.

The used approach is aimed to provide in the real industrial environment:

- revealing the statistically valuable industrial technology factors effecting each performance index of a product or process considered;
- on-line, semi-quantitative characterizing the factors separate and collective effects to solve possible actual technology problems;
- corresponding adequate regression models specifying;
- comprehensive on-line computer control of an industrial technology process;
- off-line computer investigations of the collective and separate effects for the revealed industrial factors with possible novel synergetic phenomena discovering;

- specifying the fields of possible further industrial technology improvements, deep specific laboratory applied or fundamental researches .

3 Methodologies

As input data arrays for the statistical investigations performed in the work the industrially obtained results of the quality inspection tests for thick sheets made of the ASTM A514 steel were used which specify the steel chemical element concentrations and the sheet standard mechanical properties indexes.

As the components of the approach the following techniques were consistently used:

- C&RT procedure resulting in the dendrogram building for each product performance index depicting the responsible technology factors and their effects with the 98 % confidence probability;
- based on the revealed variables construction of the multiple regression model for every control (dependent) characteristic;
- computer experiment workability verification for each built regression model with the possible model coefficients correction to achieve the highest adequacy;
- MC simulations of the traditional pair regression scatter plots obtained under the simultaneous change conditions for all responsible factors, which are typically built in the course of the conventional industrial quality analysis using unsorted, raw experimental data, as an additional workability verification tool for each regression model;
- MC simulations of the separate effects for each regressor under the constant values of the rest ones in a regression model, for the research purposes of the manufacturing technologies improvements.
- Multi-purpose optimization of the revealed technology parameters with applying a MC extremum searching technique.

4 Results and Discussion

Some results obtained by the above approach industrial application are considered below. The final goal of the conducted research was to specify chemical composition and heat treating technology parameters values providing for ASTM A514 steel the standard mechanical properties combination which exceeds the technical requirements with 98% confidence probability. Actuality of the researches is caused by the extreme performance instabilities for the thick sheets made of the Boron containing steel due to its complex alloying and heat treating technologies interactions.

According to the methodology proposed, the first step of the investigations was C&RT analysis, resulting in the dendrograms building for each steel performance index. As it follows from the dendrogram for the steel yield stress shown on Fig. 1, the following statistically valuable technology factors effect the static steel strength:

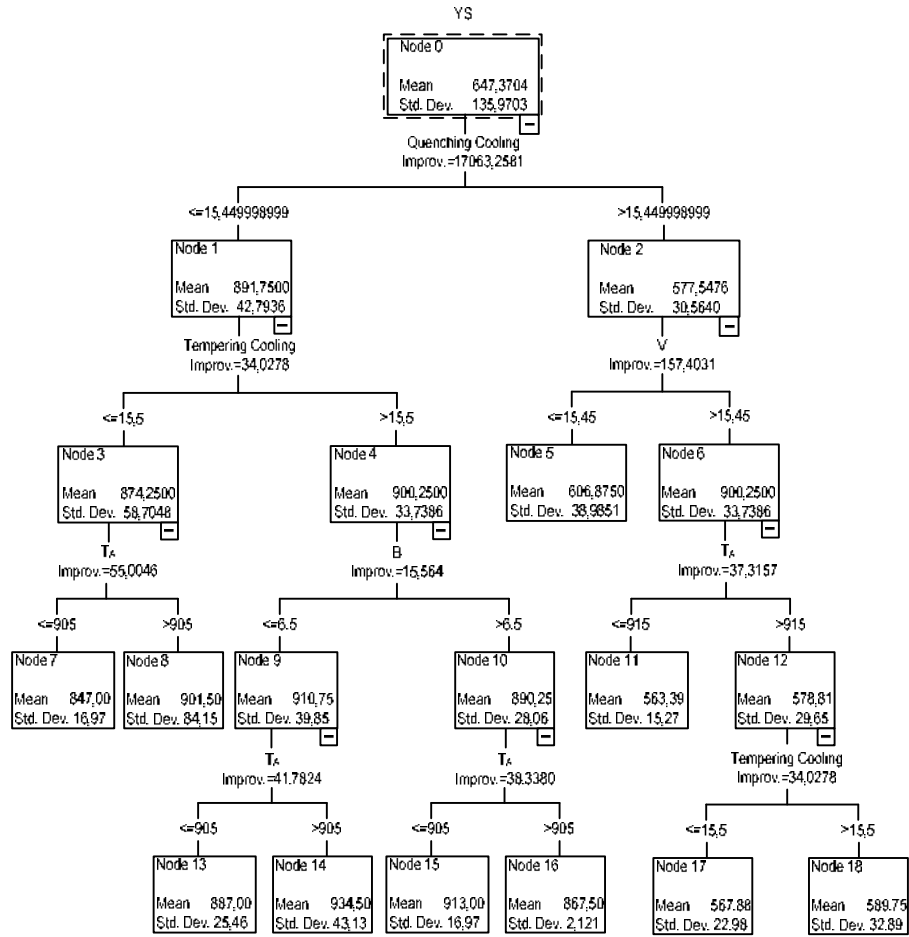


Fig. 1. C&RT dendrogram showing the statistically valuable industrial factors and their effects on yield stress of ASTM A514 steel in thermally improved state

cooling duration at the steel quenching t_{cool}^Q and tempering t_{cool}^{temp} , holding temperatures at austenitizing T_A and tempering T_{temp} together with the following chemical elements concentrations: V and B. Semi-quantitative characterization the factors separate and collective effects on the static steel strength may be also visually obtained from the dendrogram.

Further step of the approach was to elaborate the multiple regression models, based on the corresponding dendrogram related data. The model developed for the steel yield stress based on the dendrogram shown on Fig. 1 is as follows:

$$YS = 400 + 63(B - 4) + 15(V - 35) + 0.03T_A \left\{ 1 + 25 \cdot 10^{-5} \cdot T_{temp} \cdot t_{cool}^Q \cdot [14 - (0.12V + B)] \right\} \quad (1)$$

where $B = \%B \cdot 10^5$ and $V = \%V \cdot 10^3$;

$\%B$ – boron concentration, wt. %;

$\%V$ – vanadium concentration, wt. %.

The necessary in such a case adequacy verification for the obtained regression models was conducted by the use of MC technique to build the frequency distributions for each performance index considered in the current investigation.

Taking into account the statistical comprehensiveness of such a mathematic description of a measured quantity, the procedure employed may be considered as a regression model workability verification. The MC simulated frequency distributions

built using the constructed regression models were compared with the real experimental ones. The simulation results for the considered above performance index obtained using the corresponding regression model 1 are shown on Fig. 2 as the frequency distribution line. As it seen, good correspondence of the simulated (line) and real experiment (histogram) distributions had been reached.

An important formal advanced feature of the regression models like the shown

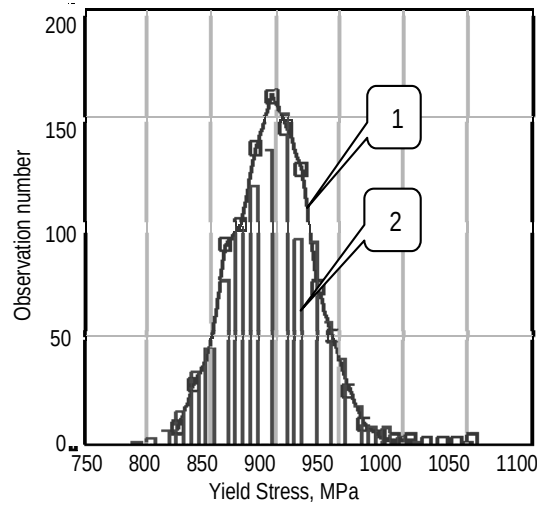


Fig. 2. Frequency distributions for yield stress of ASTM A514 steel in thermally improved state according to results of MC simulations (1) and real experiments (2)

above one should be outlined. Namely, the obtained models provide high workability in the performance descriptions by taking into account only real values of the technology parameters and their multiplications without using the terms of two or higher power. It allows to propose a real, physically grounded interpretation of such equations in terms of the industrial factors separate effects and their interactions. In turn, such conclusions may be further used for the corresponding phenomena mechanisms investigations. Workability of the models was also verified by the MC simulations of two dimensional scatter plots corresponding to pair regression relations of a performance index vs. an industrial factor, under the conditions of all the factors simultaneous variations. Such effect of the factors should evidently be considered as collective one caused by interactions of all the factors simultaneously changed.

Some examples of the MC simulated and real experimental scatter plots for the steel currently studied are shown on Fig. 3, 4. As it seen, good agreement of the simulated and experimental data is provided that is an additional confirmation of the regression model high workability.

An important role in control of multi-factor phenomena and systems of complex physical-social-economical nature such as industrial technological processes, finished products etc. is played by specifying the sole effects of each valuable manufacturing factor on the performance indexes. As a rule, such information is unavailable under

real industrial environment due to simultaneous variations of numerous effecting manufacturing factors. Such research methodology restrictions can be avoided by analyzing currently available industrial data using the considered approach. An example of the revealed separate effects of the valuable manufacturing factors on the yield stress of ASTM A514 steel is shown on Fig. 5. The corresponding shown regression dependences were simulated for each valuable manufacturing factor which varies under some different constant values of the rest variables. These constant values for each accompanying variable were chosen randomly within the intervals of its possible

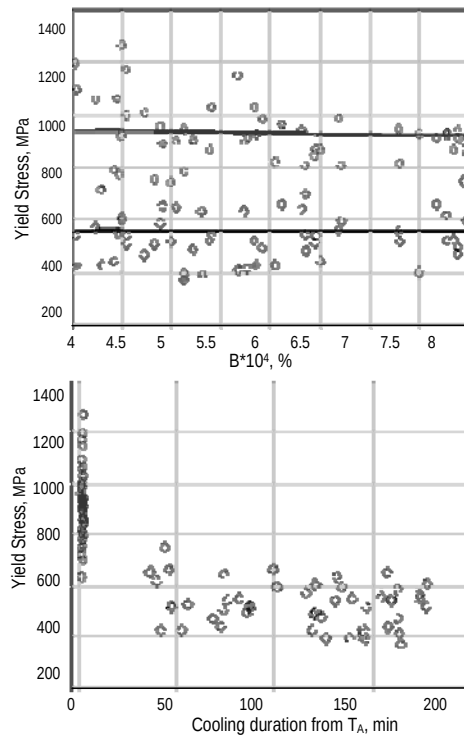


Fig. 3. MC simulation results for combined effects of some industrial factors on yield stress for ASTM A514 steel

variation in the steel.

As it seen, varying the values of accompanying variables considerably influencing the steel yield stress levels or even the general character of its dependence from a considered factor. Particularly, the boron dependence of the steel yield stress changes from decreasing to increasing type under simultaneous transition to the parameters values: $> \sim 0.5 \% V$; cooling duration from austenite temperature $< \sim 45$ min; cooling duration from subcritical temperatures $\neq \sim 10$ min; austenitizing temperature $< 900^{\circ}C$ or $> 930^{\circ}C$. It should be additionally noted that the above results are in a good accordance with the known specific Material Science data concerned with the considered

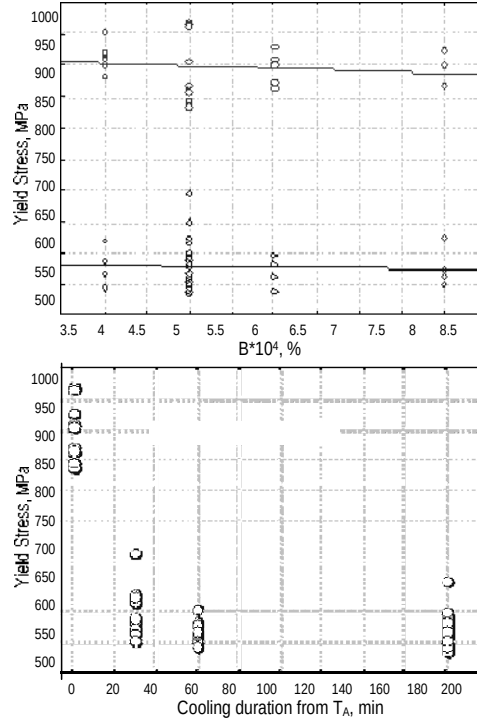


Fig. 4. Results of real experiments for combined effects of some industrial factors on yield stress for ASTM A514 steel

factors effects on structure and properties of corresponding steels and allow to explain the discrepancies often observed for boron containing steels in the literature.

Based on the results obtained using the applied integrated approach, some predictions were made aimed to improve of the industrial manufacturing technologies for the steel, particularly, its chemical composition and heat treatment technology.

The technology parameters thus predicted provide guaranteed exceeding the technical requirements to the steel performance indexes with 98% confidence probability. The adequacy of the corresponding technology recommendations was verified in real industrial conditions: the following combination of the performance indexes for thick sheets made of the researched steel was provided: $YS = 950 \pm 40$ MPa, $\psi = 56.5 \pm 4\%$, $KV = 44 \pm 2$ J. It should be noted the considerably low standard deviations for the just given standard mechanical properties characteristics, that is in violent contrast with the previously obtained industrial data, particularly shown in Fig. 1 and Fig. 2. So, the results of the developed integrated statistical investigation approach employment show considerable improvement of the finished industrial product reliability comparing with the same product made in traditional industrial conditions.

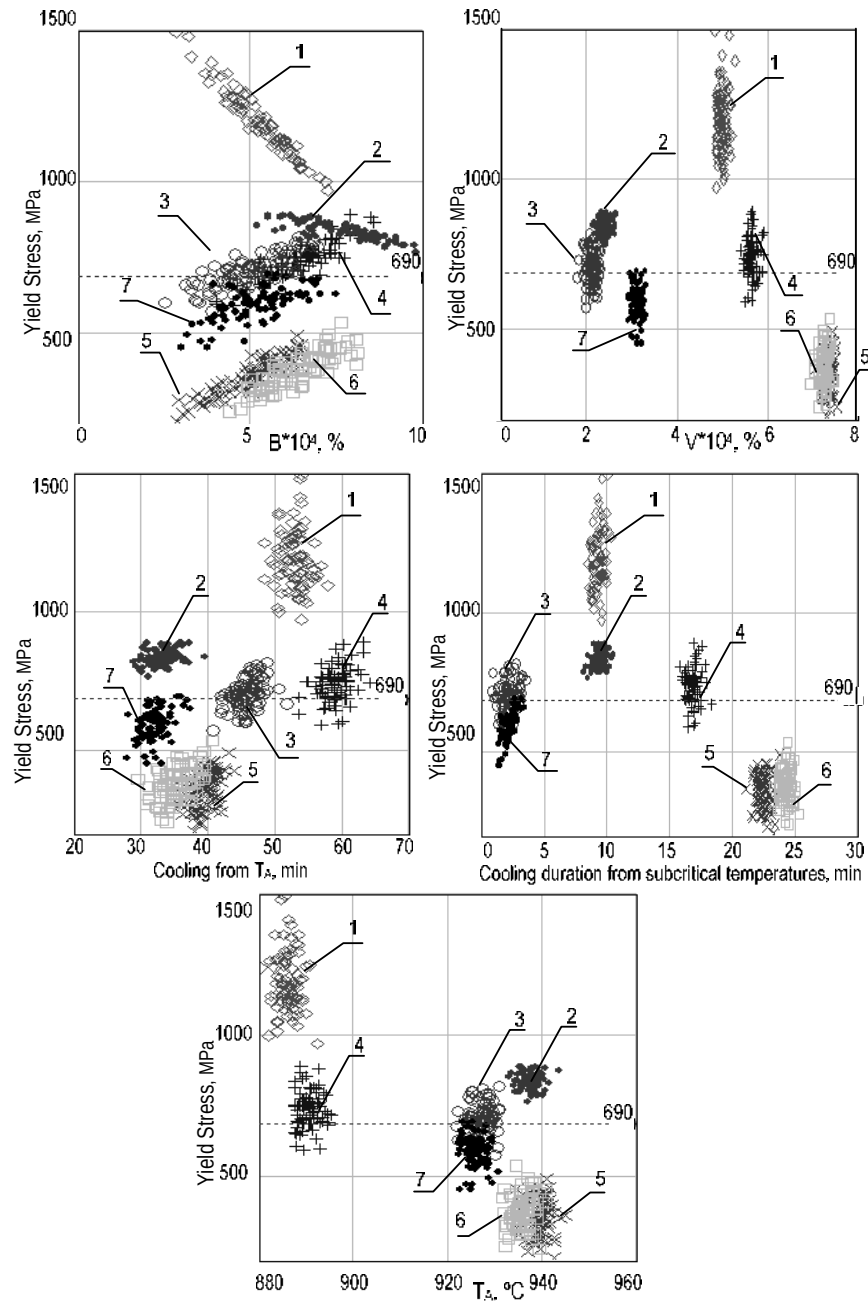


Fig. 5. Computer simulated scatter plots showing separate effects of the revealed valuable manufacturing factors on yield stress of ASTM A514 steel. Numerals on the plots correspond consecutive numbers of randomly chosen combinations of the variables values used in Eq(1)

5 Conclusions

1. In view of the Industry 4.0 needs, an approach to the industrial statistical investigations was developed aimed to improve effectiveness of big data, multi-dimensional array analysis and its results practical applications.
2. The developed approach integrates: C&RT technique, as a Data Mining procedure allowing the obtained results further physical interpretation and mathematical treatment; multiple regression models building to express the revealed draft regularities in a rigorous mathematical form; Monte-Carlo simulations to verify the regression models, to conduct computer investigations and outgoing product quality index predictions, to specify further research areas.
3. The approach developed was applied to solving some industrial quality and reliability problems for thick sheets made of boron-containing ASTM A514 steel concerned with its typical low and unstable yield stress and impact resistance on the levels: $YS = 900 \pm 100 \text{ MPa}$, $KV = 35 \pm 12 \text{ J}$.
4. The main performance indexes values obtained in industry for the steel with the confidence probability 98%, as a result of the approach application, are as follows: $YS = 950 \pm 40 \text{ MPa}$, $KV = 44 \pm 2 \text{ J}$.
5. As a result of the approach application the following technology advantages have been reached providing the guaranteed finished industrial product performance improvements:
 - specification of industrial technology factors having valuable effects on the product quality and reliability;
 - development of regression models providing the statistically comprehensive description of the revealed effects;
 - determination of separate effects for each of the revealed factors and conditions of the effects realization;
 - determination of the multipurposely optimized industrial technology parameters providing increase and stabilizing a combination of the finished product quality indexes.
6. In view of the demonstrated high effectiveness of the developed approach in solving the task having been considered together with generality of its background principles, successful application of the approach to solving analogous multi-factor problems in various technical and social environment should be expected.

References

1. Erboz, G.: How To Define Industry 4.0: Main Pillars Of Industry 4.0. In: Proceedings of 7th International Conference on Management (ICoM), pp. 245 – 249 (2017)
2. Hermann, M., Pentek, T., Otto, B.: Design Principles for Industry 4.0 Scenarios. Proceedings of 49th Hawaii International Conference on System Sciences (HICSS), pp. 3928-3937 (2016) doi: 10.1109/HICSS.2016.488

3. Lee, J., Bagheri, B., Kao, H.: Recent Advances and Trends of Cyber-Physical Systems and Big Data Analytics in Industrial Informatics. Proceeding of Int. Conference on Industrial Informatics (INDIN) (2014) doi: 10.13140/2.1.1464.1920.
4. Lee, J., Lapira, E., Bagheri, B., Kao, H.: Recent advances and trends in predictive manufacturing systems in big data environment. Manufacturing Letters. 1: pp. 38-41 (2013) doi: 10.1016/j.mfglet.2013.09.005
5. Witten, I., Frank, E., Hall, M.: Data Mining: Practical Machine Learning Tools and Techniques. 3rd Edition, Morgan Kaufmann, Massachusetts (2011)
6. Bethge, M., Ecker, A., Gatys, L.: A Neural Algorithm of Artistic Style. (2015) arXiv: 1508.06576
7. Ojha, V., Abraham, A., Snášel, V.: Metaheuristic design of feedforward neural networks: A review of two decades of research. Engineering Applications of Artificial Intelligence 60: pp. 97–116 (2017) doi: 10.1016/j.engappai.2017.01.013
8. Wang, F., Rudin, C.: Falling Rule Lists. Artificial Intelligence and Statistics. 1013-1022 (2015)
9. Gareth, J., Witten, D., Hastie, T., Tibshirani, R.: An Introduction to Statistical Learning. Springer, New York (2015)
10. Suzuki, K.: Artificial Neural Networks: Industrial and Control Engineering Applications. BoD – Books on Demand (2011)
11. Bisi, M., Goyal, N. Artificial Neural Network Applications for Software Reliability Prediction. Wiley – Scrivener Publishing (2017)
12. Guo, Z., Malinov, S., Sha, W.: Modeling beta transus temperature of titanium alloys using artificial neural network. Computation Material Science, 32(1): 1-12 (2005) doi: 10.1016/j.commatsci.2004.05.004
13. Tkachenko, I. Machine parts service efficiency forecasting based on the multipurpose optimization of material performance [Прогнозирование эксплуатационной эффективности деталей на основе многоцелевой оптимизации свойств материалов] Metal and casting of Ukraine. 7-8: 72-77 (2005)
14. McBride, J., Malinov, S., Shaa, W.: Modeling tensile properties of gamma-based titanium aluminides using artificial neural network. Materials Science and Engineering. 384(1): 129-137 (2004) doi: 10.1016/j.msea.2004.05.072
15. Tkachenko, I.: Multi-purpose optimization of heat strengthening technology for thick sheets made of high strength weldable steels using the Data Mining computer technology [Многоцелевая оптимизация технологии термического упрочнения проката высокопрочных свариваемых сталей с использованием компьютерной технологии "Data Mining"] Reporter of Priazovskyi State Technical University: Collection of scientific. Vol. 14: pp. 111-117 (2004)
16. Landau D., Binder K.: A Guide to Monte-Carlo Simulations in Statistical Physics. Cambridge University Press, Cambridge (2000)

Information Technology in the Modeling of Dry Gas Seal for Centrifugal Compressors

Lyudmyla Rozova¹[0000-0002-0781-7473] and Gennadii Martynenko²[0000-0001-5309-3608]

National Technical University “Kharkiv Polytechnic Institute”, Department of Dynamics and Strength of Machines, NTU “KhPI”, 2, Kyrpychova str., 61002, Kharkiv, Ukraine
¹ luda.rozova@gmail.com, ² gmartynenko@ukr.net

Abstract. The developed specialized software package for computer modeling of dry gas seals for the rotors of centrifugal compressors is presented. This software package implements the mathematical model that describes improved methodology for seals modeling and solving under steady-state compressor operating conditions. It allows to make a model of working gap between the seal rings, which changes under the influence of gas pressure and temperature. This improved computer modeling can provide the reliable operation of dry gas seals. The developed software package has been tested on real constructions of dry gas seals with different types of grooves. The results of test solutions were compared with available in literature and experimental studies.

Keywords: Computer Simulation, Engineering Software, Compressor Gas Seals

1 Introduction

The issues of computer modeling and information technology are very relevant in the study of engineering objects. Increasingly, they are the basis of a computer experiment, which in most cases can be used to replace the implementation of expensive real experimental studies during finishing work and in the process of servicing working structures. Computer technology capabilities allow for rather accurate modeling and numerical research of engineering structures, using existing technical software systems for general engineering solutions. Particular difficulty in this case is the need to solve problems from different areas of physics within the solution of general technical problem [1].

Another way is to create specialized software products using different programming languages and environments. This allows to use more accurate mathematical models for analysis, taking into account processes and phenomena of different physical nature. The use of such information technology for modeling improves quality of design work when creating prototypes and industrial designs [2].

The need to create a specialized software package also arose for the modeling of dry gas seals for the rotors of centrifugal compressors, designed to maintain gas pressure in main gas pipelines [3]. Such compressor are shown in Figure 1.

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

This type of seals is shown in Figure 2. It has been widely applied in view of a number of advantages. The main ones are: the possibility of operation at high rotation speeds for compressor rotor; the absence of an expensive oil supply system; minimal gas leakage during operation [4, 5].

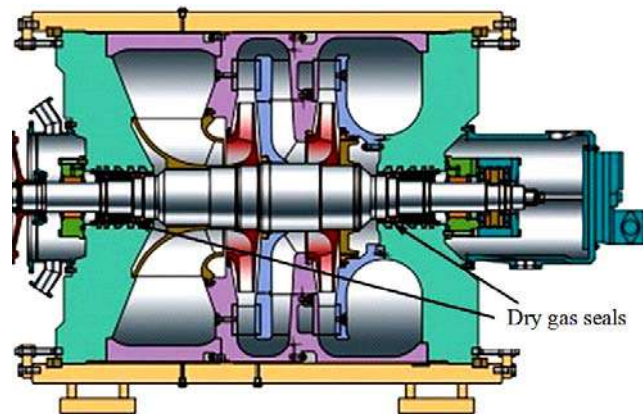


Fig. 1. Centrifugal compressor

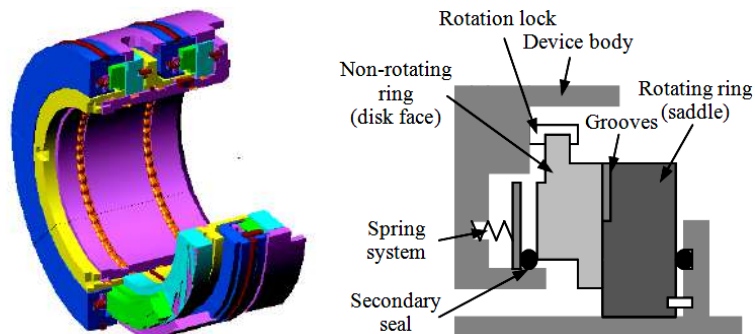


Fig. 2. Dry gas seals

Dry gas seals can work stably and reliably with various gases and under various operation conditions only in the case of it very accurate modeling and solving. An improved methodology of gas seals analysis under steady-state operation conditions should include solving of gasdynamics, heat transfer and thermoelasticity problems [5].

The seals system is one of the most critical centrifugal compressor systems. That is why the presence of improved methodology for gas seals modeling and software product that implements this methodology makes a huge contribution to the reliable operation of the system as a whole.

This paper is devoted to the issues of using information technology for computer modeling of processes that take place during the operation of dry gas seals and the creation of specialized software package for modeling and solving.

2 Formal problem statement

The aim of this work is to present the developed specialized software package for computer modeling and solving of dry gas seals for the rotors of centrifugal compressors. This software package implements the mathematical model that describes improved methodology for seals modeling under steady-state compressor operating conditions.

3 Literature review

An analysis of the existing researches, dedicated to the modeling of gas face seals, has shown the relevance of this problem. Despite the large number of publications on this topic, many issues of modeling and solving remain open.

The main operation condition of dry gas seals is the existence of the necessary gap between the seal rings with a thickness of 3-4 μm (micrometer) in steady state operations, supported by the gas pressure (Fig. 2). The necessary gas pressure between the rings in working condition is created and maintained due to the special type of micro-grooves on the rotating seal ring, with a depth of 6-7 μm .

The type of grooves is selected, taking into account the operation features of seals. More often spiral grooves are used. Considering such small gaps between the sealing rings and the shallow depth of the grooves, much attention is paid to computer modeling and solving of gas seals.

Nowadays, the existing universal engineering software packages based on the finite element method are widely used for modeling and solving of dry gas seals. So, in studies [6, 7, 8, 9, 10], well-known general technical software packages are used to analyze the gas pressure between the rings. These works are devoted to the selection of the optimal type of microgrooves for direct and reverse seal operation, by calculating the gas pressure in the gas layer between the rings. The influence of the surface treatment of the rings on the working pressure of the gas in the gap is studied [11].

Existing software packages are also used for modeling and visualization of calculation results.

However, it is necessary to take into account the temperature distribution in the seal rings and the appearance of temperature strains in rings.

Taking into account these strains is necessary in view of small gap between the rings. Temperature strains can lead to redistribution of gas pressure between the seals rings and incorrect seals operation. The literature review on solving of heat-transfer and thermoelasticity tasks for dry gas seals showed the existence of a few number of papers devoted to this problem [12, 13, 14].

So, in studies [12, 14], the temperature distribution in the rings is taken into account. The appearance of temperature strains is taken into account in [13]. These problems are solved by using of general technical software packages. Either an alternative to this are developed programs, which are based on simplified engineering solution methods. The above tasks are solved separately, because of their rather difficult simultaneous analysis.

Existing software packages allow to create accurate model research object, visualize the results of the solution. However, the existing general technical software systems are designed to solve a wide class of problems and they, of course, cannot fully take into account the aspects of a highly specialized task in solving a specific problem for a particular object. The presence of a specialized program allows more accurately and adequately simulate processes, and take into account the relationship between phenomena of different physical nature.

Therefore, the creation of specialized software package for computer modeling and solving of dry gas seals is a relevant problem today.

4 Specialized software package GasDin

In this paper, the developed specialized software package GasDin is presented that implements an improved methodology of modeling of dry gas seals for the rotors of centrifugal compressors. This methodology is based on iterative solution of related gasdynamic, heat transfer and thermoelasticity tasks.

4.1 Theoretical basis

The two-dimensional distribution of gas pressure in the gap between the seal rings is described by the nonlinear gas lubrication equation. After some conversions, taking into account the temperature changes, this equation has the form [15]:

$$\frac{\partial}{\partial x} \left[\frac{1}{\mu T_{av}} \left(h^3 \frac{\partial p}{\partial x} - 12\mu h \omega \sqrt{p} z \right) \right] + \frac{\partial}{\partial z} \left[\frac{1}{\mu T_{av}} \left(h^3 \frac{\partial p}{\partial z} + 12\mu h \omega \sqrt{p} x \right) \right] = 0, \quad (1)$$

where μ – dynamic coefficient of viscosity for gas; $T_{av} = T_{av}(x, z)$ – an average integral function of changing the gas temperature over the gap thickness (along the y coordinate); h – the thickness of the gas layer; $p = P^2$ – is the square of the gas pressure; ω – the angular velocity of rotating ring.

The temperature distribution in the gas film between the rings is described by the heat equation, taking into account the heat generation in the gas layer due to viscosity and convective heat transfer [15]:

$$\frac{\partial}{\partial x} \left[k_T \frac{\partial T}{\partial x} \right] + \frac{\partial}{\partial y} \left[k_T \frac{\partial T}{\partial y} \right] + \frac{\partial}{\partial z} \left[k_T \frac{\partial T}{\partial z} \right] + \mu \left[\left(\frac{\partial v_x}{\partial y} \right)^2 + \left(\frac{\partial v_z}{\partial y} \right)^2 \right] - \frac{PC_v}{RT} \left(v_x \frac{\partial T}{\partial x} + v_z \frac{\partial T}{\partial z} \right) = 0, \quad (2)$$

where k_T – thermal conductivity of the gas film; $T = T(x, y, z)$ – gas temperature; v_x and v_z – gas velocity components along the gap; C_v – specific gas heat intensity ratio at a constant volume; R – universal gas constant.

The boundary conditions for equations (1), (2) are the set values of temperature and pressure at the inlet and outlet of the seal.

The temperature distribution in working rings is:

$$\nabla(k_{T1,2} \nabla T) = 0, \quad (3)$$

where k_{T1} and k_{T2} – thermal conductivities of rotating and axially movable seal rings.

4.2 Computer modeling methodology

An algorithm based on the use of the Bubnov-Galerkin method in combination with the finite element method has been developed to solve the gasdynamic task for dry gas seals. The finite element method in variational formulation is used for solving the heat transfer and thermoelasticity tasks. Because of non-linearity of formulated tasks, an iterative algorithm of their simultaneous solution has been developed.

The conversion of cumbersome analytical expressions for resolving system of non-linear equations, which formed the basis of the GasDin software package, were made using package of computer mathematics. These expressions were automatically translated into the codes of C++ programming language. Thus, it was possible to avoid numerous mistakes. Analytical expressions were also used in testing to compare the results of calculations.

This software package has been developed by use of C++ programming language.

The main window of the developed software package is shown in Figure 3.

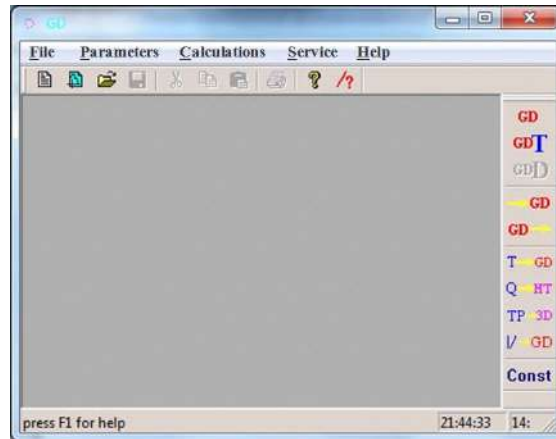


Fig. 3. The main window of the software package GasDin

The main requirement for the service part of the software package was the creation of an on-screen menu with corresponding “buttons”, allowing to manage standard set of work with files and program modules, various forms of data input and output, visualization of results of calculations.

4.3 Structure and implementation of the software package GasDin based on the use of integrated technologies

Block scheme of software package GasDin is presented on Figure 4.

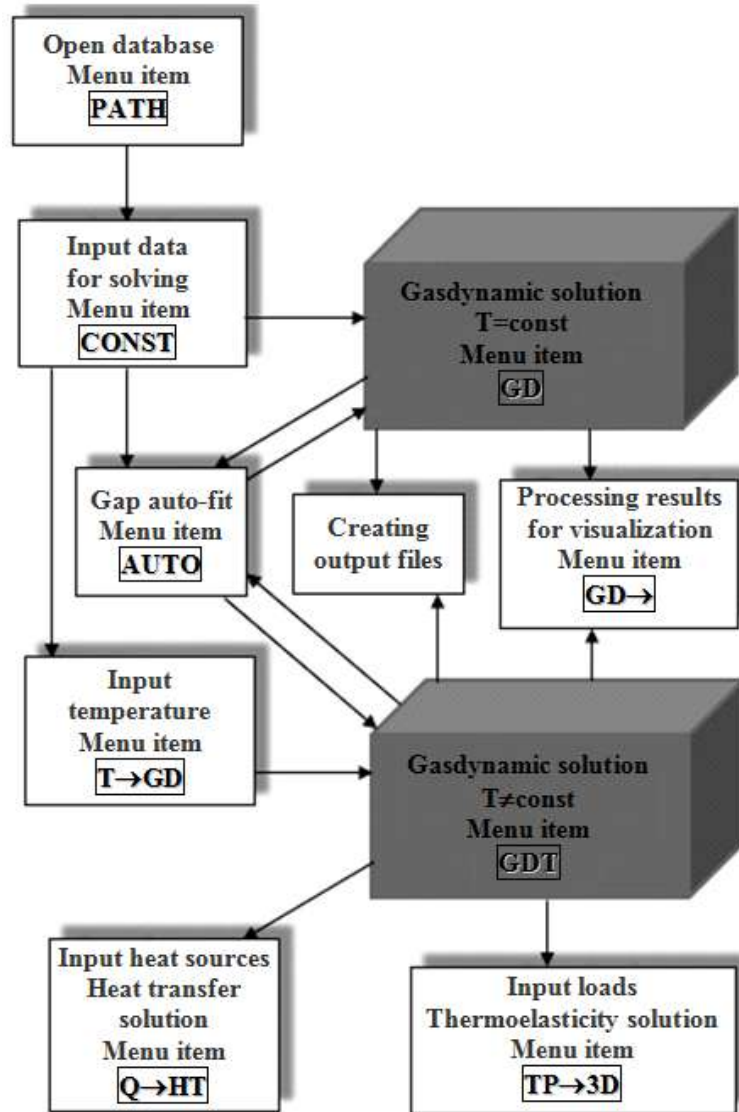


Fig. 4. Block scheme of software package GasDin

It should be noted that the solution of heat transfer and thermoelasticity tasks, the visualization of the obtained pressure fields and gas velocities in the gap is carried out using universal finite element software package. Data exchange between the software packages during iterative process is carried out by specially designed programs.

The developed software package allows solving for databases located in any directory. The initial data for the gasdynamic solution are: the gas pressure at the inlet and outlet of the seal P_2 and P_0 , the external and internal radius of the flow region r_2 and r_0 , the working gap δ , the depth of the grooves, the specific gas heat intensity ratio at

a constant volume C_v , universal gas constant R , gas temperature at the inlet and outlet of the seal T_2 and T_0 , the angular velocity of rotation ω , the accuracy of the gasdynamic calculation ε .

The software package GasDin allows to perform two types of gasdynamic solutions: at a constant temperature of the gas layer, and taking into account heat generation and temperature changes in the gas layer. The non-linear differential equation of gas lubrication is solved (1). The system of nonlinear algebraic equations for unknown pressure and temperature at the nodes of the finite element grid is solved by the method of simple iterations, where the linearized system is solved at each iteration using the Gauss method. The end of the iterative process is determined by the user-specified accuracy of solution. Setting of gas temperature field at each iteration is carried out by specially created programs for automatic transfer of the temperature distribution to the software package GasDin, obtained by solving the heat transfer task for gas film and working rings.

The results obtained by gasdynamic calculation are recorded to special files. They contain the values of the pressure in the nodes, the coordinates of the nodes. The value of gas leakage through the seal, resulting gasdynamic force, are also calculated. The developed software package also allows for the initial selection of the working gap between the seals by aligning the gasdynamic and gasostatic resulting forces. The gasostatic force is calculated during the preparation of the initial data after thorough study of the working drawings, the method of sealing of non-rotating ring and setting the boundary conditions.

After gas-dynamic solution, the software package calculates the intensity of heat sources in the gas layer and exports them into the database to solve the heat transfer problem. Further, the temperature fields of rings and pressures in the gas layer are exported into the corresponding databases for solving the thermoelasticity problem in three-dimensional and axisymmetric simulations, taking into account the possible mismatch of grids. Recalculation of the working gap, taking into account the obtained strains of seal rings, is also performed automatically.

4.4 Testing and practical use of software package GasDin

Testing of the created software package GasDin was carried out by solving test tasks for special cases which have an analytical solution, such as the distribution of gas pressure in the gap without grooves. Also, test solutions were compared with solutions in literature and experimental studies [15].

To determine the gas pressure in the gap between the rings, the region of the gas layer is considered, taking into account the cyclic periodicity conditions, containing a groove and a land. The results of the gasdynamic solution were checked in two stages. It is known, that for gas layer with constant thickness and temperature, the gas lubrication equation (1) has an analytical solution.

$$P^2 = -\frac{1-P_0^2}{\ln r_0} \ln r + 1; P = \sqrt{1 - \frac{1-P_0^2}{\ln r_0} \ln r} . \quad (4)$$

The solving results agreed with the results of the analytical solution up to the accurate to five decimal places and are shown in Figure 5. The exact solution values are marked with squares. For solutions, the following initial data were used: $r_0=53.6$ mm; $r_1=68$ mm; $r_2=80$ mm; $P_0=1$ atm; $P_2=53$ atm; the thickness of the gas layer outside the groove $h_1=3$ μm ; in the groove $h_2=11$ μm .

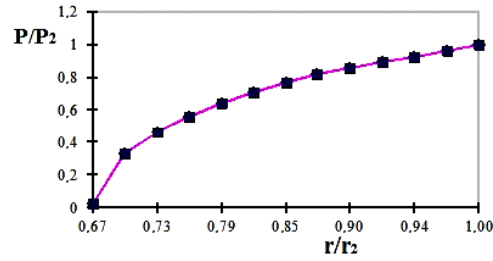


Fig. 5. Pressure distribution in the gas layer without grooves

Testing of the iterative solution algorithm and the created software package was also carried out for the real designs of dry gas seals used at the Joint Stock Company “Sumy Machine-Building Science-and-Production Association”, Sumy, Ukraine. To determine gas pressure in the gap between the rings, the region of gas layer is considered taking into account the cyclic periodicity conditions. Solution model contains a groove and a land. Geometric and operating parameters of the seal with spiral grooves: $r_0=101$ mm; $r_1=112.1$ mm; $r_2=125$ mm; $P_0=1.3$ atm; $P_2=60.8$ atm; the thickness of the gas layer outside the groove $h_1=3$ μm ; in the groove $h_2=10$ μm .

Figure 6a shows the distribution of gas pressure in the gap, taking into account temperature changes; Figure 6b the distribution of gas velocities in the gap.

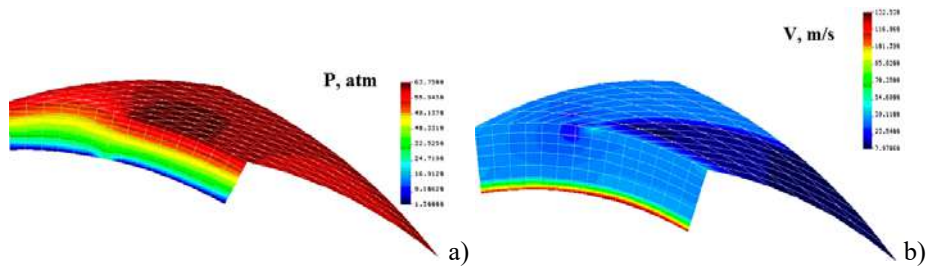


Fig. 6. For seals with spiral grooves: a) gas pressure distribution in the gap, taking into account temperature changes; b) the distribution of gas velocities in the gap

The developed solution algorithm and software package allows to find the gas pressure distribution for any type of grooves. Figure 7 shows the gas pressure distribution for seals with T-grooves under direct (Fig. 7a) and reverse (Fig. 7b) operation modes. The results of the gas pressure distribution in gap are in a good agreement with available in literature for seals with spiral grooves and T-grooves [6, 7, 8].

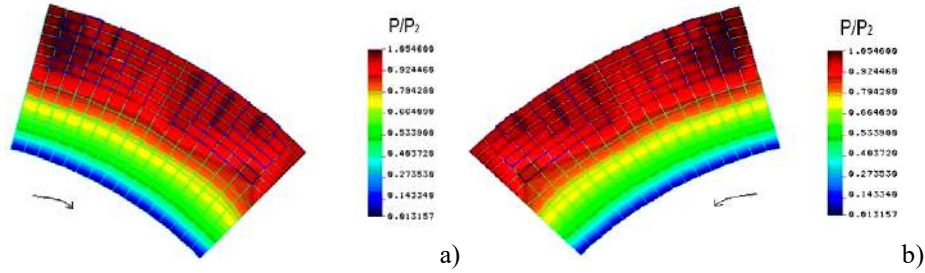


Fig. 7. Gas pressure distribution for seals with T-grooves: a) direct operation mode; b) reverse operation mode

An axisymmetric model of seal rings and gas layer is used for heat transfer solution. To determine the strain state of seal working rings, three-dimensional ring models are used taking into account the conditions of cyclic symmetry shown in Figure 8.

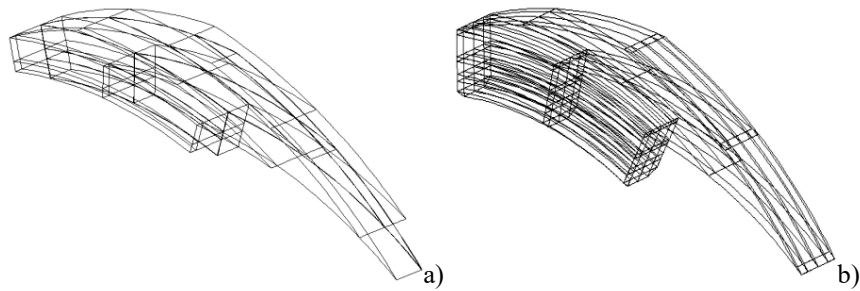


Fig. 8. The working rings models with spiral grooves: a) axially movable ring; b) rotating ring

It should be noted, that the Joint Stock Company “Sumy Machine-Building Science-and-Production Association” has created an experimental stand for dry gas seals. It determines the inlet and outlet pressure in seals, determines the gas velocities and temperature around seal rings, which can be used as the initial data for solutions. The gas temperature at the inlet and outlet of seals and the flow characteristics of seals are also determined.

As mentioned above, the heat transfer modeling for seal rings is done in axisymmetric setting. The resultant temperature distribution is shown in Figure 9.

Experiments for studied seals showed that gas in the gap heats up on average to 40-50°C, in the absence of contact of surfaces. As a result of computer modeling studies, the gas at the outlet is heated up to 46°C (319 K) for these seals. That is in a good agreement with experimental data.

The developed software package allows to determine the gas leakage through the seal, according to the integral expression for the gas leakage:

$$Q_r = \int_0^{2\pi} \frac{1}{12\mu RT} h^3 P \frac{\partial P}{\partial r} d\varphi \quad (5)$$

The gas leakage obtained as a result of computer modeling of studied gas seals using developed software package GasDin is 1.03×10^{-3} kg/s. The changes of working gap caused by straining of seal rings are taken into account.

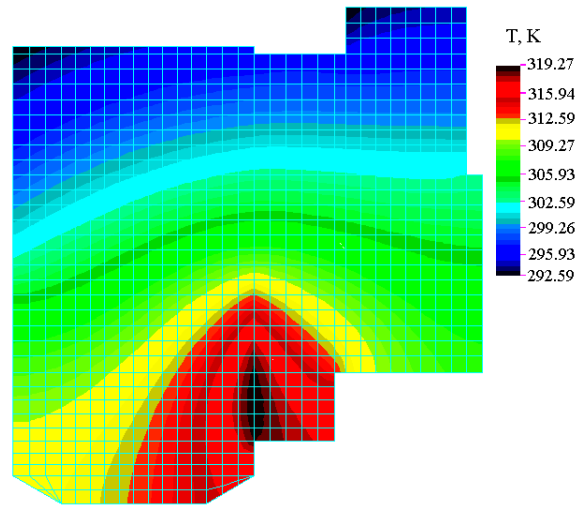


Fig. 9. Temperature distribution in gas seals

The calculated gas leakage differs from the experimental value for this seal design by 9%. It can be considered a good result.

The test modeling and solving confirm the reliability of results obtained using the developed software package GasDin.

5 Conclusion

Information technologies are widely used for computer modeling of various engineering facilities. Their use allows to improve the accuracy of designed facilities, reduce the service costs. Presented in this paper specialized software package allows to carry out computer modeling and solving of gas seals of centrifugal compressors.

The advantage of the developed software package GasDin is the ability to find the distribution of gas pressure in the gap between the seal rings at a constant temperature and taking into account of temperature changes. It is an advantage over existing general engineering software packages. Seeing that, the distribution of gas pressure in the gap is described by a non-linear equation related to particular part of gasdynamic, the theory of gas lubricant.

Also, the software package GasDin implements an iterative algorithm for modeling and solving of gas seals. It allows to make a model of working gap between the seal rings, which changes under the influence of gas pressure and temperature. This improved computer modeling can provide the reliable operation of dry gas seals.

References

1. Rusanov, A., Martynenko, G., Avramov, K., Martynenko, V.: Detection of accident causes on turbine-generator sets by means of numerical simulations. In: 2018 IEEE 3rd International Conference on Intelligent Energy and Power Systems (IEPS 2018), pp. 51-54. IEEE, Kharkiv, Ukraine (2018). doi: 10.1109/IEPS.2018.8559546
2. Martynenko, G.Y., Marusenko, O.M., Ulyanov, Y.M., Rozova, L.V.: The Use of Information Technology for the Design of a Prototype Engine with Rotor in Magnetic Bearings. In: International Scientific and Technical Conference on Integrated Computer Technologies in Mechanical Engineering – Synergetic Engineering (ICTM 2019). Kharkiv, Ukraine (2019). Advances in Intelligent Systems and Computing. 1113, 2020, 301-309 (2020). doi: 10.1007/978-3-030-37618-5_26
3. Robert, F.: Seals and Sealing Handbook. Sixth Edition. Elsevier Ltd. (2014). doi:10.1016/C2012-0-03302-9
4. Bloch, H.P., Carmody, C.: Advances in dry gas seal technology for compressors. Hydrocarbon processing **12**, 69-73 (2015)
5. Huebner, M., Barker, J.: Considerations for dual pressurized gas seals in pump applications. In: 43rd Turbomachinery & 30th Pump Users Symposia (Pump & Turbo 2014), pp. 23-25. Houston, USA (2014)
6. Shahina, I., Gadalab, M., Alqaradawib, M., Badr, O.: Centrifugal compressor spiral dry gas seal simulation working at reverse rotation. Procedia Engineering **68**, 285-292 (2013). doi: 10.1016/j.proeng.2013.12.181
7. Jing, X., Xudong, P., Shaoxian, B., Xiangkai, M.: CFD Simulation of micro scale flow field in spiral groove dry gas seal. In: 2012 IEEE/ASME 8th IEEE/ASME International Conference on Mechatronic and Embedded Systems and Applications, pp. 211-217. IEEE, China, Suzhou (2012). doi: 10.1109/MESA.2012.6275564
8. Gao, L.: Steady simulation of T-groove and spiral groove dry gas seals. International Journal of Heat and Technology **37**(3), 839-845 (2019). doi: 10.18280/ijht.370321
9. Su, H., Rahmani, R., Rahnejat, H.: Performance evaluation of bidirectional dry gas seals with special groove geometry. Tribology Transactions **60**, 58-69 (2017)
10. Li, X., Wang, K., Du, B.: The structure optimization of T-slots dry gas seal faces. In: 2016 International Conference on Advanced Materials Science and Environmental Engineering (AMSEE 2016). Advances in Engineering Research, vol. 52, pp. 189-192. Atlantis Press (2016). doi: 10.2991/amsee-16.2016.51
11. Wang, Y., Hu, Q., Sun, J., Wang, D., Zheng, X.: Numerical analysis of T-groove dry gas seal with orientation texture at the groove bottom. Advances in Mechanical Engineering **11**(1), 1-12 (2019). doi:10.1177/1687814018821775
12. Su, H., Rahmani, R., Rahnejat, H.: Thermohydrodynamics of bidirectional groove dry gas seals with slip flow. International Journal of Thermal Sciences **110**, 270-284 (2016)
13. Thomas, S., Brunetière, N., Tournier, B.: Thermoelastohydrodynamic behavior of mechanical gas face seals operating at high pressure. Journal of Tribology **129**, 841-850. (2007)
14. Chen, W., Song, P.Y., Mao, W.Y., Jiao, F.: Numerical analysis of temperature field of gas film inspiral groove dry gas seal. Journal of Drainage and Irrigation Machinery Engineering **33**(5), 422-428 (2015). doi: 10.3969/j.issn.1674-8530.15.0013
15. Rogovoy, Y.D., Rozova L.V., Kiselev A.M.: A combined solution to the problem of designing of gasdynamic mechanical seals. Bulletin of the National Technical University «Kharkiv Polytechnic Institute» **3**(8), 61-66. Kharkiv, Ukraine (2003). (in Russian)

Simulation Modeling of External Perturbations Affecting Wheeled Vehicles of Special Purpose

Oleg Nikonov^[0000-0002-8878-4318], Igor Kyrychenko^[0000-0002-2128-3500],

Vladyslav Shuliakov^[0000-0002-6654-6749]

Kharkiv National Automobile and Highway University,
Str. Yaroslava Mudrogo 25, Kharkiv, 61002, Ukraine
nikonov.oj@gmail.com, igk160450@gmail.com, jason07@ukr.net

Abstract. External perturbations affecting wheeled vehicles of special purpose on the basis of 3D terrain model are estimated. The mathematical model of the system of sprinkling for special purpose wheeled vehicles was developed on the basis of the use of continuous "color" and "fractional" noise as external perturbations, affecting the system, which allows to determine the influence of the impact of the rough terrain on wheeled vehicles of special purpose with integrated information and control telematics systems. A comparative analysis of the algorithms for the formation of external perturbations on the basis of "color" and "fractional" noise was carried out. 3D models of terrain were developed on the basis of the actual normal Markov field, which gives an opportunity to raise a qualitative level mathematical models of special purpose wheeled vehicles with integrated information and control telematics systems. It was generated an algorithm of a two-dimensional fractional field. The proposed approach allows you to simulate 3D terrain with given characteristics and analyze the work of special purpose wheeled vehicles systems on different operating modes.

Keywords. Wheeled Vehicles, External Perturbations, Simulation Modeling

1 Introduction

Solving various problems related to the automated design of systems and units of special purpose wheeled vehicles (WV) the dynamic process should be simulated in the process of its movement in rough terrain.

In real conditions, the WV of special purpose does not move along the microprofile of the road, but on the surface; therefore, the replacement of the movement of a three-dimensional object by the surface with the movement of a two-dimensional object along the road profile is an excessive simplification of the model, which may lead to significant differences between the simulation model and reality.

Secondly, the use of ICTS, GIS systems, satellite communications requires data processing and visualization in 3D format. Therefore, when simulating the terrain by a probable method, it is necessary to use 3D terrain models with given characteristics.

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

2 Formulation of the problem

The aim of the research is to develop the mathematical model of the system of sprinkling for special purpose wheeled vehicles on the basis of the use of continuous "color" and "fractional" noise as external perturbations, affecting the system, which allows to determine the influence of specific conditions on the system. A comparative analysis of the algorithms for the formation of external perturbations on the basis of "color" and "fractional" noise should be carried out. It is necessary to develop 3D models of terrain on the basis of the actual normal Markov field, what can give an opportunity to raise a qualitative level mathematical models of special purpose wheeled vehicles with integrated information and control telematics systems.

The algorithm of a two-dimensional fractional field should be generated. To consider an approach which allows to simulate 3D terrain with given characteristics and analyze the work of special purpose wheeled vehicles systems on different operating modes.

3 Literature Review

One of the important issues that arises in these problems is the simulation of random surfaces, in which the process of movement of vehicles takes place [1]. At present, for the aforementioned tasks, it is accepted to use complex functional mathematical models (CFMM) of special purpose WV in cooperation with the terrain and the driver [2, 3]. In such CFMM the terrain can be represented by a probable method or determinative one. Usually a probable representation method is used, because the determinative method in the case of a detailed description of a rather large area of the terrain is cumbersome and excessive [4-6]. The purpose of the work is to estimate external perturbations affecting wheeled vehicles of special purpose on the basis of 3D terrain model.

Let's consider in more detail the modeling of the oscillations of the body of wheeled vehicles of special purpose. The experience in the operation of automatic devices and systems shows that automation tools operating on vehicles and other machines are effective when their elements are properly tuned [7, 8]. An efficient approach to vehicle design is to simulate dynamic processes in real time [9-11]. It should be noted that the simulation as itself must be carried out taking into account the real properties of the environment, which is especially important for wheeled vehicles of special purpose.

As a rule, in simulation modeling, stochastic modeling based on the use of the "white" noise process is used as a perturbation factor associated with the profile properties [7]. The cross-country profile on which wheeled vehicles of special purpose move contains obstacles such as pits, ditches, logs, reinforced concrete structures, etc. To describe profiles of this kind in transport problems of simulation modeling, it is necessary to use stochastic processes, whose properties are close to the indicated perturbations, and thus different from the properties of white noise. The known process of "fractional" noise, it seems, can be used in these tasks.

4 Research Methodology

We consider an integral quadratic function based on the valid normal Markov two-dimensional field (NMD field) $H(x, y)$ [6]

$$J[H] = \int_0^a \int_0^b h^2(x, y) dx dy \quad (1)$$

where $h = h(x, y)$ – the realization of a Gaussian two-dimensional field $H(x, y)$ in a rectangular region $\{x \in [0, a], y \in [0, b]\}$ on surface. The determining property of a stationary NMD field is its correlator

$$K_{XY}(x, y; x', y') = E_H[h(x, y) h(x', y')] = pq\sigma_H,$$

$$p = \exp(-\nu |x - x'|), \quad q = \exp(-\mu |y - y'|),$$

where $E_H[\cdot]$ – the operator of mathematical expectation;

$\sigma_H = E_H[h^2(x, y)]$ – field intensity;

ν and μ – decrements of the field fading of the axes X and Y respectively.

The generalization of known constructions - transition probabilities density for the normal Markov process of Ornstein-Uhlenbeck (OU-process) – can be the following transition probability distribution density for the NMD field

$$f_H(h(x, y) | h(x', y), h(x, y'), h(x', y')) = \frac{1}{\sqrt{2\pi (1-p^2) (1-q^2) \sigma_H}} \times$$

$$\times \exp \left\{ -\frac{[h(x, y) - ph(x', y) - qh(x, y') + pqh(x', y')]^2}{2 (1-p^2) (1-q^2) \sigma_H} \right\}. \quad (2)$$

Steering $x' \rightarrow -\infty$ and $y' \rightarrow -\infty$, we obtain limited transitions probability distribution density, which serve as a transition for partial OU processes

$$f_H(h(x, y) | h(x', y)) = \frac{1}{\sqrt{2\pi (1-p^2) \sigma_H}} \exp \left\{ -\frac{[h(x, y) - ph(x', y)]^2}{2 (1-p^2) \sigma_H} \right\} \quad (3)$$

$$f_H(h(x, y) \mid h(x, y')) = \frac{1}{\sqrt{2\pi(1-q^2)\sigma_H}} \exp \left\{ -\frac{[h(x, y) - qh(x, y')]^2}{2(1-q^2)\sigma_H} \right\} \quad (4)$$

The peak density distribution of probabilities of an equilibrium type for a random variable - the realization of the NMD field at the coordinate point (x, y) has this form

$$f_H(x, y) = \frac{1}{\sqrt{2\pi\sigma_H}} \exp \left\{ -\frac{h^2(x, y)}{2\sigma_H} \right\} \quad (5)$$

To consider the equation of movement for the amplitude of the NMD field [6]. We place on the surface Cartesian coordinate system with the beginning at the point $(0,0)$. The dynamics of a random field $H(x, y)$ in a rectangle $\{x \in [0, a], y \in [0, b]\}$ with a vertex in $(0,0)$ it can be described with the help of a generalized Langevin equation for the OU-process

$$\left(\frac{\partial}{\partial x} + \nu \right) \left(\frac{\partial}{\partial y} + \mu \right) h(x, y) = \sqrt{\sigma_H} u(x, y) \quad (6)$$

where $u(x, y)$ – is a random field with properties of Gaussian two-dimensional noise with unit intensity. As the limited conditions to (6), we use two normal stochastic processes, which are described by the Langevin equations

$$\begin{aligned} \left(\frac{\partial}{\partial x} + \nu \right) h(x, 0) &= \sqrt{\sigma_H} u(x, 0); \\ \left(\frac{\partial}{\partial y} + \mu \right) h(0, y) &= \sqrt{\sigma_H} u(0, y), \end{aligned} \quad (7)$$

which are realized along axes X and Y accordingly, and the initial condition to them will be the peak values of the random variable

$$h(0, 0) = \sqrt{\sigma_H} u(0, 0) \quad (8)$$

The solution of equation (6) with conditions (7) and (8) is as follows

$$\begin{aligned}
h(x, 0) &= h(0, 0) \exp(-\nu x) + \sqrt{\sigma_H} \int_0^x \exp[-\nu(x-x')] u(x', 0) dx'; \\
h(0, y) &= h(0, 0) \exp(-\mu y) + \sqrt{\sigma_H} \int_0^y \exp[-\mu(y-y')] u(0, y') dy'; \\
h(x, y) &= h(0, 0) \exp(-\nu x - \mu y) + \sqrt{2\nu\sigma_H} \exp(-\mu y) \times \\
&\times \int_0^x \exp[-\nu(x-x')] u(x', 0) dx' + \sqrt{2\mu\sigma_H} \exp(-\nu x) \times \\
&\times \int_0^y \exp[-\mu(y-y')] u(0, y') dy' + \sqrt{4\nu\mu\sigma_H} \times \\
&\times \int_0^x \int_0^y \exp[-\nu(x-x') - \mu(y-y')] u(x', y') dx' dy'
\end{aligned} \tag{9}$$

From the obtained solution it is obvious that the Markov property is performed along the axes X and Y .

Using the known rules of discretization OU process in the nodes of a rectangular lattice and renormalizing the generating field $u(x, y)$, it is possible to construct a numerical algorithm for generation of a NMD field based on the solution (9). This hierarchical algorithm for generating the reference in the nodes of a random normal stationary Markov field in a rectangular surface can be given by the following four steps:

1. Generation of the countdown at the vertex

$$h_{0,0} = \sqrt{\sigma_H} u_{0,0}.$$

2. Generation of the countdown of the process along the X-limit rectangle ($j > 0$)

$$h_{j+1,0} = ph_{j,0} + \sqrt{(1-p^2) \sigma_H} u_{j+1,0}.$$

3. Generation of the countdown of the process along the Y-limit rectangle ($k > 0$)

$$h_{0,k+1} = qh_{0,k} + \sqrt{(1-q^2) \sigma_H} u_{0,k+1}.$$

4. Sequential filling of the countdown internal nodes of the rectangle ($j > 0, k > 0$)

$$h_{j+1,k+1} = ph_{j,k+1} + qh_{j+1,k} - pqh_{j,k} + \sqrt{(1-p^2)(1-q^2) \sigma_H} u_{j+1,k+1}.$$

In the presented algorithm

$$p = \exp(-\nu\Delta_x); \quad q = \exp(-\mu\Delta_y),$$

where ν and μ – partial decrements;

Δ_x and Δ_y – steps of nodes on the axes X and Y respectively.

To note that for selected Δ_x , Δ_y (i.e., for given dimensions of the rectangle a , b and the corresponding number of steps $N_x = a/\Delta_x$, $N_y = b/\Delta_y$) the intensity in this algorithm needs to be renormalized so that the energy of the NMD field per unit area coincides with the given in any number of steps.

Thus, the aforementioned algorithm for generating random countdown in a rectangle on a surface is stationary [8, 9].

In conducting many experiments, the following results were obtained. The simulation was carried out in a rectangular area $\{x \in [0, 10], y \in [0, 10]\}$. It is shown three realizations of a random NMD field (fig. 1), which differ only in the value of the field intensity (case a, b, c). The picture shows the dynamics of the formation of fluctuations of NMD-fields on both coordinates. In fig. 2 it is shown the original output profile $s(x, y) = [\sin(x) \cdot \cos(y)] / 2$. When the additive overlay of a random field (fig. 1b) on a regular profile (fig. 2) is obtained the surface $p(x, y) = s(x, y) + h(x, y)$, shown in fig. 3.

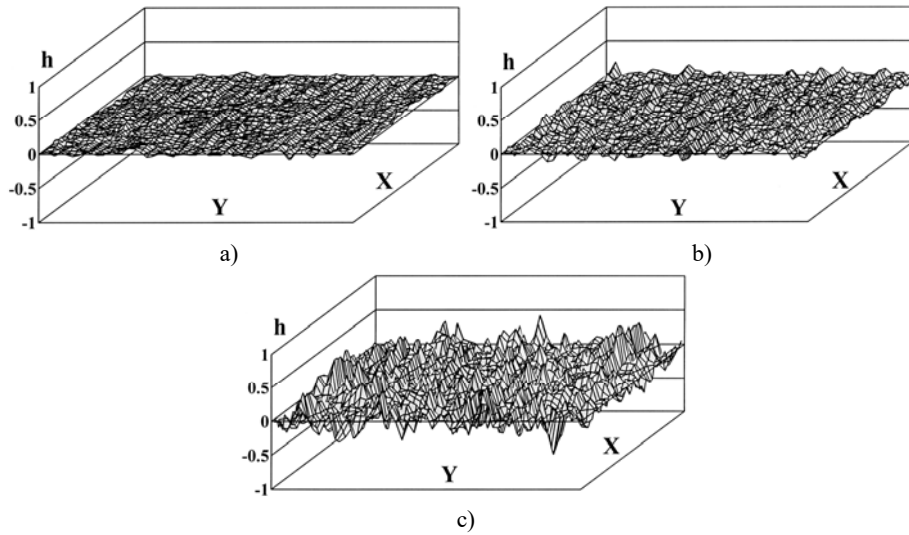


Fig. 1. Three realizations of a random NMD field in a rectangular area: a) small intensity; b) medium intensity; c) high intensity $\{x \in [0, 10], y \in [0, 10]\}$

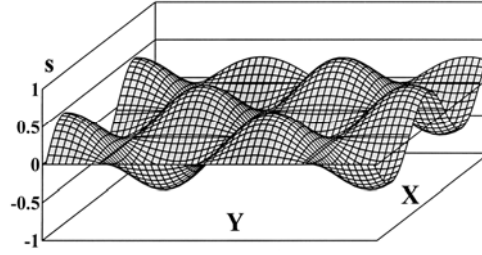


Fig. 2. The original output profile $s(x, y) = [\sin(x) \cdot \cos(y)] / 2$

In the study of spatial images on a given rectangle $[(0, a) \times (0, b)]$ one of the main questions of the theory of evaluation is the information on the distribution of random values η of the functional $J[H]$.

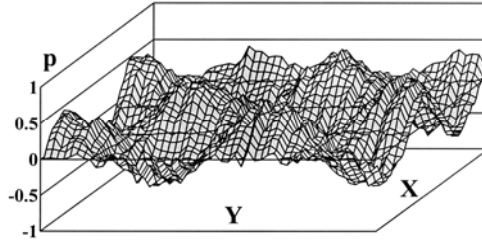


Fig. 3. Additive overlay of a random field (fig 1b) on outgoing original output profile (fig 2)

Because of the positive definiteness and additivity of the integral on a rectangle $[(0, a) \times (0, b)]$ it is convenient to describe the properties of a functional $J[H]$ using a function in the form of the next mathematical expectation

$$Q_{XY}(\lambda) = E_H [\exp(-\lambda J[H])] = E_H \left[\exp \left(-\lambda \int_0^a \int_0^b h^2(x, y) dx dy \right) \right] \quad (10)$$

where λ – random parameter.

The desired function can be represented as a completely convergent result [6]

$$Q_{XY}(\lambda) = \prod_{n=1}^{\infty} \prod_{m=1}^{\infty} \frac{1}{\sqrt{1 + \frac{\lambda}{\lambda_{X,n} \lambda_{Y,m}}}} \quad (11)$$

where $\{\lambda_X\}$ and $\{\lambda_Y\}$ – sets of proper values of partial correlators K_X and K_Y , associated with solutions of integral equations corresponding to the limits of a rectangle

$$\phi_X(x) = \lambda_X (K_X, \phi_X(x)) = \lambda_X \sigma_H \int_0^a \exp(-\nu |x - x'|) \phi_X(x') dx' \quad (12)$$

$$\phi_Y(y) = \lambda_Y(K_Y, \phi_Y(y)) = \lambda_Y \sigma_H \int_0^b \exp(-\mu |y - y'|) \phi_Y(y') dy' \quad (13)$$

In the considered case, the NMD field for partial functions $Q_X(\lambda)$ and $Q_Y(\lambda)$ is well known [6], that

$$Q_X(\lambda) = \sqrt{\frac{4\nu r_X \exp(\nu a)}{(r_X + \nu)^2 \exp(r_X a) - (r_X - \nu)^2 \exp(-r_X a)}} \quad (14)$$

$$Q_Y(\lambda) = \sqrt{\frac{4\mu r_Y \exp(\mu b)}{(r_Y + \mu)^2 \exp(r_Y b) - (r_Y - \mu)^2 \exp(-r_Y b)}} \quad (15)$$

where $r_X = r_X(\lambda) = \sqrt{\nu^2 + 2\lambda\nu\sigma_H}$; $r_Y = r_Y(\lambda) = \sqrt{\mu^2 + 2\lambda\mu\sigma_H}$.

The simple zeros of the denominators of the given partial functions $Q_X(\lambda)$ and $Q_Y(\lambda)$ the essence of the negatively defined sets of numbers $\{\lambda_{X,n}\}_{n=1}^{\infty}$ and $\{\lambda_{Y,m}\}_{m=1}^{\infty}$. Thus, on the basis of these sets, it is necessary to construct a formula for the desired function. In other words, it is necessary to reduce the result (11) and to obtain a constructive analytical expression for $Q_{XY}(\lambda)$, without the necessity to find the specific values of the sets $\{\lambda_X\}$ and $\{\lambda_Y\}$.

The main analytical result is the following representation for the function $Q_{XY}(\lambda)$ of the functional (1)

$$Q_{XY}(\lambda) = \exp \left\{ -\frac{1}{(2\pi i)^2} \iint_C \left(\ln Q_X(x) \frac{d}{dx} \right) \left(\ln Q_Y(y) \frac{d}{dy} \right) \ln \left(1 + \frac{\lambda}{xy} \frac{\langle \eta \rangle_{XY}}{\langle \eta \rangle_X \langle \eta \rangle_Y} \right) dx dy \right\} \quad (16)$$

where $\langle \eta \rangle_{XY} = ab\sigma_H$ – the average value of the functional (1);

$\langle \eta \rangle_X = a\sigma_H$ and $\langle \eta \rangle_Y = b\sigma_H$ – partial mathematical expectations;

C – the contour of integration in the complex surface passing through the imaginary axis bypassing the point $(0,0)$ on the left is a fairly small radius of a semicircle.

The density of the probability distribution $f_J(\eta)$ of random values η of a functional $J[H]$ is determined, therefore, on the basis of the inverse Laplace transformation of the function $Q_{XY}(\lambda)$. Such a transformation for functions (14), (15) or (16) can only be performed numerically, so this procedure will be implemented the better, the more accurate the analytical information about the function $Q_{XY}(\lambda)$.

To reason the results of the simulation, families of histograms corresponding to the distribution of the integral quadratic functional $J[H]$ for a particular field were obtained. The volume of the statistical sample in the simulation was 10^4 , the relative

sample dispersion did not exceed 3%. The obtained histograms were the same with the given accuracy. A number of similar experiments showed the preservation of the type of histograms in numerical simulation for any region.

Also the algorithm was developed for generation of two-dimensional fractional field based on the description of fractional noise. In fig. 4 it is shown four realization of random fractional fields for different values of the number of strokes K .

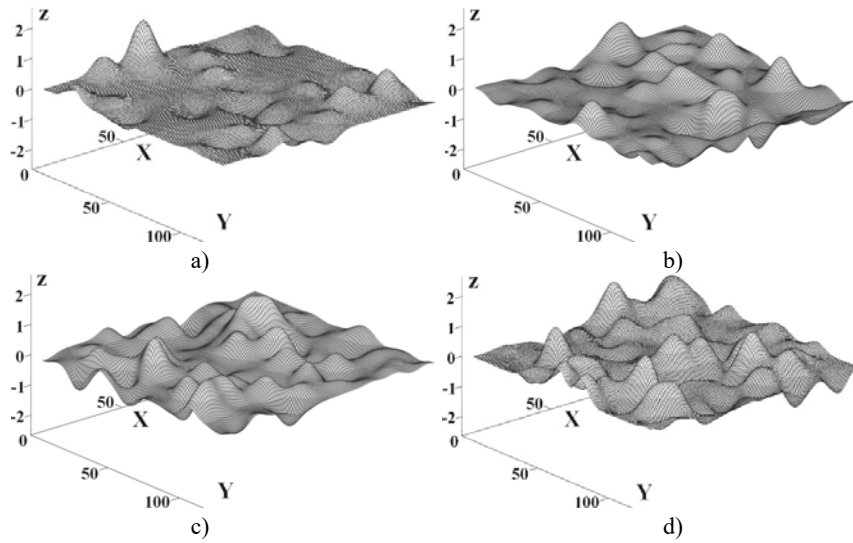


Fig. 4. Realization of a random fractional field in a rectangular region at: a) $K = 50$; b) $K = 100$; c) $K = 150$; d) $K = 200$

The picture shows well the dynamics of the formation of fluctuations on both coordinates, which agrees with the results of 2D fractional noise modeling. The calculation of the fluctuations of the body of special purpose wheeled vehicles and its main systems and aggregates, taking into account the movement of the 3D field, is carried out according to the known calculation formulas.

5 Conclusions

The mathematical model of the system of sprinkling for special purpose wheeled vehicles was developed on the basis of the use of continuous "color" and "fractional" noise as external perturbations, affecting the system, which allows to determine the influence of the impact of the rough terrain on wheeled vehicles of special purpose with integrated information and control telematics systems. It is proved that for the simulation of the terrain with the help of CFMM it is most appropriate to use the probable representation method of the terrain, because the determinative method in the case of a detailed description of a rather large area of the terrain is cumbersome and excessive. A comparative analysis of the algorithms for the formation of external

perturbations on the basis of "color" and "fractional" noise was carried out. On the basis of many experiments it has been proved that in the problems of modeling the movement of special purpose wheeled vehicles, it is possible to use the external perturbations based on the use of "fractional" noise and "color" noise, depending on the type of terrain. 3D models of terrain were developed on the basis of the actual normal Markov field, which gives an opportunity to raise a qualitative level mathematical models of special purpose wheeled vehicles with integrated information and control telematics systems. The proposed approach allows you to simulate virtually any 3D terrain with given characteristics and analyze the operation of wheeled vehicles control systems at different operating modes, as well as to use calculated 3D terrain models for geoinformation programs of information and control telematics systems. It was generated an algorithm of a two-dimensional fractional field. The proposed approach allows you to simulate 3D terrain with given characteristics and analyze the work of special purpose wheeled vehicles systems on different operating modes.

References

1. Xu, G., Ma, Z., Lu, F., Hou, P.: Kinematic Analysis of Hydraulic Excavator Working Device Based on DH Method. In: International Conference on Applied Mechanics, Mechanical and Materials Engineering, pp. 8 (2016). doi: 10.12783/dtmse/ammme2016/6857
2. Xue-wen, Chen, Yue, Zhou: Modelling and Analysis of Automobile Vibration System Based on Fuzzy Theory under Different Road Excitation Information. In: Complexity, Jinzhou, 9 pages (2018) <https://doi.org/10.1155/2018/2381568>
3. Shahgholi, G., Chiyaneh, H., Gundoshmian T.: Modeling of Soil Compaction Beneath the Tractor Tire using Multilayer Perceptron Neural Networks. Journal of Agricultural Machinery, vol.8, pp. 105-118 (2018)
4. Liu Y., Tong S.: Adaptive fuzzy control for a class of unknown nonlinear dynamical systems. In: Fuzzy Sets and Systems, vol. 263, pp. 49-70 (2015)
5. Gao, Y.: Adaptive fuzzy optimal control using direct heuristic dynamic programming for chaotic discrete-time system. Journal of Vibration and Control, vol. 22, pp. 595-603 (2016)
6. Mazmanishvili, A., Rafalovich, O.: Construction example of regional phase maps of error probability for chosen satellite communication network configuration. In: Engineering Simulation, vol. 15, pp. 715-719 (1998) (in Russian)
7. Aleksandrov, E.: Dynamics of transport and wheeled and tracked vehicles. Kharkiv, KHDATU (2001) (in Russian)
8. Aleksandrov, E.: To the task of optimizing the parameters of the vehicle suspension system. Mechanics and engineering. Kharkiv, KHDPU, pp. 35-41 (2000) (in Russian)
9. Uspensky, B., Avramov, K., Liubarskyi, B., Andrieiev, Y., Nikonov, O.: Nonlinear torsional vibrations of electromechanical coupling of diesel engine gear system and electric generator. Journal of Sound and Vibration, vol. 460, 114877 (2019) doi:10.1016/j.jsv.2019.114877
10. Gurko, A., Kyrychenko, I., Yaryzhko, A.: Trajectories Planning and Simulation of a Backhoe Manipulator Movement. CMIS, pp. 771-785 (2019)
11. Uspensky, B., Avramov, K., Nikonov, O.: Nonlinear modes of piecewise linear systems forced vibrations close to superharmonic resonances. Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science, vol. 233, Issue 23-24, pp. 7489-7497 (2019) doi: 10.1177/0954406219869967

Parallel simulation of rock deformation processes

Yevhen Tiurin¹[0000-0001-8362-9223], Antonina Andrieieva¹[0000-0002-0361-5436]

¹Donetsk National Technical University, Shybankova sq., 2, Pokrovsk, 85300, Ukraine
yevhen.tiurin@donntu.edu.ua, antonina.andrieieva@donntu.edu.ua

Abstract. This paper enlightens the simulation of rock deformation processes on multicore processing. For mines rock deformation processes affect gas emissions and its prediction is an important research and modeling task. In this paper, the parallel architectures OpenMP and MPI are used. The results show the modeling time of OpenMP and MPI are better than the serial programming when the number of grids is given more than 400 points.

Keywords. Parallel simulation, SIMD, MIMD, deformation process, rock deformation.

1 Introduction

With the transition of mining operations to the region of great depths (significantly lower than the gas weathering zone), external and internal factors related to the coal layer, which cause the occurrence of GDH are likely to change. In this aspect, the most careful attention should be paid to the geomechanical processes occurring in the near-bottom part of the coal layer.

In various studies of the stress state of a rock mass, as a rule, the mass is considered as an elastic medium. The stress concentration in the edge of the layer, located at depths of more than about 300 - 400 m, leads to the formation of an inelastic deformation zone [1, 2, 3, 4, 5]. However, taking into account the physical and mechanical properties of the coal layer, the geomechanical, chemical and gas-dynamic processes occurring in this zone, besides plastic deformation, brittle destruction of the rock and deformation of the genetic return also occur. It will be generally fair to consider this zone as a zone of inelastic deformations.

In the process of coal mining in the bottom, the boundary of the zone of inelastic deformation moves deep into the massif and the bottom-hole part of the coal layer is extracted into the mine.

Over time and under the influence of existing stresses, the equilibrium state of the bottom-hole part of the coal layer can be disturbed, and often it happens. The interaction of these factors, and the physicommechanical properties of the layer and host rocks, leads to a critical state of coal in the near-bottom part of the layer and its natural softening and destruction.

It should be noted that the destruction of the edge of the coal layer occurs during unloading due to excavation of the next strip (part) of coal in the lava. The condition

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

on the destruction of a solids not with an increase or prolonged exposure to normal or shear stresses, but with unloading, was noted back in the 40-50s of the last century by the American scientist P. Bridgman [6, 7].

Over time, numerous experimental studies have proved the fundamental impossibility of the origin and occurrence of GDH in the unloaded part of the outburst hazardous layer below the gas weathering zone, at great depths.

The formation and development of the zone of inelastic deformations in the boundary part of the layer always occurs during stress redistribution. This part of the layer is characterized by reduced tension and, consequently, an increased tendency to gas recovery - gas evolution. With an increase in the depth of mining operations, and consequently, an increase in rock pressure, the processes of destruction of the edge part are undoubtedly intensified.

The conditions on the non-outburst hazard of the edge of the outburst layer formed the basis of many methods for predicting, preventing GDH and monitoring their effectiveness, which were included in the new standard of the «Ministry of Industry of Ukraine» [8].

Substantial growth below the gas weathering zones of the extraction strains was determined, which allows, in this regard, to produce in it a more productive, controlled extraction of coal from the outburst hazardous layer without causing sudden outbursts.

On this basis, rock deformation processes affect gas emissions and its prediction is an important research and modeling task.

2 Mathematical description

To simulate the deformation process, a system of equations of mechanics including the equations:

- continuity

$$\rho + \rho u_{i,j} = 0 \quad (1)$$

- movement

$$\sigma_{ji,i} + \rho F_j = \rho \dot{u}_j. \quad (2)$$

Here:

ρ – the density of the material;

u_j – the components of the velocity vector;

$\sigma_{j,i}$ – the components of the Cauchy stress tensor;

F_j – mass forces.

Here ρ is the density of the material; u_j are the components of the velocity vector; $\sigma_{j,i}$ are the components of the Cauchy stress tensor; F_j - mass forces; the dot above means the time derivative, the index after the decimal point means the corresponding coordinate derivative. The system of equations is closed by defining relations that specify the behavior of the medium during deformation. When writing the defining relations, the decomposition of the stress tensor into the spherical and deviator parts is used:

$$\sigma_{j,i} = -\sigma\delta_{j,i} + S_{j,i} \quad (3)$$

Where:

– σ – average pressure;

$S_{j,i}$ – stress tensor deviator components:

$\delta_{j,i}$ – Kronecker symbol.

Before the onset of a plastic state, the rates of stress and strain changes are related by a hypoelastic law:

$$\frac{DS_{j,i}}{D_t} = 2\mu_0(\dot{\epsilon}_{j,i} - \frac{1}{3}\dot{\epsilon}_{nn}\delta_{j,i}), \quad (4)$$

$$\frac{DS_{j,i}}{D_t} = \dot{S}_{j,i} - S_{ni}\dot{\omega}_{nj} - S_{nj}\dot{\omega}_{n,j}, \quad (5)$$

$$\dot{\sigma} = -K\frac{\dot{V}}{V}. \quad (6)$$

where K и μ – compression and shear modules, respectively. The components of the strain rate tensor $\dot{\epsilon}_{j,i}$ rotational speed tensor components $\dot{\omega}_{j,i}$ are determined from the relations:

$$\dot{\epsilon}_{j,i} = \frac{1}{2}(u_{j,i} + u_{i,j}), \quad (7)$$

$$\dot{\omega}_{j,i} = \frac{1}{2}(u_{j,i} - u_{i,j}). \quad (8)$$

It is accepted that the deformation consists of elastic e and plastic p parts. For strain rate, this decomposition is written as:

$$\dot{\epsilon} = \dot{\epsilon}^e + \dot{\epsilon}^p. \quad (9)$$

Plastic deformation is determined in accordance with the equation of the limiting surface and plastic potential:

$$f(\sigma_{j,i}, \varepsilon_{j,i}^p) = 0, \quad (10)$$

$$g(\sigma_{j,i}, \varepsilon_{j,i}^p) = 0, \quad (11)$$

$$d\varepsilon_{j,i}^p = d\lambda \frac{\partial g}{\partial \sigma_{j,i}}, \quad (12)$$

Where:

f – load surface (function) equation;

g – plastic potential;

$d\lambda$ - determined during deformation from the condition of plasticity;

$\varepsilon_{j,i}^p$ – components of plastic (inelastic) deformation.

Limiting surface (fig. 1) in the area of shear deformation in the pressure range $\sigma_j \leq \sigma \leq \sigma_0$ described by the equation

$$f_1(\sigma, \tau) = \tau - \alpha\sigma - c \quad (13)$$

and with pressures $\sigma > \sigma_0$ – by the equation

$$f_2(\sigma, \tau) = \frac{(\sigma - \sigma_0)^2}{a^2} + \frac{\tau^2}{b^2} - 1 = 0. \quad (14)$$

Here $\alpha = \alpha(e^n, \sigma)$, $c = c(e^n, \sigma)$ – coefficients of internal friction and adhesion, $\tau = (S_{j,i} S_{i,j} / 2)^{\frac{1}{2}}$ – second stress deviator invariant, σ_t – tear strength, σ_0 – threshold pressure at which material compaction begins, $a = \sigma_1 - \sigma_0$, $b = c + \alpha\sigma_0$.

The normal to the plastic potential $g = \tau - \Lambda\sigma$ determines the direction of plastic strain increments, where Λ – is the dilatancy coefficient.

It is accepted that during the development of inelastic deformation, a change in the surface of the limiting state occurs, as well as damage accumulation, which macroscopically manifests itself in a change in volume, i.e., in medium dilatancy. The hardening of the medium is described by the relation

$$c(\gamma^p) = c_0[1 + h(A(\gamma^p) - D(\gamma^p))] \quad (15)$$

where h – hardening coefficient, $d\gamma^p = 2((de_{j,i})^{2p} / 2)^{\frac{1}{2}}$ shear plastic strain rate, γ^* – critical deformation, after which degradation of the material predominates. To account for hardening, a linear relationship $A\gamma(\gamma^p) = 2\gamma^2 / \gamma$ and quadratic - to take into account softening (accumulation of damage)

$$D(\gamma^p) = (\gamma^p / \gamma) \quad (16)$$

The effect of pressure on the limiting strain that the material can withstand before softening begins is taken into account by the expression:

$$\gamma^* = \gamma_0^* (1 + w\sigma / \sigma^*) \quad (17)$$

where γ_0^* – plastic deformation of the onset of fracture in the absence of compressive (restraining) pressure; w and σ^* – parameters. This method made it possible to describe the transition from brittle to “viscous” behavior with increasing pressure [9].

Two-dimensional deformation processes are considered for conditions of plane deformation when $u_z = 0, \dot{\epsilon}_{xz} = \dot{\epsilon}_{yz} = \dot{\epsilon}_{zz} = 0$.

To simulate the process of rock deformation, the approach [10] was used, which is based on solving the equations of dynamics of an elastic-plastic medium using an explicit numerical scheme [11].

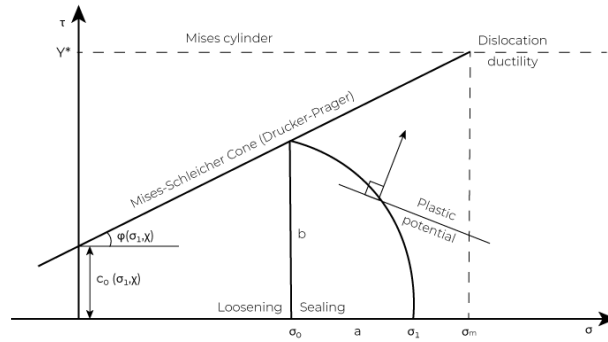


Fig. 1. Surface view of the limiting state of rocks

The main difficulty in describing the deformation beyond the elastic limit lies in the fact that the limiting surface is not fixed, it changes during deformation. On the loading diagrams, a hardening section is observed, where the increment of plastic deformation occurs with increasing stresses, followed by stress relief, and destruction occurs. In the course of deformation, not only the limiting surface changes, but also the ratio between the increments of the shear and volume parts of plastic deformation,

i.e., the direction of the plastic deformation vector. Thus, parameters describing the behavior of the rock beyond the elastic limit become accumulated plastic deformation and pressure functions.

It should be noted that the hardening section on the loading diagram in such media is very difficult to interpret. Formally, strength is described by two parameters: adhesion and internal friction. However, geometric causes contribute to effective hardening. Dilatancy hardening [12, 13] is especially evident in cramped conditions of deformation, when an increase in volume leads to an increase in pressure. This phenomenon is of particular importance in geomechanical processes. It was shown in [14] that an increase in the effective strength of the sample can be observed even with a slight decrease in the adhesion.

3 Parallel architecture

There are two different approach in multicore parallel computing - Shared memory parallelization and Distributed memory parallelization:

- Shared memory parallelization(SIMD) - architecture where each processors using the main memory for communicating of data. OpenMP is a main platform shared memory architecture. OpenMP is a simple instruction to divide a process into some subprocess which are called threads. OpenMP using in many programming language such as C/C++ and Fortran.
- Distributed memory parallelization(MIMD) - architecture where each processors has own local memory which are called nodes. All nodes connected in one communication network. Each node will execute the code and working with data using own local memory. Then, final result should be distributed to the main(master) node to gather all results. To work with this architecture is used MPI (Message Passing Interface). MPI have implementation in programming languages C/C++ and Fortran and other.

In addition to the two main platforms OpenMP and MPI, there is a hybrid platform - OpenMPI, which uses the capabilities of both platforms. In addition to OpenMPI, there is also a full-fledged language for parallel programming of hybrid architectures - Unified Parallel C.

Also in OpenMP, in addition to the built-in functions for parallelizing calculations, the built-in capabilities of the processor for parallelization, the so-called Math Kernel Library(MKL), are used. MKL can compute operations of addition, subtraction and multiplication over vectors and matrices in parallel.

The parallel speedup can be obtained by:

$$S(p) = \frac{T_1}{T_p} \quad (18)$$

Where:

- $S(p)$ is speedup by using parallel programming using p processor,
- T_1 the CPU time by using single processor,
- T_p CPU time by using p processor.

The parallel efficiency can be obtained by:

$$E = \frac{S(p)}{p} \times 100\% \quad (19)$$

In the process of coal mining in the bottom, the boundary of the zone of inelastic deformations moves deep into the massif, which increases the number of calculations during the simulation, and the acceleration of simulation of this process will make it possible to obtain data much faster and more accurately due to an increase in the size grid.

The solution of the system of equations is carried out on the computational grid, which covers the studied region of the medium or the entire sample, the discretization of the computational region is carried out. On the constructed computational grid, all functions are approximated and a system of equations is solved. Of course, difference schemes that have been successfully used for many years, either for problems in the deformation of the geomedium, are described in detail in [15, 16]. The standard features of numerical calculation are used, which are implemented in original software packages.

The OpenMP program is simple, for instance in C++, the instruction only add syntax `"#pragma omp parallel for"` above the looping codes, then, the single core will automatically divide the loop into several block.

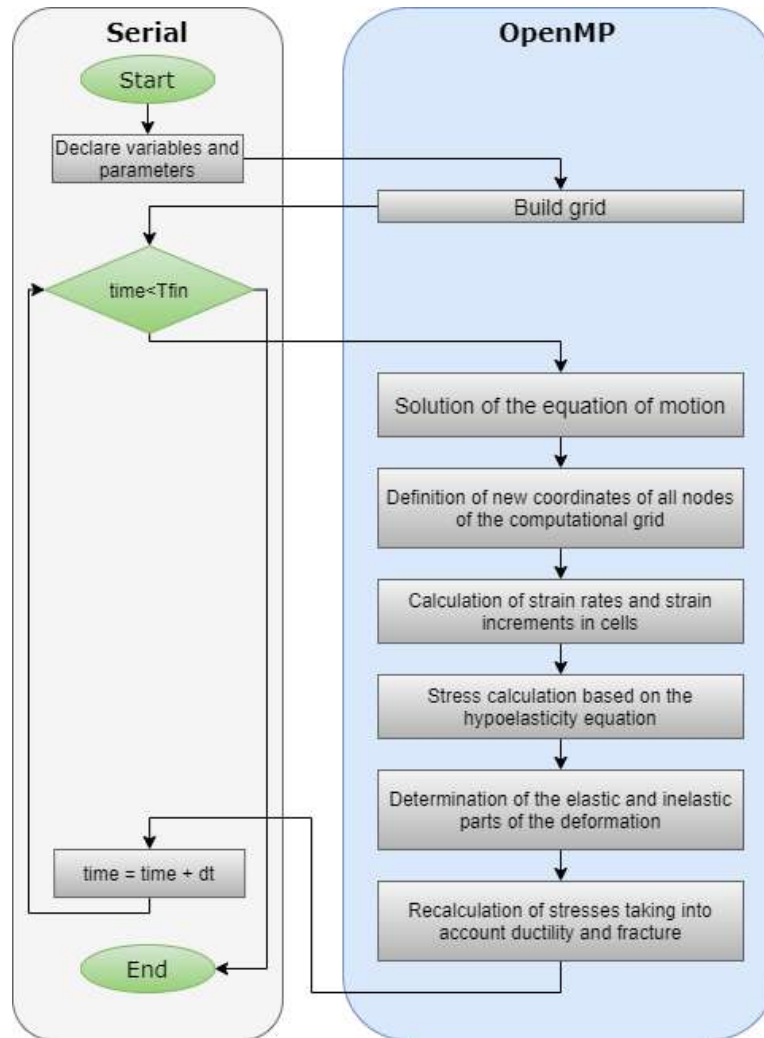


Fig. 2. Block process of OpenMP architecture

The MPI program in Fig. 3 has two domains of computation. The difference part from block process of OpenMP is located on looping time. Other processors will be executed time loop similar to the master.

In MPI, each processors need communication to each other for transferring data from master to slave. MPI have commands “MPI_Send()” and “MPI_Recv()” to send and receive data. But MPI (MIMD architecture) needs more communication time than the SIMD architecture and MPI could not guarantee produces best execution time.

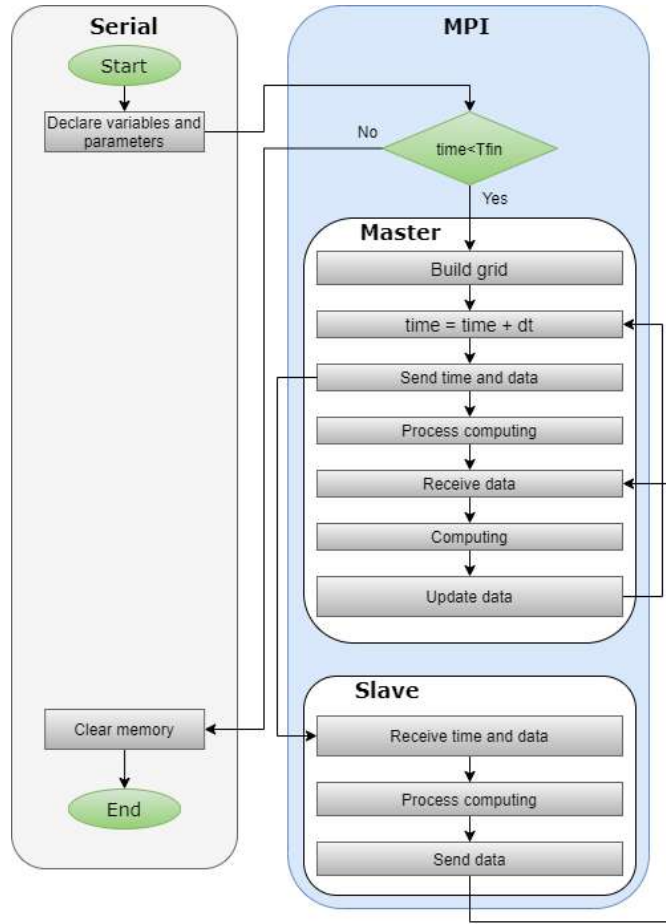


Fig. 3. Block process of MPI architecture

4 Experiments and results

To obtain simulation results and parallel performance, the following computer specifications are used.

Table 1. The Computer Specifications

Name	Type
Operating system	CentOS 6
Processors	Intel Xeon X5560 2.8 GHz
RAM	16GB
CPU(s)	16

The performance of MPI and OpenMP will be elaborated. The runtime for MPI and OpenMP for simulation is done for 10 seconds of the process.

Table 2. Computing time

Number of grid	Serial (s)	Time (s) of OpenMP		Time (s) of MPI	
		8 cores	16 cores	8 cores	16 cores
200	27.8096	49.5230	61.9437	36.8255	54.3485
400	94.8357	110.2547	99.3012	82.2647	135.8475
800	370.5342	245.9861	240.1843	237.1254	330.1293
1600	1490.5422	690.5362	580.1232	740.9543	997.3753

Table 2 shows the comparison of modeling time in serial parallel using 8 and 16 cores for each type. Form the Table 2, the modeling time in serial for number of grids 200 and 400 is smaller than the modeling time in parallel except MPI type for 8 cores. This means the parallel processing is not less efficient since the number of grid points is small, in this case using single core is already enough. However, using the large number of points (from 800 and greater), the modeling time in parallel is better than the serial.

From Table 2, the speedup (18) and efficiency (19) performances of OpenMP and MPI using 8 and 16 cores can be achieved. The speedup and efficiency profiles are shown in Fig. 2 and 3 respectively. In Fig. 2, the speedup of OpenMP is better than MPI for different number of cores. However, the speedup using MPI shown slightly better in number of grids 800 8 cores. The MPI fails to get the good speedup performance due to the time for changing messages in processors is dominant than the computing time.

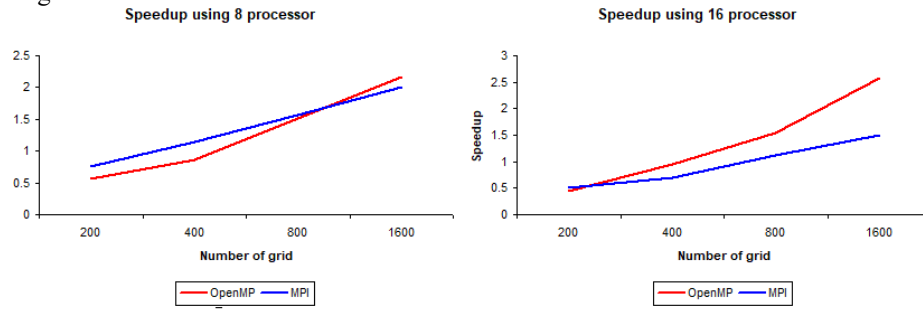


Fig. 4. Speedup of OpenMP and MPI using 8 (left) and 16 (right) cores.

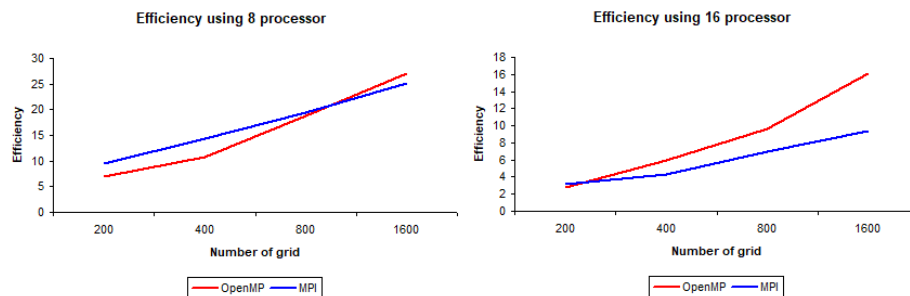


Fig. 5. Efficiency of OpenMP and MPI using 8 (left) and 16 (right) cores.

The efficiency profiles in Fig. 3 are given. The efficiency of parallel processing using 8 cores is better than 16 cores in both platforms.

5 Conclusion

Two parallel architectures OpenMP and MPI have been implemented for simulating the 2D deformation process. The results show the modeling time of OpenMP and MPI are better than the serial programming when the number of grids is more than 200 points.

References

1. Bobrov, I.V., Krichevskiy, R.M., Mihaylov, B.I.: Vnezapnyie vyibrosyi uglya i gaza na shahtah Donbassa, Ugletekhizdat (1954)
2. Nikolin, V.I., Mordasov, V.I., Savchenko, P.I.: Obosnovanie minimalnogo chisla neschastnyih sluchaev, neobhodimyyh dlya dostovernogo analiza travmatizma. In: Izvestiya Gornogo institute, vol. 2, pp. 41-44 (1999)
3. Ruppeneyt, K.V.: Nekotoryie voprosyi mehaniki gornyyh porod, Ugletekhizdat (1954)
4. Bridgmen, P. W.J.: Appl. Phys (1960)
5. Buzer, G.D., Hiller, K.H., Serdengekti, S.: Vliyanie porovoy zhidkosti na deformatsionnoe povedenie gornyyh porod pri trehosnom szhatii. In: Mehanika gornyyh porod, pp. 372-406 (1966)
6. Zabigaylo, V.E., Nikolin, V.I.: Vliyanie katageneza gornyyh porod i metamorfizma ugley na ih vyibrosoopasnost, Naukova dumka (1990)
7. Bridzhmen, P.V.: Issledovanie bolshih plasticheskikh deformatsiy razryivov, Izd-vo inostrannoy literaturyi (1955)
8. Levkin, N.B.: Predotvraschenie avariyy i travmatizma v ugolnyih shahtah Ukrainyi, Donbass (2002)
9. Stefanov, Yu.P.: Chislennoe modelirovanie deformirovaniya i razrusheniya gornyyh porod na primere rascheta povedeniya obraztsov peschanika. In: Fiziko-tehnicheskie problemy razrabotki poleznyih iskopaemyih, vol. 1, pp. 73-83 (2008)
10. Stefanov, Yu.P.: Nekotoryie osobennosti chislennogo modelirovaniya povedeniya uprugohrupkoplastichnyih materialov. In: Fizicheskaya mezomehanika, vol. 8, pp. 129-142 (2005)

11. Wilkins, M.L.: Raschet uprugoplasticheskikh techeniy. In: Vyichislitelnyie metodyi v gidrodinamike, pp. 212-263 (1967)
12. Nikolaevskiy, V.N.: Mehanicheskie svoystva gruntov i teoriya plastichnosti. In: Mehanika tverdyih deformiruemyih tel. Itogi nauki i tehniki, vol. 6, pp. 5-85 (1972)
13. Garagash, I.A., Nikolaevskiy, V.N.: Neassotsiirovannyye zakonyi techeniya i lokalizatsii plasticheskoy deformatsii. In: Uspehi mehaniki, vol. 12, pp. 131-183 (1989)
14. Stefanov, Yu.P.: Numerical investigation of deformation localization and crack formation in elastic brittle-plastic materials. In: J.Fract., vol. 128, pp. 345-352 (2004)
15. Wilkins, M.L.: Computer Simulation of Fracture. Lawrence Livermore Laboratory, Rept. UCRL-75246 (1972)
16. Wilkins, M.L.: Computer Simulation of Dynamic Phenomena, Berlin–Heidelberg–New York: Springer-Verlag (1999)

Irregular layout problem for additive production

Andrey Chugay^[0000-0002-4079-5632], Aleksandr Pankratov^[0000-0002-2958-8923],
Tetyana Romanova^[0000-0002-8618-4917]

A. Pidgorny Institute of Mechanical Engineering Problems of the National Academy of
Sciences of Ukraine, 2/10 Pozharsky St., 61046 Kharkiv, Ukraine

chugay.andrey80@gmail.com, pankratov2001@yahoo.com,
tarom27@yahoo.com

Abstract. One of the interesting applications of optimization layout problems is additive production. The problem of layout of 3D objects (parts) inside a container (a working chamber of a 3D printer) to minimize the container height is studied. It is aimed to reduce printing costs by minimizing the number of 3D-printing layers while reducing the number of the printer starts. A mathematical model of the layout problem is provided in the form of nonlinear programming problem using the phi-function technique. A solution algorithm to search for optimized layouts is proposed. Computational results demonstrate the efficiency of our approach.

Keywords: additive production, packing, mathematical modeling, phi-function, quasi phi-function, nonlinear optimization.

1 Introduction

Optimization 3D layout problems have a wide spectrum of real-word applications, including transportation, logistics, chemical and aerospace engineering, shipbuilding, robotics, additive manufacturing, materials science. In this paper the smart technique to optimize the 3D-printing process for selective laser sintering (SLS) additive manufacturing [1] is developed. The SLS technology uses high power laser sintering for small particles of plastic, ceramic, glass or metal flour in three-dimensional structure.

This technology empowers the fast, flexible, cost-efficient, and easy manufacture of prototypes for various application of required shape and size by using powder based material. A physical prototype is an important for design confirmation and operational examination by creating the prototype unswervingly from CAD data. The main feature of this technology is the use of powder, consisting of particles of metal coated polymer. After the sintering process piece is placed in a high temperature kiln to burn plastic and fusible took the bronze. The advantages of the technology include no need for material support. Parts immersed into a powder, which works on as a support [2].

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

Recently 3D-prototyping technologies are evolving rapidly. The purpose of the research is development of smart technology to improve 3D-printing process for advanced additive production. We propose the approach for accelerating printing cycle due to the simultaneous printing of several parts providing dense filling the entire volume of the working chamber 3D printer using SLS technology.

One of the important problems arising in the process of creating new prototypes (final products) is reducing the time and cost production. For each start of SLS printer requires time and energy for heating and maintaining temperature. In [3] data on what savings can be achieved by optimizing the layout of objects to be created are provided.

Our approach allows optimizing the process of 3D printing for the following factors:

- printing of several prototypes (products) providing dense filling the volume of the 3D printer working chamber [4];
- minimizing the time and cost of 3D parts production by reducing printing cycle.

In this paper the optimization layout problem of irregular 3D objects into optimized cuboid is studied.

Our approach is based on the mathematical modelling of relations between irregular geometric objects by means of the phi-function technique. It allows us reducing the layout problem to nonlinear programming model.

2 Literature review

The list of publications related to the layout problem of irregular 3D objects, taking into account the minimum allowable distances is very scarce within the field of Packing and Cutting. Arbitrary shaped objects in most cases are approximated by sets of cuboids or spheres. To solve the layout problems heuristic and meta-heuristic algorithms are used that resulting in the loss of optimal solutions.

3D object layout problems is NP-hard. In order to find feasible solutions some researchers use different techniques, including heuristics (based on different approximation rules heuristics [5], genetic algorithms [6], simulated annealing algorithms [7], artificial bee colony algorithms [8]), extended pattern search [8], traditional optimization methods [9, 10], nonlinear mathematical programming [11].

In the majority of papers, either orientation of 3D objects is fixed or only discrete rotations (by 45 or 90 degrees) are allowed. In particular, paper [2] uses the parallel translation algorithm for packing convex polytopes. The authors of [12] propose the HAPE3D algorithm which can be applied to arbitrarily shaped polyhedra that can be rotated around each coordinate axis at eight different angles. In [13] the issue is discussed that for 3D packing problems making calculations of 0 to 360 degrees orientations of objects with respect to each axis is impossible. Analysis of irregular three-dimensional packing problems in additive manufacturing is provided in [14]. The paper [15-22] presents an intelligent layout planning for rapid prototyping. Only few works consider continuous rotations of 3D objects (see, e.g. [16-22]).

3 Problem statement

In order to minimize the time of 3D parts production using SLS-technology the number of layers should be minimized. The problem of minimizing layers can be formulated as a problem of layout (packing) of parts in the container of minimum height (fig.1).

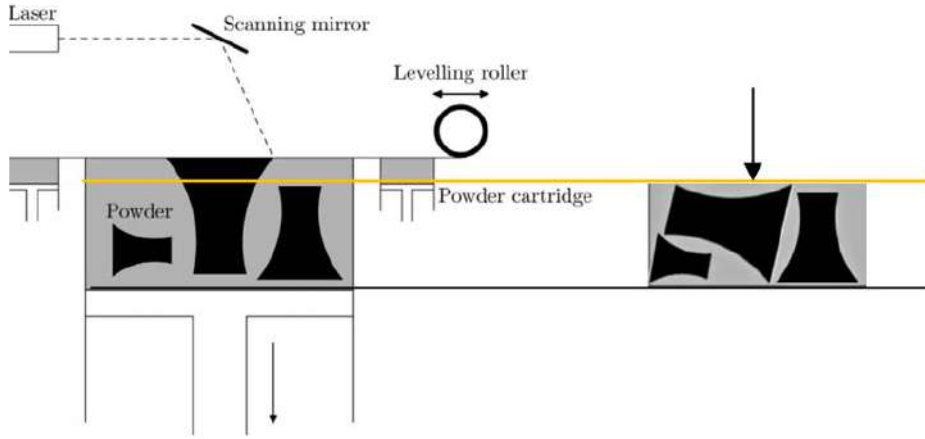


Fig. 1. Minimizing of the height of the occupied part of the 3D printer working chamber

Let the set of irregular 3D objects $T_i, i \in I_n = \{1, 2, \dots, n\}$, and container $\Omega = \{(x, y, z) \in R^3, 0 \leq w_1 \leq x \leq w_2, 0 \leq l_1 \leq y \leq l_2, 0 \leq h_1 \leq z \leq h_2\}$ be given. Here h_1 and h_2 are variable. Denote the container Ω of variable sizes by $\Omega(h_1, h_2)$.

Each object T_i is presented by a union of convex polyhedra

$$T_i = \bigcup_{k=1}^{n_i} T_{ik}, i \in I_n,$$

where T_{ik} is defined by the collection of vertices $\{p_{ik}\}$.

Layout of T_i in R^3 determined by the translation vector $v_i = (x_i, y_i, z_i)$ and the vector of rotation angles $\theta_i = (\alpha_i, \beta_i, \gamma_i)$, $i \in I_n$. Thus, vector $u_i = (v_i, \theta_i)$ determines placement of P_i in the three-dimensional space R^3 .

Further object T_i , translated on the vector v_i and rotated by angles $\alpha_i, \beta_i, \gamma_i$ is denoted by $T_i(u_i)$.

Optimization layout problem. Find vector $u = (u_1, \dots, u_n)$ that provides layout of

objects $T_i(u_i)$, $i \in I$, inside the container $\Omega(h_1, h_2)$ so that the height $H = h_2 - h_1$ will reach the minimum value.

4 Mathematical model and its properties

Using the phi-function technique [16-22] a mathematical model of the optimization layout problem can be presented as the following nonlinear programming problem:

$$\min_{X \in W} H, \quad (1)$$

$$W = \{X \in R^m : \Phi'_{ij}(u_i, u_j, u'_{ij}) \geq 0, i < j \in I_n, \Phi_i(u_i, h_1, h_2) \geq 0, i \in I_n, h_2 - h_1 \geq 0\}, \quad (2)$$

where $X = (h_1, h_2, u, u')$, $\Phi'_{ij}(u_i, u_j, u'_{ij})$ is the quasi phi-function for polyhedra T_i and T_j [18, 21], $u' = (u'_{ij}, i < j \in I_n)$, u'_{ij} is the vector auxiliary variables for the quasi phi-function $\Phi'_{ij}(u_i, u_j, u'_{ij})$, $\Phi_i(u_i, h_1, h_2)$ is the phi-function for objects T_i and $\Omega^* = R^3 \setminus \text{int}\Omega$.

The inequality $\Phi'_{ij}(u_i, u_j, u'_{ij}) \geq 0$ provides non-overlapping T_i and T_j and inequality $\Phi_i(u_i, h_1, h_2) \geq 0$ provides containment of T_i into Ω .

The problem (1)-(2) is an exact formulation of the optimization layout problem of 3D objects.

The feasible region W of the problem (1)-(2), in the general case, is a disconnected set, and each of its connected components is a multiply connected.

5 Solution approach

Our solution approach is addressed to the placement of non-convex continuously rotated objects. To construct feasible starting points the clustering algorithm is proposed. Local optimization is performed using the IPOPT code combined with the decomposition strategy. To search for local extrema, a multistart strategy is used.

Firstly we solve the problem of clustering of pairs of 3D objects into optimized containing spheres or cuboids. Then depending on the shape of clusters auxiliary subproblems of packing cuboids or spheres are solved, employing the clusters homothetic transformations. This allows constructing fast feasible starting points.

The reduction of computational costs is also facilitated by the fact that the process of finding a local extremum of the problem is divided into two stages: solving NLP subproblems by fixing the rotation angles and solving NLP subproblems allowing free object rotations. In addition, the strategy of finding an approximation to the global extremum is used.

As an approximation to the global minimum of the optimization layout problem (1)-(2) the best local minimum found by our approach is considered.

5.1 Generation of feasible starting points

In order to generate a feasible starting point for problem (1) - (2) we use the following algorithm. Firstly, pairs of non-overlapping objects are placed into containing regions (cuboids or spheres) of the minimum volume. Then we solve the problem of packing the set of the obtained clusters into the container (cuboid) of minimum height. This algorithm returns feasible placement parameters for each polyhedron. To compute rotation angles of each of polyhedra the following algorithm is proposed.

The set of objects T_i , $i \in I_n$, is divided into k groups. Each group involves l_k identical polyhedra.

Each object T_i is contained into the sphere S_i of minimum radius r_i^* , using the following NLP subproblem:

$$r_i^* = \min_{(v_i, r_i) \in D_i \subset R^4} r_i, \quad i \in I_n,$$

$$D_i = \{(v_i, r_i) \in R^4 : \Psi_{ij} = r_i^2 - (x'_{ij} - x_i)^2 - (y'_{ij} - y_i)^2 - (z'_{ij} - z_i)^2 \geq 0, j \in J_i\}.$$

Denote a local minimum point of the subproblem by (v_i^*, r_i^*) . Then each object T_i is translated by the vector v_i^* .

Further $C_n^2 + n$ subproblems of packing the objects T_i , $i \in I_n$, into cuboid Ω_{ij} of minimum volume D_{ij}^C are solved:

$$r_i^* = \min_{(u_i, u_j, h_1, h_2) \in W_{ij} \subset R^{18}} D_{ij}^C(h_1, h_2), \quad (3)$$

$$\begin{aligned} W_{ij} = \{ & (u_i, u_j, h_1, h_2) \in R^{18} : \Phi_{ij}(u_i, u_j) \geq 0, \Phi_i(u_i, h_1, h_2) \geq 0, \\ & \Phi_j(u_j, h_1, h_2) \geq 0, F(h_1, h_2) \geq 0 \} \end{aligned} \quad (4)$$

where $i < j \in I_n$,

$$D_{ij}^C(h_1, h_2) = (h_2 - h_1)(w_2 - w_1)(l_2 - l_1), \quad F(h_1, h_2) = \min\{h_2 - h_1, w_2 - w_1, l_2 - l_1\}.$$

The inequality $\Phi_{ij}(u_i, u_j) \geq 0$ implies that $\text{int } T_i \cap \text{int } T_j = \emptyset$, while the inequalities $\Phi_i(u_i, h_1, h_2) \geq 0$ and $\Phi_j(u_j, h_1, h_2) \geq 0$ guarantee the arrangement of T_i and T_j fully inside containing region Ω_{ij} .

Next we solve the layout problem of subset of clusters Q_i , $i \in M$, inside the cuboid Ω of minimum height.

Now the problem (1)-(2) is reduced to the following NLP model:

$$\min_{(\tilde{u}, h_1, h_2) \in \tilde{W} \subset R^{6\mu+6}} H(h_1, h_2), \quad (5)$$

$$\begin{aligned} \tilde{W} = \{(\tilde{u}, h_1, h_2) \in R^{6\mu+6} : \Phi_{ij}(\tilde{u}_i, \tilde{u}_j) \geq 0, i < j \in M, \Phi_i(\tilde{u}_i, h_1, h_2) \geq 0, \\ i \in M, h_2 - h_1 \geq 0\}, \end{aligned} \quad (6)$$

where $H(h_1, h_2) = h_2 - h_1$.

Let the point $(\tilde{u}^*, h_1^*, h_2^*) \in R^{6\mu+6}$ be an approximation to the global minimum point of the problem (5) - (6). The point corresponds to packing clusters $Q_i(u_i^*)$, $i \in M$ into cuboid $\Omega(h_1^*, h_2^*)$. Each cluster Q_i contains the pair of polyhedra T_{k_i} and T_{l_i} with placement parameters $u_{k_i}^Q$ and $u_{l_i}^Q$ in the local coordinate system of the cluster Q_i .

In order to construct a feasible point $(u^0, h_1^0, h_2^0) \in W$ of the problem (1) - (2) regarding the arrangement of clusters Q_i , $i \in M$, we set the arrangement of object T_i using the equation $v_i^0 = \tilde{v}_i^* + v_i^Q$ for $i \in I_n$.

To define the rotation angles θ_i^0 of polyhedra T_i , $i \in I_n$, we solve the sequence of n subproblems of the following form:

$$r_{13}^{i*} = \min_{R_i \in D_i \subset R^9} r_{13}^i, \quad (7)$$

$$\begin{aligned} D_i = \{R_i \in R^9 : V_1^i R^i = \tilde{V}_1^i, V_2^i R^i = \tilde{V}_2^i, V_3^i R^i = \tilde{V}_3^i, \sum_i r_{ij} r_{ik} = \delta_{jk}, \\ \sum_i r_{ji} r_{ki} = \delta_{jk}, i = 1, 2, 3\} \end{aligned} \quad (8)$$

where V_1^i, V_2^i, V_3^i are vectors of initial coordinates of the first three vertices of the polyhedron P_i , $\tilde{V}_j^i = \tilde{R}_i^*(R_i^Q V_j^i + v_i^Q)$, $j = 1, 2, 3$, R_i is the rotation matrix, $i \in I_n$. Let r_i^* be a solution of the problem (7) - (8). Then the angles of T_i can be derived in the form: $\beta_i = \arcsin(r_{13}^{i*})$, $\alpha_i = \arcsin(-r_{23}^{i*} / \cos \beta_i)$, $\gamma_i = \arccos(-r_{12}^{i*} / \cos \beta_i)$.

5.2 Local optimization

To find a local extremum of the problem (1)-(2) the following algorithm is used. This algorithm allows reducing CPU.

The feasible region of the problem (1)-(2) can be always represented by a union of subregions (see e.g. [21]). It enables to search for a local minimum of the problem (1)-(2) by solving a collection of NLP subproblems with a considerably smaller number of inequalities.

The key idea of the proposed algorithm is based on the decomposition strategy (see, e.g. [23]). The large scale problem (1)-(2) is reduced to a sequence of subproblems of smaller dimension. The following stages are performed:

- generating feasible subregions of the feasible region (2) related to the appropriate starting points;
- forming the system of ε – active constraints;
- searching for local extrema of the subproblems generated at the first step, employing state-of-the-art NLP-solvers;
- replacing subregions.

Now we consider the algorithm in detail.

Let the point $X^\bullet \in W$ be a starting point. Then we select an appropriate subregion W_0 , such that $X^\bullet \in W_0 \subset W$ and substitute the point $X^\bullet$ in the inequality system (2). Each quasi phi-function has the form

$$\Phi'_{ij}(u_i, u_j, u') = \max \left\{ \Psi_{ij}^s(u_i, u_j, u'), s = 1, \dots, \wp_{ij} \right\}.$$

Then we select one of the functions $\Psi_{ij}^{a_{ij}}(u_i, u_j, u')$, $a_{ij} \in \{1, \dots, \wp_{ij}\}$, $i < j \in I$, such that

$$\Phi'_{ij}(u_i^\bullet, u_j^\bullet, u') = \Psi_{ij}^{a_{ij}}(u_i^\bullet, u_j^\bullet, u') = \chi_{ij}^\bullet.$$

Similarly we choose $\Phi_i(u_i, u_\Omega) \geq 0, i \in I$. It results in the system of inequalities $\Upsilon^0(X) \geq 0$ describing the subregion W_0 . Then the subproblem

$$F(u_\Omega^{0*}) = \min_{X \in W_0 \subset R^3} F(u_\Omega)$$

is solved. The inequality system $\Upsilon^0(X^{0*}) \geq 0$ distinguishes the active inequality

$\varsigma_{j0}(\xi_j^{0*}) \geq 0$, $j \in \Gamma_0 = \{1, \dots, \mu_0\} \subset \Gamma = \{1, \dots, \mu\}$. Denote the subsystem by

$\Psi_{ij}^a(u_i, u_j) \geq 0$, $i \in I_{0\eta_1} \subset I$, $j \in I_{0\eta_2} \subset I$. This allows choosing quasi phi-functions

$\Phi'_{ij}(u_i, u_j, u')$ that involve functions $\Psi_{ij}^a(u_i, u_j)$, for $i \in I_{0\eta_1}$, $j \in I_{0\eta_2}$.

Then we calculate the values of the functions at the point X^{0*} .

Let

$$\Phi'_{ij}(u_i^{0*}, u_j^{0*}) = \Psi_{ij}^c(u_i^{0*}, u_j^{0*}) = \chi_{ij}^0, \quad i \in I_{0\eta_1}, \quad j \in I_{0\eta_2}.$$

If $\chi_{ij}^0 > 0$, $i \in I_{0\eta_1}$, $j \in I_{0\eta_2}$ then replace subsystems $\Psi_{ij}^a(u_i, u_j) \geq 0$ by systems $\Psi_{ij}^c(u_i, u_j) \geq 0$, $i \in I_{0\eta_1}$, $j \in I_{0\eta_2}$. Thus a new subsystem of inequalities defining a

new subregion $W_1 \subset W$ is generated. Obviously, $X^{0*} \in W_1$.

Taking the starting point X^{0*} , we solve the problem

$$F(u_\Omega^{1*}) = \min_{X \in W_1 \subset R^m} F(u_\Omega),$$

and search for a local minimum point X^{1*} .

The computational process is repeated until $F(u_\Omega^{(\rho-1)*}) = F(u_\Omega^{\rho*})$.

The search for a local minimum of the problem (1) - (2) can be divided into two stages: optimization of the system with linear constraints and nonlinear optimization. The first stage is realized by fixing the rotation angles θ_i^0 of objects T_i , $i \in I_n$ at the feasible starting point $(u^0, u_\Omega^0) \in W$. Fixing rotation angles significantly reduces the dimension of the problem (1) - (2) switching to the linear constraints to describe the feasible region.

Figure 3 depicts layout of irregular 3D objects that corresponds to a) a feasible starting point and b) the appropriate local minimum found by our algorithm.

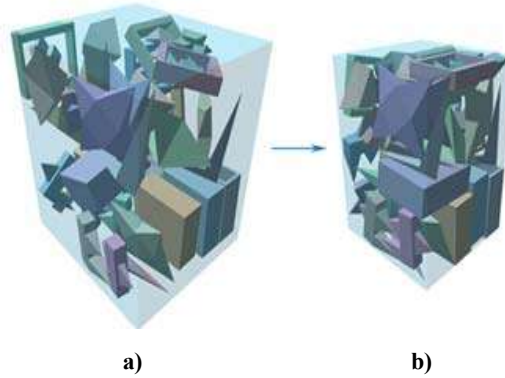


Fig. 1. Example of layouts of irregular objects corresponding: a) a starting point; b) a local minimum point

6 Computation experiments

We present some examples to demonstrate the efficiency of our methodology. We have run all experiments on an Intel I5 2320 computer, programming language C++, Windows 10 OS. To solve NLP problems IPOPT [24] is used, which is available at an open access software depository (<https://projects.coin-or.org/Ipopt>).

Figure 2 demonstrates some benchmark examples of irregular layouts obtained by our approach.

In order to show the efficiency of our approach a number of benchmarks instances

given in [12] are tested. The results are shown in Table 1.

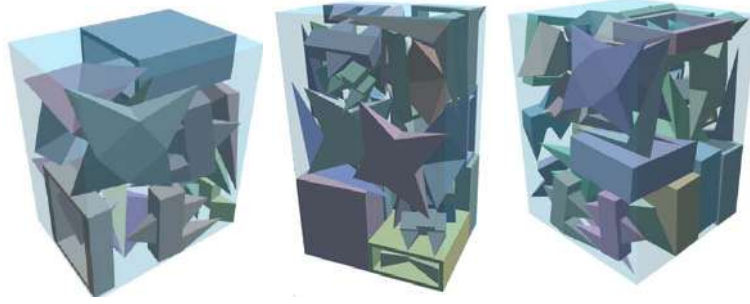


Fig. 2. Examples of 3D irregular object layouts

Table 1. Comparison of our results with those published in [12]

<i>Approach</i>	<i>HAPE3D</i>	<i>Our algorithm</i>
The result of packing 20 irregular 3D objects		
Volume	32550	28500
Runtime (sec)	26202	6656
The result of packing 30 irregular 3D objects		
Volume	12480	10720
Runtime (sec)	9637	4789
The result of packing 36 irregular 3D objects		
Volume	48300	42450
Runtime (sec)	53741	9543
The result of packing 40 irregular 3D objects		
Volume	61950	56012
Runtime (sec)	99952	24543
The result of packing 50 irregular 3D objects		
Volume	77280	71800
Runtime (sec)	125210	36543

7 Conclusions

The 3D-printing procedure using SLS technology takes a long time (many hours or even days) and requires a great financial cost associated with: the printer running, the

camera heating and the temperature stabilization. Development of the optimization techniques allowing saving time and material is of paramount importance.

The optimization problem of layout of irregular 3D objects into cuboid of minimum height is formulated. The mathematical model is constructed, using the phi-function technique. The solution strategy is proposed. To demonstrate the efficiency of our methodology some instances are provided. Obtainment of optimized layouts of 3D objects makes possible reducing the printing cost by minimizing the number of layers of 3D printing and therefore reducing the number of the printer starts.

Acknowledgments: The authors would like to thank anonymous referees for careful reading the paper and constructive comments. The authors were partially supported by the “Program for the State Priority Scientific Research and Technological (Experimental) Development of the Department of Physical and Technical Problems of Energy of the National Academy of Sciences of Ukraine” (#6541230).

References

1. Ramya, A., Sai Vanapalli: 3D printing technologies in various applications. *International Journal of Mechanical Engineering and Technology* 7(3), 396-409 (2016).
2. Egeblad, J., Nielse, B.K., Brazil, M.: Translational packing of arbitrary polytopes. *Computational Geometry* 42, 269–288 (2009).
3. Baumann, M., Dickens, P., Tuck, C., Hague, R.M.: The cost of additive manufacturing: machine productivity, economies of scale and technology-push. *Technological Forecasting & Social Change*, 102, 193-201 (2016).
4. Romanova, T., Stoyan, Y., Pankratov, A., Litvinchev, I., Avramov, K., Chernobryvko, M., Yanchevskiy, I., Mozhova, I., Bennell, J.: Optimal layout of ellipses and its application for additive manufacturing, *International Journal of Production Research*, 27 pages, (2019).
5. Verkhoturov, M., Petunin, A., Verkhoturova, G., Danilov, K., Kurenkov, D.: The 3D Object Packing Problem into a Parallelepiped Container Based on Discrete-Logical Representation. *IFAC-PapersOnLine* 49(12), 1-5 (2016).
6. Karabulut, K., İnceoğlu, M.: A Hybrid Genetic Algorithm for Packing in 3D with Deepest Bottom Left with Fill Method. *Advances in Information Systems* 3261, 441-450 (2004).
7. Cao, P., Fan, Z., Gao, R., Tang, J.: Complex Housing: Modeling and Optimization Using an Improved Multi-Objective Simulated Annealing Algorithm. *Proceedings of the ASME*, 60563, V02BT03A034 (2016).
8. Guangqiang, L., Fengqiang, Z., Rubo, Z., Du, Jialu, Du., Chen, G., Yiran, Z.: A Parallel Particle Bee Colony Algorithm Approach to Layout Optimization. *Journal of Computational and Theoretical Nanoscience* 13(7), 4151-4157 (2016).
9. Torczon, V., Trosset, M.: From evolutionary operation to parallel direct search: Pattern search algorithms for numerical optimization. *Computing Science and Statistics* 29, 396-401 (1998).
10. Birgin, E.G., Lobato, R.D., Martinez, J.M.: Packing ellipsoids by nonlinear optimization. *Journal of Global Optimization* 65, 709-743 (2016).
11. Fasano, G.: A global optimization point of view to handle non-standard object packing problems. *Journal of Global Optimization* 55(2), 279 –299 (2013).
12. Liu, X., Liu, J., Cao, A., Yao, Z.: HAPE3D - a new constructive algorithm for the 3D

irregular packing problem. *Frontiers of Information Technology & Electronic Engineering* 16(5), 380-390 (2015).

13. Youn-Kyoung, Joung, Sang Do Noh.: Intelligent 3D packing using a grouping algorithm for automotive container engineering. *Journal of Computational Design and Engineering* 1(2), 140-151 (2014).
14. Araújo, L. J. P., Özcan, E., Atkin, J. A. D., & Baumann, M. (2019). Analysis of irregular three-dimensional packing problems in additive manufacturing: a new taxonomy and dataset. *International Journal of Production Research*, 57(18), 5920–5934.
15. Gogate, A. S., & Pande, S. S. (2008). Intelligent layout planning for rapid prototyping. *International Journal of Production Research*, 46(20), 5607–5631.
16. Stoyan, Y.G., Chugay, A.M.: Packing different cuboids with rotations and spheres into a cuboid. *Advances in Decision Sciences*. Available at <https://www.hindawi.com/journals/ads/2014/571743> (2014).
17. Stoyan, Y.G., Semkin, V.V., Chugay, A.M.: Modeling Close Packing of 3D Objects. *Cybernetics and Systems Analysis* 52 (2), 296-304 (2016).
18. Stoyan, Y.G., Chugay, A.M.: Mathematical modeling of the interaction of non-oriented convex polytopes. *Cybernetic and Systems Analysis* 48, 837-845 (2012).
19. Grebennik, I.V., Pankratov, A.V., Chugay, A.M., Baranov, A.V.: Packing n-dimensional parallelepipeds with the feasibility of changing their orthogonal orientation in an n-dimensional parallelepiped. *Cybernetics and Systems Analysis* 46(5), 793-802 (2010).
20. Stoyan, Y., Pankratov, A., Romanova, T., Fasano, G., Pintér, J.D., Stoian, Y.E., Chugay, A.: *Optimized Packings in Space Engineering Applications*, Springer Optimization and Its Applications, 144, 478 (2019).
21. Romanova, T., Bennell, J., Stoyan, Y., Pankratov, A.: Packing of concave polyhedra with continuous rotations using nonlinear optimisation, *European Journal of Operational Research*, 268 (1), 37-53, (2018).
22. Kovalenko, A.A., Romanova, T.E., & Stetsyuk, P.I. (2015). Balance Layout Problem for 3D-Objects: Mathematical Model and Solution Methods. *Cybernetics and Systems Analysis*, 51(4), 556–565.
23. Romanova, T., Stoyan, Y., Pankratov, A., Litvinchev, I., Marmolejo, J.A.: Decomposition Algorithm for Irregular Placement Problems. In: Vasant P., Zelinka I., Weber GW. (eds) *Intelligent Computing and Optimization*. *Advances in Intelligent Systems and Computing*, 1072, 214-221 (2019).
24. Wachter, A., Biegler, L.T.: On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Math. Program.* 106 (1), 25-57 (2006).

The Data Errors Control in the Modular Number System Based on the Nullification Procedure

Victor Krasnobayev^{1[0000-0001-5192-9918]}, Alexandr Kuznetsov^{1[0000-0003-2331-6326]},
Alina Yanko^{2[0000-0003-2876-9316]} and Kateryna Kuznetsova^{1[0000-0002-5605-9293]}

¹ V. N. Karazin Kharkiv National University, Svobody sq., 4, Kharkiv, 61022, Ukraine
v.a.krasnobayev@gmail.com, kuznetsov@karazin.ua,
kate.kuznetsova.2000@gmail.com

² Poltava National Technical Yuri Kondratyuk University, Poltava, Ukraine
al9_yanko@ukr.net

Abstract. A method for error control in the modular number system (MNS) based on the use of the zeroing procedure is proposed. Error control in the MNS is a non-positional operation and requires the development of special methods, designed to increase the efficiency of this procedure. This method is designed to verify the correct implementation of the computing process of computer systems and components. It is assumed that the error in one module remainder does not affect the residual values corresponding to other modules (bases) of the MNS. The essence of the proposed method is that, when performing the procedure of zeroing in the MNS, the operation of determining is combined in time, in accordance with the digits $a_i^{(i-1)}$ and $a_{n-i+1}^{(i-1)}$ of the number $A^{(i-1)}$, the zeroing constant $ZC^{(i)}$ and the calculation operation for the values of $a_i^{(i-1)}$ and $a_{n-i+1}^{(i-1)}$ of the following digits $a_{i+1}^{(i)}$ and $a_{n-i}^{(i)}$ of the number $A^{(i)}$. This makes it possible to increase the efficiency of information control, presented in the modular number system.

Keywords: Automated Design Systems; Computer System and Components; Method for Error Control; Modular Number System; Pair Number Zeroing with Preliminary Selection of Digits; Positional Number Systems; Zeroing Block; Zeroing Procedure.

1 Introduction

The main direction of modern science and technology is the development and use of new advanced information technologies based on the extensive use of computer systems and components (CSC). Information technologies are increasingly invading our lives, penetrating all processes (social, economic, political). Scales and complexity of the tasks solved by modern computer systems impose qualitatively new requirements to their main characteristics: productivity, reliability and efficiency of systems that causes need of improvement existing and creations of new means of information processing. In modern computer systems, an improvement in one group of quality

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

indicators, for example, an increase in productivity, leads to a deterioration in others – a complication of structure, an increase in cost, a decrease in reliability, etc. [1-7].

In connection with the constant complication of scientific and technical problems of processing integer data, the trend of development of CSC is aimed at increasing the speed (productivity) and reliability of the implementation of integer arithmetic operations [3, 7-9]. The results of recent years in the field of information technology by various groups of the scientists and engineers of methods for increasing the productivity, reliability, survivability and also reliability of calculations of computer systems have shown that it is practically impossible to achieve this within the limits of the positional number systems (PNS) [9-13].

This is due to the main disadvantage of modern CSC, functioning in the PNS: the presence of inter-digit relations between the processed numbers. These relations negatively affect the architecture of the CSC and the methods of implementing arithmetic operations, complicate the equipment, they limit the speed and reliability of performing arithmetic operations. In this regard, in the PNS, the increase in the performance of the CSC is achieved by increasing the clock frequency, as well as through the use of methods and tools for parallel data processing, and also by use of different types of reservation [14-18].

Application of the basic methods of increasing the productivity of the CSC, based on the parallelization of computations, by using some properties of solvable tasks and algorithms cannot increase the productivity of CSC in each and every case. The scope of their application is limited to a class of tasks to be solved. In addition, the process of artificial dismemberment of the algorithm itself, the determination and allocation of independent computing branches and related operations requires large labor costs, and it is not always possible to parallelize arbitrary algorithms in general. It should be noted that all existing methods of increasing productivity in PNS have a general disadvantage: the impossibility of parsing the maximum algorithms that are solved at the level of elementary operations.

However, this approach does not always solve the problem of cardinal increase in speed and reliability of performing arithmetic operations in the PNS.

To date, there has been a gap between the increasing requirements for improving the performance of real-time computer systems, on the one hand, and the impossibility of satisfying these requests based on the use of existing PNS, on the other hand.

This fact led to the need to find ways to increase productivity, for example, based on the use of new structural solutions in the creation of CSC.

Scientific researches were conducted in recent years, identify promising ways to improve the performance of computer systems, which are based on the use of the modular number system (MNS) [7-11]. However, in existing researches little attention is paid to issues devoted to the implementation of positional operations in the MNS [13-15]. This article focuses on solving this problem.

2 Research methodology and analysis of results

2.1 Search of ways of increase in reliability

Currently, intensive searches are underway to improve the efficiency of arithmetic operations through the development and implementation of reliable and fast real-time CSC.

The results of the studies devoted to the improvement of the characteristics of CSC showed that one really practical direction is the approach based on the use of MNS codes [3, 10, 12]. Ascending from the known Chinese remainder theorem (the task of restoring the original number A_k by the aggregating of its remains (deductions) $\{a_i\}$ by dividing it into a series of natural numbers $m_1, m_2, \dots, m_n$ (modules) of MNS), which was previously interpreted as a structural theorem of abstract algebra, guaranteed the specified parallelism in the calculations over integers, under the conditions that the result of ring operations belongs to the range of integers, defined by modules product of MNS. Having its ideological roots of the classical works of Euler, Gauss and Chebyshev on the theory of comparisons, MNS introduced new ideas in the development of creation methods of highly-productive and ultra reliable CSC.

At present again interest in use of MNS as tool for increase in productivity, reliability, survivability and also reliability of calculations of computer systems increases. It is caused primarily by the following circumstances:

- the emergence of the numerous scientific and theoretical publications devoted to the theory and practice of the computer systems and components creating in MNS;
- wide distribution of mobile processors that require high speed data processing at low energy consumption; the lack of inter-bits transfers during arithmetic operations of addition and multiplication of numbers in MNS allows to reduce energy consumption;
- strong interest to MNS is being shown by the banking structures, where it is necessary in real time to handle large amount of data safely and reliably, i.e. they are required highly-productive means for highly reliable computing with errors self-correction, that is typical to the MNS codes;
- the elements density increasing on a single chip doesn't always allow to perform a complete and qualitative testing; in this case there is an increasing importance of providing failover operation of CSC;
- the need for the use of the specialized CSC to perform a large number of operations on vectors, which require high-speed performance of integer addition and multiplication operations (matrix multiplication problems, the problems of the scalar product of vectors, Fourier transformation, etc.);
- the widespread introduction of microelectronics into all spheres of human activity significantly increased relevance and importance of previously rare, and now so massive scientific and practical problems, as a digital signal and image processing, image recognition, cryptography, multi-bit data processing and storage, etc.; this circumstance requires enormous computing resources being in excess of the existing possibilities;

– the current level of microelectronics development is coming to its limits from the point of view of productive provision and reliability of existing and future computer systems and components of large data sets processing in real time;

– taking it over nanoelectronics, molecular electronics, micromechanics, bioelectronics, optical, optoelectronic and photonic computers and others are still rather far from the real industrial production and employment.

– the modern development of integrated circuit technology allows to have a fresh look at the principles of devices construction with modular arithmetic employment and provides wide opportunities to use new design techniques (such as the methodology of systems design on a chip-SoC) both in the development of individual computing units, and computer systems in general; integral technology enables more flexible design of computer systems and components and allows us to implement MNS -based devices as effectively as on the basis of the binary system; furthermore at present in order to improve the effectiveness of computer devices development, automated design systems (ADS) are widely used; in this respect, the design of computer systems and components based on MNS does not differ from the working with the help of ADS data of binary data-blocks in PNS;

– unfortunately, Ukraine today in contrast to the theoretical development, technologically is behind the foreign microelectronics of some leading countries; in this case, it is advisable to use the existing theoretical achievements and practical experience in the creation of effective computer systems and components in MNS [10, 12, 14, 16-18].

One of the disadvantages of MNS is that there are no simple signs of the output of the result of operations outside the operating range $[0, M)$, where:

$$M = \prod_{i=1}^n m_i$$

– operating range; m_i – i -th MNS base; n – number of operating bases of MNS.

This requires additional time to implement the error correction process. This circumstance reduces the effectiveness of the use of MNS in the CSC.

At the heart of the majority of control methods of data are based on the analysis of information, that is on comparison of data. Therefore researches and development of mathematical models, methods and algorithms of comparison of numbers in MNS is an important and relevant task. Now it is possible to allocate three groups of methods of comparison of numbers in MNS [10-12].

The first group includes methods of direct comparison, based on the conversion of numbers A_{MNS} and B_{MNS} from a code MNS at PNS:

$$A_{PNS} = \overline{\alpha_{\rho-1}\alpha_{\rho-2}\dots\alpha_0}$$

and

$$B_{PNS} = \overline{\beta_{\rho-1}\beta_{\rho-2}\dots\beta_0}$$

(ρ – digits number (number length) A_{MNS} and B_{MNS}) and their further comparison on the basis of use of binary position adders.

The second group of methods includes methods based on the principle of zeroing. The procedure for the zeroing process consists in transition from initial number:

$$A_{MNS} = (a_1, a_2, \dots, a_{i-1}, a_i, a_{i+1}, \dots, a_n)$$

presented to MNS to the number of species:

$$A_{MNS}^{(z)} = (0, 0, \dots, 0, \gamma_n^{(A)}).$$

Then, on value $\gamma_n^{(A)}$ the interval $[jm_i, (j+1)m_i)$ of hit of number is defined A_{MNS} . The number zeroing is performed in the same way:

$$B_{MNS} = (b_1, b_2, \dots, b_{i-1}, b_i, b_{i+1}, \dots, b_n)$$

from where we receive values $\gamma_n^{(B)}$. Position comparison of the received values $\gamma_n^{(A)}$ and $\gamma_n^{(B)}$ defines result of comparison of numbers A_{MNS} and B_{MNS} [10-14].

To the third group of methods, we will assign the methods based on the definition (allocation) or the formation of special features, the so-called positional features of the non-positional code.

To detect errors in MNS, the most commonly used procedure is zeroing. The essence of the procedure consists in the successive subtraction from the initial number:

$$A = (a_1, a_2, \dots, a_n, a_{n+1})$$

of certain minimum numbers $ZC(i)$ – zeroing constants such that the number A is successively transformed into a number of type:

$$A^{(n)} = (0, 0, \dots, 0, \gamma_{n+1})$$

in n cycles. If the obtained value of the remainder on the control basis $\gamma_{n+1} \neq 0$, then it is assumed that the number A is erroneous. In this case, the zeroing constants must be chosen in such a way that in the subtractions such as $A - ZC^{(i)}$ the output of the number outside the operating $[0, M)$ range [10, 12] would not take place. A significant disadvantage of methods of error detection in MNS is the need for significant time and hardware costs in the implementation, which causes significant unproductive computing costs [14-16].

The purpose of this article is the development and research of the error control method in MNS based on the application of the zeroing procedure.

2.2 Method of errors control

In general, the essence of the procedure of the process of zeroing consists of the sequence of the following operations.

Stage 1. Initial checked number:

$$A = A^{(0)} = (a_1^{(0)}, a_2^{(0)}, \dots, a_i^{(0)}, a_{i+1}^{(0)}, \dots, a_n^{(0)}, a_{n+1}^{(0)}) \quad (1)$$

Expression (1) is successively reduced to the form:

$$A^{(H)} = (0, 0, \dots, 0, 0, \gamma_{n+1})$$

by means of a subtraction operation sequence that does not result in the output of a numerical value of the $A^{(0)}$ number outside of the operating range $[0, M)$ of MNS. As noted earlier, this operation in MNS is called zeroing, and consists from successive subtraction (from one of the MNS bases) from the initial number $A^{(0)}$ of minimum numbers, the so-called zeroing constants ($ZC^{(i)}$) of the form:

$$\begin{aligned} ZC^{(1)} &= (t_{1,1}, t_{2,1}, t_{3,1}, \dots, t_{n,1}, t_{n+1,1}), t_{1,1} = \overline{1, m_1 - 1}; \\ ZC^{(2)} &= (0, t_{2,2}, t_{3,2}, \dots, t_{n,2}, t_{n+1,2}), t_{2,2} = \overline{1, m_2 - 1}; \\ ZC^{(3)} &= (0, 0, t_{3,3}, \dots, t_{n,3}, t_{n+1,3}), t_{3,3} = \overline{1, m_3 - 1}; \\ ZC^{(i)} &= (0, 0, \dots, 0, t_{i,i}, t_{i+1,i}, \dots, t_{n,i}, t_{n+1,i}), t_{i,i} = \overline{1, m_i - 1}; \\ ZC^{(n)} &= (0, 0, \dots, 0, t_{n,n}, t_{n+1,n}), t_{n,n} = \overline{1, m_n - 1}. \end{aligned} \quad (2)$$

Next, the initial checked number:

$$A = A^{(0)} = (a_1^{(0)}, a_2^{(0)}, \dots, a_i^{(0)}, a_{i+1}^{(0)}, \dots, a_n^{(0)}, a_{n+1}^{(0)})$$

is successively reduced to the form $A^{(H)}$, that is,

$$A = A^{(0)} = (a_1^{(0)}, a_2^{(0)}, \dots, a_i^{(0)}, a_{i+1}^{(0)}, \dots, a_n^{(0)}, a_{n+1}^{(0)})$$

$$A^{(1)} = (0, a_2^{(1)}, a_3^{(1)}, \dots, a_n^{(1)}, a_{n+1}^{(1)}),$$

$$A^{(2)} = (0, 0, a_3^{(2)}, \dots, a_n^{(2)}, a_{n+1}^{(2)}),$$

$$A^{(3)} = (0, 0, 0, a_4^{(3)}, \dots, a_n^{(3)}, a_{n+1}^{(3)})$$

and so on.

Repeating the subtraction n times we get the value $A^{(H)} = (0, 0, \dots, 0, a_{n+1}^{(n)})$, or $A^{(H)} = (0, 0, \dots, 0, \gamma_{n+1})$, where $\gamma_{n+1} = a_{n+1}^{(n)}$. The general scheme of subtraction $A^{(i)} = A^{(i-1)} - ZC^{(i)}$ involving zeroing constants (2) is presented in the following form:

$$\begin{array}{r} A^{(i-1)} = (0, 0, \dots, 0, a_i^{(i-1)}, a_{i+1}^{(i-1)}, \dots, a_n^{(i-1)}, a_{n+1}^{(i-1)}) \\ - \\ ZC^{(i)} = (0, 0, \dots, 0, a_i^{(i-1)}, t_{i+1,i}, \dots, t_{n,i}, t_{n+1,i}) \\ \hline A^{(i)} = [0, \dots, 0, [a_i^{(i-1)} - a_i^{(i-1)}] \bmod m_i, \\ [a_{i+1}^{(i-1)} - t_{i+1,i}] \bmod m_{i+1}, \dots, [a_{n+1}^{(i-1)} - t_{n+1,i}] \bmod m_{n+1}] \end{array}$$

where:

$$a_{i+1}^{(i)} = (a_{i+1}^{(i-1)} - t_{i+1,i}) \bmod m_{i+1}.$$

Denoting the sampling time ZC from the corresponding zeroing block (ZB) CSC as t_1 , and the subtraction time from the number $A^{(i-1)}$ of constant $ZC^{(i)}$, that is, performing operation $A^{(i)} = A^{(i-1)} - ZC^{(i)}$ – after t_2 , we get the total time for performing the operation of zeroing in the form $T_{H1} = n(t_1 + t_2)$. When presenting ZB in the tabular form, we can assume that practically $t_1 = t_2 = \tau_{add}$. In this case, the zeroing time is equal to the value $T_{H1} = 2n\tau_{add}$, where: τ_{add} – subtraction time from number $A^{(i-1)}$ of zeroing constant $KH^{(i)}$; n – number of information bases of MNS.

Stage 2. After finding the value γ_{n+1} in the first step, the second stage compares γ_{n+1} with zero. If $\gamma_{n+1} = 0$ (number A is in range $[0, M)$), then the conclusion is drawn that the number A is not distorted (correct), i.e. there are no errors. If $\gamma_{n+1} \neq 0$ (number A isn't in range $[0, M)$), then the conclusion is drawn that the number A is distorted (wrong), i.e. there is an error on one of the bases (modules) m_i of MNS. Total time T_1 of error detection is defined as $T_1 = T_{Z1} + T_{C1}$, where T_{C1} – time of comparing γ_{n+1} with zero. Practically time T_{C1} comparison is performed in one clock cycle, in this case it can be assumed that $T_1 \approx T_{Z1} = 2n\tau_{add}$.

The essence of the method of information error detection in MNS proposed in the article is based on the implementation of the procedure of pair number zeroing with preliminary selection of digits (PNZPSD). The PNZPSD procedure is that the zeroing operation in the ZB is combined in time with the BZC selection operation by digits $a_i^{(i-1)}$ and $a_{n-i+1}^{(i-1)}$ of number $A^{(i-1)}$ of the constant $ZC^{(i)}$ and creation operation on values $a_i^{(i-1)}$ and $a_{n-i+1}^{(i-1)}$ of numbers $a_{i+1}^{(i)}$ and $a_{n-i}^{(i)}$. At the same time, the subtraction

operation from the number $A^{(i-1)}$ of the zeroing constant $ZC^{(i)}$ (i.e., operation $A^{(i-1)} - ZC^{(i)}$) and the operation of selecting the next zeroing constant:

$$ZC^{(i+1)} = (0, \dots, 0, t_{i+1,i+1}, t_{i+2,i+1}, \dots, t_{n-1,i+1}, 0, \dots, 0, t_{n+1,i+1}).$$

According to the values of $a_{i+1}^{(i)}$ and $a_{n-i}^{(i)}$ in the next stage of zeroing, on the bases of m_{i+1} and m_{n-i} , we will refer to the BZC for the next zeroing constant:

$$ZC^{(i+1)} = (0, \dots, 0, t_{i+1,i+1}, t_{i+2,i+1}, \dots, t_{n-1,i+1}, 0, \dots, 0, t_{n+1,i+1}).$$

Indeed, the values of Δa_{i+1} and Δa_{n-i} , which will be subtracted from $a_{i+1}^{(i)}$ and $a_{n-i}^{(i)}$, respectively, in order to obtain $a_{i+1}^{(i+1)}$ and $a_{n-i-1}^{(i+1)}$, are determined only by the values of $a_i^{(i-1)}$ and $a_{n-i+1}^{(i-1)}$. The number of clock cycles that are free from addition, during which the reference is made to the BZC CSC and the formation of the next address is equal to the value $\lceil (n+1)/2 \rceil$, ($\lceil x \rceil$ is the integer closest to x , but not exceeding it). At the same time, zeroing is carried out simultaneously on two information bases of MNS $a_1, a_n; a_2, a_{n-1}$, etc. After every two subtractions, one additional time step is required to form the next address and access the accumulator of zeroing constants. In this regard, for every two addition clock cycles ($\tau_{add} = \tau_0$) there is one clock cycle that is free from addition. Let's compare the effectiveness of the method of error detection in the MNS proposed in the article with the existing method based on the procedure of ordinary zeroing.

To quantify the effectiveness of the proposed method, we introduce the notion of an efficiency coefficient:

$$K_{j\text{ ef}}^{(n)} = \frac{T_{Z1}/\tau_{add} - T_{Zj}/\tau_{add}}{T_{Z1}/\tau_{add}} \cdot 100\% \quad (3)$$

where j – number of the zeroing method ($j = 2$, for pairwise zeroing; $j = 3$, for pairwise zeroing with prefetching of digits; $j = 4$, for pairwise number zeroing with prefetching of digits).

Expression (3) can also be represented in the form (4):

$$K_{j\text{ ef}}^{(n)} = \frac{T_{Z1} - T_{Zj}}{T_{Z1}} \cdot 100\%. \quad (4)$$

In accordance with the expression (4), we define the quantitative value $K_{j\text{ ef}}^{(n)}$ for $j = \overline{2,4}$ while $n = 4, n = 6, n = 8, n = 10$ and $n = 16$, i.e. for l -byte machine words ($l = 1, 2, 3, 4, 8$) of CSC.

The resulting calculated data will be placed in Table 1.

Table 1. The value of efficiency coefficient

$l(n)$	1(4)	2(6)	3(8)	4(10)	8(16)
K_{ef}					
$K_{ef}^{(n)}, [\%]$	62	66	62	65	62

Table 1 shows the calculated data $\frac{T}{\tau_{add}}$ of the relative error detection time of information in the MNS for the value of the number n of bases. The number of information bases of the MNS $\overline{n=1,16}$ provides a range of representation of numbers in modern CSC, which makes it possible to use the data obtained when designing them.

Here is an example of a specific technical implementation of the error detection operation in the CSC, which functions in the MNS. Let MNS be given by the bases $m_1 = 3$, $m_2 = 4$, $m_3 = 5$, $m_4 = 7$, $m_5 = 11$ ($n = 4$), i.e. one-byte ($l = 1$) CSC is considered.

In this case, the working numerical range is:

$$M = \prod_{i=1}^4 m_i = 3 \cdot 4 \cdot 5 \cdot 7 = 420 ,$$

and the full range is:

$$M_1 = M \cdot m_{n+1} = 420 \cdot 11 = 4620 .$$

The error distribution intervals are shown in Table 2.

Table 2. The error distribution intervals

$[0, M_i),$ $i = \overline{0,10}$	γ_{n+1}	$[0, M_i),$ $i = \overline{0,10}$	γ_{n+1}
$0 \div 419$	0	$2520 \div 2939$	1
$420 \div 839$	2	$2940 \div 3359$	3
$840 \div 1259$	4	$3360 \div 3779$	5
$1260 \div 1679$	6	$3780 \div 4199$	7
$1680 \div 2099$	8	$4200 \div 4619$	9
$2100 \div 2519$	10		

Suppose it is necessary to carry out a control (check the fact of presence or absence of an error) of the number:

$$A = A^{(0)} = (a_1^{(0)}, a_2^{(0)}, a_3^{(0)}, a_4^{(0)}, a_5^{(0)}) = (1, 0, 0, 1, 4),$$

represented in the MNS.

To do this, from the values of the digits $a_1^{(0)} = 1$ and $a_4^{(0)} = 1$ of the number A we choose from the ZB (see Table 3) the zeroing constant in the form:

$$ZC^{(1)} = (t_{1,1}, t_{2,1}, t_{3,1}, t_{4,1}, t_{5,1}),$$

where $t_{1,1} = a_1^{(0)} = 1$ and $t_{4,1} = a_4^{(0)} = 1$. In this case with ZB we choose $ZC^{(1)} = (1, 1, 1, 1, 1)$, Table 3.

Further, in accordance with the proposed method of PNZPSD, we perform an operation $A^{(1)} = A^{(0)} - ZC^{(1)}$:

$$\begin{array}{r} A^{(0)} = (1, 0, 0, 1, 4) \\ - \\ ZC^{(1)} = (1, 1, 1, 1, 1) \\ \hline A^{(1)} = (0, 3, 4, 0, 3) \end{array}$$

and, simultaneously, for number:

$$A^{(1)} = (0, 3, 4, 0, 3)$$

with ZB we choose:

$$ZC^{(2)} = (0, t_{2,2}, t_{3,2}, 0, t_{5,2}),$$

of form $a_2^{(1)} = t_{2,2} = 3$ and $a_3^{(1)} = t_{3,2} = 4$. In this case (see Table 4) $ZC^{(2)}$ is defined as:

$$ZC^{(2)} = (0, 3, 4, 0, 3).$$

Next, we define the difference $A^{(1)} - ZC^{(2)}$:

$$\begin{array}{r} A^{(1)} = (0, 3, 4, 0, 3) \\ - \\ ZC^{(2)} = (0, 3, 4, 0, 3) \\ \hline A^{(2)} = (0, 0, 0, 0, 0). \end{array}$$

Thus, a zeroed number is obtained

$$A^{(2)} = A^{(Z)} = (0, 0, \dots, 0, \dots, 0, \gamma_{n+1}) = (0, 0, 0, 0, \gamma_5),$$

where $\gamma_5 = 0$. Conclusion: the number $A^{(0)} = (1, 0, 0, 1, 4)$ has no errors (see Table 2).

Verification: the number $A^{(0)}$ in the PNS is $A^{(0)} = 400$, i.e. is within the working numerical range $[0, 419)$.

Table 3. The value of ZC

PNS	$m_1 = 3,$ $m_4 = 7$
1	1,1,1,1,1
2	2,2,2,2,2
3	0,3,3,3,3
4	1,0,4,4,4
5	2,1,0,5,5
6	0,2,1,6,6
7	1,3,2,0,7
8	2,0,3,1,8
9	0,1,4,2,9
10	1,2,0,3,10
11	2,3,1,4,0
12	0,0,2,5,1
13	1,1,3,6,0
14	2,2,4,0,3
15	0,3,0,1,4
16	1,0,1,2,5
17	2,1,2,3,6
18	0,2,3,4,7
19	1,3,4,5,8
20	2,0,0,6,9

Table 4. The value of ZC

PNS	$m_2 = 4,$ $m_3 = 5$
21	0,1,1,0,10
84	0,0,4,0,7
105	0,1,0,0,6
42	0,2,2,0,9
63	0,3,3,0,8
126	0,2,1,0,5
147	0,3,2,0,4
168	0,0,3,0,3
189	0,1,4,0,2
252	0,0,2,0,10
273	0,1,3,0,9
210	0,2,0,0,1
231	0,3,1,0,0
294	0,2,4,0,8
315	0,3,0,0,7
336	0,0,1,0,6
357	0,1,2,0,5
378	0,2,3,0,4
399	0,3,4,0,3

3 Conclusion

In the modern world rapid growth of volumes of information and increase in complexity of the set scientific and technical tasks, connected with achievement of appropriate level of quality and reliability of transmitted data is observed. Therefore, the

main objective of scientists in the field is development of theoretical bases for construction of high-speed and reliable CSC.

In PNS the problem of increase in reliability and productivity can't be effectively solved without deterioration some key technical and economic indicators of CSC. At the same time, there are positive results of researches which have shown efficiency of application of MNS for increase in speed of realization of integer arithmetic operations, reduction of time of error detection and as a result increase the productivity and reliability of CSC. The methodological basis for building a CSC in the MNS involves a comprehensive solution to the problem of increasing the productivity and integrity of the processing of integer data, as well as providing information security, impedance, performance and durability of the functioning of CSC. Existing data comparison method in MNS don't provide the maximum accuracy of comparison of numbers. Thus, there is a problem of improvement of a method of the fast comparison of data based on the application of the zeroing procedure.

It is known that considerable time of control of data reduces overall effectiveness of application of CSC in MNS, at realization of integer arithmetic and other modular operations. Results of a research of control methods of the data in MNS which are carried out in article have shown that the existing control methods of data in MNS based on use of application of the zeroing procedure reduce control time.

Applications of this method provides obtaining reliable result of control of data in MNS. By the accuracy of the control data in the MNS, we understand the probability of obtaining the true result of the control operation data presented in the MNS.

The essence of the method of error control is to use the procedure of pair number zeroing with the preliminary selection of digits. This makes it possible to increase the efficiency of the procedure for data zeroing in comparison with other control methods up to 30%. The practical significance of the results obtained is that, in comparison with the existing methods of error control in MNS, the error detection time is more than halved. This circumstance makes it possible to increase the overall efficiency of the use of MNS in the creation of CSC [19-24].

References

1. Shu, S., Wang, Y., Wang, Y.: A research of architecture-based reliability with fault propagation for software-intensive systems. In: 2016 Annual Reliability and Maintainability Symposium (RAMS). IEEE (2016). doi:10.1109/rams.2016.7447984
2. Gokhale, S.S., Lyu, M.R., Trivedi, K.S.: Reliability simulation of component-based software systems. In: Proceedings Ninth International Symposium on Software Reliability Engineering (Cat. No.98TB100257). IEEE Comput. Soc (0). doi:10.1109/issre.1998.730882
3. Krasnobayev, V., Kuznetsov, A., Koshman, S., Moroz, S.: Improved Method of Determining the Alternative Set of Numbers in Residue Number System. In: Advances in Intelligent Systems and Computing. pp. 319–328. Springer International Publishing (2018). doi: 10.1007/978-3-319-97885-7_31
4. Tiwari, A., Tomko, K.A.: Enhanced Reliability of Finite-State Machines in FPGA Through Efficient Fault Detection and Correction. IEEE Transactions on Reliability. 54, 459–467 (2005). doi:10.1109/tr.2005.853438

5. Reddy, C.M., Nalini, N.: FT2R2Cloud: Fault tolerance using time-out and retransmission of requests for cloud applications. In: 2014 International Conference on Advances in Electronics Computers and Communications. IEEE (2014). doi:10.1109/icaecc.2014.7002396
6. Braun, C., Wunderlich, H.-J.: Algorithm-based fault tolerance for many-core architectures. In: 2010 15th IEEE European Test Symposium. IEEE (2010). doi:10.1109/etsym.2010.5512738
7. Krasnobayev, V., Kuznetsov, A., Kononchenko, A., Kuznetsova, T.: Method of data control in the residue classes. In Proceedings of the Second International Workshop on Computer Modeling and Intelligent Systems (CMIS-2019). Zaporizhzhia, Ukraine, April 15-19, pp. 241–252 (2019)
8. Popov, D.I., Gapochkin, A.V.: Development of Algorithm for Control and Correction of Errors of Digital Signals, Represented in System of Residual Classes. In: 2018 International Russian Automation Conference (RusAutoCon). IEEE (2018). doi:10.1109/rusautocon.2018.8501826
9. Kocherov, Y.N., Samoylenko, D.V., Koldaev, A.I.: Development of an Antinoise Method of Data Sharing Based on the Application of a Two-Step-Up System of Residual Classes. In: 2018 International Multi-Conference on Industrial Engineering and Modern Technologies (FarEastCon). IEEE (2018). doi:10.1109/fareastcon.2018.8602764
10. Krasnobayev, V., Kuznetsov, A., Zub, M., Kuznetsova, K.: Methods for comparing numbers in non-positional notation of residual classes. In Proceedings of the Second International Workshop on Computer Modeling and Intelligent Systems (CMIS-2019). Zaporizhzhia, Ukraine, April 15-19, pp. 581–595 (2019)
11. Kasianchuk, M., Yakymenko, I., Pazdriy, I., Melnyk, A., Ivasiev, S.: Rabin's modified method of encryption using various forms of system of residual classes. In: 2017 14th International Conference The Experience of Designing and Application of CAD Systems in Microelectronics (CADSM). IEEE (2017)
12. Krasnobayev, V., Kuznetsov, A., Lokotkova, I., Dyachenko, A.: The Method of Single Errors Correction in the Residue Class. In: 2019 3rd International Conference on Advanced Information and Communications Technologies (AICT). IEEE (2019). doi:10.1109/AIACT.2019.8847845
13. Dubrova, E.: Fault-Tolerant Design. Springer New York (2013). doi:10.1007/978-1-4614-2113-9
14. Krasnobayev, V., Kuznetsov, A., Babenko, V., Denysenko, M., Zub, M., Hryhorenko, V.: The Method of Raising Numbers, Represented in the System of Residual Classes to an Arbitrary Power of a Natural Number. In: 2019 IEEE 2nd Ukraine Conference on Electrical and Computer Engineering (UKRCON). IEEE (2019). doi:10.1109/UKRCON.2019.8879793
15. Radu, M.: Reliability and fault tolerance analysis of FPGA platforms. In: IEEE Long Island Systems, Applications and Technology (LISAT) Conference 2014. IEEE (2014). doi:10.1109/lisat.2014.6845211
16. Yanko, A., Koshman, S., Krasnobayev, V.: Algorithms of data processing in the residual classes system. In: 2017 4th International Scientific-Practical Conference Problems of Infocommunications. Science and Technology (PIC S&T). IEEE (2017). doi:10.1109/infocommst.2017.8246363
17. Krasnobayev, V.A., Yanko, A.S., Koshman, S.A.: A Method for Arithmetic Comparison of Data Represented in a Residue Number System. Cybernetics and Systems Analysis. 52, 145–150 (2016). doi: 10.1007/s10559-016-9809-2

18. Kasianchuk, M., Yakymenko, I., Pazdriy, I., Zastavnyy, O.: Algorithms of findings of perfect shape modules of remaining classes system. In: The Experience of Designing and Application of CAD Systems in Microelectronics. IEEE (2015)
19. Chornei, R.K., Daduna V.M., H., Knopov, P.S.: Controlled Markov Fields with Finite State Space on Graphs. Stochastic Models. 21, 847–874 (2005). doi:10.1080/15326340500294520
20. Ponochovniy, Y., Bulba, E., Yanko, A., Hozbenko, E.: Influence of diagnostics errors on safety: Indicators and requirements. In: 2018 IEEE 9th International Conference on Dependable Systems, Services and Technologies (DESSERT). IEEE (2018). doi:10.1109/dessert.2018.8409098
21. Runovski, K., Schmeisser, H.: On the convergence of fourier means and interpolation means. Journal of Computational Analysis and Applications. 6(3), 211-227 (2004)
22. Tkach, B.P., Urmancheva, L.B.: Numerical-analytic method for finding solutions of systems with distributed parameters and integral condition. Nonlinear Oscillations. 12, 113–122 (2009). doi:10.1007/s11072-009-0064-6
23. Bondarenko, S., Liliya, B., Oksana, K., & Inna, G.: Modelling instruments in risk management. International Journal of Civil Engineering and Technology. 10(1), 1561-1568 (2019)
24. Krasnobayev, V., Koshman, S., Yanko, A., Martynenko, A.: Method of Error Control of the Information Presented in the Modular Number System. In: 2018 International Scientific-Practical Conference Problems of Infocommunications. Science and Technology (PIC S&T). IEEE (2018). doi:10.1109/infocommst.2018.8632049

The Analysis of the Methods of Data Diagnostic in a Residue Number System

Victor Krasnobayev¹[0000-0001-5192-9918], Alexandr Kuznetsov¹[0000-0003-2331-6326],
Alina Yanko²[0000-0003-2876-9316] and Tetiana Kuznetsova¹[0000-0001-6154-7139]

¹ V. N. Karazin Kharkiv National University, Svobody sq., 4, Kharkiv, 61022, Ukraine
v.a.krasnobayev@gmail.com, kuznetsov@karazin.ua,
kuznetsova.tatiana17@gmail.com

² Poltava National Technical Yuri Kondratyuk University, Poltava, Ukraine
al9_yanko@ukr.net

Abstract. The article presents the results of the analysis of the methods of data diagnostic presented in residue number system (RNS). Two practical methods of data diagnostic in RNS are investigated. Their advantages and disadvantages are shown. The main disadvantage of these methods is the lack of the efficiency in data diagnostic in RNS. The third method of the efficient diagnostic in RNS, which eliminates the above-mentioned disadvantage, has been reviewed in the article. The usage of this method can significantly increase the efficiency of data diagnostic in RNS. The main drawback of this method is a significant amount of equipment required to implement the process of data diagnostic in RNS. The method of the efficient diagnostic has been improved in terms of reducing the amount of equipment required for implementing the process of data diagnostic in RNS. The application of the improved method of the efficient diagnostics allows reducing the amount of equipment for the implementation of a diagnostic data procedure in RNS without increasing the diagnostic time. Examples of practical use of the improved method of data diagnostic in RNS are presented.

Keywords: Alternative Set of Numbers; Data Diagnostic; Diagnostic Efficiency; Error Control and Correction; Residue Number System; Zeroisation Procedure.

1 Introduction

Data diagnostic in residue number system (RNS) is the process of determining the distorted residues in redundant non-positional code structure (NCS) presented in the following form $A_{RNS} = (a_1 \parallel a_2 \parallel \dots \parallel a_{i-1} \parallel a_i \parallel \dots \parallel a_n \parallel \dots \parallel a_{n+k})$ where n and k are the number of, respectively, informational and control bases $m_i (i = \overline{1, n+k})$ of ordered $(m_i < m_{i+1})$ RNS. The diagnostic is carried out after data control, if it is necessary for the subsequent error correction. Some methods, algorithms and devices for data diagnostic in RNS have already been presented [1-3]. To monitor, diagnose and

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

correct errors, the certain information redundancy must be introduced. Power R of the information redundancy, which as in positional number system (PNS), determines the corrective abilities of the code, is estimated by the valued $d_{\min}^{(RNS)}$ of a minimum code distance (MCD). In RNS the value of MCD is determined by the ratio $d_{\min}^{(RNS)} = k + 1$ [4-7]. For one control base, the value of MCD is equal to $d_{\min}^{(RNS)} = 2$. In accordance with the general coding theory, in RNS with a minimum code distance $d_{\min}^{(RNS)} = 2$ the distortion of only one of the residues can be reliably established (one-time error) in NCS. For example, to correct a one-time error (in one residue) and determine double errors (in two residues) it is necessary to ensure that $d_{\min}^{(RNS)} = 3$ [1, 8-12]. Due to the influence of RNS properties on the data processing it is possible, in some cases, to correct one-time data errors (in one NCS residue) when introducing the minimal ($k = 1$) information code redundancy. So, the property of the independence of the residues of NCS allows us to correct not intermediate calculation results, but final one. A typical example for this case is the possibility of implementing the data error correction procedure with one control base without stopping the intermediate computing process (during the computational process). To implement such procedure, it becomes necessary to diagnose intermediate results of calculations based on the use of the concept of an alternative set of numbers (AS) in RNS [13-19].

The purpose of the article is to study the methods of data diagnostic, presented in non-positional residue number system with one control base.

Main part. Let us consider the method of data diagnostic in RNS based on the concept of AS numbers in RNS.

The first method of diagnosis. The alternative set $W(\tilde{A}) = \{m_{l_1}, m_{l_2}, \dots, m_{l_p}\}$ of incorrect number $\tilde{A}_{RNS} = (a_1 \parallel a_2 \parallel \dots \parallel a_{i-1} \parallel \tilde{a}_i \parallel a_{i+1} \parallel \dots \parallel a_n \parallel a_{n+1})$ can be determined by a sequential testing of each base $m_i (i = \overline{1, n})$ RNS. We determine the set of numbers, that have the same residues for all bases of RNS, as number $\tilde{A}$, except one certain residue (base), and differ only in values of possible residues on this base. In this set there may be no correct numbers or there may be only one correct number. In the last case, the number is a part of AS of number $\tilde{A}$.

The proposed method involves carrying out similar verifications for each of the information base of RNS (a control base always is a part of a set of bases of AS). The result of such sequential verifications completely and reliably determines the AS $W(\tilde{A}) = \{m_{l_1}, m_{l_2}, \dots, m_{l_p}\}$ of the incorrect number $\tilde{A}$. The disadvantage of the method is the low efficiency in determining AS. This is due to the considerable time of consecutive executions of data diagnostic stages in RNS.

The second method of diagnosis. This method is also based on the determination of AS $W(\tilde{A}) = \{m_{l_1}, m_{l_2}, \dots, m_{l_p}\}$. In this case, the whole procedure of diagnosing

NCS is carried out by simultaneous and parallel calculation of all possible projections $\tilde{A}_{i\ RNS} = (a_1 \| a_2 \| \dots \| a_{i-1} \| \tilde{a}_i \| a_{i+1} \| \dots \| a_n \| a_{n+1})$ of the incorrect number $\tilde{A}_{RNS} = (a_1 \| a_2 \| \dots \| a_{i-1} \| \tilde{a}_i \| a_{i+1} \| \dots \| a_n \| a_{n+1})$, and their subsequent comparison with the value of $M = \prod_{i=1}^n m_i$ without the redundant numeric information interval (information volume of code words) $0 \div M - 1$ given in RNS. It is proved in [1, 7, 8], that the necessary and sufficient condition of the entry of the bases of RNS in AS $W(\tilde{A}) = \{m_{l_1}, m_{l_2}, \dots, m_{l_p}\}$ of number $\tilde{A}_{RNS} = (a_1 \| a_2 \| \dots \| a_{i-1} \| \tilde{a}_i \| a_{i+1} \| \dots \| a_n \| a_{n+1})$ is correctness of $(\tilde{A}_{i\ RNS} < M)$ for its projection $\tilde{A}_{i\ RNS} = (a_1 \| a_2 \| \dots \| a_{i-1} \| \tilde{a}_i \| a_{i+1} \| \dots \| a_n \| a_{n+1})$. Parallelization of the procedure of calculating all possible projections $\tilde{A}_{i\ RNS} = (a_1 \| a_2 \| \dots \| a_{i-1} \| \tilde{a}_i \| a_{i+1} \| \dots \| a_n \| a_{n+1})$ of the incorrect number $\tilde{A}_{RNS} = (a_1 \| a_2 \| \dots \| a_{i-1} \| \tilde{a}_i \| a_{i+1} \| \dots \| a_n \| a_{n+1})$ reduces the time of AS determination and increases the efficiency of diagnosing data in RNS.

Let us consider the following example of data diagnostic based on the usage of the second method.

Example 1. Let us determine the AS of the number $\tilde{A}_{RNS} = (0 \| 0 \| 0 \| 0 \| 5)$, which is defined in RNS by the information $m_1 = 3$, $m_2 = 4$, $m_3 = 5$, $m_4 = 7$ and control bases $m_k = m_5 = 11$. Wherein $M = \prod_{i=1}^n m_i = \prod_{i=1}^4 m_i = 420$ and the full range $0 \div M_0 - 1$ of coded words equals to $M_0 = M \cdot m_{n+1} = 420 \cdot 11 = 4620$ (Table 1).

At first, the procedure of controlling number $A_{RNS} = (0 \| 0 \| 0 \| 0 \| 5)$ is carried out by the known method [1, 18, 19]. According to the standard control procedure we determine the value of the original number in PNS. In the end of the control it is determined that $A_{PNS} = 3360 > M = 420$. In this case, assuming the occurrence of only one-time (in one residue number) errors, it can be concluded that the considered number $\tilde{A}_{3360} = (0 \| 0 \| 0 \| 0 \| 5)$ is incorrect, i.e., one of the number residues is distorted. Then the procedure of determining AS $\tilde{A}_{3360} = (0 \| 0 \| 0 \| 0 \| 5)$ is realized (Table 1). For the number $A_{RNS} = (0 \| 0 \| 0 \| 0 \| 5)$ not distorted residues have been determined. They are $a_2 = 0$ and $a_3 = 0$. The values of residues on the bases m_1 , m_4 and m_5 , i.e., residues $a_1 = 0$, $a_4 = 0$ and $a_5 = 5$ may be incorrect. In this case, for the number $A_{RNS} = (0 \| 0 \| 0 \| 0 \| 5)$ AS will be equal to the set of RNS bases $W(\tilde{A}) = \{m_1, m_4, m_5\}$.

Table 1. Table of code words

A in PNS	A in RNS					A in PNS	A in RNS				
0	0	0	0	0	0	2310					
1	1	1	1	1	1	2311					
2	2	2	2	2	2	2312					
3	0	3	3	3	3	2313					
⋮						⋮					
418						2728					
419						2729					
420						2730					
						⋮					
						⋮					
						⋮					
						3360	0	0	0	0	5
						⋮					
						⋮					
2308						4618					
2309						4619					

The use of the second method of data diagnostic in RNS allows us to speed up the process of determining AS $W(\tilde{A}) = \{m_{l_1}, m_{l_2}, \dots, m_{l_p}\}$ of the number $\tilde{A}_{RNS} = (a_1 \parallel a_2 \parallel \dots \parallel a_{i-1} \parallel \tilde{a}_l \parallel a_{i+1} \parallel \dots \parallel a_n \parallel a_{n+1})$, due to the possibility of parallel determination of projections $\tilde{A}_j$ of incorrect number $\tilde{A}_{RNS} = (a_1 \parallel a_2 \parallel \dots \parallel a_{i-1} \parallel \tilde{a}_l \parallel a_{i+1} \parallel \dots \parallel a_n \parallel a_{n+1})$. It should be noted, that for the second method the procedure of determining the number of AS includes such basic operations as transferring $\tilde{A}_{RNS} = (a_1 \parallel a_2 \parallel \dots \parallel a_{i-1} \parallel \tilde{a}_l \parallel a_{i+1} \parallel \dots \parallel a_n \parallel a_{n+1})$ from RNS to PNS; converting

projections $\tilde{A}_{i\ RNS} = (a_1 \parallel a_2 \parallel \dots \parallel a_{i-1} \parallel a_{i+1} \parallel \dots \parallel a_n \parallel a_{n+1})$ of the incorrect number $\tilde{A}_{RNS}$ from RNS to PNS and the operation of comparing them with the value M . In RNS the listed operations refer to non-positional operations, the implementation of which is very consuming both in time and hardware.

The known methods of diagnosing in RNS have the common drawback, that is the low efficiency of data diagnostic. This reduces the effectiveness of RNS usage for rapid implementation of integer-valued operations.

The third recent designed method of data diagnosis is presented in [2, 7, 8]. Its usage allows increasing the efficiency of diagnosing in RNS. The essence of the developed method of improving the efficiency of diagnosing data in RNS is that AS $W(\tilde{A}) = \{m_{l_1}, m_{l_2}, \dots, m_{l_p}\}$ of the number $\tilde{A}_{RNS}$ is determined not in the whole interval

$[jM, (j+1)M)$, which contains the incorrect number $\tilde{A}_{RNS}$, but only in a small numerical interval $\Delta A^{(H)} = (\tilde{A}_{RNS} - \tilde{A}_{RNS}^{(H)}) < M$, where

$\tilde{A}_{RNS}^{(H)} = (0 \parallel 0 \parallel \dots \parallel 0 \parallel \gamma_{n+1})$ is a number reduced to zero in RNS. The essence of reducing to zero in RNS is to replace the original number $\tilde{A}_{RNS} = (a_1 \parallel a_2 \parallel \dots \parallel a_{i-1} \parallel \tilde{a}_l \parallel a_{i+1} \parallel \dots \parallel a_n \parallel a_{n+1})$ with the number $\tilde{A}_{RNS}^{(H)} = (0 \parallel 0 \parallel \dots \parallel 0 \parallel \gamma_{n+1})$, by using a sequence of transformations, by which any intermediate number does not go beyond the working range $0 \div M-1$.

zeroisation procedure can be implemented by various methods. The essence of all these methods is that some minimum $ZC^{(i)}$ numbers, so called zeroisation constants (ZC), are sequentially subtracted from the initial number

$\tilde{A}_{RNS} = (a_1 \parallel a_2 \parallel \dots \parallel a_{i-1} \parallel \tilde{a}_l \parallel a_{i+1} \parallel \dots \parallel a_n \parallel \dots \parallel a_{n+k})$ until the number $\tilde{A}_{RNS}$ is converted into the number $\tilde{A}_{RNS}^{(H)} = (0 \parallel 0 \parallel \dots \parallel 0 \parallel \gamma_{n+1})$ and the value of the number

$\tilde{A}_{RNS}$ does not go beyond the range $[0, M)$. Geometrically, zeroisation procedure corresponds to the offset of the original number

$\tilde{A}_{RNS} = (a_1 \parallel a_2 \parallel \dots \parallel a_{i-1} \parallel \tilde{a}_l \parallel a_{i+1} \parallel \dots \parallel a_n \parallel a_{n+1})$ to the left edge jM of its numeric range $[jM, (j+1)M)$. Thus, to eliminate the redundancy of AS

$W(\tilde{A}) = \{m_{l_1}, m_{l_2}, \dots, m_{l_p}\}$, by reducing the interval range of the number $\tilde{A}_{RNS}$, the values $A_{RNS}^{(H)} = (0 \parallel 0 \parallel \dots \parallel 0 \parallel \gamma_{n+1})$ and $\Delta A^{(H)} = (\tilde{A}_{RNS} - \tilde{A}_{RNS}^{(H)}) \bmod M$ have to be pre-defined. It can be conveniently demonstrated for particular RNS.

As an example, for RNS defined by the bases $m_1 = 2, m_2 = 3, m_3 = m_{n+1} = 5$ ($M = 2 \cdot 3 = 6; M_0 = 2 \cdot 3 \cdot 5 = 30$) (Table 2), in accordance with the distribution of errors in the intervals of the working range $[0, M)$ [1], for each interval

$[jM, (j+1)M)$ two-entry tables are preliminarily compiled. Tables 3 of the correspondence of $\bar{W}(\tilde{A}) = \Phi(\gamma_{n+1}; \Delta A^{(H)})$.

Table 2. Code Words in RNS

A to PNS	A in RNS			A to PNS	A in RNS		
0	0	0	0	15	1	0	0
1	1	1	1	16	0	1	1
2	0	2	2	17	1	2	2
3	1	0	3	18	0	0	3
4	0	1	4	19	1	1	4
5	1	2	0	20	0	2	0
6	0	0	1	21	1	0	1
7	1	1	2	22	0	1	2
8	0	2	3	23	1	2	3
9	1	0	4	24	0	0	4
10	0	1	0	25	1	1	0
11	1	2	1	26	0	2	1
12	0	0	2	27	1	0	2
13	1	1	3	28	0	1	3
14	0	2	4	29	1	2	4

As it was noted above AS $W(\tilde{A}) = \{m_{l_1}, m_{l_2}, \dots, m_{l_p}\}$ numbers are determined not on the whole range $[jM, (j+1)M)$, which contains the incorrect number $\tilde{A}$, but only on the numerical range $\Delta A^{(H)}$. The method of on-line data diagnostic in RNS is presented in Fig. 1.

The considered method allows reducing the time of data diagnostic in RNS. The time to diagnose data is reduced, firstly, by eliminating non-positional operations such as converting numbers from RNS to PNS and comparing numbers, and, secondly, by using a single-entry tabular sampling of AS value. The proposed method of the rapid

diagnostic of data errors improves the overall efficiency of using non-positional code structures in RNS.

Table 3. Table of values AS $W(\tilde{A}) = \Phi(\gamma_{n+1}; \Delta A^{(H)})$

ΔA		γ_{n+1}			
		Z_1		Z_2	
		1	2	3	4
Z_3	0	m_3	m_2, m_3	m_1, m_3	m_2, m_3
	1	m_3	m_2, m_3	m_1, m_3	m_2, m_3
	2	m_3	m_2, m_3	m_1, m_2, m_3	m_3
Z_4	3	m_3	m_1, m_2, m_3	m_2, m_3	m_3
	4	m_2, m_3	m_1, m_3	m_2, m_3	m_3
	5	m_2, m_3	m_1, m_3	m_2, m_3	m_3

The drawback of the considered method of rapid data diagnostics in RNS is the considerable amount of equipment required for its implementation due to the large volumes ($\Delta \tilde{A} \times (\gamma_{n+1} - 1)$ is a memory unit) of the memory (MMU) realizing function $\Phi(\gamma_{n+1}; \Delta A^{(H)})$. We propose the following improvements in order to reduce the amount of the necessary equipment to implement the method of rapid diagnostic.

The essence of the improvements is to decrease in half the amount of the required equipment for the implementation of MMU content. This allows reducing the total amount of the required equipment for the implementation of the procedure for error diagnosing in NCS presented in RNS [20-22].

This is done by using the symmetry properties of the numerical data of the complete MMU table (Table 7) relative to the point with coordinate $\frac{M_0 + M - 1}{2}$, that corresponds to the value m_2, m_3 and is analytically expressed (1) in the following way:

$$\bar{W}(\tilde{A}) = \Phi_1(\gamma_{n+1}; \Delta A^{(H)}) = \Phi_2\{[m_{n+1} - \gamma_{n+1}], [(M-1) - \Delta A^{(H)}]\} \quad (1)$$

1. For a given RNS, a two-entry (two-coordinate) table of AS $\bar{W}(\tilde{A}) = \Phi(\gamma_{n+1}; \Delta A^{(H)})$ values contained in MMU is compiled. There $1 \leq \gamma_{n+1} \leq m_{n+1} - 1$. Each pair of values γ_{n+1} and $\Delta A^{(H)}$ corresponds to a specific set of AS bases.

2. By means of a set of reduction to zero constants $ZC^{(i)}$ initial incorrect $\tilde{A}_{RNS} = (a_1 \parallel a_2 \parallel \dots \parallel a_{i-1} \parallel \tilde{a}_i \parallel a_{i+1} \parallel \dots \parallel a_n \parallel a_{n+1})$ number converted (reduced to zero) to $A^{(H)} = (0 \parallel 0 \parallel \dots \parallel 0 \parallel \gamma_{n+1})$ number. We obtain value γ_{n+1} that corresponds to the first coordinate in the lookup table $\bar{W}(\tilde{A}) = \Phi(\gamma_{n+1}; \Delta A^{(H)})$.

3. $\Delta A^{(H)} = (\tilde{A}_{RNS} - \tilde{A}_{RNS}^{(H)})$ is determined. Therefore we obtain value $\Delta A^{(H)}$ of the second coordinate in the lookup table $\bar{W}(\tilde{A}) = \Phi(\gamma_{n+1}; \Delta A^{(H)})$.

4. According to obtained values of two coordinates $\Delta A^{(H)}$ and γ_{n+1} we refer to the two-entry lookup table $\bar{W}(\tilde{A}) = \Phi(\gamma_{n+1}; \Delta A^{(H)})$ from which the specific value of AS $W(\tilde{A}) = \{m_{l_1}, m_{l_2}, \dots, m_{l_p}\}$ of incorrect number $\tilde{A}_{RNS} = (a_1 \parallel a_2 \parallel \dots \parallel a_{i-1} \parallel \tilde{a}_i \parallel a_{i+1} \parallel \dots \parallel a_n \parallel a_{n+1})$ in RNS is determined.

Fig. 1. Method of on-line data diagnostic in RNS

The correctness of (1) can be easily shown by using the results of the lemma on the distribution of the terms of number sequence $A_{is} = (a_1, a_2, \dots, a_{i-1}, s, a_{i+1}, \dots, a_n, a_{n+1})$ in the numerical range $(0, M_0)$, where $s = 0, 1, \dots, m_{i-1} (i = \overline{1, n+1})$ [1, 7, 8]. Basing on (1), the content of MMU for the proposed method of data diagnostic in RNS is presented in Table 4. Table 5 presents the characteristics Z_i of quadrant numbers from the completed Table 3 of MMU data and Table 6 presents the attributes of quadrant numbers of the shortened Table 4 of MMU data. In Table 7 there are the values of numerical ranges for finding the MMU input numbers and the correspondent data attributes formed by the group of decoders.

When implementing this method of data diagnostic in RNS [21, 22], in the diagnostic scheme the module of determining characteristics is intended for to form and use the characteristics $Z_1 \div Z_4$ of quadrant numbers $\tilde{\Delta A} \times (\gamma_{n+1} - 1)$ of the completed data table MMU $W(\tilde{A}) = \{m_{l_1}, m_{l_2}, \dots, m_{l_p}\}$ (Table 3). The characteristics are formed

by means of a group of decoders (Table 4) and a combination of OR elements. Using the values $Z_1 \div Z_4$, according to input data γ_{n+1} and $\Delta\tilde{A}$, the AS $W(\tilde{A}) = \{m_{l_1}, m_{l_2}, \dots, m_{l_p}\}$ is determined by shortened table $\Delta\tilde{A} \times \left(\frac{\gamma_{n+1}-1}{2}\right)$ of MMU data (Table 4).

Table 4. AS $W(\tilde{A})$ values of shortened MMU

ΔA		γ_{n+1}	
		$Z_i Z_1$	
		1	2
Z_3	0	m_3	m_2, m_3
	1	m_3	m_2, m_3
	2	m_3	m_2, m_3
Z_4	3	m_3	m_1, m_2, m_3
	4	m_2, m_3	m_1, m_3
	5	m_2, m_3	m_1, m_3

Table 5. Characteristics $Z_i (i = \overline{1,4})$ of quadrant numbers $\Delta\tilde{A} \times (\gamma_{n+1} - 1)$ of the completed table AS data $W(\tilde{A}) = \{m_{l_1}, m_{l_2}, \dots, m_{l_p}\}$

II ($Z_1 Z_3$)	I ($Z_2 Z_3$)
III ($Z_1 Z_4$)	IV ($Z_2 Z_4$)

The characteristics $Z_1 \div Z_4$ are applied as follows (Table 7): Z_1 and Z_2 are the characteristics of finding a distorted $\tilde{A}_{RNS}$ number in the numerical ranges $1 \div \frac{m_{n+1}-1}{2}$ and

$\frac{m_{n+1} + 1}{2} \div m_{n+1} - 1$ respectively; Z_3 and Z_4 - characteristics of finding a distorted number $\tilde{A}_{RNS}$ in the numerical ranges $0 \div \frac{(M-1)-1}{2}$ and $\frac{M}{2} \div M-1$ respectively. For the second (II) and the third (III) quadrants, shortened Table 6, AS $W(\tilde{A})$ values are determined by formula $W(\tilde{A}) = F_1(\gamma_{n+1}; \Delta A^{(H)})$.

Table 6. Characteristics of quadrant numbers $\Delta \tilde{A} \times \left(\frac{\gamma_{n+1} - 1}{2} \right)$ of the table of the data

$$W(\tilde{A}) = \{ m_{l_1}, m_{l_2}, \dots, m_{l_p} \}$$

$$\begin{matrix} \text{II} \\ (Z_1 Z_3) \end{matrix}$$

$$\begin{matrix} \text{III} \\ (Z_1 Z_4) \end{matrix}$$

Table 7. The value of numerical ranges and their correspondence to the data attributes

Decoder Group Outputs	Numerical range	Numerical range attribute
The group of the first decoder outputs (the first group of MMU inputs)	$1 \div \frac{m_{n+1} - 1}{2}$	Z_1
The group of the second decoder outputs (the second group of MMU inputs)	$0 \div M - 1$	Z_1, Z_4
The first group of the third decoder outputs	$1 \div \frac{m_{n+1} - 1}{2}$	Z_1
The second group of the third decoder outputs	$\frac{m_{n+1} + 1}{2} \div m_{n+1}$	Z_2
The first group of the fourth decoder outputs	$0 \div \frac{(M-1)-1}{2}$	Z_3
The second group of the fourth decoder outputs	$\frac{M}{2} \div M - 1$	Z_4

For the first (I) and the fourth (IV) quadrants of the completed Table 3, according to the values of the shortened Table 4, AS $W(\tilde{A})$ values are determined by formula (2):

$$W(\tilde{A}) = F_2 \{ [m_{n+1} - \gamma_{n+1}] [(M-1) - \Delta A] \} \quad (2)$$

The method of rapid data diagnostic in RNS is presented in Fig. 2.

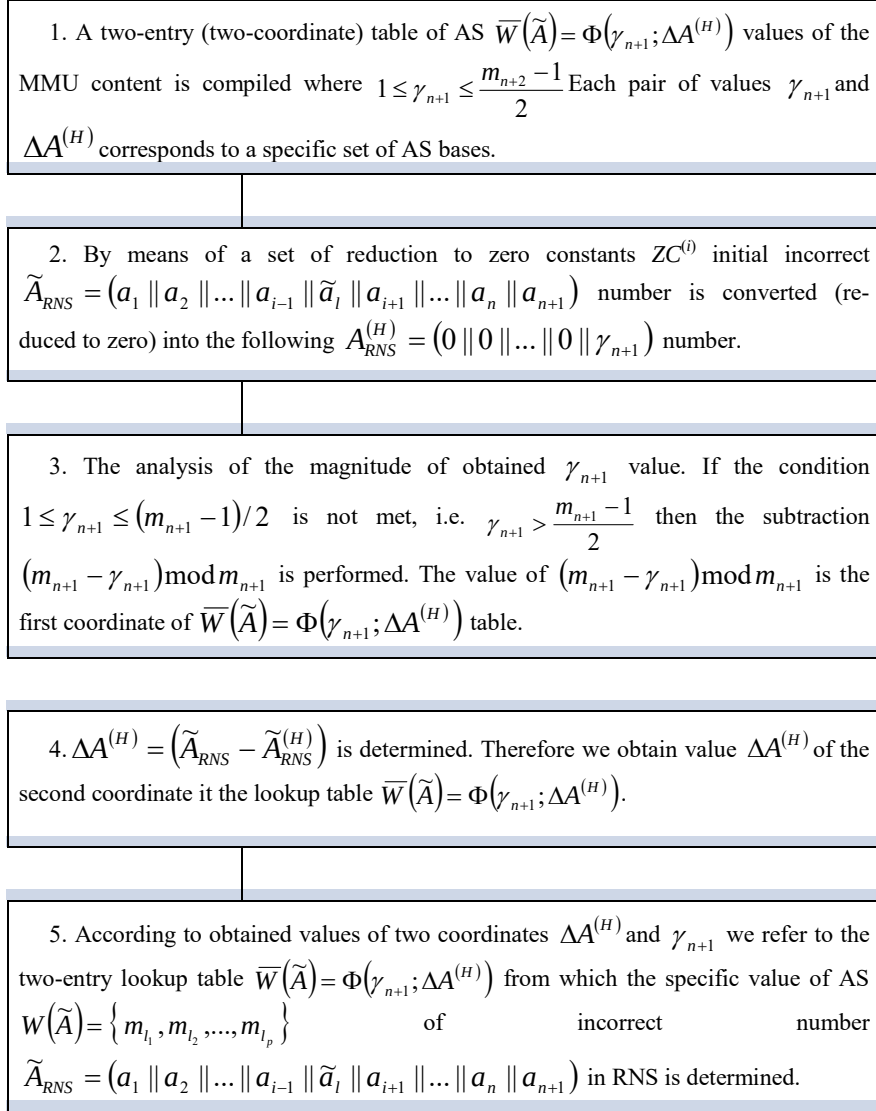


Fig. 2. Method of on-line data diagnostic in RNS.

2 Examples of using the method of rapid data diagnostic in RNS

In accordance with Fig. 2, let us present the examples 2-4 [21, 22] of using the method of on-line data diagnostic in RNS determined by bases $m_1 = 2, m_2 = 3, m_3 = m_{n+1} = 5$; $M = 2 \cdot 3 = 6$; $M_0 = 2 \cdot 3 \cdot 5 = 30$ (Table 2). Tables 8 and 9 present some zeroisation constants for the corresponding RNS basis.

Table 8. The reduction to zero constants for the first base of RNS

a_1	ZC
0	$(0 \parallel 0 \parallel 0)$
1	$(1 \parallel 1 \parallel 1)$

To check the obtained diagnostic result $W(\tilde{A}) = F_{RES}[\gamma_{n+1}; \Delta \tilde{A}]$, which is determined by the shortened Table 4 of $W(\tilde{A})$ MMU of the dimension $\Delta \tilde{A} \times \left(\frac{\gamma_{n+1}-1}{2}\right)$, the values $W(\tilde{A}) = F_{TEST}[\gamma_{n+1}; \Delta \tilde{A}]$ are used, which are determined by the completed Table 3 of MMU data of the dimension $\Delta \tilde{A} \times (\gamma_{n+1} - 1)$.

Table 9. The reduction to zero constants for the second base of RNS

a_2	ZC
0	$(0 \parallel 0 \parallel 0)$
1	$(0 \parallel 1 \parallel 4)$
2	$(0 \parallel 2 \parallel 2)$

Example 3. It is assumed to determine AS $W(\tilde{A})$ of the number $\tilde{A} = (1 \parallel 1 \parallel 2)$. The value of the zeroisation number is represented as $\tilde{A}^{(H)} = \tilde{A} - KH = (1 \parallel 1 \parallel 2) - (1 \parallel 1 \parallel 1) = (0 \parallel 0 \parallel 1)$ (Table 8). Thus, we have the value $\gamma_{n+1} = 1$ (in binary code 001) and also determine that $\tilde{A}^{(H)} = \tilde{A} - A^{(H)} = (1 \parallel 1 \parallel 2) - (0 \parallel 0 \parallel 1) = (1 \parallel 1 \parallel 1)$. The value $\gamma_{n+1} = 1$ (in binary code 001) is fed to the input of the decoder, from the output of which the value $\gamma_{n+1} = 1$ is fed to the input of the corresponding element OR in the unitary code. The

value $\Delta\tilde{A}^{(H)} = 1$ (in binary code 001) is fed to the input of the fourth decoder, from the output of which the value $\Delta\tilde{A}^{(H)} = 1$ is fed to the input of the corresponding OR element in the unitary code (Table 7). The value $\gamma_{n+1} = 1$ (in the binary code 001) is fed to the decoder, from the output of which the value 1 in a unitary code, through a corresponding OR element, is fed to the first input of the first groups of MMU inputs. At the same time, the value $\Delta\tilde{A}^{(H)} = 1$ (in binary code 001) is fed to the input of the second decoder, from the output of which value 1 in the unitary code is fed to the first input of the second group of MMU inputs (Table 4). In accordance with the $W(\tilde{A})$ data of MMU (Table 4), we obtain $W(\tilde{A}) = \{m_3\}$ as the result of the procedure. Therefore $W(\tilde{A}) = F_{RES}(\gamma_{n+1}; \Delta\tilde{A}) = F_{RES}(1; 1) = \{m_3\}$.

Check (Table 3): $W(\tilde{A}) = F_{TEST}(\gamma_{n+1}; \Delta\tilde{A}) = F_{TEST}(1; 1) = \{m_3\}$.

Example 4. Number $\tilde{A} = (0 \parallel 0 \parallel 4)$ is assumed to be diagnosed (AS $W(\tilde{A})$ of $\tilde{A} = (0 \parallel 0 \parallel 4)$ number must be determined). First $\gamma_{n+1} = 4 \neq 0$ is determined. Then we obtain $\tilde{A}^{(H)} = \tilde{A} - A^{(H)} = (0 \parallel 0 \parallel 4) - (0 \parallel 0 \parallel 4) = (0 \parallel 0 \parallel 0)$ and therefore $\Delta\tilde{A} = 0$. Value $\gamma_{n+1} = 4$ is fed to the input of the decoder, from the output of which value $\gamma_{n+1} = 4$ is fed to the input of the OR element in the unitary code (Table 7). The value $\Delta\tilde{A} = 0$ is fed to the input of the fourth decoder, from the output of which value $\Delta\tilde{A} = 0$ is fed to the input of the OR element in the unitary code. The value $\gamma_{n+1} = 4$ (in the binary code 100) is fed to the inverter from the output of which the value $m_{n+1} - \gamma_{n+1} = 5 - 4 = 1$ (in the binary code 001) is fed to the first decoder from the output of which the value 1 in a unitary code, through the corresponding OR element, is fed to a first input of the first group of MMU inputs (Table 4). Simultaneously, the value $\Delta\tilde{A} = 0$ is fed to the inverter in binary cod, from the output of which the value $(M - 1) - \Delta\tilde{A} = (6 - 1) - 0 = 5$ (in the binary code 101), through the OR element, is fed to the decoder input from the output of which the value 5 is fed to the fifth input of the second group of MMU inputs in a unitary code (Table 4). In accordance with the $W(\tilde{A})$ data of MMU (Table 4), the result of the diagnosing is determined by the value γ_{n+1} that equals 1, and by the value $\Delta\tilde{A}$ that equals 5. We obtain $W(\tilde{A}) = \{m_2, m_3\}$ as the result of the procedure. Therefore $W(\tilde{A}) = F_{RES}[(m_{n+1} - \gamma_{n+1}); [(M - 1) - \Delta\tilde{A}]] = F_{RES}(1; 5) = \{m_2, m_3\}$.

Check (Table 3): $W(\tilde{A}) = F_{TEST}(\gamma_{n+1}; \Delta\tilde{A}) = F_{TEST}(4; 0) = \{m_2, m_3\}$.

3 Conclusion

According to the results of studying the methods of data diagnostic in RNS the improved method of rapid diagnostic is proposed for the practical implementation. Application of this method allows reducing the amount of the equipment required for implementing data diagnostic procedures in RNS without increasing the time of diagnosis. This is achieved by reducing the amount of equipment for completed table $\Delta\tilde{A} \times (\gamma_{n+1} - 1)$ of MMU, by forming and using numerical characteristics $Z_1 \div Z_4$ which show the belonging of the input numbers γ_{n+1} and $\Delta\tilde{A}$ of the table of MMU to each of the four quadrants of the completed data table AS $w(\tilde{A})$ of the numbers $\tilde{A}$ in RNS. This makes it possible to perform reliable diagnostic of the distorted number $\tilde{A}$ in RNS, i.e., precisely determine those bases of RNS where the residues of the correct number A have been distorted. The values of only a half (the second and the third quadrants) of the completed data table AS $w(\tilde{A})$ of MMU are used. The examples of the practical usage of the method of diagnosis have been presented. The verification of the diagnosis of numbers in RNS, carried out by the developed method confirms the validity of the stated goal and the practical feasibility of diagnosing data in RNS. Based on the proposed diagnostic method, an algorithm of its implementation has been developed and the patentable device has been produced. A device for monitoring and diagnosing data presented in RNS has been patented in Ukraine. It should be noted that by increasing the length of the discharge grid of the calculator in RNS, the efficiency of the proposed method also increases. This can be used to solve various applied problems of computer science [25-30].

References

1. Akushsky, I.Ya., Yuditsky, D.I.: Machine arithmetic in residual classes. Moscow, Sov. Radio. 440 p. (1968) (in Russian)
2. Torgashov, V.A.: System of residual classes and the reliability of a computer. Moscow, Sov. Radio. 118 p. (1973) (in Russian)
3. Sasaki, A.: The Basis for Implementation of Arithmetic Operations in the Residue Number System. IEEE Transactions on Computers. C-17, 1066–1073 (1968). doi:10.1109/TC.1968.226466
4. Givaki, K., Hojabr, R., Najafi, M.H., Khonsari, A., Gholamrezayi, M.H., Gorgin, S., Rahmati, D.: Using Residue Number Systems to Accelerate Deterministic Bit-stream Multiplication. In: 2019 IEEE 30th International Conference on Application-specific Systems, Architectures and Processors (ASAP). IEEE (2019). doi: 10.1109/ASAP.2019.00-33
5. Hariri, A., Navi, K., Rastegar, R.: A Simplified Modulo $(2n-1)$ Squaring Scheme for Residue Number System. In: EUROCON 2005 - The International Conference on "Computer as a Tool." IEEE (2005). doi: 10.1109/EURCON.2005.1630004
6. Singh, T.: Residue number system for fault detection in communication networks. In: 2014 International Conference on Medical Imaging, m-Health and Emerging Communication Systems (MedCom). IEEE (2014). doi: 10.1109/MedCom.2014.7005995

7. Krasnobayev, V., Kuznetsov, A., Kononchenko, A., Kuznetsova, T.: Method of data control in the residue classes. In Proceedings of the Second International Workshop on Computer Modeling and Intelligent Systems (CMIS-2019). Zaporizhzhia, Ukraine, April 15-19, pp. 241–252 (2019)
8. Banerjee, S., Chakraborty, S., Dey, N., Kumar Pal, A., Ray, R.: High Payload Watermarking using Residue Number System. *International Journal of Image, Graphics and Signal Processing*. 7, 1–8 (2015). doi: 10.5815/ijigsp.2015.03.01
9. Krasnobayev, V., Kuznetsov, A., Zub, M., Kuznetsova, K.: Methods for comparing numbers in non-positional notation of residual classes. In Proceedings of the Second International Workshop on Computer Modeling and Intelligent Systems (CMIS-2019). Zaporizhzhia, Ukraine, April 15-19, pp. 581–595 (2019)
10. Gbolagade, K.A., Cotofana, S.D.: Residue Number System operands to decimal conversion for 3-moduli sets. In: 2008 51st Midwest Symposium on Circuits and Systems. IEEE (2008). doi: 10.1109/MWSCAS.2008.4616918
11. Parhami, B.: On equivalences and fair comparisons among residue number systems with special moduli. In: 2010 Conference Record of the Forty Fourth Asilomar Conference on Signals, Systems and Computers. IEEE (2010). doi: 10.1109/ACSSC.2010.5757827
12. Krasnobayev, V., Kuznetsov, A., Koshman, S., Moroz, S.: Improved Method of Determining the Alternative Set of Numbers in Residue Number System. In: *Advances in Intelligent Systems and Computing*. pp. 319–328. Springer International Publishing (2018). doi: 10.1007/978-3-319-97885-7_31
13. Phalakarn, K., Surarerks, A.: Alternative Redundant Residue Number System Construction with Redundant Residue Representations. In: 2018 3rd International Conference on Computer and Communication Systems (ICCCS). IEEE (2018). doi:10.1109/CCOMS.2018.8463305
14. Younes, D., Steffan, P.: Efficient image processing application using residue number system. *Proceedings of the 20th International Conference Mixed Design of Integrated Circuits and Systems - MIXDES 2013*, Gdynia, pp. 468-472 (2013)
15. Yatskiv, V., Tsavolyk, T., Yatskiv, N.: The correcting codes formation method based on the residue number system. In: 2017 14th International Conference The Experience of Designing and Application of CAD Systems in Microelectronics (CADSM). IEEE (2017). doi: 10.1109/CADSM.2017.7916124
16. Popov, D.I., Gapochkin, A.V.: Development of Algorithm for Control and Correction of Errors of Digital Signals, Represented in System of Residual Classes. In: 2018 International Russian Automation Conference (RusAutoCon). IEEE (2018). doi:10.1109/rusautocon.2018.8501826
17. Kocherov, Y.N., Samoylenko, D.V., Koldaev, A.I.: Development of an Antinoise Method of Data Sharing Based on the Application of a Two-Step-Up System of Residual Classes. In: 2018 International Multi-Conference on Industrial Engineering and Modern Technologies (FarEastCon). IEEE (2018). doi:10.1109/fareastcon.2018.8602764
18. Kasianchuk, M., Yakymenko, I., Pazdriy, I., Melnyk, A., Ivasiev, S.: Rabin's modified method of encryption using various forms of system of residual classes. In: 2017 14th International Conference The Experience of Designing and Application of CAD Systems in Microelectronics (CADSM). IEEE (2017)
19. Krasnobayev, V., Kuznetsov, A., Lokotkova, I., Dyachenko, A.: The Method of Single Errors Correction in the Residue Class. In: 2019 3rd International Conference on Advanced Information and Communications Technologies (AICT). IEEE (2019). doi:10.1109/AIACT.2019.8847845

20. Roshanzadeh, M., Saqaeean, S.: Error Detection & Correction in Wireless Sensor Networks By Using Residue Number Systems. *International Journal of Computer Network and Information Security*. 4, 29–35 (2012). doi: 10.5815/ijcnis.2012.02.05
21. Krasnobaev, V., Kuznetsov, A., Babenko, V., Denysenko, M., Zub, M., Hryhorenko, V.: The Method of Raising Numbers, Represented in the System of Residual Classes to an Arbitrary Power of a Natural Number. In: 2019 IEEE 2nd Ukraine Conference on Electrical and Computer Engineering (UKRCON). IEEE (2019). doi:10.1109/UKRCON.2019.8879793
22. Kasianchuk, M., Yakymenko, I., Pazdriy, I., Zastavnyy, O.: Algorithms of findings of perfect shape modules of remaining classes system. In: *The Experience of Designing and Application of CAD Systems in Microelectronics*. IEEE (2015)
23. Krasnobayev, V.A., Koshman, S.A.: A Method for Operational Diagnosis of Data Represented in a Residue Number System. *Cybernetics and Systems Analysis*. 54, 336–344 (2018). doi:10.1007/s10559-018-0035-y
24. Patent for invention No. 112731, Ukraine, IPC G 06 F 11/08 (2006.01). Signs of monitoring and diagnostics Presented in system standards. (2016)
25. Krasnobayev, V., Koshman, S., Yanko, A., Martynenko, A.: Method of Error Control of the Information Presented in the Modular Number System. In: 2018 International Scientific-Practical Conference Problems of Infocommunications. Science and Technology (PIC S&T). IEEE (2018). doi:10.1109/infocommst.2018.8632049
26. Chornei, R.K., Daduna V.M., H., Knopov, P.S.: Controlled Markov Fields with Finite State Space on Graphs. *Stochastic Models*. 21, 847–874 (2005). doi:10.1080/15326340500294520
27. Runovski, K., Schmeisser, H.: On the convergence of fourier means and interpolation means. *Journal of Computational Analysis and Applications*. 6(3), 211–227 (2004)
28. Tkach, B.P., Urmancheva, L.B.: Numerical-analytic method for finding solutions of systems with distributed parameters and integral condition. *Nonlinear Oscillations*. 12, 113–122 (2009). doi:10.1007/s11072-009-0064-6
29. Bondarenko, S., Liliya, B., Oksana, K., & Inna, G.: Modelling instruments in risk management. *International Journal of Civil Engineering and Technology*. 10(1), 1561–1568 (2019)
30. Ponochovniy, Y., Bulba, E., Yanko, A., Hozbenko, E.: Influence of diagnostics errors on safety: Indicators and requirements. In: 2018 IEEE 9th International Conference on Dependable Systems, Services and Technologies (DESSERT). IEEE (2018). doi:10.1109/dessert.2018.8409098

Extrapolation to Calculate the Probability of a Double Spending Attack

Nikolay Poluyanenko¹ [0000-0001-9386-2547], Alexandr Kuznetsov¹ [0000-0003-2331-6326],
Elizaveta Lazareva¹ [0000-0002-6212-8048] and Andrey Marakushyn² [0000-0002-9060-5120]

¹ V. N. Karazin Kharkiv National University, Svobody sq., 4, Kharkiv, 61022, Ukraine
nlfsr01@gmail.com, kuznetsov@karazin.ua,
lazareva15elizaveta@gmail.com

² Simon Kuznets Kharkiv National University of Economics, Nauky avenue 9a, Kharkiv,
61166, Ukraine
Andrey.marakushyn@hneu.net

Abstract. One of the important aspects of the efficiency of modern distributed networks built using blockchain technologies is the study of the security of consensus protocols. In particular, the most common cryptocurrencies and blockchain systems with probabilistic consensus protocols are subjects to so-called double-spending attack. The basis of such an attack is the use of the attacker's computing capabilities to form alternative blockchains. If the generated sequence is longer than the public chain of blocks, the attacker can present it as the proof of work and thus disrupt the correct functioning of the network. In this article we explore the probability of a successful double-spending attack, derive formulas for evaluating the corresponding events, consisting of the formation by the attacker of an alternative sequence of blocks. These formulas are extremely cumbersome and difficult to calculate. The paper proposes simplified analytical expressions to quickly assess the probability of a successful double-spending attack. For this we use the extrapolation of intermediate calculations using the Lagrange interpolation formulas, as well as binomial approximation. The simulation results show that the use of simplified expressions allows us to provide acceptable accuracy of calculations.

Keywords: blockchain technology, consensus protocols, proof-of-work, double-spending attack, polynomial interpolation, binomial approximation

1 Introduction

Modern decentralized systems are a real alternative to the classical centralized approach in creating of complex information and telecommunication systems and networks [1-4]. For example, it is advisable to use decentralization in complex systems of corporate relations, in the provision of state and administrative functions, to build independent cryptocurrencies, as well as to create secure registers, cadastres, etc. [1-

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

3]. These and many other aspects show the clear advantage of decentralized systems built on blockchain technology [5, 6].

At the same time, for the most developed and studied blockchain networks with probabilistic consensus protocols, there is a risk of a double-spending attack. The essence of the attack is to use the features of consensus building protocols. For example, in the proof-of-work protocols, each record in the protected registry (each block of records) can be made only by the participant who was the first to solve the exhaustive task of finding the preimage of the cryptographic hash function. The block chain stores all the important records, and each following record (block of records) is connected with the previous ones through a one-way hash function. This is a reliable and safe way to save critical data, protect it from falsification, alteration, intentional or accidental modification. At the same time, an attacker who has sufficient computing resources may try to form an alternative chain of blocks (with other records in previous blocks). This will give him the opportunity to disrupt the proper functioning of the blockchain network, for example, by spending the same assets twice (cryptocurrencies, tokens, etc.). The probability of such unauthorized interference is extremely small (with the small computing capabilities of an individual attacker it is almost impossible). However, when building a critical information infrastructure, an extremely important aspect is the assessment of various risks and threats to information security, including the probabilities of successful implementation of various attacks. Therefore, assessing the security of blockchain systems and, in particular, assessing the likelihood of a successful implementation of a double-spending attack in systems with probabilistic consensus protocols, is an urgent and practically important scientific task.

In this article, we consider consensus protocols with probabilistic completion (such as proof-of-work), and assess the probability of a successful double-spending attack. For this, we use the model of independent players (as in our previous works [4]). This model differs from the gambler's ruin model adopted in [2, 3] by both the number of elementary outcomes and general expressions for assessing the probability of a successful attack. To simplify computational calculations, we use polynomial interpolation, as well as binomial coefficients to extrapolate intermediate calculations. The results of numerical simulation show that simplified formulas can significantly reduce computational costs and obtain approximate estimates with acceptable accuracy of calculations. These results can be used to calculate the security indicators of blockchain networks with probabilistic consensus protocols.

2 The probabilities of forming a chain of blocks under the same initial conditions

To obtain the formula for the probability of a successful double-spending attack, let us consider the following possible probabilities and combinations in which the attacker succeeds:

1. Success is not possible on the first attempt ($t = N + j + k = 1 + 0 + 0 = 1$). The attacker may form no more than one block in one attempt, and in order to succeed, he needs to wait for the generation of the block by the honest network and extend the chain of blocks to one more. Thus, the attacker needs to form at least two blocks. The probability of success of the attacker is defined as: $PI_{N=1, j=0, k=0} = 0$
2. The attacker may succeed on the second attempt ($t = N + 0 + 1 = 2$) if he managed to form on both attempts per block, and the honest network will form only one block (no matter on the first or second attempt). The total probability of this event:

$$PI_{N=1, j=0, k=1} = p \cdot (1-p) \cdot [q \cdot q] + (1-p) \cdot p \cdot [q \cdot q] = 2 \cdot p \cdot (1-p) \cdot q^2$$
3. On the third attempt ($t = N + 0 + 2 = 3$). If the honest network forms a block on the first or second attempt, then the second block should be formed by the attacker only on the third attempt, otherwise (if the second block is formed on the second attempt) there is a previous case (as for $PI_{N=1, j=0, k=1}$). If the honest network forms the second block on the third attempt, there are no restrictions on the formation of blocks for the attacker. The total probability of the event will be calculated similarly to the events described above. Thus, the total probability will be:

$$PI_{N=1, j=0, k=2} = 2 \cdot p^1 \cdot (1-p)^2 \left[2 \cdot q^2 \cdot (1-q)^1 \right] + p^1 \cdot (1-p)^2 \left[3 \cdot q^2 \cdot (1-q)^1 + q^3 \right].$$

4. On the fourth attempt ($t = N + 0 + 3 = 4$) similar to the situation described in the previous paragraph, the total probability for the attacker to succeed will be equal to:

$$PI_{N=1, j=0, k=3} = 3 \cdot p^1 \cdot (1-p)^3 \left[3 \cdot q^2 \cdot (1-q)^2 \right] + p^1 \cdot (1-p)^3 \left[6 \cdot q^2 \cdot (1-q)^2 + 4 \cdot q^3 \cdot (1-q)^1 + q^4 \right].$$

5. On the t -th attempt ($t = N + 0 + k$) the probability of success of the attacker is equal to:

$$PI_{N=1, j=0, k=k} = p^1 \cdot (1-p)^k \cdot \left\{ k^2 \cdot q^2 \cdot (1-q)^{k-1} + \left[1 - \binom{k+1}{0} \cdot q^0 \cdot (1-q)^{k+1} - \binom{k+1}{1} \cdot q^1 \cdot (1-q)^k \right] \right\}$$

Carrying out similar constructions and summing all the $k = 0, 1, 2, \dots$, we find the probability of a successful double-spending attack, provided that the honest network is formed of no more than $N = 1$ blocks:

$$PI_{N=1, j=0} = \sum_{k=1}^{\infty} \left\{ p \cdot (1-p)^k \cdot \left[k \cdot q^2 \cdot (1-q)^{k-1} + \right. \right. \\ \left. \left. + \left(1 - \binom{k+1}{0} \cdot q^0 \cdot (1-q)^{k+1} - \binom{k+1}{1} \cdot q^1 \cdot (1-q)^k \right) \right] \right\}.$$

Let us consider the case when $N = 2, j = 0$.

If $k = 0$, as well as in the previous case, success is impossible.

If $k = 1$ and, therefore, $t = N + 0 + 1 = 3$, the probability of success of the attacker is equal to:

$$PI_{N=2, j=0, k=1} = 3 \cdot p^2 \cdot (1-p) \cdot q^3.$$

If $k = 2$ ($t = 4$), the probability of successful attack is equal to:

$$PI_{N=2, j=0, k=2} = 3 \cdot p^2 \cdot (1-p)^2 \cdot [3 \cdot q^3 \cdot (1-q)] + \\ + 3 \cdot p^2 \cdot (1-p)^2 \cdot [4 \cdot q^3 \cdot (1-q) + q^4].$$

If $k = 3$ ($t = 5$), the probability of success of the attacker is equal to:

$$PI_{N=2, j=0, k=3} = 6 \cdot p^2 \cdot (1-p)^3 \cdot [6 \cdot q^3 \cdot (1-q)^2] + \\ + 4 \cdot p^2 \cdot (1-p)^3 \cdot [10 \cdot q^3 \cdot (1-q)^2 + 5 \cdot q^4 \cdot (1-q)^1 + 1 \cdot q^5].$$

The above results allow us to obtain expressions for arbitrary values of N and k :

$$PI_{N=N, j=0} = \sum_{k=1}^{\infty} p^N \cdot (1-p)^k \cdot \left\{ \binom{t-1}{N} \cdot q^{N+1} \cdot (1-q)^{k-1} + \right. \\ \left. + \binom{t-1}{N-1} \cdot \left[1 - \sum_{i=0}^N \binom{t}{i} \cdot q^i \cdot (1-q)^{t-i} \right] \right\}.$$

Let us consider the case for $N = 1, j = 1$.

If $k = 0$ attacker's success is impossible

If $k = 1$ the probability of success of the attacker is equal to:

$$PI_{N=1, j=1, k=1} = p \cdot p \cdot (1-p) \cdot [q \cdot q \cdot q].$$

If $k = 2$ ($t = 4$) the probability of the attacker's success is equal to:

$$PI_{N=1, j=1, k=2} = p^2 \cdot (1-p)^2 \cdot [q^3 \cdot (1-q) \cdot \{3 + 2 + 2\}].$$

If $k = 3$ ($t = 5$) the probability for the attacer to succeed is equal to:

$$PI_{N=1, j=1, k=3} = p^{N+j} \cdot (1-p)^k \cdot q^{N+j+1} \cdot (1-q)^{k-1} \cdot \left\{ \binom{k}{1} \cdot \binom{t-1}{2} + \binom{k-1}{1} \cdot \binom{t-2}{2} + \binom{k-2}{1} \cdot \binom{t-3}{2} \right\}.$$

Thus, summarizing the results for arbitrary initial values, we obtain the probability of a successful double-spending attack (PI) on blockchain systems using the consensus algorithm Proof of Work based on a hash function (without the attacker's advantage in one pre-formed block):

$$PI = p^N \cdot \sum_{k=1}^{\infty} (1-p)^k \cdot \left\{ \binom{t-1}{N} \cdot q^{(N+1)} \cdot (1-q)^{(k-1)} + \binom{t-1}{N-1} \cdot \left[1 - \sum_{i=0}^N \binom{t}{i} \cdot q^i \cdot (1-q)^{(t-i)} \right] \right\} + \sum_{j=1}^{\infty} \left\{ p^{(N+j)} \cdot q^{(N+j+1)} \cdot \sum_{k=1}^{\infty} \left[sum_p \cdot (1-p)^k \cdot (1-q)^{(k-1)} \right] \right\}, \quad (1)$$

where

$$t = N + j + k ;$$

$$sum_p = \sum_{ip_{(N+j)}=1}^k \left(\sum_{ip_{(N+j-1)}=ip_{(N+j)}+1}^{k+1} \left(\dots \sum_{ip_2=ip_3+1}^{t-2} \left(\sum_{ip_1=ip_2+1}^{t-1} (sum_q) \right) \right) \right);$$

$$sum_q =$$

$$= \sum_{iq_{(N+j)}=1}^k \left(\sum_{iq_{(N+j-1)}=iq_{(N+j)}+1}^{k+1} \left(\dots \sum_{iq_{(j+1)}=iq_{(j+2)}+1}^{t-j-1} \left(\sum_{iq_{(j)}=\max(iq_{(j+1)}+1, ip_{(j)})}^{t-j} \left(\dots \sum_{iq_2=\max(iq_3+1, ip_2)}^{t-2} \left(\sum_{iq_1=\max(iq_2+1, ip_1)}^{t-1} (1) \right) \right) \right) \right) \right).$$

Although the expression (1) provides an accurate quantitative result on the probability of successful double-spending attacks, it also has limitations on its use, which is related to the polynomial complexity of computational calculations. In this paper, we propose using an approximate formula to calculate the probability of a successful attack. For this, simplified expressions based on polynomial interpolation can be used. In addition, we explore other approximation methods, including using binomial coefficients.

3 Extrapolation of the sum in the formula of calculation of the probability of success of the attacker

The value of sums (sum_p and sum_q) is increasing very fast and even at $j = 10$ and $k = 15$ the calculation becomes a very difficult task. For example, if $j = 20$ and $k = 7$ the sum $sum_p = 16\,570\,275\,123$ (this value was calculated by the computer for several hours). On the other hand, these sums for each value N could be calculated once and used for different probabilities. Tables 1, 2, as an example, provide calculated values of sum_p for $N = 1, 5$ and some values of j and k .

Table 1. Values of sum_p in the expression(1) for $N = 1$

j	k						
	1	2	3	4	5	6	7
1	1	7	25	65	140	266	462
2	1	11	58	210	602	1470	3192
3	1	16	117	563	2073	6327	16797
4	1	22	213	1314	6041	22528	71775
5	1	29	359	2761	15495	69305	260923
6	1	37	570	5345	35950	189909	833918
7	1	46	863	9690	76927	473768	2399565
8	1	56	1257	16648	154007	1093596	6327475
9	1	67	1773	27349	291592	2364642	15498742
10	1	79	2434	43256	526520	4835606	35639160
11	1	92	3265	66225	912695	9423549	77586723
12	1	106	4293	98570	1526907	17608428	161007165
13	1	121	5547	143133	2476031	31706737	320288355
14	1	137	7058	203359	3905808	55248173	613629478
15	1	154	8859	283376	6011425	93484314	1136709035
16	1	172	10985	388080	9050125	154064036	2042783757
17	1	191	13473	523225	13356092	247916850	3571657702
18	1	211	16362	695518	19357870	390392550	6090688552
19	1	232	19693	912719	27598589	602713571	10151890353
20	1	254	23509	1183746	38759285	913805304	16570275123

However, given the limited number of calculated coefficients of sum_p , the expression (1) gives a good match with the experimental data (setting up the computational experiment is given in [4]) when the probabilities q and p significantly (twice or more) differ from each other. However, there is no need to calculate the large number of values in the sum by j , that allows to be limited to some small pre-calculated set of values of sum_p . Also, at $q, p > 0,2$ the blocks will be formed with a relatively

high probability, that makes it possible to significantly reduce the sum by k . In addition, for smaller values N , the sum sum_p grows more slowly, making it possible to calculate sum_p for more values j and k .

However, in order to improve the accuracy of the calculation (increasing the calculated coefficients j and k), it is possible to extrapolate the values of sum_p using polynomial approximation. Thus, for $j=1$ the value sum_p is very well approximated (within known values of k) and extrapolated (beyond calculated values of k) by a polynomial:

$$sum_p(N=1, j=1, k) = 0,125 \cdot k^4 + 0,4167 \cdot k^3 + 0,375 \cdot k^2 + 0,0833 \cdot k,$$

for $j=2$:

$$sum_p(N=1, j=2, k) = 0,0097 \cdot k^6 + 0,0792 \cdot k^5 + 0,2431 \cdot k^4 + 0,3542 \cdot k^3 + 0,2464 \cdot k^2 + 0,0947 \cdot k - 0,2158.$$

Extrapolation is also well done by j , and the value of k is fixed.

Table 2. Values of sum_p in the expression (1) for $N=5$.

j	k						
	1	2	3	4	5	6	7
1	1	43	631	5335	31795	148219	575107
2	1	51	900	9100	64215	350709	1578214
3	1	60	1265	15185	125925	799834	4145505
4	1	70	1745	24600	237279	1736315	10277050
5	1	81	2361	38661	429387	3587388	24053848
6	1	93	3136	59045	748230	7079128	53381664
7	1	106	4095	87850	1259860	13400268	112900788
8	1	120	5265	127660	2056860	24434838	228674250
9	1	135	6675	181615	3266253	43084995	445514295
10	1	151	8356	253486	5059063	73710049	838128214
11	1	168	10341	347755	7661745	122712954	1527675457
12	1	186	12665	469700	11369715	199311469	2705845884
13	1	205	15365	625485	16563225	316537844	4669213780
14	1	225	18480	822255	23725842	492518292	7867415920
15	1	246	22051	1068236	33465804	752091712	12969669012
16	1	268	26121	1372840	46540540	1128836172	
17	1	291	30735	1746775	63884655	1667581587	
18	1	315	35940	2202160	86641695	2427497877	
19	1	340	41785	2752645	116200021	3485859706	
20	1	366	48321	3413536	154233135	4942601727	

As established experimentally, it is desirable to use the Lagrangian interpolation polynomial for computational methods (see, for example, [7] or [8]). The essence of which is as follows. We know some values of sum_p at the first values of x_i (we assume x_i to be k_i or j_i), where $i = 1, 2, \dots, n$, and n is the number of points by which the Lagrangian polynomial is constructed (in our case, for $N = 1$, n must be chosen as $n = 3 + 2 \cdot j$ for extrapolations by k and as $n = 2 \cdot k \pm 1$ for extrapolations by j), then we can extrapolate sum_p by Lagrange interpolation polynomial:

$$sum_p = \sum_{i=0}^n \left(sum_p(x_i) \cdot \prod_{\substack{s=0, \\ s \neq i}}^n \frac{x - x_s}{x_i - x_s} \right).$$

The results of the extrapolation of value of sum_p are illustrated in Figure 1, which shows a graph of the dependence of precisely calculated values of sum_p (solid line) by formula (1) and its approximation and extrapolation (dashed line).

Extrapolation using a polynomial gives a very good accuracy, but with increasing k (j) increases the number of first values of sum_p that must be known and whose values are taken into account in the Lagrangian interpolation polynomial. Given that we are aware of a rather limited number of sum_p (for example, in Table 2 for $k = 7$ this is only for $j \leq 15$) polynomial extrapolation will also have some limits.

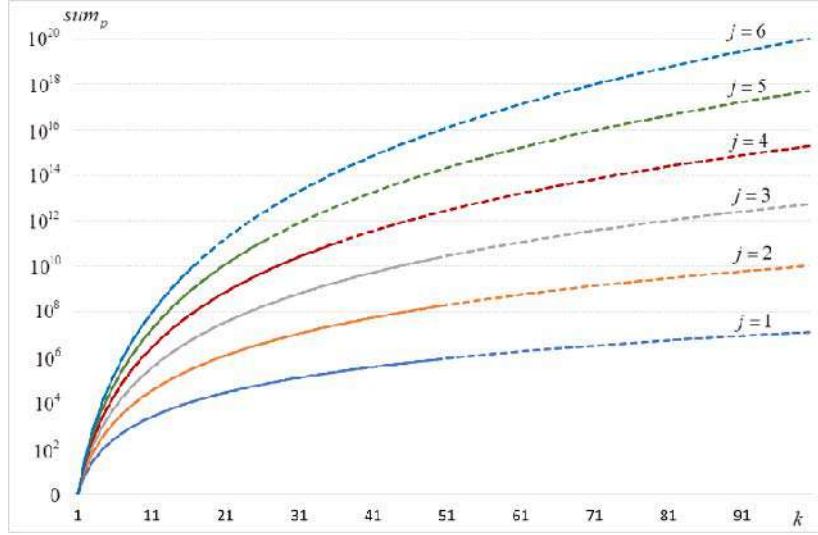
The next extrapolation method we have applied is binomial extrapolation, that is, extrapolation using binomial coefficients of the form

$$sum_p = d \cdot \binom{N + j + k_i - 1}{N + j}^2$$

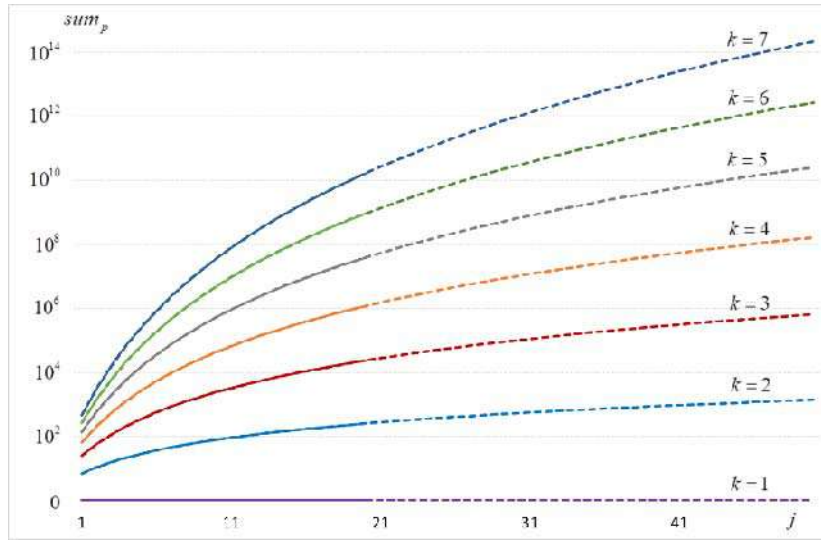
where d – some coefficient $0 < d < 1$, which is selected experimentally.

Binomial extrapolation gives much less precision than polynomial, but better than neglecting additives in general. This is especially noticeable when $q \approx p$ and a significant number of coefficients j and k must be considered.

Thus, the results obtained can significantly reduce the complexity of the calculations when calculating the probability of a successful double-spending attack. In particular, to calculate the probabilities by the formula (1), it is necessary to calculate an infinite number of sums sum_p , each of which in turn is formed by summing a large number of terms. We were able to simplify these intermediate sums by introducing interpolation polynomial equations. By extrapolating the sums to a larger data range, we can replace complex and cumbersome calculations with simple and convenient calculations based on interpolation polynomials. The simulation results show that the values calculated in this way almost 100% repeat the results obtained by exact expressions using formula (1).



a)



b)

Fig. 1. Dependencies of values of sum_p (solid line) calculated by formula (1) and its approximation and extrapolation (dashed line) for $N = 1$ a) extrapolation by k b) extrapolation by j

These results can be useful for assessing the security of blockchain networks with probabilistic consensus building [9-11]. In particular, practical recommendations for constructing asset transfer protocols in decentralized systems can be justified through restrictions on the probability of corresponding risks. For example, given the restrictions on the admissible probability of loss of an asset as a result of a double-spending

attack, our formulas allow us to calculate the minimum number of confirmed transactions (length of a chain of blocks) and these calculations can be performed very quickly even in the conditions of a rapid change in the share of controlled computing resources. These and other studies are our promising areas of further work.

4 Conclusions

This paper address in detail one of the main vulnerabilities of blockchain systems built by consensus with probabilistic completeness - the double-spending attack.

Basing on the model of "independent players", we have obtained the analytical expression of the probability of a successful double-spending attack on a blockchain system which uses the consensus algorithm Proof of Work (PoW) based on a hash function depending on the number of confirmations used and the number of attempts and hash rate of both the honest network and the attacker.

The adopted model of "independent players" and the resulting formula (1) eliminates the significant disadvantages inherent in other work in this field, namely:

- the race between two participants of the network does not have to be endless, it is enough to be limited by some fixed number of attempts;
- uses a more adequate, in the authors' view, model of "independent players", which includes a space of four elementary events, instead of two, used in the model of "gambler's ruin";
- the probability of forming a block with the honest network and the attacker are independent quantities that are directly determined by the capacities possessed by the participants, and the mentioned probabilities are independent on each other, that is, the requirement $p + q = 1$ is optional;
- it is calculated the probability for the attacker to be ahead of the honest network, and not only the probability of catching up with it, that is, when the attacker does not have an advantage in one pre-formed block.

The quantitative values obtained by the expression (1) of the probability of a successful attack for different opportunities of the attacker (the probability of forming a block), the different number of formed blocks after which the agreement is considered confirmed, the different duration of the race (the number of blocks during which the attacker continues to catch up the network) are given. To simplify the numerical calculations, it is proposed to use polynomial approximation. In particular, the intermediate sums in formula (1) were able to be approximated by Lagrange interpolation polynomials. In addition, we used binomial extrapolation. However, the simulation results show that polynomial interpolation provides greater accuracy. The formulas obtained make it possible to quickly assess the probability of a successful double-spending attack using these simplified formulas.

This study can be useful for modeling and evaluating the effectiveness of blockchain networks, as well as in other practically important applications [12-16].

References

1. Zaghloul, E., Li, T., Mutka, M.W., Ren, J.: Bitcoin and Blockchain: Security and Privacy. ArXiv, abs/1904.11435 (2019)
2. Nakamoto, S.: Bitcoin: A Peer-to-Peer Electronic Cash System (2009)
3. Rosenfeld, M.: Analysis of hashrate-based double-spending (2014)
4. Poluyanenko, N.A., Kuznetsov, A.A.: Simulation of double spend attack on the “Proof of Work” consensus protocol. *Radiotekhnika*. 3, 146–161 (2019). doi: 10.30837/rt.2019.3.198.11 (in Russian)
5. Zaghloul, E., Li, T., Mutka, M.W., Ren, J.: Bitcoin and Blockchain: Security and Privacy (2019)
6. Ozisik, A.P., Levine, B.N.: An Explanation of Nakamoto's Analysis of Double-spend Attacks (2017)
7. Turchak, L.I., Plotnikov, P.V.: Fundamentals of numerical methods. Moscow, Fizmat-lit. 304 p. (2003) (in Russian)
8. Interpolating Polynomial, <http://dx.doi.org/10.3840/001276>, (2007)
9. Pilkington, M.: Blockchain technology: principles and applications, <http://dx.doi.org/10.4337/9781784717766.00019>, (0)
10. Salman, T., Jain, R., Gupta, L.: Probabilistic Blockchains: A Blockchain Paradigm for Collaborative Decision-Making. In: 2018 9th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON). IEEE (2018). doi: 10.1109/UEMCON.2018.8796512
11. Bhat, M., Vijayal, S.: A Probabilistic Analysis on Crypto-Currencies Based on Blockchain. In: 2017 International Conference on Next Generation Computing and Information Systems (ICNGCIS). IEEE (2017). doi: 10.1109/ICNGCIS.2017.37
12. Kuznetsov, A., Shekhanin, K., Kolhatin, A., Kovalchuk, D., Babenko, V., Perevozova, I.: Performance of Hash Algorithms on GPUs for Use in Blockchain. In: 2019 IEEE International Conference on Advanced Trends in Information Theory (ATIT). IEEE (2019). doi: 10.1109/ATIT49449.2019.9030442
13. Bondarenko, S., Liliya, B., Oksana, K., & Inna, G.: Modelling instruments in risk management. *International Journal of Civil Engineering and Technology*. 10(1), 1561-1568 (2019)
14. Runovski, K., Schmeisser, H.: On the convergence of fourier means and interpolation means. *Journal of Computational Analysis and Applications*. 6(3), 211-227 (2004)
15. Tkach, B.P., Urmancheva, L.B.: Numerical-analytic method for finding solutions of systems with distributed parameters and integral condition. *Nonlinear Oscillations*. 12, 113–122 (2009). doi:10.1007/s11072-009-0064-6
16. Chornei, R.K., Daduna V.M., H., Knopov, P.S.: Controlled Markov Fields with Finite State Space on Graphs. *Stochastic Models*. 21, 847–874 (2005). doi:10.1080/15326340500294520

On parallel algebraic multilevel preconditioner with approximate inversion by basis matrix method

V.O. Bohaienko¹[0000-0002-3317-9022]

¹ V.M. Glushkov Institute of Cybernetics of NAS of Ukraine, Glushkov ave., 40, Kyiv, 03187, Ukraine

sewab@ukr.net

Abstract. The paper presents a parallel calculation scheme for distributed memory systems for a modification of the AMLI preconditioner that performs incomplete inversions of matrix blocks by a technique that produces approximate inverse of a matrix by an orthogonalization procedure, referred to as the incomplete basis matrix method. Theoretical estimates of algorithm's performance and the results of its experimental testing are presented. The results of experimental analysis of preconditioner parameters influence on the convergence of the iterative algorithm are given. The characteristics of the proposed algorithm are compared with those of known algorithms implemented in the *hype* software package. The tests on the matrices obtained from finite element discretization of soil stress-strain state modelling problem showed that the proposed algorithm has better performance for significantly ill-conditioned linear systems when solving them on a small number of computational resources.

Keywords. parallel algorithm, distributed memory systems, linear algebraic systems, preconditioner, algebraic multilevel iteration method

1 Introduction

The need to solve linear algebraic systems arises in mathematical modelling of most physical processes, in particular when modelling soil stress-strain state. In many cases, linear systems have ill-conditioned sparse matrices of large size. In these situations, iterative methods, such as conjugate gradient method (CG) or biconjugate gradient stabilized method (BiCGstab), are most commonly used [1].

The main method used to improve the convergence and the accuracy while solving ill-conditioned linear systems is the use of preconditioners [1, 2] - matrices, whose multiplication on linear system matrix leads to the decrease of condition number. The most commonly used methods for constructing preconditioners are incomplete matrix decomposition (e.g. ILU decomposition [1]), incomplete matrix inversion (e.g. polynomial preconditioners [3]), Geometric and Algebraic Multigrid (AMG) [4] methods, the Algebraic Multilevel Iteration method (AMLI) [5,6], and wavelet preconditioners [7,8].

The so-called basis matrix method [9, 10] is one of the methods developed to ana-

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

lyse ill-conditioned and rectangular linear systems. This method constructs iteratively the inverse of a given matrix using some orthogonalization procedure. Based on this method, an incomplete inversion algorithm is created, that can be used as a preconditioner. Further, it is combined [11] with the AMLI preconditioner and applied for solving elasticity problems.

The speed of the AMLI preconditioner, similarly to the multigrid ones, in particular the AMG, is low comparing to the incomplete decomposition and the incomplete inversion preconditioners and its application takes most of the time spent to obtain solutions of linear systems. This makes crucial the development of parallel schemes for these algorithms. While the AMLI is shown to have good scalability on vector and parallel-vector machines [12, 13], its parallel implementations on cluster systems are poorly studied.

2 Preconditioner on the base of the AMLI and the basis matrix method

The AMLI preconditioner [5] is based on the representation of an input matrix as having 2x2 blocks with its subsequent LU factorization:

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} = \begin{bmatrix} I & \\ A_{21}A_{11}^{-1} & I \end{bmatrix} \begin{bmatrix} A_{11} & A_{12} \\ S & \end{bmatrix} \quad (1)$$

where $S = A_{22} - A_{21}A_{11}^{-1}A_{12}$ is the Schur complement of A .

After approximating the Schur complement with a polynomial and the recursive application of the expansion (1) to the matrix A_{22} , the following algorithm for constructing a preconditioner that approximates the matrix A^{-1} is obtained:

$$\begin{cases} S^{(k-1)} \approx A_{22}^{(k-1)} \left[I - P_v \left(\left(M^{(k-1)} \right)^{-1} A_{22}^{(k-1)} \right) \right]^{-1} \\ M^{(k)} = \begin{bmatrix} I & \\ A_{21}^{(k)} \left(A_{11}^{(k)} \right)^{-1} & I \end{bmatrix} \begin{bmatrix} A_{11}^{(k)} & A_{12}^{(k)} \\ S^{(k-1)} & \end{bmatrix} \end{cases} \quad (2)$$

For the matrix A of the size $\dim A = [K * L, K * L]$ that is divided into $L * L$ blocks $\bar{A}_{ij}, i, j = 1, \dots, L$ of the size $\dim \bar{A}_{ij} = [K, K]$ we have

$$\begin{aligned} A_{11}^{(k)} &= \bar{A}_{L-k, L-k}, \dim A_{11}^{(k)} = [K, K], \\ A_{12}^{(k)} &= (\bar{A}_{L-k, L-k+1}, \dots, \bar{A}_{L-k, L}), \dim A_{12}^{(k)} = [K, kK], \\ A_{21}^{(k)} &= (\bar{A}_{L-k+1, L-k}, \dots, \bar{A}_{L, L-k}), \dim A_{21}^{(k)} = [kK, K], \end{aligned}$$

$$A_{22}^{(k)} = \begin{bmatrix} \bar{A}_{L-k+1,L-k+1} & \cdots & \bar{A}_{L-k+1,L} \\ \cdots & \cdots & \cdots \\ \bar{A}_{L,L-k+1} & \cdots & \bar{A}_{LL} \end{bmatrix}, \dim A_{22}^{(k)} = [kK, kK],$$

$$\dim S^{(k-1)} = \dim A_{22}^{(k)}, S^{(0)} = \bar{A}_{LL}^{-1}, \dim M^{(k)} = [(k+1)K, (k+1)K], k = 1, \dots, L-1$$

and $M^{(L-1)}$ is an approximate inverse of A .

The AMLI algorithm can be optimally stabilized by choosing the degree of an approximating polynomial [12]. However, the applicability of this theory while solving stress-strain problems described by hyperbolic differential equations wasn't proved. As such problems are of our specific interest, we consider the simplest case of $P_v(t) = P_l(t) = 1-t$. Then, the Schur complement approximation becomes $S^{(k-1)} \approx M^{(k-1)}$ and the recursive scheme (2) simplifies to

$$M^{(k)} = \begin{bmatrix} I & \\ A_{21}^{(k)} (A_{11}^{(k)})^{-1} & I \end{bmatrix} \begin{bmatrix} A_{11}^{(k)} & A_{12}^{(k)} \\ & M^{(k-1)} \end{bmatrix}. \quad (3)$$

In the scheme (3), the matrices $(A_{11}^{(k)})^{-1}$ can be approximated by incomplete decomposition methods such as the ILU factorization [14] or incomplete Gram-Schmidt method [15]. To make the scheme more efficient while solving ill-conditioned linear systems, we propose to approximate them by the incomplete basis matrix method that is specifically developed for this case and is described in detail in [16].

Summarizing, the sequential computational scheme for the preconditioning procedure (3) is as follows. The matrix A is divided into $(N+1) \times (N+1)$ blocks the following way:

$$A = \begin{bmatrix} A_{11}^{(0)} & A_{21}^{(0)} & \cdots & A_{12N}^{(0)} \\ A_{211}^{(0)} & \cdots & \cdots & \cdots \\ \cdots & \cdots & A_{11}^{(N-1)} & A_{12N}^{(N-1)} \\ A_{21N}^{(0)} & \cdots & A_{21N}^{(P-1)} & A_{11}^{(N)} \end{bmatrix},$$

$$A_{12}^{(i)} = (A_{12,i+1}^{(i)}, \dots, A_{12N}^{(i)}), A_{21}^{(i)} = (A_{21,i+1}^{(i)}, \dots, A_{21N}^{(i)}).$$

Considering the application of the preconditioner (3) as a solution of the linear system $M^{(N)}x = b, x = (x_0, \dots, x_N), b = (b_0, \dots, b_N)$, it can be performed by a two-stage procedure, V-cycle, consisting of a "direct" and a "reverse" part.

On the "direct" part of the procedure, values of the following vectors are sequentially computed:

$$w_0 = \tilde{A}_{11}^{(0)} b_0, w_i = b_i - A_{21}^{(i)} \tilde{A}_{11}^{(i)} w_{i-1}, i = 1, \dots, N. \quad (4)$$

The "reverse" part consists in the sequential computation of the resulting vector

$$x_N = \tilde{A}_{11}^{(N)} w_N, x_i = w_i - \tilde{A}_{11}^{(i)} A_{12}^{(i)}(x_{i+1}, \dots, x_N), i = 0, \dots, N-1. \quad (5)$$

In (4),(5), matrices $\tilde{A}_{11}^{(i)}$ are the approximates to the corresponding inverse matrices obtained by the incomplete basis matrix method [16].

As the initial matrix is sparse, we use the compressed row storage (CSR) scheme for all the matrices.

The approximation accuracy is controlled on the iterations of the algorithm by calculating $\varepsilon_1 = \max_i (A\tilde{A})_{ii}$ for all the processed rows. When ε_1 becomes bigger than a given threshold value, we perform complete inversion. As the resulting matrices aren't always dense, we also use the CSR storage scheme doing complete inversion.

3 Parallel implementation of the preconditioner

To make the AMLI algorithm combined with matrix blocks incomplete inversions using the basis matrix method run in parallel on distributed memory systems, the following data partitioning scheme can be applied. Each of i row blocks

$$(A_{21i}^{(0)}, \dots, A_{21i}^{(i-1)}, A_{11}^{(i)}, A_{12,i+1}^{(i)}, \dots, A_{12N}^{(i)})$$

is distributed on a system of P processes by blocks of rows. Thus, each process stores blocks of rows of the submatrices $A_{11}^{(i)}, A_{12}^{(i)}, A_{21j}^{(i)}$ and can perform multiplications that arise while applying the preconditioner (3) in parallel.

The computational scheme that we use for multiplying a sparse matrix A of size $M \times M$ by a vector v has the following form:

1. The process P stores the block
 $B_p = [r_{p0}, r_{p1}], \bigcup_{i=0}^P B_i = [0, \dots, M], B_i \cap B_j = \emptyset, \forall i, j$ of the matrix A rows and the corresponding elements of the vector v ;
2. Denote a set of columns of the matrix A in which non-zero elements are present in the row block B_p as C_p ;
3. The process P sends values of the elements of the vector v from the index set $C_i \cap B_p$ to the processes $i, C_i \cap B_p \neq \emptyset$ and receives the values of the elements from the index set $C_p \cap B_j$ from the processes $j, C_p \cap B_j \neq \emptyset$;
4. After the synchronization is performed on the step 3, each of the processes independently multiplies the block of rows of the matrix A by the vector v . The result of this multiplication has the same partitioning as the vector v has.

Performing the "direct" part (4) of the AMLI algorithm, when vector multiplications by the matrices $A_{21}^{(i)}$ are organized as successive multiplications by the submatrices $A_{21j}^{(i)}$, the synchronizations are done once according to the fill-in pattern of the matrix $A_{21}^{(i)}$. Similarly, when the non-changeable component x_i of the result vector x is multiplied by the submatrices $A_{12i}^{(j)}, j < i$ during the "reverse" part (5), the syn-

chronization can be performed only once according to the combined fill-in pattern of the submatrices $A_{21i}^{(i)}$.

The matrices $\tilde{A}_{11}^{(i)}$ can be obtained in parallel the following way:

- The process P handles the sequence $k_{p2} > k \geq k_{p1}$ of the matrices $A_{11}^{(k)}$. The sequences are constructed to make the processes evenly loaded as far as it is possible. This is ensured by approximately the same number of non-zero elements in the matrices $A_{11}^{(k)}$ that are processed by each of the processes.
- The processes exchange the values of the elements of the matrices $A_{11}^{(i)}$ to make the matrices $A_{11}^{(k)}, k_{p2} > k \geq k_{p1}$ completely stored in the memory of the process P . The process P sends row blocks B_p of the matrices $A_{11}^{(k)}, k_{i2} > k \geq k_{i1}$ to the processes $i, i \neq P$ and receives row blocks B_i of the matrices $A_{11}^{(k)}, k_{p2} > k \geq k_{p1}$ from them.
- Each of the processes independently performs a complete or incomplete inversion of the matrices.
- The processes exchange the values of the elements of the matrices $\tilde{A}_{11}^{(i)}$ to restore the initial data partitioning. The process P sends row blocks B_i of the matrices $\tilde{A}_{11}^{(k)}, k_{p2} > k \geq k_{p1}$ to the processes $i, i \neq P$ and receives row blocks B_p of $\tilde{A}_{11}^{(k)}, k_{i2} > k \geq k_{i1}$ matrices from them.

Since the data partitioning of the blocks of rows within the matrices $A_{12}^{(i)}, A_{21i}^{(j)}$ may differ for the different values of i , an additional synchronization should be performed when vectors are multiplied by these matrices. Denote a block of rows from the submatrices $A_{12}^{(i)}, A_{21i}^{(j)}$ that is handled by the process P as B_{pi} . Then, before multiplying a vector v with the partitioning $B_i = (B_{0i}, \dots, B_{pi})$ by the matrix $A_{21}^{(i)}$, we need to synchronize its elements to make each of the processes store the block $B_{pi} \cup \dots \cup B_{pN}$. To do so, the process P sends the block $B_{pi} \cap B_{rj}$ of the elements of the vector to the process r if $B_{pi} \cap B_{rj} \neq \emptyset, j \neq i$. Then, the process P receives a corresponding block from the process r if $B_{pj} \cap B_{ri} \neq \emptyset, j \neq i$.

Similar procedures should be performed when a vector $v = (v_{i+1}, \dots, v_N)$ is multiplied by the matrix $A_{12}^{(i)}$ for each of its components. When the non-changeable component x_j is multiplied by the matrices $A_{12}^{(i)}$ during the "reverse" part (5) of the AMLI algorithm, the synchronization can be performed once before calculating the result of the first multiplication operation.

When the preconditioner is used with an iterative method for solving a linear alge-

braic system, we perform a calculation of the matrices $\tilde{A}_{11}$ and the sets of the elements that need synchronization once on the first iteration. The execution of these operations will be further denoted as "initialization procedure".

4 Estimation of algorithms performance

We estimate the performance of the considered algorithms on the base of the following assumptions:

- the matrices $A_{11}^{(i)}$ have the same fill-in factor k_{11} , and the matrices $A_{12j}^{(i)}, A_{21j}^{(i)}$ - the same fill-in factor k_2 which does not dependent on the block size;
- A percentage k_f of the matrices $A_{11}^{(i)}$ for which the full inversion procedure is applied does not depend on the block size;
- The inverse matrices of $A_{11}^{(i)}$ are completely filled;
- The number of elements that each of the processes must synchronize before performing a parallel multiplication is considered proportional (with the coefficient of proportionality equal to k_s) to the number of non-zero elements in the block of matrix rows. We assume that each of the processes performs only one pair of data exchange operations doing the synchronization;
- The submatrices are considered evenly distributed across the system.

Under these assumptions, the time sequential algorithm spend to perform calculations on (4), (5) can be estimated as

$$T_1(n, N) = k_p \left[\frac{n^2}{N} (2(k_f + (1 - k_f)k_{11}) + (N - 1)k_2) + 4n + n(N - 1) \right] \quad (6)$$

where k_p is the performance coefficient.

The minimal time is achieved here for the number of blocks

$$N = N_m = \sqrt{(2(k_f + (1 - k_f)k_{11}) - k_2)n}. \quad (7)$$

The time spent on computation of the matrices $\tilde{A}_{11}^{(k)}$ with an assumption that the computational complexity of one step of the incomplete basis matrix method is proportional to the maximal number of non-zero elements in the rows of the matrices $A_{11}^{(k)}$ can be estimated as follows:

$$T_{i1}(n, N) = Nk_{p2} \left(k_f \frac{n^3}{N^3} + k_{11} \frac{n^2}{N^2} \right) \quad (8)$$

where k_{p2} is the performance coefficient.

We evaluate the computation time of the parallel algorithm using the linear model of a pairwise data exchange operation:

$$T_s(n) = n/k_e + k_a \quad (9)$$

where k_e is the bandwidth, n is the data size, k_a is the latency.

Taking into account (9), the total time spent by the parallel algorithm on performing calculations on (4), (5) can be estimated as

$$T_p(n, N) = \frac{1}{P} T_1(n, N) + 2 \frac{n}{k_e} \left[\frac{k_s}{P} ((1-k_f)k_{11} + \frac{N-1}{N} k_2) + \frac{P-1}{P} k_f \right] + (2N(2+k_f P - 2k_f) - 2)k_a \quad (10)$$

where P is the number of processes.

The minimal time spent on data exchanges for a fixed N is achieved when the number of processes is equal to

$$P_m(N) = \sqrt{\frac{2nk_s((1-k_f)k_{11} + \frac{N-1}{N} k_2)}{2k_e k_a k_f N}}. \quad (11)$$

When doing parallel computations, the time spent to calculate the matrices $\tilde{A}_1^{(k)}$ can be estimated with the assumption that each of the processes performs incomplete

or complete inversion of $\frac{N}{P}$ submatrices (we consider this number to be integer) as

$$T_{ip}(n, N) = \frac{1}{P} T_{il}(n, N) + \frac{N}{P} (P-1) \left(\left(\frac{k_{11} n^2}{N^2 P k_e} + k_a \right) + \left((1-k_f)k_{11} + k_f \right) \frac{n^2}{N^2 P k_e} + k_a \right). \quad (12)$$

5 Experimental analysis of the algorithm's performance

We apply the proposed parallel modification of the preconditioner to solve linear systems arising while modelling stress-strain state of the soil under a dam using the model of dynamic consolidation of saturated soils as described in [11].

The two-dimensional model has the form of the hyperbolic differential equations' system with respect to 8 unknown function - the horizontal and the vertical components u_w, v_w, u_{sk}, v_{sk} of water and soil skeleton displacement vectors and their first derivatives. The used boundary and the initial conditions, solution domain, and parameter values are given in [11]. The problem was discretised by the Galerkin finite element method on a rectangular mesh with respect to the space variables and by the Crank-Nicolson scheme with respect to the time variable.

While forming linear systems, their rows were ordered to form blocks related to a single unknown function. To minimize a number of synchronizations we use the

Cuthill-McKee ordering of mesh nodes. Performing the diagonal scaling of rows we obtain matrices that have the following properties: non-symmetry; loosely coupled block structure; high fill-in factor; high condition number due to different speeds of the simulated processes. As in the considered model of soil stress-strain state we have 8 unknown functions, the matrices can be easily represented as 2x2 block matrices. Such properties of matrices make them an appropriate testing case for the AMLI preconditioner that is built upon the idea of block matrix subdivision and thus can be efficient working with matrices that have loosely coupled diagonal blocks in their structure. The usage of the incomplete basis matrix method here is beneficial because matrices contain significantly ill-conditioned blocks and other methods like incomplete LU decomposition can be unstable in this case.

The performance and the convergence of the considered algorithm were experimentally studied on the SCIT-3 cluster of the Institute of Cybernetics of National Academy of Sciences of Ukraine. The eight-core nodes based on Intel Xeon 5345 processors with 2GB RAM per core were used.

The algorithm was tested on the matrix which arises on the first time step with a length $\tau = 2\text{hours}$. This matrix of this linear system, further denoted as SD, has the size equal to 107856×107856 with an average number of non-zero elements in a row equal to 23.42. Its condition number estimated by the LSMR algorithm [17] has an order of 10^{12} .

Linear system was solved using the BiCGStab [1] algorithm with the proposed modification of the AMLI preconditioner. The BiCGStab was used because of non-symmetric nature of the matrix. The submatrices $A_{11}^{(k)}$ were completely inverted when $\varepsilon_1 < 0.05$. This resulted in the complete inversion of : 50% of the submatrices. The acceptable accuracy of the calculations (here and further - the value below which the residual has to be reduced on BiCGStab iterations) was equal to 10^{-9} .

The use of the BiCGStab algorithm itself, or with the procedure of a diagonal scaling of rows, did not allow obtaining solutions of the required accuracy. The computation time when the BiCGStab was used without the usage of any preconditioner was equal to 140 ms per iteration.

To experimentally assess the accuracy of the estimates (6) and (7) of single preconditioner application execution time on a single processor core without the initialization procedure, the linear system SD was solved with the variable size of blocks on which the matrix is split. The values of the estimates' coefficients were, here and further, found using the least squares method on the base of the measured execution time. The block size which minimizes the computation time was between 125 and 500, which corresponds to the calculated value of the theoretical estimate (7) equal to 179. The obtained experimental data confirm the properties of the algorithm reflected in the estimate (6) whose accuracy was not worse than 25%.

The performance estimate (10) in the similar case for the parallel algorithm was verified measuring the time spent on data exchanges between processes. The estimation error here remains within 30% for the block size ranging from 250 to 3000 and the number of processes ranging from 12 to 24. In other cases, the error increases. For a fixed block size, the optimal number of processes, that ensures the minimal time

spent on exchanges, in most cases, was slightly lower than the values calculated by the estimate (11).

The time spent on the calculations of the matrices $\tilde{A}_{11}^{(k)}$ (initialization procedure) depending on the block size in the case of the sequential algorithm was measured for the block size ranging from 125 to 4000. Then, the time was assessed by the estimate (8). The accuracy of the estimate was within 12% for the block size less than 3000, while the block size increase leads to the increase of the error. These inaccuracies may be caused by a change in the percentage of fully inverted submatrices. In the case of the parallel algorithm, the corresponding time was measured and estimated by (12). The general behaviour of the estimate (12) coincides with the measured time (Fig.1) but the estimation errors were high.

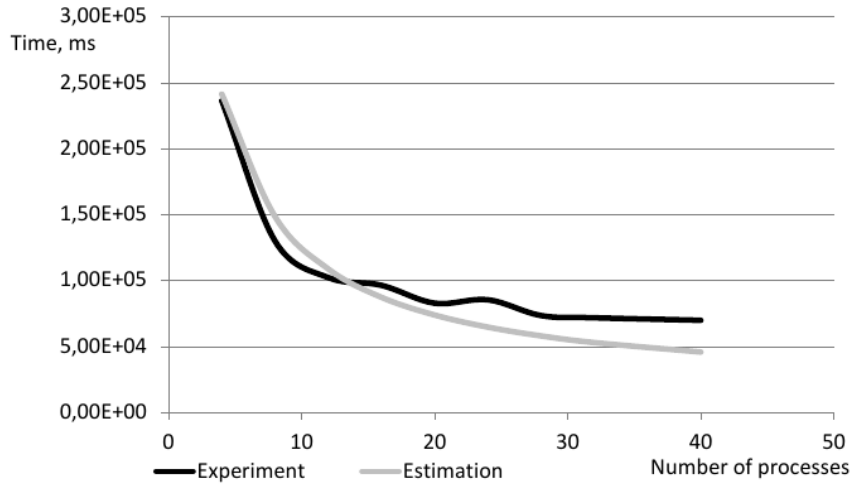


Fig. 1. Time, ms, spent on the calculations of $\tilde{A}_{11}^{(k)}$ matrices and its estimate in the case of the parallel algorithm

Further, we studied an influence of the size of blocks on which the matrix is split on the iterative method's convergence. The number of iterations required to achieve 10^{-9} order of accuracy was measured along with the overall time spent by the BiCG-stab algorithm and the proposed preconditioner running on a single processor core. Here, an increase of the block size leads to a decrease of the fill-in factor of the matrices $A_{11}^{(k)}$. As the incomplete basis matrix method performs better in terms of accuracy on highly filled sparse matrices, the decrease of the fill-in factor results in the increase of the number of completely inverted matrices. This accelerates the convergence of the iterative algorithm (Fig. 2) along with the significant increase of its running time. The block size for which the required accuracy is reached in the minimal time was in the range from 75 to 125 elements.

The decrease of the number of fully inverted submatrices $A_{11}^{(k)}$ reduces the time spent on preconditioner initialization and thus significantly decreases the running time the parallel algorithm in the case when a small number of computational resources are

used. On the other hand, the speed-up of the initialization operation is higher for the higher values of k_f , so the time spent on initialization becomes close to equal for the different values of k_f when the number of processes increase. The time spent on the parallel calculations on (4), (5) behaves likewise. The smaller number of completely inverted submatrices, in general, accelerates this operation, but the speed-up decreases with the increase of the number of computing resources involved. All this, given that convergence is faster for a larger number of completely inverted submatrices, leads to the situation when k_f reduction is effective only when a small number of processes are involved.

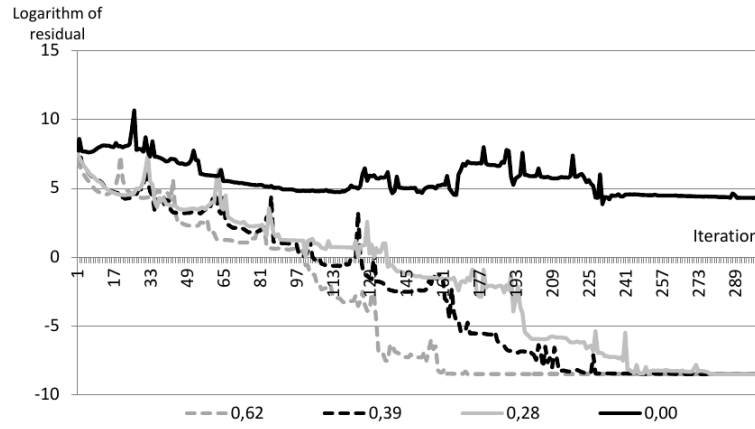


Fig. 2. The logarithm of residual at the iterations of the algorithm for different percentages of completely inverted submatrices

The efficiency of the proposed parallel preconditioner was compared with the efficiency of known algorithms implemented in the hypr software package v.2.10.0b [18]. The only combination of an iterative method and a preconditioner implemented in the hypr which allowed obtaining solutions of 10^{-9} accuracy order was the Gmres [1] and the AMG preconditioner [3] (we will further denote this pair as the Gmres+AMG). The total time needed to obtain the solution of the linear system SD with 10^{-9} order of accuracy is given in Fig. 3. Here the proposed parallel preconditioner was used with the block size equal to 250 and 500. It was faster than the Gmres+AMG when the number of processes was less than 20 with the block size equal to 500. However, the scalability of the proposed parallel algorithm was lower comparing with the Gmres+AMG.

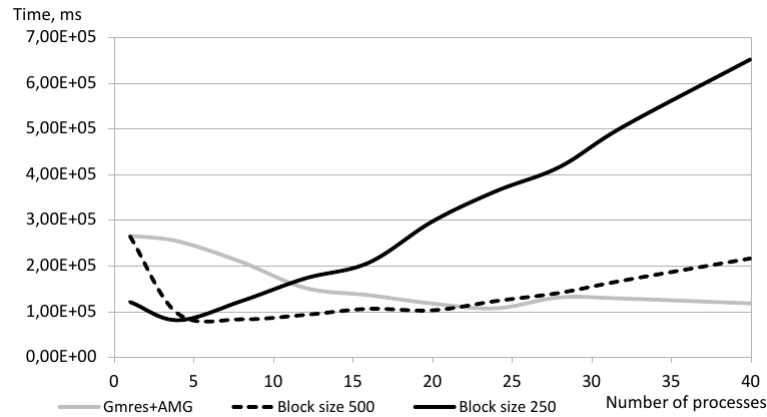


Fig. 3. Computation time, ms, subject to the number of processes

6 Conclusions

The parallel computation scheme was constructed for the AMLI preconditioner with incomplete inversion of matrix blocks using the so-called basis matrix method and the theoretical estimates of its performance were obtained and experimentally verified.

For the linear system obtained while modelling stress-strain state of soil which has the size equal to 107856×107856 , the condition number of 10^{12} order, the maximal achieved speed-up was 24% when scaling the algorithm on 8 processes for the block size equal to 500 elements. The low scalability of the algorithm here can be explained by the large number of data exchange operations that must be performed before each matrix-vector multiplication.

We compared the characteristics of the proposed preconditioner with the AMG used as preconditioner in the Gmres algorithm implemented in the hypre software package. The scalability of the proposed algorithm was generally worse, but it had better performance for the considered significantly ill-conditioned linear system when solving it on less than 20 processor cores.

References

1. Saad, Y.: Iterative methods for sparse linear systems, 2 edition. Society for Industrial and Applied Mathematics (2003). doi: 10.1137/1.9780898718003
2. Chen, K.: Matrix Preconditioning Techniques and Applications. Cambridge University Press (2005). doi: 10.1017/CBO9780511543258
3. Saad, Y.: Practical use of polynomial preconditionings for the conjugate gradient method. SIAM J. Sci. Stat. Comp. **6**, 865-881 (1985). doi:10.1137/0906059
4. Shapira, Y.: Matrix-Based Multigrid: Theory and Applications. Springer Science & Business Media (2003).
5. Mense, C., Nabben, R.: On algebraic multilevel methods for non-symmetric systems - convergence results. Electron. Trans. Numer. Anal. **30**, 323-345 (2008).

6. Bayramov, N.R., Kraus, J.K.: Multigrid methods for convection–diffusion problems discretized by a monotone scheme. *Comput. Method. Appl. M.* **317**, 723-745 (2017). doi:10.1016/j.cma.2017.01.004
7. Chan, T., Tang, W., Wan, W.: Wavelet sparse approximate inverse preconditioners. *BIT Numerical Mathematics* **37**, 644-660 (1997). doi:10.1007/BF02510244
8. Bianchi, D., Buccini, A.: Generalized Structure Preserving Preconditioners for Frame-Based Image Deblurring. *Mathematics* **8(4)**, 468 (2020). doi:10.3390/math8040468
9. Kudin, V.I., Liashko, S.I., Kharytonenko, N.V., Iatsenko, Iu.P.: Analysis of the properties of a linear system using the method of artificial basis matrices. *Cybernetics and Systems Analysis* **43(4)**, 563-570 (2007). doi:10.1007/s10559-007-0081-3
10. Bogaienko, V.O., Kudin, V.I., Skopetsky, V.V.: Aspects of the organization of computations using the basis-matrix method. *Cybernetics and Systems Analysis* **48(4)**, 30-34 (2012). doi:10.1007/s10559-012-9441-8
11. Bohaienko, V.O.: Using Basis Matrix Method with AMLI Preconditioner for Solving Elasticity Problems. In: *Proceedings of the 4th International Scientific Conference of Students and Young Scientists "Theoretical and Applied Aspects of Cybernetics"*; Kyiv, Ukraine, pp.24-28 (2014).
12. Axelsson, O., Barker, V.A., Neytcheva, M., Polman, B.: Solving the Stokes Problem on a Massively Parallel Computer. *Math Model Anal.* **6**, 7-27 (2001).
13. Li, Zh., Wu, Sh., Xu, J., Zhang, Ch.: Toward Cost-Effective Reservoir Simulation Solvers on GPUs. *Adv. Appl. Math. Mech.* **8(6)**, 971-991 (2016). doi:10.4208/aamm.2015.m1138
14. Saad, Y.: Schur Complement Preconditioners for Distributed General Sparse Linear Systems. In: *Domain Decomposition Methods in Science and Engineering XVI, Lecture Notes in Computational Science and Engineering*, 55, pp. 127-138 (2007).
15. Liu, Q.H., Zhang, F.D.: Incomplete hyperbolic Gram-Schmidt-based preconditioners for the solution of large indefinite least squares problems. *J. Comput. Appl. Math.* **250**, 210-216 (2013). doi:10.1016/j.cam.2013.02.016
16. Bohaienko, V.O., Kudin, V.I.: Building preconditioners using basis matrix method. *International Journal "Information Content and Processing"* **1(2)**, 182-187 (2014).
17. Fong, D.C.-L., Saunders, M.: LSMR: An Iterative Algorithm for Sparse Least-Squares Problems. *SIAM J. Sci. Comput.* **33(5)**, 2950-2971 (2011). doi:10.1137/10079687X
18. Falgout, R.D., Yang, U.M.: hypre: A Library of High Performance Preconditioners. In: Sloot PMA, Hoekstra AG, Tan CJK, Dongarra JJ (editors). *Lecture Notes in Computer Science*, vol 2331, Wien: Springer Verlag, pp. 632-641 (2002).

Managing Multifractal Properties of The Binary Sequence Generated With The Markov Chains

Hanna Drieieva¹[0000-0002-8557-3443], Oleksii Smirnov¹[0000-0001-9543-874X] and
Oleksandr Drieiev¹[0000-0001-6951-2002] Volodymyr Simakhin²[0000-0003-4497-0925],
Serhiy Bondar²[0000-0003-4140-7985] and Roman Odarchenko^{3,4}[0000-0003-4140-7985]

¹ Central Ukrainian National Technical University, Ukraine, Kropyvnytskyi

² International Research and Training Center for Information Technologies and Systems,
Ukraine, Kyiv

³ National Aviation University, Kyiv, Ukraine, 03058

⁴Yessenov University, Aktau, Kazakhstan
dr.SmirnovOA@gmail.com

Abstract. With the development of modern telecommunications systems and the exponential growth in the demand for information transmission, there is a constant lack of bandwidth for the available telecommunication channels under the management of routers and communicators, because it is necessary to distribute the load of the network segments taking into account their self-similar nature of the traffic. Now, it is not possible to analytically build criteria and algorithms for optimal traffic management to ensure QoS measurements. The correspondence of the practical results with the theoretical ones was experimentally confirmed, although non-standard metric of the coverage measure expression was used in the derivation of analytical dependencies. That's why it was decided to develop theoretical method and provide experimental confirmation using conventional methods of estimating the time series' fractal dimension. The actual possibilities of adjusting the Hearst index on a given time scaling were established as a result of the numerical experiment. The obtained time series with the help of a cascade binary sequence generator have multifractal properties. That is, the cascade generator has more possibilities for receiving traffic, which will correspond to the examples of binary traffic in real telecommunication networks. However, the cascade generator requires further theoretical study to analytically express the coefficients to obtain the desired fractal characteristics and flow intensity.

Keywords: multifractal properties, binary sequence, Markov Chains, traffic modeling, Hurst exponent.

1 Introduction

With the development of modern computer telecommunications systems and the exponential growth in the demand for information transmission, there is a constant lack of bandwidth for the available telecommunication channels under the management of

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

routers and communicators, because it is necessary to distribute the load of the network segments taking into account their self-similar nature. Unfortunately, it is not possible to analytically build criteria and algorithms for optimal traffic management to ensure QoS measurements. Therefore, to select suitable modes of work for real telecommunication equipment the simulation systems (in particular: ModelNet, Ohio Network Emulator, ENDE, EMPOWERb OPNET, Emulab, NISTNET, DummyNet, NS, GTNeS, NSE, NETWARS) are commonly used [1,2]. As a result, such systems should have telecommunication traffic simulation sources with self-similar properties that would meet the actual characteristics determined experimentally at the simulation preparation stage.

The analysis of imitation methods for telecommunication traffic sources distinguishes the following: Poisson processes; fractal Brownian motion-based generators; fractal Gaussian noise with discrete wavelet transform generation (where wavelet coefficients are independent random variables with normal distribution); Levi's fractal movement (a generalized Brownian movement, which is self-similar and forms distributions with "heavy tails"); autoregressive models (which assume that the current value of the process is the sum of a constant weighted sum of the previous values and the error of the model); neural network models (that are trained with experimental traffic to predict a new element); Markov chains (that allow you to create relatively simple models of discrete traffic generation with a wide range of properties) [3-5].

In most cases, the choice of traffic generation method should be based on the traffic properties obtained experimentally from the telecommunication system, for which traffic management algorithms will be implemented.

In [6], the authors analyzed Markov chain-based generation methods that are characterized by low computational complexity and a wide range of applications. An analytical expression of the generator parameters was solved to ensure the intensity of the simulated flow and its fractal dimension. However, the generated traffic, as shown in further studies in this article, contains multifractal properties that did not meet the real indicators, which led to the goal of improving the traffic generation simulator for telecommunication networks to replicate and manage multifractal properties.

2 Literature Analysis and Problem Statement

As a result of the previous study [6], a traffic generator G (Fig. 1) was used, which contains states "1", "0" and the probabilities to keep the state are $p("0" \rightarrow "0") = p_0$ and $p("1" \rightarrow "1") = p_1$. The generator outputs the value of the current state discretely upon an event of keeping or changing the state.

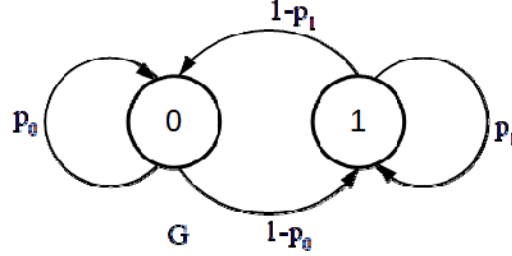


Fig. 1. Fractal Binary Sequence Generator

For such a generator flow intensity is determined as [6]:

$$\lambda = \frac{1 - p_0}{2 - (p_0 + p_1)} \quad (1)$$

Note that with $p_0 = p_1$ flow is generated with an intensity of 0.5; but as the probabilities to keep the next generated value is changed, flow's fractal dimension changes as well. A number of simulation experiments must be performed to confirm the theoretical conclusions.

The theoretical estimation of fractal dimension of the generated binary traffic was carried out [6], which depends not only on the probabilities of the generation parameters, but also on the size of the sub-band, i.e. scaling over time. To abstract from the size of the sub-band, a limit was found when the length of the sub-band tends to one. However, the obtained results are theoretical and experimental confirmation using conventional methods of estimating the time series' fractal dimension is required.

3 Evaluation of the Hurst exponent

The experimental simulation involves performing R/S analysis on the generated binary sequence. One of the R/S analysis implementations is described in [7]. The measurement is performed 200 times on different implementations of the sequence due to low accuracy of the method. The result of the experiment is Hurst exponent values, which are shown in Fig. 2 and Fig. 3, where the probabilities of keeping the state unchanged are 0.1 and 0.9, respectively:

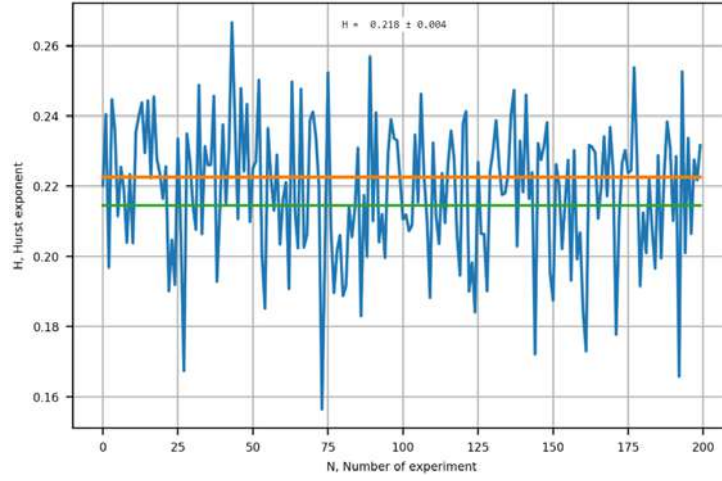


Fig. 2. Hurst exponent measurement series at 5, 10, 15, 20 samples, $p_0 = p_1 = 0.1$

Fig. 2 and Fig. 3 show that the probability of keeping the next state value equal to the previous does not change the intensity of the flow of "1", but the probability of obtaining long chains with values "0" and "1" is greatly changed. With a high probability of keeping the next value equal to the previous ($p_0 = p_1 = 0.9$), a persistent series (Fig. 3) with Hurst exponent value $H = 0.947 \pm 0.002$ is obtained. In the case of high probability to keep the next value opposite to the previous one ($p_0 = p_1 = 0.1$) – an antipersistent series with Hurst exponent $H = 0.218 \pm 0.004$ is obtained.

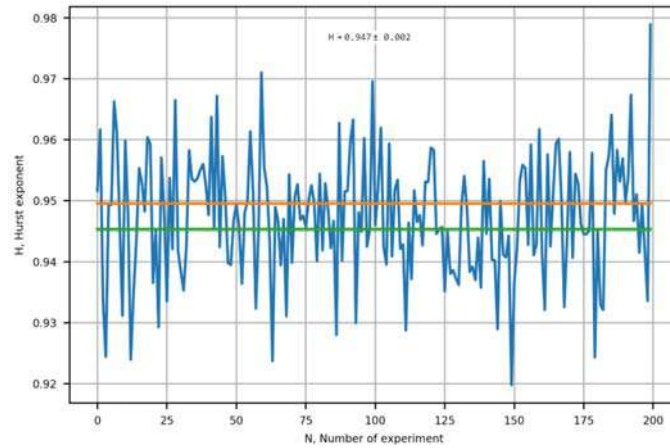


Fig. 3. Hurst exponent measurement series at 5, 10, 15, 20 samples, $p_0 = p_1 = 0.9$

In [6] the authors have obtained an estimate of the fractal dimension of a series produced by the traffic generator G (Fig. 1) on another metric, when the coverage was considered zero if the series contains only "1" or "0" elements. The fractal dimension of the series relating to the Hurst exponent as (2):

$$H(p) = \frac{p \log(p)}{1-p}, \quad (2)$$

where $p = p_0 = p_1$. The dependence of the Hurst exponent $H(p)$ on the probability of the next value change for p is shown in Fig. 4 where theoretical dependence is upper red curve and practically measured dependence is blue curve.

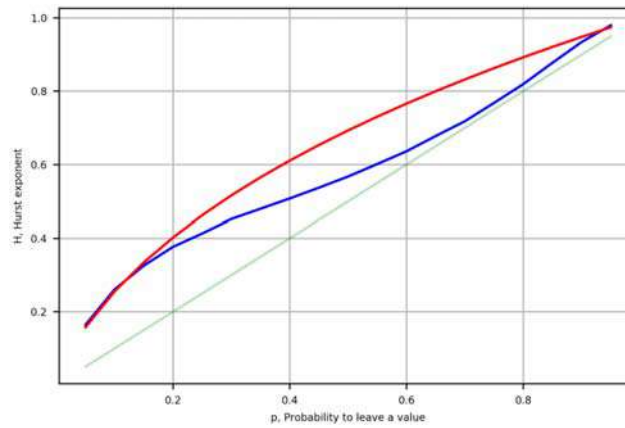


Fig. 4. The dependence of the Hurst exponent on the probability of re-generation by the traffic generator G

There are procedures of conformity assessment, formation and implementation of control actions, efficiency estimation during regular mode. In case of intelligence based mode the additional procedures are executed when efficiency decreases.

The authors assumed that the problem of efficiency decreasing can be solved based on artificial intelligence principles utilization.

Let's consider the algorithms for data processing. The list of algorithms can be presented as:

- A_1 is an algorithm of data collecting;
- A_2 is an algorithm of conformity assessment;
- A_3 is an algorithm of parameters stability assessment in case of changepoint;
- A_4 is an algorithm of preventive and corrective actions formation;
- A_5 is an algorithm of preventive and corrective actions implementation;
- A_6 is an algorithm of decision making about efficiency providing after control actions implementation;

- A_7 is an algorithm of decision making about additional data processing procedures;
- A_8 is an algorithm of decision making about usage of intelligence based procedures;
- A_9 is an algorithm of statistical processing for diagnostic variables;
- A_{10} is an algorithm of statistical processing for reliability parameters.

Algorithms A_7 , A_8 , A_9 , A_{10} have complex structure and contain the set of procedures.

All algorithms are generalized. For detailed description of algorithms it is necessary to solve synthesis and analysis problems, to choose best option for criterion of maximum efficiency, etc. Initial information for synthesis and analysis problems is measured data trends model.

Considered algorithms contain detection, estimation, filtration, extrapolation, interpolation, and other procedures. There are algorithms with known sample size, and sequential algorithms. The sequential procedures have advantages in duration of decision making [3].

Usage of adaptability principles is based on the following approaches:

- logic based solution finding;
- fuzzy logic;
- Bayesian network;
- adaptable learning after observation;
- semantic network;
- neural network, etc [4].

More over, during diagnostic variables measuring, expert evaluation and subjective probability based estimates can be used [1].

There are different types of adaptation:

- 1) adaptation to the models and models parameters;
- 2) adaptation to the external conditions;
- 3) adaptation to internal changes in operation system (OS);
- 4) adaptation to the new requirements of regulatory and normative documents;
- 5) adaptation to OS aims, etc.

4 Estimation of the Hurst exponent at different scales over time

For the binary series, the fractal dimension depends on the length of the sequence for which rescaled range (R/S) analysis is performed [8, 9]. Estimation value of the Hurst exponent can be calculated by estimating the dependence of the R/S on the time span n (where n is set of time span intervals, i.e. $n = N, N/2, N/4, \dots$). For example, under the same conditions as in Fig. 2 and 3, an increase in scaling (when the range and dispersion are calculated not at intervals of 5-20 samples, but in 450-600 samples) results the Hurst exponent values $H(0.1) \approx H(0.9) \approx 0.5$. Fig. 5 and 6 shows the results of the conducted experiment, which displays the "degeneration" of the Hurst exponent up to 0.5.

In Fig. 5 the generator G was used with the probabilities of keeping the previous state unchanged as $p_0 = p_1 = 0.1$, which should result in a visible antipersistence. However, it is obvious from that graph that aggregating cumulative sums of 450/500/550/600 samples keeps the Hurst exponent close to 0.5. Scilicet, on the specified time scale the generated sequence does not differ from a random time series in which the next element is independent of the previous one. This shows that the central limit theorem is valid, and therefore, as the quantity of samples increases for the number series' cumulative sums, these sums are getting closer to the normal distribution.

A similar impact of the cumulative sums values distribution approximation to the normal distribution is observed for the persistent series with high probability to keep the previous state $p_0 = p_1 = 0.9$, which can be seen from Fig. 6 as a decrease in the Hurst exponent values from 0.95 to 0.55.

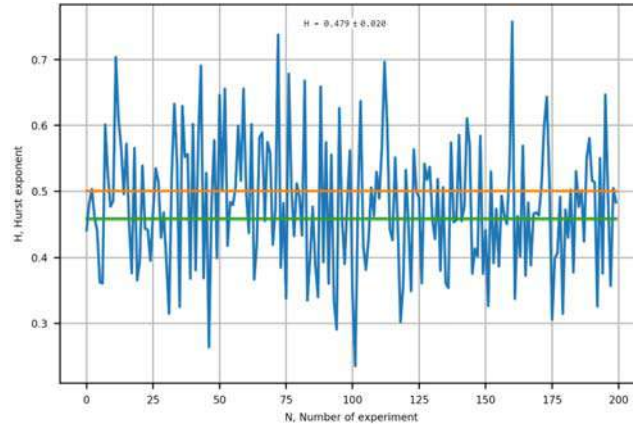


Fig. 5. Hurst exponent values in the series at 450, 500, 550, 600 samples, $p_0 = p_1 = 0.1$

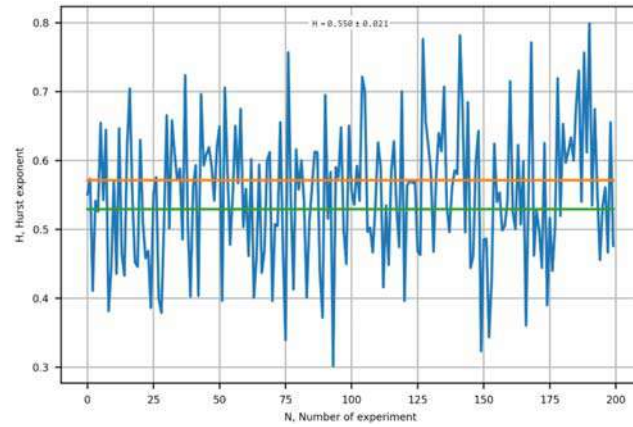


Fig. 6. Hurst exponent values in the series at 450, 500, 550, 600 samples, $p_0 = p_1 = 0.9$

As shown in [10] (Fig. 7), the change in fractal dimension with time scaling of real traffic data is relevant for quite a long time, as evidenced by publications on the development of an effective multifractal traffic generator [11, 12]. In this case, the fractal dimension can both increase and decrease with changing time scale on the real data.

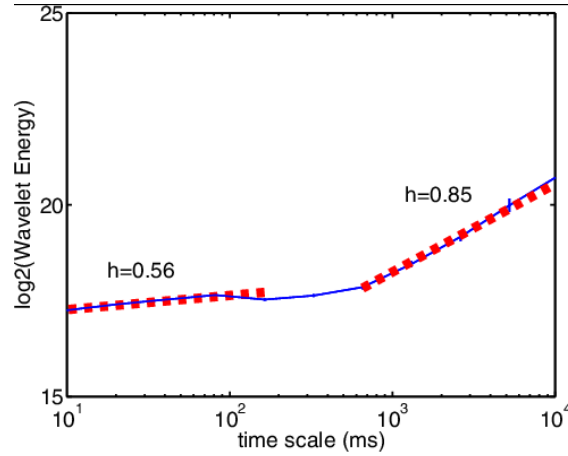


Fig. 7. Change of the Hurst exponent from time scale [10, Fig. 1.a]

In order to reveal the dependence of the Hurst exponent on the length of partial (cumulative) sums in R/S analysis, the dependence graph of the Hurst exponent on the length of these sums (Fig. 8) is shown.

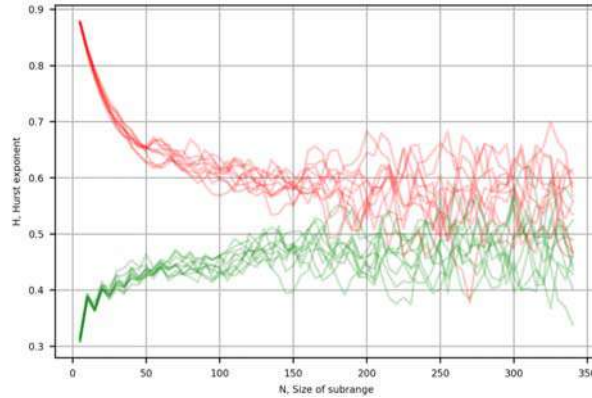


Fig. 8. Dependence of the Hurst exponent on the N interval's length in case of R/S analysis

Fig. 8 contains two bundles of curves, where each curve corresponds to one experimental determination of the Hurst exponent with the specified scaling. The upper bundle (red curves) corresponds to the experiment of measuring the Hurst exponent

with the generator G parameters set to $p_0 = p_1 = 0.9$, and the lower bundle (green curves) parameters are $p_0 = p_1 = 0.1$. It is obvious that both bundles approach $H = 0.5$.

Practical implementations of the communication networks' multifractal traffic may vary at different scales, so the purpose of this work is to add to the binary traffic generator based on stochastic automaton G a mechanism that would be able to control the fractal dimension of the binary series at different scales. This will greatly increase the relevance of the generated data to actual network traffic and extend the applicability of the generator.

Adjustment of the fractal dimension of the generated traffic based on the generator G can be performed with the proposed model (Fig. 9), which is based on a cascade binary traffic generator [13, 14]:

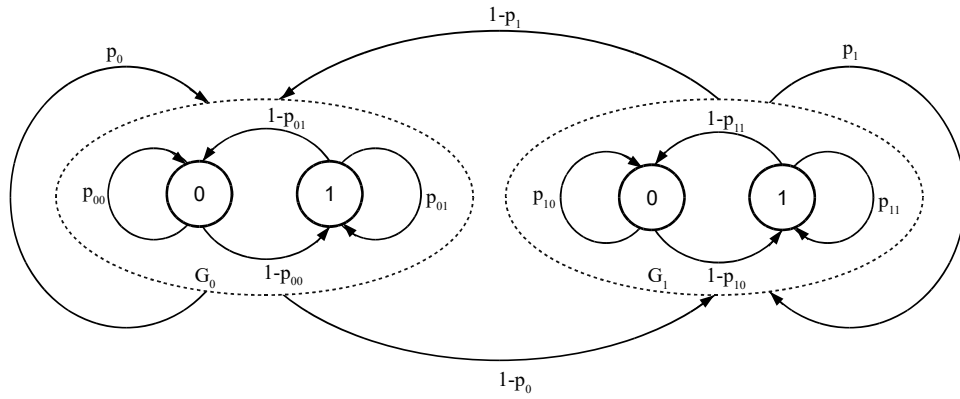


Fig. 9. Multifractal binary sequence generator

The proposed generator is cascaded and contains an external generator G , which aggregates the generators G_0 and G_1 . Such generator G is characterized by the following properties: p_{00}, p_{01} – the probability of keeping the next value unchanged for the generator G_0 ; p_{10}, p_{11} – the probability of keeping the next value unchanged for generator G_1 ; p_0, p_1 – the probability of keeping the next set of d values for the same G_1 generator as in the previous series.

According to the formation of cumulative sums, the generators G_0 and G_1 must differ in a sufficiently long period, which is not possible at the same intensity of the generated flow according to the limit theorem. Therefore, the generators are configured to generate sequences with the same fractal dimension but with different flow intensities. Therefore, on a large scale, the generator G is responsible for balancing the total flow intensity, which in this example is symmetrical and thus averages the flows from the generators G_0 and G_1 .

In order to visually evaluate the formed sequence, a binary series that is aggregated by 25 counts of the generator G cycles is shown in Fig. 10. The result is a char-

acteristic of high-intensity pulsating flow $\tau = 0.5$. This series is obtained with the following generation parameters (2):

$$\begin{aligned} d = 10; p_0 = 0.9; p_1 = 0.9; p_{00} = 0.95; p_{01} = 0.15; \\ p_{10} = 0.15; p_{11} = 0.95 \end{aligned} \quad (2)$$

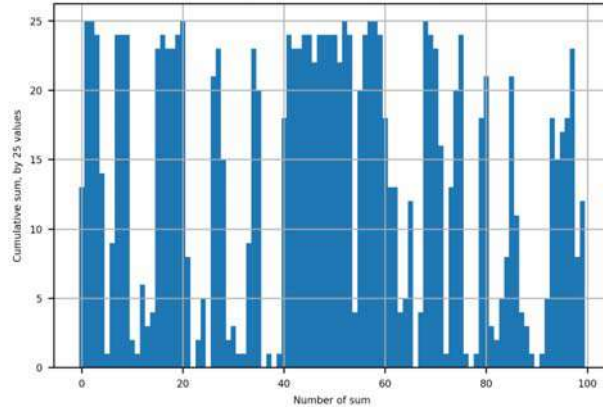


Fig. 10. An example of a cascaded traffic generator that has persistence on a wider time scale

The Fig. 11 is similarly constructed, but with the following parameters (3):

$$\begin{aligned} d = 10; p_0 = 0.05; p_1 = 0.05; p_{00} = 0.95; p_{01} = 0.15; \\ p_{10} = 0.15; p_{11} = 0.95 \end{aligned} \quad (3)$$

It should be emphasized that the changes occurred only for the traffic generator G.

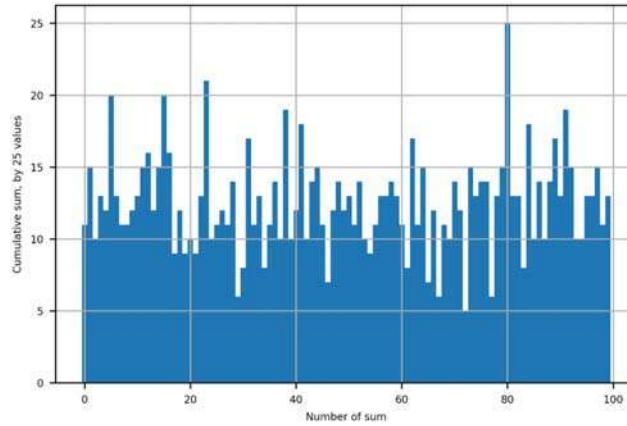


Fig. 11. An example of a cascaded traffic generator that is persistent and antipersistent on a different time scale

For the cascade generator G , a graph for the dependence of the Hurst exponent on time scale is plotted (Fig. 12). The graph contains three bundles of curves that correspond to the parameter of generator G described as (2) as (3), which have already been shown in Fig. 8, and the parameter set described below (4):

$$d = 10; p_0 = 0.0; p_1 = 1.0; p_{00} = 0.9; p_{01} = 0.9; p_{10} = 0.9; p_{11} = 0.9 \quad (4)$$

Set of parameters (4) means that the modulator switches to the operation only of the generator G_1 , with parameters $p_{10} = 0.90; p_{11} = 0.90$, and it corresponds to the operation of the generator G , without correction of the fractal dimension in large time scales. In Fig. 12 parameters set (4) is shown by the green bundle of curves, which is equivalent to the upper bundle of curves in the graph. 8. Red and blue bundles are equivalent to sets (2) and (3) respectively.

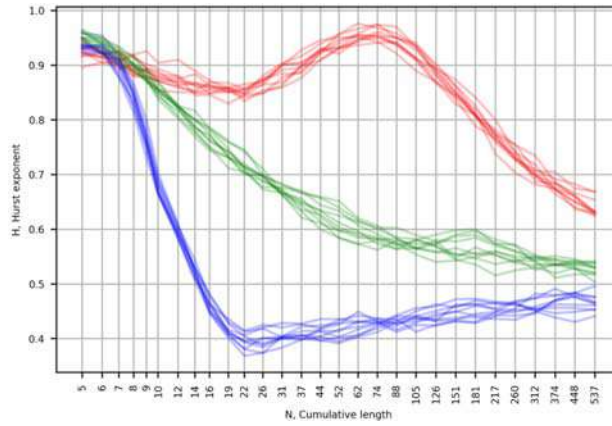


Fig. 12. Comparison of the dependence of the Hurst exponent on time scale for the cascade and standard generators G

The most important property for using a cascade generator is its ease of setup, which actually means being able to get the coefficients $d, p_0, p_1, p_{00}, p_{01}, p_{10}, p_{11}$ from the real traffic sample. In [6] the authors have already obtained the following ratio for the generator G :

$$G(0) = \frac{1 - p_1}{2 - (p_0 + p_1)}; G(1) = \frac{1 - p_0}{2 - (p_0 + p_1)} \quad (5)$$

where $G(1)$ is the probability for the modulator to use the first generator and $G(0)$ – to use a zero generator.

The corresponding flow intensities for the dependent generators can be expressed from (1) as (6):

$$\lambda_0 = \frac{1-p_{00}}{2-(p_{00}+p_{01})}; \lambda_1 = \frac{1-p_{10}}{2-(p_{10}+p_{11})} \quad (6)$$

The total flow intensity for the cascade generator can be expressed as (7) due to the known "portions" of the work of the both generators G_0 and G_1 :

$$\lambda = \frac{(1-p_1)}{2-(p_0+p_1)} \cdot \frac{(1-p_{00})}{2-(p_{00}+p_{01})} + \frac{(1-p_0)}{2-(p_0+p_1)} \cdot \frac{(1-p_{10})}{2-(p_{10}+p_{11})} \quad (7)$$

According to (7), the expression of flow intensity in traffic depends on six parameters, which indicates the possibility to change the properties of the generated traffic within wide limits. However, in practice, this dependency is not applicable, because the fractal dimension of traffic on two temporal scales needs to be linked here, but this problem has not yet been solved.

5 Conclusion

The correspondence of the practical results with the theoretical ones was experimentally confirmed, although non-standard metric of the coverage measure expression was used in the derivation of analytical dependencies.

The actual possibilities of adjusting the Hearst index on a given time scaling were established as a result of the numerical experiment. The obtained time series with the help of a cascade binary sequence generator have multifractal properties. That is, the cascade generator has more possibilities for receiving traffic, which will correspond to the examples of binary traffic in real telecommunication networks.

However, the cascade generator requires further theoretical study to analytically express the coefficients $d, p_0, p_1, p_{00}, p_{01}, p_{10}, p_{11}$ to obtain the desired fractal characteristics and flow intensity.

The generalized diagram of processing procedure is considered in the paper. This diagram suggests two modes of OS: regular mode and adaptable mode. The adaptability principles utilization expands the possibilities of flexible control of operations. Three strategies of OS components inspection were analyzed. The numerical example showed advantages of catchall inspection.

References

1. Srivastava, S., Anmulwar, S., Sapkal, A.M., Batra, T., Gupta, A.K. and Kumar, V.: Comparative study of various traffic generator tools. In: Recent Advances in Engineering and Computational Sciences (RAECS), University Institute of Engineering and Technology, Panjab University, Chandigarh, pp. 1-6 (2014)
2. Avallone, S., Pescape, A. and Ventre, G.: Analysis and experimentation of internet traffic generator. In: Next Generation Teletraffic and Wired/Wireless Advanced Networking, February 02-06, 2004, St. Petersburg, Russia, pp 70-75 (2004)
3. Robert, S. and Le Boudec, J.Y.: New models for pseudo self-similar traffic. In: Performance Evaluation, vol. 30 (1-2), pp 57-68 (1997)
4. Horn, G., Kvalbein, A., Blomskøld, J. and Nilsen, E.: An empirical comparison of generators for self similar simulated traffic. In: Performance Evaluation, vol. 64 (2), pp 162-190 (2007)
5. Sobh, T., Elleithy, K. and Mahmood, A. eds.: Novel algorithms and techniques in telecommunications and networking. In: Springer, pp 41-46 (2010)
6. Drieieva H.M., Smirnov O.A., Drieiev O.M., Smirnova T.V.; A fractal analysis of a Markov chain based self-similar traffic generator. In: Central Ukrainian Scientific Bulletin. Engineering sciences. vol. 2 (33), pp 161-172 (2019)
7. GitHub - Mottl/hurst: Hurst exponent evaluation and R/S-analysis in Python. <https://github.com/Mottl/hurst>
8. Nazarychev, S.A., Zagretidinov, A.R., Ziganshin, S.G. and Vankov, Y.V.: Classification of time series using the Hurst exponent. In: Journal of Physics: Conference Series, vol. 1328, pp 66-73. (2019)
9. Fontugne, R., Abry, P., Fukuda, K., Veitch, D., Cho, K., Borgnat, P. and Wendt, H.: Scaling in internet traffic: a 14 year and 3 day longitudinal study, with multiscale analyses and random projections. In: IEEE/ACM Transactions on Networking, vol. 25 (4), pp 2152-2165 (2017)
10. Vinay J. Ribeiro, Zhi-Li Zhang, Sue Moon, and Christophe Diot: Small-time scaling behavior of Internet backbone traffic. In: Computer Networks, vol. 48, pp.315–334 (2005)
11. G. Millána, G. Lefranc.: A Fast Multifractal Model for Self-Similar Traffic Flows in High-Speed Computer Networks. In: Information Technology and Quantitative Management. ITQM 2013 Procedia Computer Science, vol. 3, pp 420 – 425 (2013)
12. Areström, E. and Carlsson, N.: Early Online Classification of Encrypted Traffic Streams using Multi-fractal Features. In: IEEE INFOCOM 2019-IEEE Conference on Computer Communications Workshops, pp 84-89 (2019)
13. Kant, K. On aggregate traffic generation with multifractal properties. In: Seamless Interconnection for Universal Services, Global Telecommunications Conference. GLOBECOM'99. vol. 2, pp 1179-1183 (1999)
14. Anderson, D., Cleveland, W.S. and Xi, B.: Multifractal and Gaussian fractional sum-difference models for Internet traffic. In: Performance Evaluation, vol. 107, pp 1-33 (2017)

Hiding data in images using a pseudo-random sequence

Alexandr Kuznetsov¹ [0000-0003-2331-6326], Oleksii Smirnov² [0000-0001-9543-874X],
Ludmila Gorbacheva¹ [0000-0002-6053-7235] and Vitalina Babenko¹ [0000-0002-4816-4579]

¹V. N. Karazin Kharkiv National University, Svobody sq., 4, Kharkiv, 61022, Ukraine
kuznetsov@karazin.ua,

²Central Ukrainian National Technical University, avenue University, 8, Kropivnitskiy, 25006,
Ukraine, dr.smirnova@gmail.com

Abstract. In this article are discussed techniques of hiding information messages in cover image using direct spectrum spreading technology. This technology is based on the use of poorly correlated pseudorandom (noise) sequences. Modulating the information data with such signals, the message is presented as a noise-like form, which makes it very difficult to detect. Hiding means adding a modulated message to the cover image. If this image is interpreted as noise on the communication channel, then the task of hiding user's data is equivalent to transmitting a noise-like modulated message on the noise communication channel. At the same it is supposed that noise-like signals are poorly correlated both with each other and with the cover image (or its fragment). However, the latter assumption may not be fulfilled because a realistic image is not an implementation of a random process; its pixels have a strong correlation. Obviously, the selection of pseudo-random spreading signals must take this feature into account. We are investigating various ways of formation spreading sequences while assessing Bit Error Rate (BER) of information data as well as cover image distortion by mean squared error (MSE) and by Peak signal-to-noise ratio (PSNR). The purpose of our work is to justify the choice of extending sequences to reduce BER and MSE (increase PSNR).

Keywords: information concealment, steganography, direct spectrum spreading technology, pseudorandom sequences, spreading signals

1 Introduction

Steganographic techniques are traditionally used to hide the fact of transmission and the very existence of the information message [1-4]. With the development of computer science and digital methods of information processing, steganographic hiding of messages has become very common, it is used in image, audio, text documents processing etc. It is a very effective and reliable way to organize secret channels. For a third viewer, the covers that are transmitted (e.g. via e-mail) and contain information messages, hidden in them, are no different from ordinary user files. It gives the chance to organize a secret communication channel, without causing suspicions about the intentions, and to detect such channels it is extremely difficult [1, 2]. One of the prom-

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

ising trends in the development of modern steganography is the technique of embedding data in cover image, based on direct spectrum expansion technology [5-17]. This technology is traditionally used in communication systems to increase the latency of data transmission over a channel with noise [12-17]. Information data are modulated by expanding spectrum pseudo-random (noise) sequence. During transmission the received signals are statistically indistinguishable from natural noise, which increases communication latency.

Besides, the implemented methods of correlation reception allow providing correction of the occurred errors, which increases error correction of communication. These and many other advantages of direct spread spectrum technology allow building reliable and secure communication systems. For example, communication with significantly lower transmitter power can be arranged, which ensures environmentally friendly communication; application of large ensembles (sets) of expanding sequences allows increasing subscriber capacity of multiple access, etc. [18-21].

The same approach can be applied to the computer processing of digital images. Interpreting the image as noise in a communication channel and using the technology of direct spectrum extension, it is possible to organize hiding of information messages without visible container distortion. Such techniques are the subject of our article.

2 Literature review and research objective

In the first works of using the direct spectrum expansion technology in digital steganography the idea of using pseudo-random (noise) sequences as a "carrier" of information messages was put forward [5-11]. For example, for a binary case, a modulated message S is obtained by multiplying separate information bits b_i (represented in polar form $b_i \in \{-1, 1\}$) by an expanding noise signal φ_i :

$$S = \sum_i b_i \varphi_i, \quad (1)$$

Moreover φ_i belongs to an ensemble (set) of poorly correlated pseudo random sequences (PRS):

$$\forall \varphi_i \in \varphi = \{\varphi_0, \varphi_1, \dots, \varphi_{M-1}\}$$

This means that the correlation coefficient of two different signals (calculated as a scalar product of the sequences) is roughly equal zero:

$$\forall i \neq j : \rho(\varphi_i, \varphi_j) \approx 0$$

Expression (1), which describes the process of modulation of information bits $b_i \in \{-1, 1\}$ by expanding signals φ_i , is traditionally used in a broadband communication system with direct spectrum extension. Since the expanding signal φ_i on its statistical properties is similar to noise, then the received modulated message S is

slightly different from the noise in the communication channel, which allows making hidden transmission. Indeed, the transmitted messages get the form of noise-like sequences, and due to the high power of the set φ and the direct expansion of the frequency spectrum high secrecy and imitation resistance of organized communication channels are provided [18-21]. In systems with Code Division Multiple Access (CDMA) each signal φ_i is assigned to a separate pair of subscribers in other words the increase of cardinality M of a set $\varphi = \{\varphi_0, \varphi_1, \dots, \varphi_{M-1}\}$ allows you to increase the subscriber capacity of communication systems that makes data transfer cheaper [18, 19]. Steganography using direct spectrum extension uses these techniques in various files covers. For example, by interpreting a cover image I as natural noise in a communication channel is possible to organize the transmission of information messages "inside" the image [5-17]. In works [5-11] was suggested to use the sequences formed by PRS generators as a signals φ_i , after which the signal S formed by rule (1) is elementary summed up with the cover image I :

$$N = I + S . \quad (2)$$

Thus, the resulting stegano-cover image (2) is formed by adding a modulated message I to the original image (1). It is similar to how in communication systems the transferred modulated message S developed with natural noise.

On the receiving side, as in communication systems, the information message is restored using a correlation reception. For the binary case to extract the j -th bit calculate the correlation coefficient between the signal φ_j and the received N :

$$\rho(N, \varphi_j) = I\varphi_j + \varphi_j \sum_i b_i \varphi_i . \quad (3)$$

In communication systems, natural noise and noise signal φ_i are statistically independent (uncorrelated). Following our interpretations, it is logical to assume that the analogue of noise, the cover image I , is also uncorrelated to expansion signals, in other words to say $\rho(I, \varphi_j) = I\varphi_j \approx 0$. Different noise signals are also non-correlated to each other, i.e. $\forall j \neq i : \varphi_j \varphi_i \approx 0$. In this case $\rho(N, \varphi_j) \approx b_j \varphi_j \varphi_j$, i.e. value b_j can be symbolized $\rho(N, \varphi_j)$:

$$b_j = \text{sign}[\rho(N, \varphi_j)] . \quad (4)$$

Unfortunately, for steganographic applications the assumption $\rho(I, \varphi_j) = I\varphi_j \approx 0$ in (3) may not execute, when a cover image I is used. Indeed, if a realistic image, which is not an implementation of some random value sensor, is used to hide an information message, then a significant correlation I and φ_i can be observed. In this case, the recovery of information bits by formula (4) may be erroneous. In this paper we investi-

gate different ways of generating a set $\varphi = \{\varphi_0, \varphi_1, \dots, \varphi_{M-1}\}$ and estimate the Bit Error Rate (BER) when extracting a message from cover images N . In particular, we investigate the nonlinear method of formation of sequences with normal Gaussian distribution proposed in [7, 8, 10, 11], as well as Walsh's orthogonal sequences [22] and pseudo-random sequences with elements uniformly distributed on the interval $(-1,1)$. We also estimate cover image distortions by mean squared error (MSE) and Peak signal-to-noise (PSNR). These two important characteristics (BER and PSNR) clearly demonstrate the possibilities for reliable (error-free) and hidden (no significant cover distortion) transmission of information messages using direct spectrum extension technology.

3 Research Methodology

Several performance indicators are used to investigate various ways of hiding information in cover images.

BER [23] is used to evaluate the correctness of the recovered data (their reliability, error-free). BER is the number of bit errors N_{error} divided by the total number of transferred bits N_{total} :

$$BER = \frac{N_{error}}{N_{total}}. \quad (5)$$

BER is a unit less performance measure, often expressed as a percentage [23]. We estimated BER in absolute values, i.e., directly by (5). MSE and PSNR are used to estimate cover image distortion [23-25]. For monochrome $m \times n$ image I and its noisy approximation N distorted by error MSE value is determined by formula:

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I_{i,j} - N_{i,j}]^2. \quad (6)$$

PSNR characterizes the ratio between maximum signal power and distortion noise power. PSNR is usually expressed in logarithmic scale, i.e. in decibels:

$$\begin{aligned} PSNR &= 10 \cdot \log_{10} \left(\frac{I_{\max}^2}{MSE} \right) = 20 \cdot \log_{10} \left(\frac{I_{\max}}{\sqrt{MSE}} \right) = \\ &= 20 \cdot \log_{10} (I_{\max}) - 10 \cdot \log_{10} (MSE). \end{aligned} \quad (7)$$

where $I_{\max}$ is the maximum possible value of the image pixel.

If the pixels are encoded with m -bit values, then $I_{\max} = 2^m - 1$. For example, for the simplest case $m = 8$ we have $I_{\max} = 255$ and PSNR value is calculated by formula:

$$PSNR = 20 \cdot \log_{10}(255) - 10 \cdot \log_{10}(MSE). \quad (8)$$

For our experiments we used different 256×256 images, as in [7, 8, 10, 11], when encoding each monochrome pixel with one byte. In particular, we used Lenna's standard test image (256×256 pixels). The results given below are the averaged values obtained from several different images. For averaging results we used quadratic regression formulas with interpolation of the obtained results (we used built-in functions *regress* and *interp* of MathCad computer systems).

It should be noted that the results given here correspond to the use of different extension sequences, but without the use of error correction coding. For example, in operation [8] block codes are used to reduce BER in direct error correction mode. The same job (e.g. [8, Table 2]) provides BER estimates without using error correction codes. In this sense, our results can be compared with already available data.

4 Results of research

In our research, we have implemented several options for the formation of multiple expansion signals $\varphi = \{\varphi_0, \varphi_1, \dots, \varphi_{M-1}\}$. For each method we have realized hiding information in various cover images and evaluated BER, MSE and PSNR as (5-8). Our focus was on comparing the obtained results, in order to choose the best method for forming sequences φ_i .

4.1 Using non-linear modulation

In the works of Lisa M. Marvel et al. [7, 8, 10, 11] when using the technology of direct expansion of the spectrum it was suggested to use a nonlinear rule of forming a set $\varphi = \{\varphi_0, \varphi_1, \dots, \varphi_{M-1}\}$:

$$(\varphi_i)_j = \begin{cases} \Phi^{-1}((u_i)_j), b_i = -1; \\ \Phi^{-1}((u'_i)_j), b_i = 1, \end{cases} \quad (9)$$

where

$$(u'_i)_j = \begin{cases} (u_i)_j + 0.5, u_i < 0.5; \\ (u_i)_j - 0.5, u_i \geq 0.5, \end{cases} \quad (10)$$

$(u_i)_j$ - an uniformly distributed random variable over an interval of (0.1) and Φ^{-1} is an inverse cumulative distribution function for a standard Gaussian random variable.

Thus, the expanding spectrum sequence of $\varphi = \{\varphi_0, \varphi_1, \dots, \varphi_{M-1}\}$ is a realization of a random value distributed by a normal law with a zero mean and a single standard deviation. This random implementation is calculated by formula (9), i.e. using the inverse transformation method [26].

For practical implementation of non-linear rule (9)-(10) we used built-in functions $rnd(x)$ and $dnorm(p, \mu, \sigma)$ of MathCad computer calculation systems:

$$\Phi^{-1}(x) = dnorm(x, 0, 1); (u_i)_j = rnd(1).$$

Obviously, the rule (1) for calculate the modulated signal S at this method of forming the set $\varphi = \{\varphi_0, \varphi_1, \dots, \varphi_{M-1}\}$ should be written in such form

$$S = \sum_i \varphi_i. \quad (11)$$

Works [7, 8, 10, 11] indicate that the direct use of expanding sequences is to hide information data in the cover images leads to large bit errors in the extracted data. For this purpose, it was suggested to increase the power of expansion signals, i.e., we will write down formula (11) in form

$$S = \sum_i P \varphi_i, \quad (12)$$

where P - positive value, multiply increasing «power» of sequences φ_i . In our experiments we have realized hiding data in cover images using formulas (12), (9) and (10). The received results for different values of P are provided on the figure 1. The number of terms in (1) and (12) are determined by the number of information bits, hiding in one cover image container (or a fragment thereof). Figure 1 shows the different cases for $P = 2^i, i = 0, 1, \dots, 6$ and for different values of k . The following notation is used in the figures:

- 1) $k = 1$;
- 2) $k = 2$;
- 3) $k = 4$;
- 4) $k = 8$;
- 5) $k = 16$.

If the set of signals $\varphi = \{\varphi_0, \varphi_1, \dots, \varphi_{M-1}\}$ generate using a simplified scheme, such as

$$(\varphi_i)_j = \Phi^{-1}((u'_i)_j), (u'_i)_j = \begin{cases} (u_i)_j + 0.5, u_i < 0.5; \\ (u_i)_j - 0.5, u_i \geq 0.5, \end{cases} \quad (13)$$

then we can use the analogue of the formula (1) to hide it in the form of

$$S = \sum_i P b_i \varphi_i . \quad (14)$$

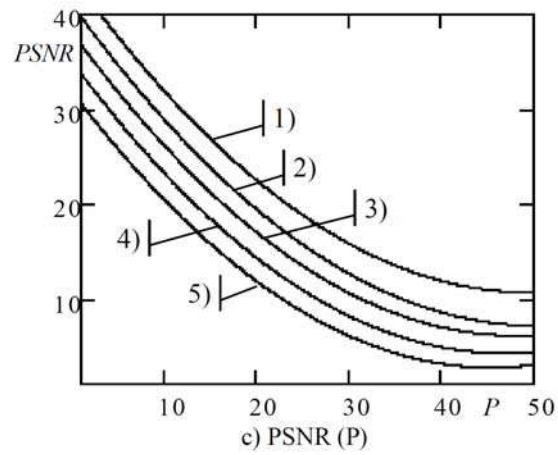
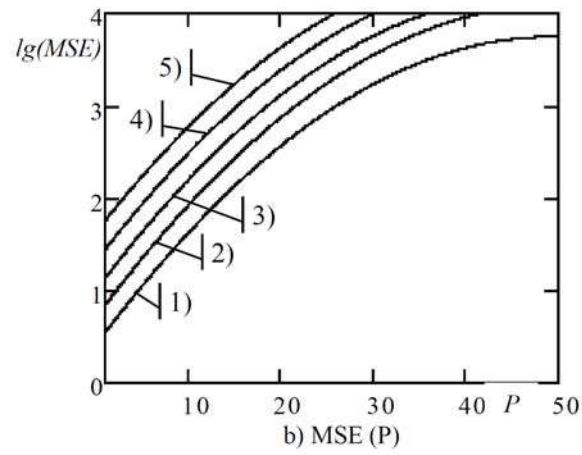
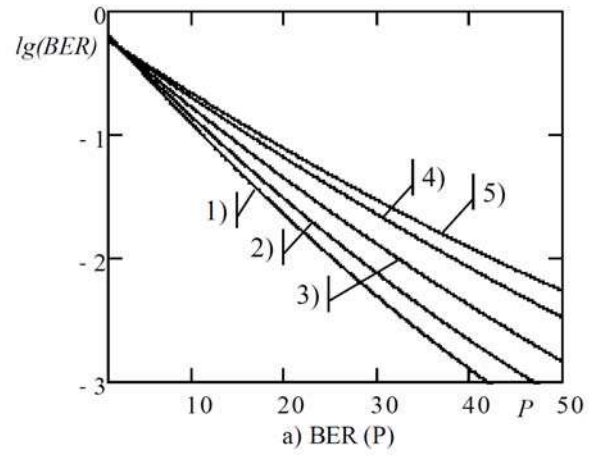


Fig. 1. Results of experimental studies in hiding data using expressions (12), (9) and (10)

We have also investigated this method of generating expansion signals; the results are shown in Figures 2.

Analyzing Figures 1 and 2, we see that both methods of forming expanding sequences (according to formulas (9), (10) and (13)) give practically equal results. In our experiments, rule (9), (10) was only slightly better in terms of PSNR (in the figures, because of the logarithmic scale, it is almost invisible). We should also note the high BER value. For example, even with "power" expansion signals, the BER value was in most cases in the range of 0.1 ... 0.01, which is on the threshold of the possible use of noise-sensitive coding.

4.2 Using random numbers uniformly distributed over an interval of (-1,1).

Another way to form a set that we investigated was to use random numbers uniformly distributed over a random values of interval of (-1,1). For this purpose we used the built-in function $rnd(x)$ of the MathCad computer system, i.e. the rule of forming sequences had the form:

$$(\varphi_i)_j = rnd(2) - 1. \quad (15)$$

The results of our studies on the effectiveness of hiding information of rule (13) for different ratios of k and P are shown in Figures 3.

The results shown in Figure 3 are practically comparable with obtained results for nonlinear modulation (9), (10), and for the simplified version by the formula (13). We observed a slight increase in BER, but PSNR also increased at the same time. In general, we can argue that the revealed differences are small and that these methods of forming expanding sequences are almost equal.

4.3 Use of Walsh orthogonal sequences

In our research we also used Walsh discrete orthogonal sequences. Such signals are generated from rows in the Hadamard matrix H_{2^i} formed by a recurrence rule:

$$H_{2^i} = \begin{bmatrix} H_{2^{i-1}} & H_{2^{i-1}} \\ H_{2^{i-1}} & -H_{2^{i-1}} \end{bmatrix}, \quad H_1 = [1]. \quad (16)$$

An iterative repetition of the rule (16) allows any Hadamard matrix H_{2^i} order 2^i , $i = 1, 2, \dots$ matrix to be formed. The rows (or columns) of formed matrices are mutually orthogonal, i.e. their scalar product is zero.

In our studies we used rule (16) and H_{2^i} matrix rows were interpreted as elements of the set $\varphi = \{\varphi_0, \varphi_1, \dots, \varphi_{M-1}\}$, the obtained results of hiding information by formula (14) are shown in Figures 4.

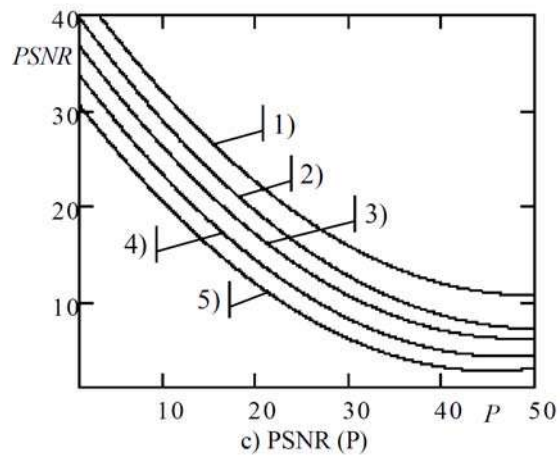
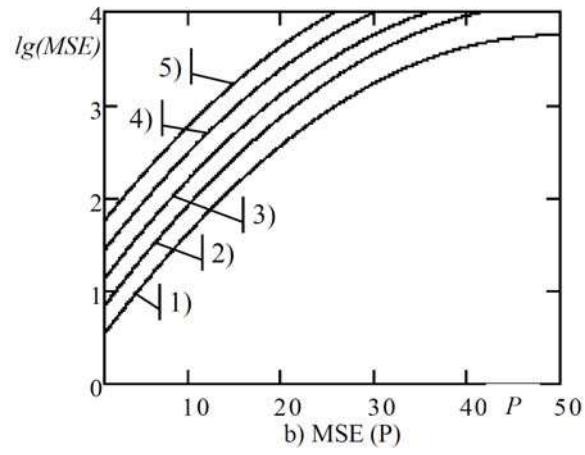
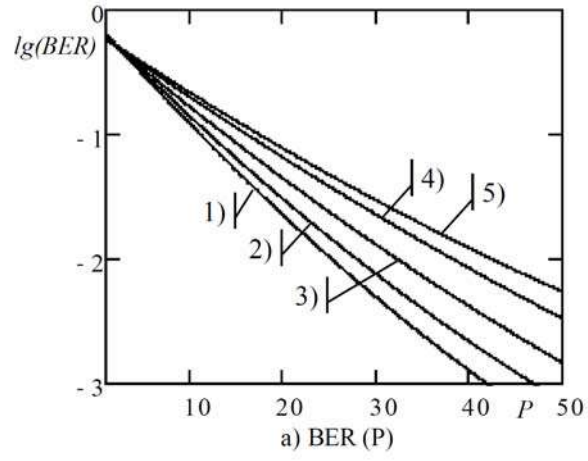


Fig. 2. Results of experimental studies in hiding data using expressions (13) and (14).

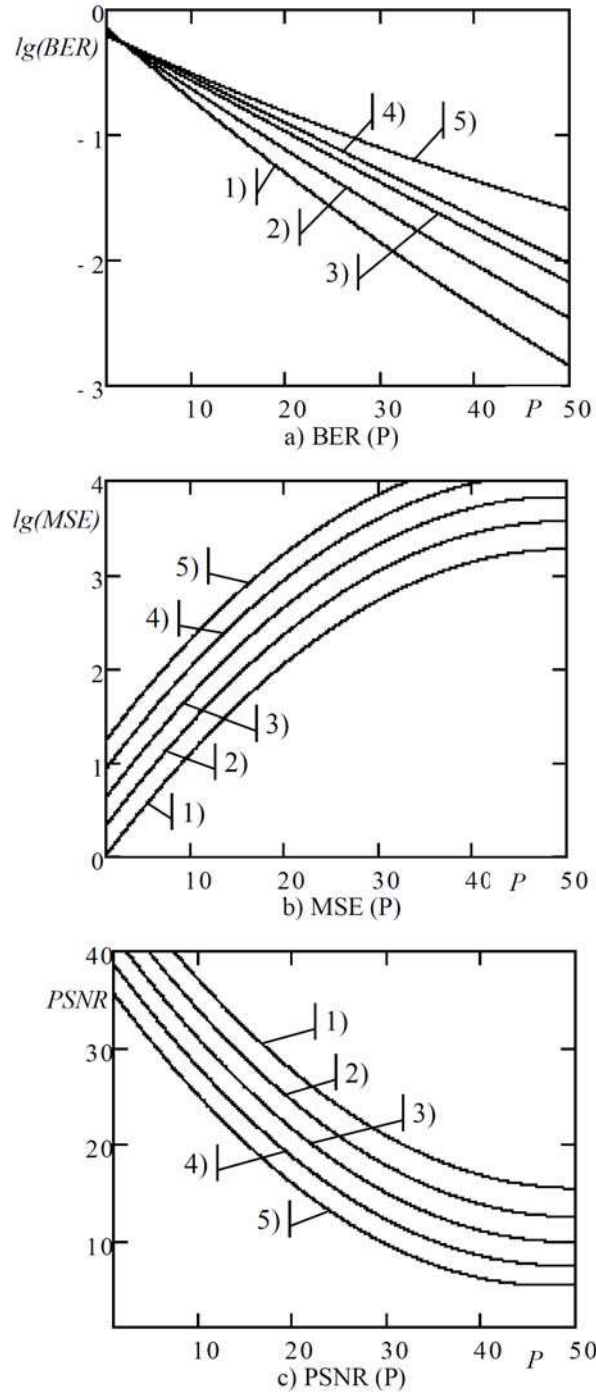


Fig. 3. Results of experimental studies in hiding data using expressions (15) and (14).

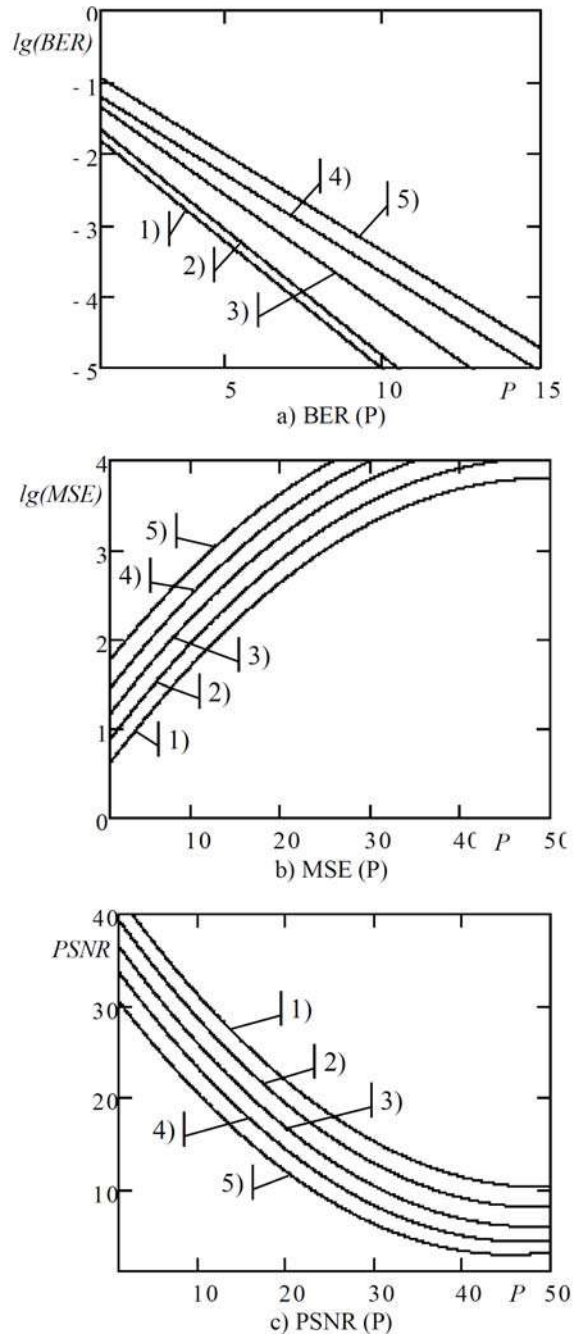


Fig. 4. Results of experimental studies in hiding data using Walsh orthogonal sequences and expressions (14)

Figure 4 clearly shows the advantage of using Walsh sequences. In fact, in our studies were obtained the lowest BER values. Even at small values $P \approx 5$, the BER was in most cases less than 0.01 and this is the best result of all considered variants of forming the expansion sequences. PSNR values when using Walsh sequences are in most cases comparable to the variants considered earlier. However, for a fixed PSNR value, using Walsh sequence leads to significantly lower BER values.

5 Discussion of the results and brief conclusions

The obtained empirical dependencies show that the use of direct spectrum extension technology can indeed be an interesting solution to the problem of hiding information messages in cover images. By interpreting an image as noise in a communication channel, it is possible to organize a hidden data channel, and image distortions may not be large. At the same time, the basic assumption about the non-correlation of expansion sequences with the cover (or its separate part) may be incorrect. In this case a high level of errors will be obtained when restoring the information data. Consequently, an important element of such steganosystem is the correct choice of expansion sequences.

In our work we have analyzed several variants of building extension sequences for hiding data in cover images. In particular, we have considered one of the first known algorithms with non-linear modulation by rules (9), (10). For this method a US patent was obtained [11], and we investigated the effectiveness of such hiding by BER, MSE and PSNR. The obtained data partially coincide with the known results from [7, 8, 10, 11], which may indirectly confirm the adequacy of our results. At the same time, we investigated other ways to form expansion sequences for hiding data in cover images. For example, we have shown that the application of the simplified rule (13) and even the use of sequences with equidistant values on the interval $(-1,1)$ does not lead to significant deterioration of the results. For example, the BER and PSNR values do not differ significantly. Finally, we studied the use of Walsh expansion sequences. As it turned out, this variant is the most successful, because a much smaller percentage of errors is achieved with comparable PSNRs. Indeed, the BER value, as it follows from our results, is much lower than for other extension sequences.

The results can be used to improve techniques for hiding information in digital images [5-12], as well as in other computer science applications [27-32]. In particular, our results suggest that the use of orthogonal discrete signals is most preferred. In our opinion, an interesting direction for further research is the use of adaptively formed discrete sequences. For example, if the rule of forming expansion signals takes into account the statistical properties of the cover, then it will be possible to significantly reduce the BER, or get an error-free transmission. Another useful result can be an increase in PSNR when the BER value is fixed (for example, ahead of the set value). In addition, we plan to use other ways to form expansion sequences in future studies. For example, discrete sequences with multilevel correlation functions were proposed in [33-35]. The use of these signals, in our opinion, will be effective in steganographic techniques with direct spectrum spreading technology.

References

1. Digital Watermarking and Steganography. Elsevier (2008). doi:10.1016/b978-0-12-372585-1.x5001-3
2. Shin, F.Y.: Digital Watermarking and Steganography. CRC Press (2017). doi:10.1201/9781315219783
3. Johnson, N.F., Jajodia, S.: Exploring steganography: Seeing the unseen. *Computer*. 31, 26–34 (1998). doi: 10.1109/MC.1998.4655281
4. Manoj, I.V.S.: Cryptography and Steganography. *International Journal of Computer Applications*. 1, 63–68 (2010). doi:10.5120/257-414
5. Tirkel, A.Z., Osborne, C.F., Van Schyndel, R.G.: Image watermarking-a spread spectrum application. In: *Proceedings of ISSSTA'95 International Symposium on Spread Spectrum Techniques and Applications*. IEEE (0). doi:10.1109/isssta.1996.563231
6. Smith, J.R., Comiskey, B.O.: Modulation and information hiding in images. In: *Information Hiding*. pp. 207–226. Springer Berlin Heidelberg (1996). doi:10.1007/3-540-61996-8_42
7. Marvel, L.M., Boncelet, C.G., Jr., R., Charles T.: Methodology of Spread-Spectrum Image Steganography. Defense Technical Information Center (1998). doi:10.21236/ada349102
8. Marvel, L.M., Boncelet, C.G., Retter, C.T.: Spread spectrum image steganography. *IEEE Transactions on Image Processing*. 8, 1075–1083 (1999). doi: 10.1109/83.777088
9. Kutter, M.: Performance Improvement of Spread Spectrum Based Image Watermarking Schemes through M-ary Modulation. In: *Information Hiding*. pp. 237–252. Springer Berlin Heidelberg (2000). doi:10.1007/10719724_17
10. Brundick, F.S., Marvel, L.M.: Implementation of Spread Spectrum Image Steganography. Defense Technical Information Center (2001). doi:10.21236/ada392155
11. Patent No.: US 6,557,103 B1, Int.Cl. G06F 11/30. Spread Spectrum Image Steganography. (2003)
12. Fan Zhang, Bin Xu, Xinhong Zhang: Digital Image Watermarking algorithm Based on CDMA Spread Spectrum. In: 2006 12th International Multi-Media Modelling Conference. IEEE (0). doi: 10.1109/MMMC.2006.1651359
13. Nguyen, T.T., Taubman, D.: Optimal linear detector for spread spectrum based multidimensional signal watermarking. In: 2009 16th IEEE International Conference on Image Processing (ICIP). IEEE (2009). doi: 10.1109/ICIP.2009.5414121
14. Nezhadarya, E., Wang, Z.J., Ward, R.K.: Image quality monitoring using spread spectrum watermarking. In: 2009 16th IEEE International Conference on Image Processing (ICIP). IEEE (2009). doi: 10.1109/ICIP.2009.5413955
15. Ghosh, S., Ray, P., Maity, S.P., Rahaman, H.: Spread Spectrum Image Watermarking with Digital Design. In: 2009 IEEE International Advance Computing Conference. IEEE (2009). doi: 10.1109/IADCC.2009.4809129
16. Altun, H.O., Orsdemir, A., Sharma, G., Bocko, M.F.: Optimal Spread Spectrum Watermark Embedding via a Multistep Feasibility Formulation. *IEEE Transactions on Image Processing*. 18, 371–387 (2009). doi: 10.1109/TIP.2008.2008222
17. Samcovic, A., Milovanovic, M.: Robust digital image watermarking based on wavelet transform and spread spectrum techniques. In: 2015 23rd Telecommunications Forum Tel-for (TELFOR). IEEE (2015). doi: 10.1109/TELFOR.2015.7377589
18. Ipatov, V.P.: Spread Spectrum and CDMA. John Wiley & Sons, Ltd (2005). doi:10.1002/0470091800
19. Introduction to CDMA Wireless Communications. Elsevier (2007). doi:10.1016/b978-0-7506-5252-0.x5001-7

20. Gerakoulis, D., Geraniotis, E.: CDMA: Access and Switching. <http://dx.doi.org/10.1002/0470841699>, (2001). doi:10.1002/0470841699
21. Hara, S., Prasad, R.: DS-CDMA, MC-CDMA and MT-CDMA for mobile multi-media communications. In: Proceedings of Vehicular Technology Conference - VTC. IEEE (0). doi: 10.1109/VETEC.1996.501483
22. Agaian, S.S., Sarukhanyan, H.G., Egiazarian, K.O., Astola, J.: Hadamard Transforms. SPIE (2011). doi:10.1117/3.890094
23. Probability Theory of Bit Error Rate. In: Optical Bit Error Rate. IEEE (2009). doi:10.1109/9780470545430.ch7
24. Korhonen, J., You, J.: Peak signal-to-noise ratio revisited: Is simple beautiful? In: 2012 Fourth International Workshop on Quality of Multimedia Experience. IEEE (2012). doi:10.1109/QoMEX.2012.6263880
25. Data Compression. Springer London (2007). doi:10.1007/978-1-84628-603-2
26. Devroye, L.: Non-Uniform Random Variate Generation. Springer New York (1986). doi:10.1007/978-1-4613-8643-8
27. Attari, A.A., Shirazi, A.A.B.: Robust and Transparent Audio Watermarking based on Spread Spectrum in Wavelet Domain. In: 2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT). IEEE (2019). doi: 10.1109/jeeit.2019.8717415
28. Runovski, K., Schmeisser, H.: On the convergence of fourier means and interpolation means. Journal of Computational Analysis and Applications. 6(3), 211-227 (2004)
29. Hua, G.: Over-Complete-Dictionary-Based Improved Spread Spectrum Watermarking Security. IEEE Signal Processing Letters. 1-1 (2020). doi: 10.1109/lsp.2020.2986154
30. Chorney, R.K., Daduna V.M., H., Knopov, P.S.: Controlled Markov Fields with Finite State Space on Graphs. Stochastic Models. 21, 847-874 (2005). doi:10.1080/15326340500294520
31. Huang, Y., Niu, B., Guan, H., Zhang, S.: Enhancing Image Watermarking With Adaptive Embedding Parameter and PSNR Guarantee. IEEE Transactions on Multimedia. 21, 2447-2460 (2019). doi:10.1109/tmm.2019.2907475
32. Bondarenko, S., Liliya, B., Oksana, K., & Inna, G.: Modelling instruments in risk management. International Journal of Civil Engineering and Technology. 10(1), 1561-1568 (2019)
33. Stasev, Y., Kuznetsov, A., Karpenko, O., Sai, V.: Discrete signals with multi-level correlation function. Telecommunications and Radio Engineering. 71, 91-98 (2012). doi: 10.1615/TelecomRadEng.v71.i1.100
34. Kuznetsov, A., Smirnov, O., Kovalchuk, D., Averchev, A., Pastukhov, M., Kuznetsova, K.: Formation of Pseudorandom Sequences with Special Correlation Properties. In: 2019 3rd International Conference on Advanced Information and Communications Technologies (AICT). IEEE (2019). doi: 10.1109/AIACT.2019.8847861
35. Kuznetsov, A., Smirnov, O., Reshetniak, O., Ivko, T., Kuznetsova, T., Katkova, T.: Generators of Pseudorandom Sequence with Multilevel Function of Correlation. In: 2019 IEEE International Scientific-Practical Conference Problems of Infocommunications, Science and Technology (PIC S&T). IEEE (2019). doi: 10.1109/PICST47496.2019.9061530

Cryptosystem Based on a Key Function of a Real Variable

Viktor Avramenko^{1[0000-0002-6317-6711]} and Volodymyr Demianenko^{2[0000-0002-1512-970X]}

^{1,2}Sumy State University, Rimsky-Korsakov st., 2, Sumy, 40007, Ukraine
¹avramenko1938@gmail.com, ²vldemyan@gmail.com

Abstract. Using the function of a real variable in cryptosystems as a key allows you to increase its cryptographic strength, because it is more difficult to pick up such key. Therefore, the development of such systems is relevant. A cryptosystem with a symmetric key is offered. This key is some function of a real variable that satisfies some restrictions. It can be either continuous or discrete.

The transmitting and receiving parties select the key-function, the first transmitted character or the first transmitted value for the analog message, the function area of the key function, and the step of changing the function argument. A Disproportion over first-order derivative is used to encrypt an analog message.

The Cauchy problem is solving for decrypting this message. Discrete messages are encrypted using the first-order disproportionality integral function. Decryption is performed by the inverse transformation of the formula for integral disproportion.

Algorithms for encrypting and decrypting messages are presented. The ability to encrypt and decrypt text information, 2D graphic images, as well as analog messages are shown. The examples show the complexity to pick up the key function and the cryptographic strength of the proposed cryptosystem.

A cryptosystem, in which the function of a real variable is used as a key and as well as disproportion functions are used, is suitable for encryption of both discrete and continuous messages. To “crack” such a system, it is required to pick up the form of the key function and to find the values of its parameters with very high accuracy. That is, the system has high cryptographic strength.

Keywords: cryptosystems, disproportion functions, real variable functions, key-function, encryption, decryption, text messages, 2D images, analog messages.

1 Introduction

It is difficult to imagine the modern world without systems of secrecy of the transmitted information in both the military and business fields. Therefore, it is not surprising that they are constantly being improved. In addition, new ways of hiding information are emerging. The most widely used encryption algorithms are divided into symmet-

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

ric and asymmetric. In symmetric systems, encryption and decryption are performed using the same key. Such systems include AES [1], GOST 28147-89 [2, 3]. Hacking them requires enumeration of possible keys. The complexity of exhaustive search is estimated as $O(2^k)$, where k is the key length in bits.

In asymmetric systems (public key systems), RSA and El-Gamal algorithms are most widely used [4, 5]. The RSA algorithm is based on the computational complexity of the integer factorization problem. The reliability of the El-Gamal algorithm is based on the difficulty of computing the discrete logarithm. Of particular interest is the El-Gamal algorithm over the group of points of an elliptic curve [6].

For asymmetric algorithms, there are cryptanalysis methods that work faster than full search. Therefore, asymmetric algorithms have to use longer keys than keys in symmetric systems.

The disadvantages of both symmetric and asymmetric systems include the fact that they are based on a set of integers. This allows you to develop various methods of hacking such cryptosystems up to the implementation of a simple selection of keys. To increase cryptographic strength, you have to gradually increase the length of the keys.

However, due to the increasing capabilities of computer technology, the way to increase the length of keys is unpromising. Currently, great efforts are being made to create quantum computers [7]. It is expected that their use will significantly affect the cryptographic strength of existing cryptosystems [8]. The well-known Grover quantum algorithm [9] for restoring the key of a symmetric encryption algorithm from the message text and cipher text. Full search on a regular computer has complexity $O(2^k)$, where k is the key length. For a quantum computer, this complexity is 2 times less and amounts to $O(2^{k/2})$ [9]. That is, the efficiency of the key length is reduced by 2 times.

Quantum algorithms also pose a threat to systems with asymmetric encryption. The cryptographic strength of the RSA encryption system is based on the super-polynomial computational complexity of factoring natural numbers. However, there is a quantum algorithm, whose complexity is polynomial $O(n^3)$ [10].

There is also a Shor quantum algorithm for computing the discrete logarithm. In [11], the Shor algorithm for a group of points of an elliptic curve over a field $GF(p)$ with complexity $O(n^3)$ is presented.

At the same time, it is known [12] that the set of real numbers has a greater power in comparison with the set of natural numbers. Therefore, the development of cryptosystems using real numbers is of practical interest. It can be expected that they may be more persistent in the event of an attempt to select a key by enumeration.

This paper presents one of the many possible options for such the cryptosystem.

2 Problem statement

It is necessary to develop algorithms for encrypting and decrypting analog and discrete messages using a function of a real variable as a key.

3 Literature review

Mostly modern cryptosystems use a block cipher. It operates with groups of bits of a fixed length - blocks. Their size can be from 64 to 256 bits.

The widely used block symmetric ciphers (BSS) are based on the use of several symmetric cryptographic transformations (elementary ciphers).

When constructing them, three main approaches are used:

- based on Feistel chains;

- IDEA-like ciphers;

- SPN (Substitution Permutation Network) structure. The AES system belongs to SPN based BSN [1]. In it, cryptographic conversions are performed in the simple substitution mode over 128-bit blocks (16 bytes), which has a block length of 128 bits. The key length has several options: 128, 192, 256 bits.

The number of cycles of conversion depends on the length of the key. It is, respectively, 10, 12 or 14 cycles.

The encryption algorithm GOST 28147-89 [2, 3] also belongs to the category of block ciphers, where two parts of the selected block of information are of equal size.

It is a classical symmetric encryption algorithm based on the Feistel network and is characterized by high cryptographic strength. However, this algorithm also has disadvantages:

- 1) In comparison with the byte-oriented algorithm, AES on 8-bit platforms GOST loses in speed by 4 times.

- 2) In the text of the standard GOST 28147-89, there are no clear criteria for the selection of replacement nodes. Quite often, fears are expressed that there are weak replacement nodes.

In 1978, three authors: Ronald Rivest, Adi Shamir, Leonard Adleman proposed the RSA algorithm [4, 5]. This algorithm was the first full-fledged public key algorithm. Encryption is carried out without the transfer of secret keys.

In addition to classical cryptosystems using integers as keys, systems based on functions of real variables were proposed [13]. Symbols from the ASCII code table are encrypted with a sum of 10 key functions with coefficients before them zero or one. The amplitudes of these functions during encryption of each new character are taken randomly. At the receiving end, using the disproportion functions [14-16], fragments of key functions that are present in the received encrypted signal are recognized. It allows to decrypt the character transmitted at the current time.

In [17, 18], a variant was proposed when, three key functions of a real variable are used to encrypt a binary codes. Next symbols are coded: "1", "0", "space", "the transition to a new row". Any other character is recognized as the transition to a new row. For unauthorized access to the intercepted message, you need to select the type and parameters of the key functions.

Both cryptosystems in the process of computer simulation showed high cryptographic strength when trying to select the parameters of key functions even if their type is known.

However, using the sum of key functions and the need to recognize them at the receiving end complicates the algorithm. In addition, the encrypted message is much

longer than the original. Therefore, the task is to develop a cryptosystem using only one key function of a real variable.

4 Mathematical statement of the problem

A message that is encrypted is a sequence of numerical character codes (or numerical values of the components of the brightness of the pixels in the case of a graphic image), which is described by a discrete function y_i , $i = 0, 1, \dots, N-1$, where N is the number of characters in the message. It is replaced by a sequence of ciphers that are calculated using the key function, which can be either discrete f_i , $i = 0, 1, \dots, M$ ($M > N$), or continuous $f(x)$. Here $x = i \cdot h$, where i is the sequence number of the character in the message, and h is the step with which the argument x changes. This step should be the same for both sides of the message transfer. Disproportion functions are used for encryption.

5 Disproportion functions

Several types of the disproportion function are known: the disproportion over n -th order derivative, the disproportion over n -th order value, relative and sequential disproportions. Virtually all of them are characteristics of numerical functions. Below is a summary of those that are used in this work.

The disproportion over n -th order derivative of the function $y(x)$ with respect to x is described by the expression:

$$@d_x^{(n)} y = \frac{y}{x^n} - \frac{1}{n!} \cdot \frac{d^n y}{dx^n} \quad (1)$$

Here, the @ symbol is chosen to indicate the operation of calculating disproportion. The symbol “ d ” means “derivative”. The order is indicated in parentheses. The left side of (1) reads “at $d n y$ with respect to x ”.

If for any value of x , the function $y(x)$ has the form $y = kx^n$, then disproportion (1) is equal to zero regardless of the value of the coefficient k .

The disproportion over 1-st order derivative ($n = 1$) has the form:

$$@d_x^{(1)} y = \frac{y}{x} - \frac{dy}{dx} \quad (2)$$

In case of a parametric specification of functions, when $x = \varphi(t)$, $y = \psi(t)$, where t is a parameter, disproportion (2) takes the form:

$$@d_{\varphi(t)}^{(1)}\psi(t) = \frac{\psi(t)}{\varphi(t)} - \frac{d\psi/dt}{d\varphi/dt} \quad (3)$$

For $\psi(t) = k\varphi(t)$ disproportion (3) is equal to zero in the entire area of existence $x = \varphi(t)$, regardless of the value of k .

If $y(x)$ is a sum of known functions taken with unknown coefficients, then the disproportion functions allow us to calculate the values of these coefficients from the data obtained for the current value of the argument. This opportunity was used both for creating cryptosystems [13, 17, 18], and in solving a number of more general problems [15, 16].

For the case when the first derivative does not exist or is equal to zero on any interval, it is proposed to use the first-order integral disproportion [19, 20]. This disproportion of the function $y(x)$ with respect to $f(x)$ has the form:

$$@I_{f(x)}^{(1)}y(x) = \frac{\int_{x-h}^x y(x)dx}{\int_{x-h}^x f(x)dx} - \frac{y(x)}{f(x)}, \quad (4)$$

where h is the preset time interval. In the discrete representation of signals, this is a time quantization step.

In this case, $y(x)$ and $f(x)$ are represented by one-dimensional arrays. If the approximate values of the integrals in (4) are calculated using the trapezoidal formula, then for the same step h for $y(x)$ and $f(x)$, disproportion (4) takes the form:

$$@I_{f_i}^{(1)}y_i = \frac{y_{i-1} + y_i}{f_{i-1} + f_i} - \frac{y_i}{f_i} \quad (5)$$

6 Encryption and decryption of text and image messages

Text messages are a sequence of character codes, for example, from an ASCII table. That is, it is a sequence of integers. When transmitting color graphic images, the brightness components of pixel-integer numbers from 0 to 255 are transmitted. As a key function, you can take any function of a real variable. It can be either a continuous or a discrete function. However, due to the fact that the message is represented by a discrete function that does not have a first derivative, in any case, integral disproportion should be used for encryption (5).

If the key function is continuous, you need to calculate an array of its values, changing the argument from the initial $x_{\min}$ to the final $x_{\max}$ values in increments of h . The parameters $x_{\min}$, $x_{\max}$, h must be the same for the transmitting and receiving sides. When encrypting the characters from the ASCII table or pixel luminance components,

their numerical representations differ by one. In these cases, the step h of changing the argument must be equal to one.

In addition, the transmission and, accordingly, the reception must begin with a certain symbol known to the transmitting and receiving sides.

Therefore, the encryption algorithm is as follows:

1. The start character that is known to both sides must be entered.
2. Enter an array of values of key function.
3. Read from the file or enter from the keyboard a sequence of message characters and convert them to the numbers. In this case, for text symbols the ASCII table can be used. As a result, the array $y_i, i = 0, 1, \dots, N-1$ will be get.
4. Using the code of the given initial character y_0 , calculate the disproportion $I_i = @I_{f_i}^{(1)} y_i, i = 1, 2, \dots, N-1$ for each next character in accordance with (5) and transmit them over the communication channel.

7 Decryption algorithm for text and image messages

1. Enter the start character y_0 that is known for the transmitting side.
2. If the key function is discrete, enter the array of its values. If the key function is continuous, calculate the array of its values in accordance with the instructions above.
3. Read from the file the accepted disproportion values for each of the symbols $I_i, i = 1, 2, \dots, N-1$.
4. Use the known initial symbol y_0 , and the accepted disproportions $I_i, i = 1, 2, \dots, N-1$, for calculating the estimated codes of the recovered message according to the formula (6).

$$y_i = \frac{(y_i - I_i \cdot (f_{i-1} + f_i)) \cdot f_i}{f_{i-1}} \quad (6)$$

Round off estimated codes to the nearest integers to get recovered codes.

5. Using the recovered codes found, reproduce the message in symbolic form (when transmitting a graphic image, reproduce the corresponding pixel attributes).

8 Examples of encryption and decryption of discrete messages

Example 1

An example of encryption and decryption of characters from the ASCII code table is provided.

The key function has the form:

$$f(x) = ae^{\beta x} + \sin(\beta x) + \ln(ax + \beta) \quad (7)$$

$x = i \cdot h$ is an argument;

i is the sequence number of the character in the encrypted message;

$h = b + c$ is the step of changing the argument.

Here $a = 0.1$, $\beta = 0.01$, $b = 0.65$, $c = 0.35$ are constants.

Since in the example the characters are encrypted from the ASCII table, in accordance with the above algorithm, the sum of the constants b and c is equal to one.

The sequence of numerical codes y_i , $i = 1, 2, \dots, N-1$, corresponding to the transmitted characters, is encrypted. When transmitting messages, let's the first character is "G". In the ASCII table, its code is 71 (that is, $y_0 = 71$).

Given the discrete nature of the message, integral disproportion (5) is used to encrypt it.

Decryption is carried out using (6) and according to the above algorithm.

Table 1 shows the transmitted characters, their ciphers, and the corresponding decrypted characters.

Table 1. Codes of transmitted characters and the results of their decryption.

Character number	Source character	Character code	Decrypted character	Character number	Source character	Character code	Decrypted character
1		-0,33745		19	@	-14,9626	@
2	C	18,56566	C	20		17,18209	
3	o	35,05402	o	21	#	-0,50861	#
4	n	23,38864	n	22		2,084013	
5	t	45,28726	t	23	\$	-1,0013	\$
6	r	75,66021	r	24		2,205811	
7	o	215,5424	o	25	%	-1,38787	%
8	l	3638,486	l	26		2,342513	
9		1392,007		27	^	-21,4272	^
10	e	-49,5932	e	28		22,22681	
11	x	36,34984	x	29	&	-1,64792	&
12	a	60,50844	a	30		2,295671	
13	m	7,687781	m	31	*	-2,84342	*
14	p	11,55305	p	32		3,393497	
15	l	13,07326	l	33	(	-2,13477	(
16	e	11,94166	e	34		2,607738	
17	!	44,92703	!	35	)	-2,34347	)
18		2,001215					

It can be seen from it that the cipher of each character is a real number with both an integer and a fractional part. Encrypted and decrypted characters match.

In order to "crack" a message, you need to select the type of the key function and the values of its parameters.

The following is an example that illustrates the cryptographic strength of the system to obtain a key, even if somehow it was possible to find out the type of key-function. Suppose that the above sequence of characters is encrypted using function

(7), and decrypted using the same type of function, but the constant a instead of the value 0.1 during decryption is chosen incorrectly and is 0.1005.

Table 2 shows the results, from which it is seen that even such a slight deviation of the parameter of the key function does not allow to decrypt transmitted characters correctly.

Table 2. Codes of transmitted characters and the results of their decryption with an incorrectly selected key function.

Char- acter num- ber	Source charac- ter	Character code	De- crypt- ed char- acter	Charac- ter number	Source charac- ter	Character code	De- crypt- ed charac- ter
1		-0,33745		19	@	-14,9626	D
2	C	18,56566	B	20		17,18209	
3	o	35,05402	n	21	#	-0,50861	M
4	n	23,38864	l	22		2,084013	Γ
5	t	45,28726	q	23	\$	-1,0013	-
6	r	75,66021	n	24		2,205811	Γ
7	o	215,5424	g	25	%	-1,38787	T
8	l	3638,486	5	26		2,342513	Π
9		1392,007	ík	27	^	-21,4272	,
10	e	-49,5932	X	28		22,22681	®
11	x	36,34984	”	29	&	-1,64792	
12	a	60,50844	f	30		2,295671	e
13	m	7,687781	(	31	*	-2,84342	s
14	p	11,55305	r	32		3,393497	ь
15	l	13,07326		33	(	-2,13477	A
16	e	11,94166	L	34		2,607738	q
17	!	44,92703	+	35	)	-2,34347	«
18		2,001215	2				

A positive feature of the proposed encryption system is that the cipher of one and the same symbol is not repeated and depends on its number in the message.

Table 3 shows the repeated characters, their ciphers, and decryption results.

Table 3. Results of decryption of the sequence of identical characters.

Char- acter num- ber	Sou- rce char- ac- ter	Character code	De- crypt- ed char- acter	Charac- ter number	Source charac- ter	Character code	De- crypt- ed char- acter
1	a	42,89534	a	23	%	-1,07004	%
2	a	55,98126	a	24	%	0,230793	%
3	a	223,1301	a	25	%	0,211779	%
4	a	-2722	a	26	%	0,195132	%
5	a	193,6388	a	27	%	0,180462	%

6	a	47,19846	a	28	%	0,167453	%
7	a	21,12258	a	29	%	0,155856	%
8	a	12,02764	a	30	%	0,145461	%
9	a	7,807816	a	31	%	0,136099	%
10	a	5,504747	a	32	%	0,127629	%
11		31,8935		33		1,189454	
12	A	-12,0762	A	34	9	-5,17547	9
13	A	1,71819	A	35	9	0,164057	9
14	A	1,413497	A	36	9	0,154966	9
15	A	1,186178	A	37	9	0,146586	9
16	A	1,011796	A	38	9	0,138837	9
17	A	0,874885	A	39	9	0,131652	9
18	A	0,765272	A	40	9	0,124967	9
19	A	0,67603	A	41	9	0,118735	9
20	A	0,602313	A	42	9	0,112905	9
21	A	0,540641	A	43	9	0,10744	9
22		9,210041		44		4,816542	

From Table 3 it can be seen that the ciphers of the same characters differ from each other, but at the same time, their decryption is error-free.

This once again shows that for hacking it is necessary to find the type of the key function, and then select the values of its parameters with high accuracy.

Example 2

The encryption of a graphic message is considered - a 2D image using the key function of a real variable

$$f(q) = ae^{\sin((b+c)q+a)}, \quad (8)$$

where q is the sequence number of the pixel;

$a = 100$, $b = 0.65$, $c = 0.35$.

The image of a seagull on the seashore is encrypted. According to the algorithm, transmission begins with a pixel known to the receiving side. Pixel brightness is also represented by integers with a step equal to one. Therefore, according to the algorithm in this case, it is required that the sum of a and b is equal to one.

Table 4 shows the disproportions of the brightness components for the first 23 pixels.

Table 4. Components brightness pixels' codes.

pixel number	disproportion of component A	component red disproportion	component green disproportion	component blue disproportion
1	-0,17260194	-0,052654658	-0,079113617	-0,109100438
2	0,249815554	0,122038436	0,147606293	0,183288106
3	-0,25138517	-0,066806224	-0,080056209	-0,073489858
4	-1,54785813	-0,671606862	-0,746100819	-0,895078346
5	-2,11017572	-1,141977451	-1,198335871	-1,339798239
6	0,638962347	0,336698458	0,322878349	0,344784243

7	0,904695026	0,493147485	0,536886675	0,603130017
8	0,35818895	0,185034305	0,203422857	0,224748673
9	-0,0619523	-0,079999226	-0,074320205	-0,069612985
10	-1,08142099	-0,6158245	-0,699207099	-0,762820098
11	-2,35243212	-1,569025352	-1,733111253	-1,953901199
12	-0,07899416	-0,025833335	-0,031099612	-0,000303777
13	1,059101779	0,630917816	0,701524602	0,778998655
14	0,482163337	0,302525074	0,337339747	0,360361396
15	0,079311391	0,042700382	0,052726876	0,071145629
16	-0,68481115	-0,484134182	-0,526938708	-0,575032293
17	-2,19903004	-1,601314543	-1,722045604	-1,879955093
18	-0,9877602	-0,663217085	-0,68687686	-0,706872244
19	1,094772111	0,607081863	0,704520639	0,846261488
20	0,632160636	0,56848388	0,574362813	0,580312445
21	0,193576492	0,084571962	0,09345652	0,077852313
22	-0,3793663	-0,320611934	-0,339514003	-0,33505087
23	-1,80013937	-1,01073232	-1,126174205	-1,185141129

It can be seen from it that the ciphers are real numbers with both an integer and a fractional part, the nature of the changes of which is difficult to predict.

Figure 1 shows the original (left) and decrypted (in the middle) images according to the disproportions received at the receiving end, known to the first pixel and the key function.



Fig. 1 Original image (left), decrypted (in the middle), decrypted with an incorrectly selected key-function parameter (right).

Figure 1 on the right shows the result when instead of $a = 100$, when decrypting the image, the value $a = 99.9999$ was mistakenly selected. Despite a slight difference from the value of this parameter during encryption, it was not possible to “crack” the image. The same result if $b + c = 0.999999$ is selected instead of unity.

This indicates the cryptographic strength of the encryption system and how difficult it is to get a key function by enumeration.

9 Examples of encryption and decryption of the analog message

The case is considered when the message $y(x)$ and the key function $f(x)$ are continuous, smooth, having a first derivative.

The transmitting and receiving parties agree on the key function, the initial value of the argument x_0 , the step of its change h and the initial value of the transmitted message $y(x_0)$.

In practice, most often, there is a need to encrypt signals from moving objects.

In this case, the signal is represented by a function of time $y(t)$. To encrypt it using the key function $f(t)$, disproportion (3) should be calculated, which in this case has the form:

$$z(t) = @ d_{f(t)}^{(1)} y(t) = \frac{y(t)}{f(t)} - \frac{dy/dt}{df/dt} \quad (8)$$

In this case the time change step h is selected from the representation condition $y(t)$ with the necessary accuracy.

The decrypted values are calculated from equation (9) obtained from (8):

$$y'(t) = f'(t) \left(\frac{y(t)}{f(t)} - z(t) \right) \quad (9)$$

The decrypted message is found from (9) by solving the Cauchy problem using the known $y(t_0)$, h and $z(t)$ values obtained via the communication channel (8).

Example 3. Analog message encryption

For simulation, suppose that an analog message is described by the following expression:

$$y(t) = k \cdot \exp((a - t) \cos(bt)) \sin^2(bx) \quad (10)$$

The key function has the form:

$$f(t) = a \cdot \exp(bt) + \sin(bt) + \ln(at + b), \quad (11)$$

where $a = 1$, $b = 10$, $k = 100$ are constants.

The time t varies from 0.1 to 4 in increments of $h = 0.0025$.

The derivatives were calculated numerically using the values of three points.

The results of the cryptosystem simulation are shown in table 5.

Table 5. Transmitted and decrypted values of a numerical function.

account number	transmitted value	decrypted value
1	0,16938	0,173057
2	0,674778	0,68059
3	1,51054	1,51856
4	2,66904	2,67926
5	4,14073	4,15308
6	5,91425	5,93181
7	7,97656	7,99589
8	10,313	10,3338
9	12,9076	12,9296
10	15,743	15,7656
11	18,8006	18,8234
12	22,061	22,0834
13	25,5038	25,5252
14	29,1082	29,1278
15	32,8525	32,8735
16	36,715	36,7328
17	40,6737	40,6874
18	44,7066	44,7188
19	48,7916	48,7979
20	52,9072	52,9066
21	57,0322	57,0268
22	61,1458	61,1315
23	65,228	65,2037
24	69,2595	69,2267
25	73,2218	73,1769
26	77,0975	77,0394
27	80,8701	80,7977
28	84,5241	84,4363
29	88,0453	87,9424
30	91,4207	91,3003

Obtained results indicate that even with a rough calculation of the derivatives, the decryption error in this example is a fraction of a percent. This error can be reduced by decreasing the step h , and by applying more accurate methods of numerical differentiation, for example, the Newton-Stirling method [21].

10 Constraints on the key-function of a real variable

When implementing the proposed cryptosystem, it is necessary to take into account the restrictions imposed on the key function.

1. The function must be defined on the set of real numbers.
2. The function should not be constant and not take zero values.
3. When using the key function, a situation should not arise when the number is divided by a number close to zero, which leads to the appearance of an unacceptable calculation error. To this end, it is recommended to test the cryptosystem for the entire alphabet of characters that will be used in messages.
4. Before sending an encrypted message, first check what it looks like after decryption. This will avoid mistakes that may occur because of not taking into account the previous paragraphs.

It should also be noted that in the system there is an effect similar to the “avalanche effect” in the AES cryptosystem. It consists in the fact that behind an incorrectly decrypted character the remaining message is decrypted incorrectly.

11 Conclusions

A cryptosystem with a symmetric key is offered. For the first time, a function of a real variable is used as a key, which satisfies some restrictions. It can be either continuous or discrete. Encryption is performed using disproportion functions. The specific examples show the possibility of encryption and decryption of text information, 2D-graphic images, as well as analog messages. Examples are also given that demonstrate the cryptographic strength of the proposed cryptosystem.

In the practical application of the proposed cryptosystem, the above limitations on the key function should be taken into account.

References

1. Federal Information Processing Standards Publication 197 November 26, 2001 Specification for the ADVANCED ENCRYPTION STANDARD (AES).
2. GOST 28147-89. Sistemy` obrabotki informaczi. Zashhita kriptograficheskaya. Algoritmy kriptograficheskogo preobrazovaniya. – Vved. 01.01.90.
3. Lebedev, A.N.: Kriptografiya s «otkry`ty`m klyuchom» i vozmozhnosti ee prakticheskogo primeneniya. Zashhita informaczi, vol. 2 (1992).
4. Rivest, R., Shamir, A., Adleman, I.: A method for obtaining digital signatures and public-key cryptosystems. Communications of the ACM, vol.21(2), pp. 120-126. (1978). doi:10.1145/359340.359342.
5. Gorbenko, I.D., Gorbenko, Yu.I.: Prykladna kryptologiya. Pidruchnyk. KhNURE, Fort, 878 p. Kharkiv (2012).
6. Hankerson, D.R., Vanstone, S.A., Menezes, A.J.: Guide to elliptic curve cryptography. XX, 311p. Springer, New York (2003).
7. Ladd, T. D., Jelezko, F., Laflamme, R., Nakamura, Y., Monroe, C., O’Brien, J. L.: Quantum Computing, Nature, vol. 464, pp. 45-53 (2010).
8. Klyucharev, P.G.: Kvantovy`j komp`yuter i kriptograficheskaya stojkost` sovremenny`kh sistem shifrovaniya. Vestnik MGEN im. N.E`.Baumana, ser. «Estestvenny`e nauki», vol. 2 (2007).

9. Grover, I.K.: Quantum Mechanics Help in Searching for a Needle in a Haystack. *Phys. Rev. Lett.*, vol. 78(2), pp. 326-328 (1997).
10. Shor, P.: Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer. *Proceedings of the 35th Annual Sympotion of Foundations of Computer Science* (1994).
11. Proos, J.A.: Shor's discrete logarithm quantum algorithm for elliptic curves. p. 35. Waterloo. Ont: Faculty of Mathematics University of Waterloo (2003).
12. Kolmogorov, A.N., Fomin, S.V.: *E'lementy' teorii funkczij i funkczional'nogo analiza*. Nauka, Moscow (1972).
13. Avramenko, V.V., Zabolotny, M.I.: A Way of Data Coding. Patent UA H041. 9/00 №42957, Ukraine (2009).
14. Avramenko, V.V.: Characteristic properties of disproportionality functions and their application to solving diagnoses problems. *Transactions of Sumy State University (SumDU)*, vol. 16, pp. 24-28, Ukraine (2000).
15. Kalashnikov, V.V., Avramenko, V.V., Kalashnykova, N.I.: Derivative disproportion functions for pattern recognition. In: Watada, J., Tan, S.C., Vasant, P., Padmanabhan, E., Jain, L.C. (eds.) *Unconventional Modelling, Simulation, and Optimization of Geoscience and Petroleum Engineering*, pp. 95–104. Springer, Heidelberg (2018). Chapter 7.
16. Kalashnikov, V.V., Avramenko, V.V., Slipushko, N.Yu., Kalashnykova, N.I., Konoplyanchenko, A.E.: Identification of quasi-stationary dynamic objects with the use of derivative disproportion functions. 108(C): 2100–2109. *Procedia Comput. Sci.* (2017).
17. Kalashnikov, V.V., Avramenko, V.V., Kalashnikova, N.I., Kalashnikov-Jr., V.V.: A cryptosystem based on sums of key functions. *International Journal of Combinatorial Optimization Problems and Informatics*, vol. 8, No.1, pp. 31-38 (2017).
18. Kalashnikova, N.I., Avramenko, V.V., Kalashnikov, V.V.: Sums of Key Functions Generating Cryptosystems. In: J. M. F. Rodrigues et al. (Eds.): *Computational Science – ICCS 2019. ICCS 2019, Chapter 23, Lecture Notes in Computer Science*, vol. 11540, pp. 293-302. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-22750-0_23 .ISBN 978-3-030-22749-4; Online ISBN 978-3-030-22750-0
19. Karpenko, A.P.: Integral'nye harakteristiki neproporcional'nosti chislovyh funkcij i ih primenienie v diagnostike. *Vestnik SumDU*, vol. 16, pp. 20-25, Ukraine (2000).
20. Avramenko, V., Moskalenko, A.: Operative Recognition of Standard Signals in the Presence of Interference with Unknown Characteristics, *Proceedings of the Second International Workshop on Computer Modeling and Intelligent Systems (CMIS-2019)*, Zaporizhzhia, Ukraine, April 15-19 (2019).
21. Rao, S. B.: *Numerical Methods: With Program in Basic, Fortran, Pascal & C++*. Hyderabad, Universities Press, (2004).

Autocorrelation Criterion for Quality Assessment of Random Number Sequences

Emil Faure¹ [0000-0002-2046-481X], Iryna Myronets² [0000-0003-2007-9943], Artem Lavdanskyi³ [0000-0002-1596-4123]

Cherkasy State Technological University, Shevchenko Blvd., 460, Cherkasy,
18006, Ukraine

¹e.faure@chdtu.edu.ua

²i.myronets@chdtu.edu.ua

³a.lavdanskyi@chdtu.edu.ua

Abstract. The authors analyze different approaches to forming estimates of autocorrelation coefficients of random and pseudorandom number sequences. An integral estimate of normalized autocorrelation coefficients is theoretically obtained. Estimates of some statistical properties of normalized autocorrelation coefficients have been improved. The autocorrelation criterion for quality assessment of time series based on simultaneous analysis of several autocorrelation coefficients has been further developed by adapting it to uniformly distributed random variables. The technique of its implementation is presented. Applying the criterion revealed statistical deviations for some pseudorandom number generators that successfully pass all TestU01 autocorrelation tests.

Keywords: Random numbers · Random number generator · Time series · Correlation · Autocorrelation · Quality assessment · Statistical criterion.

1 Introduction

Since random number sequences are multi-parameter processes, different methods and criteria are used for assessing their quality. These methods and criteria consider random processes from different standpoints using different statistical estimates.

The most famous test suits are: the Donald Knuth's statistical tests set [1]; George Marsaglia's DIEHARD tests [2]; NIST Statistical Test Suite [3]; TestU01[4]. In addition, there are other packages and tests. Among them CRYPT-X [5], NIST PUB FIPS 140-2 [6] can be distinguished.

Based on the definition of discrete white noise [7], an autocorrelation test is one of the most common tests, which allow detecting statistical irregularities of the studied sequences of numbers.

In [1], as in [8], it is recommended to use the criterion of serial (cyclic) correlation between cyclically shifted copies of the studied sequence. In [4, 9, 10], estimates of autocorrelation coefficients are formed by comparing two subsequences. The count of

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

the subsequences' elements is made from the beginning and from the end of the studied sequence. In [11], the autocorrelation coefficients are calculated for sequence successive intervals, which can overlap or stand away from each other.

Thus, there are different approaches to finding empirical autocorrelation coefficients that form an autocorrelation function (ACF) estimate. However, the main task of the autocorrelation test is to determine the correspondence of the sequence ACF estimate to the ACF of random number sequence described by the Dirac delta function [12]. Thus, estimates of the autocorrelation coefficients at non-zero points should go to zero. Their significance is most often evaluated by the Student's criterion [13].

The authors of the work [4] verify the correspondence of the distribution of autocorrelation coefficients estimates at nonzero points to the binomial law. For a large number of values, it can be approximated to normal.

In [14, 15] the statistical criteria of ACF side lobes complex estimate are proposed. Instead of testing the significance of each individual autocorrelation coefficient, these criteria check several autocorrelation coefficients to be different from zero. However, the proposed criteria are not adapted to analyze sequences of uniformly distributed random and pseudorandom numbers.

Thus, the question of estimating autocorrelation coefficients needs further study. This leads to the need for deeper analysis to identify the correlation properties inherent in sequences generated by natural sources of discrete white noise and not inherent in artificially generated pseudorandom sequences (PRS).

The purpose of this work is to develop a criterion for assessing the quality of sequences of uniformly distributed random and pseudorandom numbers, which allows detecting statistical irregularities not detected to date.

2 Formal Problem Statement

According to [16], the normalized autocorrelation coefficient $\rho_x(t', t'')$ of a random process $X(t)$ is calculated in the general form according to the expression:

$$\rho_x(t', t'') = K_x(t', t'') / \left(\sqrt{\text{Var}[X(t')]} \cdot \sqrt{\text{Var}[X(t'')]} \right),$$

where $K_x(t', t'') = E[(X(t') - E[X(t')]) \cdot (X(t'') - E[X(t'')])]$ is the correlation moment (covariance coefficient) of the intersections $X(t')$ and $X(t'')$ of the random process $X(t)$ at times t' and $t'' = t' + \tau$; $E[X(t')]$, $E[X(t'')]$, $\text{Var}[X(t')]$, $\text{Var}[X(t'')]$ are the expectations and the variances of $X(t')$, $X(t'')$. We will represent the intersections $X(t')$ and $X(t'')$ as random variables (r.v.) X' and X'' . Then $E[X(t')] = E(X')$, $E[X(t'')] = E(X'')$, $\text{Var}[X(t')] = \text{Var}(X')$, $\text{Var}[X(t'')] = \text{Var}(X'')$.

Suppose that random number generator (RNG) or pseudorandom number generator (PRNG) forms a stationary random process. For such a process, $E(X') = E(X'') = m_X = \text{const}$, $\text{Var}(X') = \text{Var}(X'') = \sigma_X^2 = \text{const}$, and an autocorrelation coefficient depends only on the lag $\tau = t'' - t'$: $K_X(t', t'') = k_X(\tau)$.

Let $\tilde{X}' = X' - E(X')$, $\tilde{X}'' = X'' - E(X'')$ be r.v. with expectations $E(\tilde{X}') = E(\tilde{X}'') = 0$ and variances $\text{Var}(\tilde{X}') = \text{Var}(\tilde{X}'') = \sigma_X^2$. Then the correlation moment $k_X(\tau) = E(\tilde{X}' \cdot \tilde{X}'')$ and the normalized autocorrelation coefficient $\rho_X(\tau) = k_X(\tau) / (\sqrt{\text{Var}(X')} \cdot \sqrt{\text{Var}(X'')}) = E(\tilde{X}' \cdot \tilde{X}'') / \sigma_X^2$.

Suppose that the r.v. $\tilde{X}'$ and $\tilde{X}''$ are uncorrelated (this corresponds to the properties of white noise intersections) and independent. Then the $\rho_X(\tau)$ value is invariant to the value of $\tau \neq 0$ and $\rho_X(\tau \neq 0) \equiv 0$ [9].

The product of r.v. $\xi = \tilde{X}' \cdot \tilde{X}''$ can be considered as a r.v. with parameters: $E(\xi) = E(\tilde{X}') \cdot E(\tilde{X}'') = 0$ and $\text{Var}(\xi) = E(\xi^2) - E^2(\xi) = E\left[\left(\tilde{X}' \cdot \tilde{X}''\right)^2\right] = E\left(\tilde{X}'^2\right) \cdot E\left(\tilde{X}''^2\right) = \text{Var}(\tilde{X}') \cdot \text{Var}(\tilde{X}'')$ or $\text{Var}(\xi) = \sigma_X^4$, and $\sigma_\xi = \sigma_X^2$. In this case $\rho_X(\tau) = E(\xi) / \sigma_\xi$.

If values of a discrete random process $\xi(t)$ form a countable set of independent values $\xi_1, \xi_2, \dots, \xi_n, \dots$ then $E(\xi) = \lim_{n \rightarrow \infty} (1/n \sum_{i=1}^n \xi_i)$.

According to the Lindeberg-Levy theorem [16], if mutually independent r.v. $\xi_1, \xi_2, \dots, \xi_n, \dots$ are equally distributed and have an expectation $E(\xi) = a$ and a variance σ_ξ^2 , then the value $(\sum_{i=1}^n \xi_i - na) / \sigma_\xi \sqrt{n}$ is normally distributed when $n \rightarrow \infty$:

$$\left(\sum_{i=1}^n \xi_i - na\right) / \sigma_\xi \sqrt{n} \rightarrow N(0; 1). \text{ Given that } E(\xi) = a = 0,$$

$$\left(\sum_{i=1}^n \xi_i - na\right) / \sigma_\xi \sqrt{n} = \sqrt{n} \cdot 1/n \cdot \sum_{i=1}^n \xi_i / \sigma_\xi = \sqrt{n} E(\xi) / \sigma_\xi = \sqrt{n} \rho_X(\tau) \rightarrow N(0; 1).$$

In other words, the marginal distribution of the normalized autocorrelation coefficient $\rho_X(\tau)$ of a random process whose intersections are independent r.v., is a normal distribution with expectation $E(\rho_X(\tau)) = 0$ and variance $\text{Var}(\rho_X(\tau)) = 1/n$.

Consider that $\text{Var}(\rho_x(\tau)) = E(\rho_x^2(\tau)) - (E(\rho_x(\tau)))^2 = \lim_{\substack{n \rightarrow \infty \\ T \rightarrow \infty}} (1/T \cdot \sum_{i=1}^T \rho_x^2(\tau))$. It follows that

$$\lim_{\substack{n \rightarrow \infty \\ T \rightarrow \infty}} \left(\sum_{\tau=1}^T \rho_x^2(\tau) \right) = T/n. \quad (1)$$

Expression (1) defines the limit value of the ACF side lobes power $W(T) = \sum_{\tau=1}^T \rho_x^2(\tau)$: $\lim_{\substack{n \rightarrow \infty \\ T \rightarrow \infty}} (W(T)) = T/n$.

In addition, since the r.v. $\rho_x(\tau) \rightarrow N(0; 1/n)$ when $n \rightarrow \infty$, the r.v. $\sum_{\tau=1}^T n \rho_x^2(\tau)$ has a distribution χ_T^2 with T degrees of freedom ($n \rightarrow \infty$):

$$n \sum_{\tau=1}^T \rho_x^2(\tau) \rightarrow \chi_T^2. \quad (2)$$

Then for side lobes power $nW(T) \rightarrow \chi_T^2$.

Thus, expression (2) is an integral estimate for autocorrelation coefficients. It creates the preconditions for constructing statistical criteria for checking the correlation properties of sequences of random and pseudorandom numbers by their empirical estimates.

Normalized autocorrelation coefficients are strictly defined by the theoretical expression (1). However, the correlation properties estimate of empirical number sequence can significantly depend on the studied sequence properties and the experiment conditions. In particular, the sequence correlation properties estimate can be performed:

- on a sequence period in the case of its periodicity, by analyzing a periodic ACF (PACF);
- on some fixed size sample;
- in real time, when sequence elements arrive to the analyzer sequentially.

In addition, the estimating correlation properties of random and pseudorandom number sequences may be conducted under conditions where the distribution law of a discrete random variable (d.r.v.) and its parameters are either fully known or empirically determined. In all these cases, it is necessary to know the first and second initial moments of normalized autocorrelation coefficients for their integral estimate.

We estimate these statistical properties of the normalized autocorrelation coefficients of a discrete random process calculated according to the defined approaches.

3 Estimate of Statistical Properties of Normalized Autocorrelation Coefficients

3.1 Periodic ACF Estimate

ACF is periodic if an original sequence is also periodic. Moreover, as it shown in [17], it is advisable for periodic signals to estimate the probabilistic moments in the minimum period. Taking into account the symmetry property of the PACF graph with respect to the axis, which is half period from the y-axis, it is allowed to reduce the number of calculated values by 2 times.

Let the number sequence $(x_0, x_1, \dots, x_{n-1})$ be repeated periodically with period n . In this case, as shown in [1] and [8], the estimate of the normalized autocorrelation coefficient is calculated according to the expression:

$$r_{PACF\ x}(\tau) = \frac{\sum_{i=0}^{n-1} [(x_i - \bar{x}) \cdot (x_{(i+\tau) \bmod n} - \bar{x})]}{\sum_{i=0}^{n-1} (x_i - \bar{x})^2} = \frac{n \sum_{i=0}^{n-1} x_i x_{(i+\tau) \bmod n} - \left(\sum_{i=0}^{n-1} x_i \right)^2}{n \sum_{i=0}^{n-1} x_i^2 - \left(\sum_{i=0}^{n-1} x_i \right)^2}, \quad (3)$$

where $\bar{x} = \sum_{i=0}^{n-1} x_i / n$ is a statistical estimate of the X expectation.

Note that the estimate (3) is performed for the whole sequence period. Therefore, the estimate of the d.r.v. X expectation coincides with the expectation: $\bar{x} = \sum_{i=0}^{n-1} x_i / n = \lim_{m \rightarrow \infty} (\bar{x} = \sum_{i=0}^m x_i / m) = E(X)$. Based on similar considerations, the variance estimate over the entire sequence period also coincides with the d.r.v. variance: $\sigma_x^2 = \sum_{i=0}^{n-1} (x_i - \bar{x})^2 / n = \lim_{m \rightarrow \infty} (\sum_{i=0}^m (x_i - E(X))^2 / m) = Var(X) = \sigma_x^2$. Then the expression (3) can be represented as $r_{PACF\ x}(\tau) = 1/n \cdot \sum_{i=0}^{n-1} [(x_i - E(X)) \cdot (x_{(i+\tau) \bmod n} - E(X))] / \sigma_x^2$.

Obviously, given the symmetry property of the PACF graph, the analysis of autocorrelation coefficients estimates is advisable to perform for $\tau \in (0; [n/2] + 1)$.

According to [1], $E(r_{PACF\ x}(\tau)) = -1/(n-1)$. The upper bound estimate of variance of $r_{PACF\ x}(\tau)$ calculated from (3) for arbitrary independent variables is:

$$Var(r_{PACF\ x}(\tau)) \leq \frac{n^2}{(n-1)^2(n-2)}. \text{ However, } Var_{norm}(r_{PACF\ x}(\tau)) = \frac{n(n-3)}{(n+1)(n-1)^2} \quad [18]$$

for the normal distribution of the initial values, and

$$Var_{unif}(r_{PACF\ x}(\tau)) = \frac{24}{5} n^{-2} + O(n^{-7/3} \log(n)) \quad [1] \text{ for their uniform distribution. In}$$

addition, as shown in [19], the value $r_{PACF\ x}(\tau)$ is distributed asymptotically normal even for sufficiently small samples ($n > 10$). In [1] it is recommended that the esti-

mate (5) should be between $E(r_{PACF\ x}(\tau)) - 2\sqrt{Var(r_{PACF\ x}(\tau))}$ and $E(r_{PACF\ x}(\tau)) + 2\sqrt{Var(r_{PACF\ x}(\tau))}$ for uncorrelated values $(x_0, x_1, \dots, x_{n-1})$.

3.2 ACF Estimate on Fixed Size Sample

ACF estimate on some fixed size sample $(x_0, x_1, \dots, x_{n-1})$ is widely used in econometrics for constructing regression models [14, 15]. In addition, a similar ACF estimate is used to investigate sequence properties in the aperiodic mode typical to information transmission in communication systems [20]. In this case, the estimate of normalized ACF is calculated according to the expression [21]:

$$r_x^*(\tau) = 1/(n-\tau) \cdot \sum_{i=0}^{n-1-\tau} [(x_i - E(X)) \cdot (x_{i+\tau} - E(X))] / \left(1/n \cdot \sum_{i=0}^{n-1} (x_i - E(X))^2 \right) \quad (4)$$

for a known a priori value of $E(X)$, or the expression [21]:

$$r_x^{**}(\tau) = 1/(n-\tau) \cdot \sum_{i=0}^{n-1-\tau} [(x_i - \bar{x}) \cdot (x_{i+\tau} - \bar{x})] / \left(1/n \cdot \sum_{i=0}^{n-1} (x_i - \bar{x})^2 \right). \quad (5)$$

In [21] it is shown that if the random vectors $(x_i, x_{i+\tau})$ are independent and equally distributed, then the distribution law of the value $r_x^*(\tau)$ (as well as $r_x^{**}(\tau)$) has an asymptotic normal distribution.

Various authors use or recommend to use for (5) the value of zero as an approximate estimate of its expectation and the value of $n^{1/2}$ [14] or $(n/\{(n+2)(n-\tau)\})^{-1/2}$ [15] as an approximate estimate of its root-mean-square deviation. However, the exact value of the $r_x^{**}(\tau)$ expectation is defined in [22] and is equal to $E(r_x^{**}(\tau)) = -(n-1)^{-1}$. The upper bound of the estimate (5) variance is defined in [23] for any law of distribution of the initial values $\{x_i\}$:

$$Var(r_x^{**}(\tau)) \leq \frac{n^4 - (\tau+7)n^3 + (7\tau+16)n^2 + 2(\tau^2 - 9\tau - 6)n - 4\tau(\tau-4)}{n(n-1)^2(n-2)(n-3)}.$$

In addition, in [23] it is also shown that $Var_{norm}(r_x^{**}(\tau)) = (n^4 - (\tau+3)n^3 + 3\tau n^2 + 2\tau(\tau+1)n - 4\tau^2) / ((n+1)n^2(n-1)^2)$ for normally distributed r.v., and the expression $Var(r_x^{**}(\tau)) = (n-\tau)/(n(n+2))$ from [15] is valid in the case of a known expectation $E(X)$, that is, for the value $r_x^*(\tau)$.

Consider the normalized ACF estimate, which is given, for example, in [9] or [10]:

$$r_x^{***}(\tau) = \sum_{i=0}^{n-1-\tau} \left[(x_i - \bar{x}) \cdot (x_{i+\tau} - \bar{x}) \right] / \sqrt{\sum_{i=0}^{n-1-\tau} (x_i - \bar{x})^2 \cdot \sum_{i=0}^{n-1-\tau} (x_{i+\tau} - \bar{x})^2}. \quad (6)$$

Let us extend and present more fully the results of estimates of expectation $E(r_x^{***}(\tau))$ and variance $\text{Var}(r_x^{***}(\tau))$ obtained in [24].

The $r_x^{***}(\tau)$ distribution law for independent and equally distributed random vectors $(x_i, x_{i+\tau})$ is an asymptotic normal distribution [25]. This statement is also confirmed by [26, 27]. They show that the distributions of correlation analysis statistics are resistant to deviations of the observed multidimensional law from normal. The empirical distributions of these statistics are well described by the boundary laws obtained from the assumption of normality of the observed values.

To find $E(r_x^{***}(\tau))$, we use the methodology described in [22].

Let $z_i = x_i - \bar{x}$. Then $\sum_{i=0}^{n-1} z_i = \sum_{i=0}^{n-1} (x_i - \bar{x}) = \sum_{i=0}^{n-1} x_i - n\bar{x} = 0$.

$$\begin{aligned} E(r_x^{***}(\tau)) &= E\left(\frac{\sum_{i=0}^{n-1-\tau} z_i \cdot z_{i+\tau}}{\sqrt{\sum_{i=0}^{n-1-\tau} z_i^2 \cdot \sum_{i=0}^{n-1-\tau} z_{i+\tau}^2}}\right) = E\left(\frac{(n-\tau) \cdot z_i \cdot z_{i+\tau}}{\sum_{i=0}^{n-1-\tau} z_i^2}\right) = n \cdot E\left(\frac{z_i \cdot z_{i+\tau}}{\sum_{i=0}^{n-1} z_i^2}\right) \\ &= \frac{1}{n-1} \cdot E\left(\frac{\sum_{i \neq j} (z_i \cdot z_j)}{\sum_{i=0}^{n-1} z_i^2}\right) = \frac{1}{n-1} \cdot E\left(\frac{\left(\sum_{i=0}^{n-1} z_i\right)^2 - \sum_{i=0}^{n-1} z_i^2}{\sum_{i=0}^{n-1} z_i^2}\right) = -(n-1)^{-1}. \end{aligned}$$

The variance $\text{Var}(r_x^{***}(\tau))$ can be calculated according to the well-known expression: $\text{Var}(r_x^{***}(\tau)) = E\left(\left(r_x^{***}(\tau)\right)^2\right) - E^2(r_x^{***}(\tau))$.

As it shown in [23], $\left(\sum_{i=0}^{n-1-\tau} z_i z_{i+\tau}\right)^2 = \sum_{i=0}^{n-1-\tau} z_i^2 z_{i+\tau}^2 + 2 \sum_{i=0}^{n-1-2\tau} z_i z_{i+\tau}^2 z_{i+2\tau} + \sum_{*} z_i z_{i+\tau} z_j z_{j+\tau} \cdot \sum_{*}$ means a summation for $i, j = 1, \dots, n-\tau$, and $i, i+\tau, j, j+\tau$ are different. In addition, for the symmetric joint distribution of values $z_0, \dots, z_{n-1}$,

$$\begin{aligned} E\left(\left(r_x^{***}(\tau)\right)^2\right) &= E\left[\left(\sum_{i=0}^{n-1-\tau} z_i^2 \cdot \sum_{i=0}^{n-1-\tau} z_{i+\tau}^2\right)^{-1} \left\{ (n-\tau) z_1^2 z_2^2 + 2(n-\tau) z_1^2 z_2 z_3 + \right. \right. \\ &\quad \left. \left. + ((n-\tau)^2 - 2(n-2\tau) - (n-\tau)) z_1 z_2 z_3 z_4 \right\}\right] = \\ &= E\left[\left(\frac{n-\tau}{n} \sum_{i=0}^{n-1} z_i^2\right)^{-2} \left\{ \frac{n-\tau}{n(n-1)} \sum_{*} z_i^2 z_j^2 + \frac{2(n-2\tau)}{n(n-1)(n-2)} \sum_{*} z_i^2 z_j z_k + \right. \right. \\ &\quad \left. \left. + \frac{(n-\tau)^2 - 2(n-2\tau) - (n-\tau)}{n(n-1)(n-2)(n-3)} \sum_{*} z_i z_j z_k z_l \right\}\right], \end{aligned}$$

where $\sum^*$ denotes a summation for all different indices from 0 to $n-1$.

Using an equality $S_r = \sum_{i=0}^{n-1} z_i^r$ for $r \geq 1$, and $\sum^* z_i^2 z_j^2 = S_2^2 - S_4$, $\sum^* z_i^2 z_j z_k = 2S_4 - S_2^2$, and $\sum^* z_i z_j z_k z_l = 3S_2^2 - 6S_4$ from [28], by analogy with [23],

$$E\left(\left(r_x^{***}(\tau)\right)^2\right) = \frac{n \left[\begin{aligned} &(-n^3 + (\tau+3)n^2 - \tau(n+6\tau))M[S_4/S_2^2] + \\ &+ (n^2(n-\tau-4) + 3(n-\tau) + 3\tau(n+\tau)) \end{aligned} \right]}{(n-1)(n-2)(n-3)(n-\tau)^2}.$$

As shown in [23, 29], $1/n \leq S_4/S_2^2 \leq 1$ for any distribution law of r.v. z_i . Then, by analogy with [23], it is possible to obtain the variance upper bound of the correlation coefficient estimate (6). To do this, replacing $E[S_4/S_2^2]$ by $1/n$ we get

$$E\left(\left(r_x^{***}(\tau)\right)^2\right) \leq \frac{n^4 - n^3(\tau+5) + n^2(4\tau+6) + n(3\tau^2 - 4\tau) - 6\tau^2}{(n-1)(n-2)(n-3)(n-\tau)^2}. \text{ Then}$$

$$\text{Var}\left(r_x^{***}(\tau)\right) \leq \frac{n^5 - n^4(\tau+7) + n^3(7\tau+16) + n^2(2\tau^2 - 18\tau - 12) - n(4\tau^2 - 16\tau)}{(n-1)^2(n-2)(n-3)(n-\tau)^2}. \quad (7)$$

According to [22], for normally distributed r.v. x_i , $E[S_4/S_2^2] = 3(n-1)/(n(n+1))$,

$$\text{Var}_{\text{norm}}\left(r_x^{***}(\tau)\right) = \frac{n^4 - (\tau+3)n^3 + 3\tau n^2 + 2\tau(\tau+1)n - 4\tau^2}{(n+1)(n-1)^2(n-\tau)^2}. \quad (8)$$

One can notice that $\text{Var}_{\text{norm}}\left(r_x^{***}(\tau)\right) \xrightarrow[n \rightarrow \infty]{n \gg \tau} 1/(n-\tau)$. In addition, in the general case

$$\text{Var}\left(r_x^{***}(\tau)\right) \leq \frac{1 - 7/n + 16/n^2 - 12/n^3 + 2\tau(\tau-1)(n-2)/(n^3(n-\tau))}{(1-1/n)^2(1-2/n)(1-3/n)(n-\tau)} \xrightarrow[n \rightarrow \infty]{n \gg \tau} \frac{1}{n-\tau}.$$

Thus, the obtained results are in full agreement with [10]. The variance of the estimate (6) is asymptotically equal to $1/(n-\tau)$ if n is large. However, the estimates (7) and (8) are more accurate.

We denote the relative error of approximation of the autocorrelation coefficient estimate variance $\text{Var}\left(r_x^{***}(\tau)\right)$ by $\delta(n, \tau) = \left| \left(\text{Var}\left(r_x^{***}(\tau)\right) - 1/(n-\tau) \right) / \text{Var}\left(r_x^{***}(\tau)\right) \right|$.

Analysis of the dependency $\delta(n, \tau)$ graph for $n \in [50; 100]$, $\tau \in [1; 25]$ (Figure 1) indicates that $\delta(n, \tau) \rightarrow \max$ for $\tau \rightarrow \max$, $n \rightarrow \min$. In particular, $\delta(100, 1) = 1.02 \cdot 10^{-4}$, $\delta(100, 25) = 1.578 \cdot 10^{-3}$, and $\delta(50, 25) = 0.02$. It is also worth noting that $\min(\delta(n, \tau)) \neq \delta(n, 1)$ for a fixed value of n .

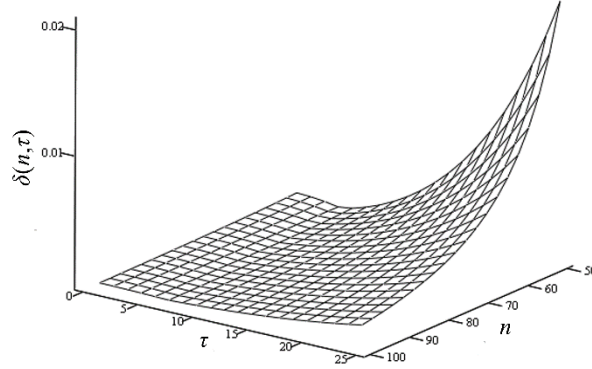


Fig. 1. The graph of dependence $\delta(n, \tau)$ for $n \in [50; 100]$, $\tau \in [1; 25]$

The graph of the relative error of approximation of the autocorrelation coefficient estimate variance $Var_{norm}(r_x^{***}(\tau))$ for normally distributed r.v. x_i $\delta_{norm}(n, \tau) = \left| \left(Var_{norm}(r_x^{***}(\tau)) - 1/(n - \tau) \right) / Var_{norm}(r_x^{***}(\tau)) \right|$ depending on $n \in [50; 100]$, $\tau \in [1; 25]$ is shown in Figure 2.

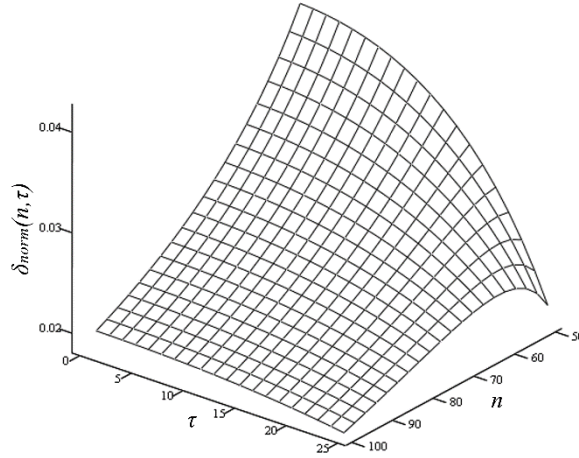


Fig. 2. The graph of dependence $\delta_{norm}(n, \tau)$ for $n \in [50; 100]$, $\tau \in [1; 25]$

For the specified definition area $\delta_{norm}(n, \tau) \rightarrow \max$ for $\tau \rightarrow \min$, $n \rightarrow \min$. In particular, $\delta_{norm}(100, 1) = 0.021$, $\delta_{norm}(100, 25) = 0.019$, and $\delta_{norm}(50, 1) = 0.042$. However, in the general case $\max(\delta_{norm}(n, \tau)) \neq \delta_{norm}(n_{\min}, \tau)$ for a fixed value of τ . In Figure 2, this situation is observed for $\tau \in [21; 25]$.

3.3 ACF Estimate for Long Period Sequences

When analyzing the correlation properties of a random number sequence, any shift of the analyzed sequence belongs to the set $\{x_i(t)\}$ of realizations of a stationary discrete random process $X(t)$. In other words, the "zero" point for the beginning of realization may be arbitrary. In this case, we can consider a set of realizations $\{x_i(t) : x_i(t) = x_{i-j}(t+j)\}$. To enable autocorrelation coefficients practical analysis, we will assume sequences of n numbers, the first of which are formed at times $t = 0, 1, 2, \dots$, by realizations of a random process $X(t)$. This approach is most effective when the correlation properties of a random number sequence are analyzed in real time, not by some fixed size sample. In this case, the sequence elements are written to a limited size buffer (this approach is similar to the sliding window method).

We denote the element of a random number sequence at a discrete time t by x_t . Then the estimate of the normalized autocorrelation coefficient of order τ is as follows:

$$r_x(\tau) = \frac{\sum_{i=0}^{n-1} [(x_{t+i} - \overline{x(t)}) \cdot (x_{t+\tau+i} - \overline{x(t+\tau)})]}{\sqrt{\sum_{i=0}^{n-1} (x_{t+i} - \overline{x(t)})^2 \cdot \sum_{i=0}^{n-1} (x_{t+\tau+i} - \overline{x(t+\tau)})^2}}, \quad (9)$$

where $\overline{x(t)} = 1/n \cdot \sum_{i=0}^{n-1} x_{t+i}$, $\overline{x(t+\tau)} = 1/n \cdot \sum_{i=0}^{n-1} x_{t+\tau+i}$ are the average sample values of random process intersections $X(t)$, $X(t+\tau)$. Note that for a stationary random process, equality $\lim_{n \rightarrow \infty} (P(|\overline{x(t)} - \overline{x(t+\tau)}| < \varepsilon)) = 1$ is satisfied for any small $\varepsilon > 0$.

The distribution law of the estimate $r = r_x(\tau)$ (9) for independent and equally distributed random vectors $(x_{t+i}, x_{t+\tau+i})$ is asymptotically normal [25] with expectation $\rho = \rho_x(\tau)$ and variance [30]

$$\text{Var}(r_x(\tau)) = \frac{\rho^2}{4n} \left(\frac{\mu_{40}}{\mu_{20}^2} + \frac{\mu_{04}}{\mu_{02}^2} + \frac{2\mu_{22}}{\mu_{20}\mu_{02}} + \frac{4\mu_{22}}{\mu_{11}^2} - \frac{4\mu_{31}}{\mu_{11}\mu_{20}} - \frac{4\mu_{13}}{\mu_{11}\mu_{02}} \right),$$

where μ_{km} are the theoretical central moments of the order k and m :

$$\mu_{km} = (x_{t+i} - E(x_{t+i}))^k (x_{t+\tau+i} - E(x_{t+\tau+i}))^m. \quad \text{According to [30],}$$

$$\text{Var}(r_x(\tau)) = (1 - \rho^2)^2 / n \text{ in the case of a normally distributed population.}$$

However, as shown in [30], applying the Fischer logarithmic transformation [31] to the sample correlation coefficients r leads to the following conclusions. The value $z = 1/2 \ln((1+r)/(1-r))$ should be considered normally distributed with an average

$1/2 \ln((1+\rho)/(1-\rho)) + \rho/(2(n-1))$ and a variance $1/(n-3)$. Therefore, the value $\lambda = \sqrt{n-3} \left(\frac{1}{2} \ln \left(\frac{1+r}{1-r} \right) - \left(\frac{1}{2} \ln \left(\frac{1+\rho}{1-\rho} \right) + \frac{\rho}{2(n-1)} \right) \right)$ is normal: $\lambda \sim N(0;1)$.

Thus, $\rho_x(\tau) \equiv 0$ for independent and equally distributed random vectors $(x_{t+i}, x_{t+\tau+i})$, the value $\lambda_x(\tau) = \sqrt{n-3}/2 \ln((1+r_x(\tau))/(1-r_x(\tau)))$ has a standard normal distribution, and $\sum_{\tau=1}^T (\lambda_x(\tau))^2 \rightarrow \chi_T^2$.

3.4 ACF Estimate for Uniformly Distributed Random/Pseudorandom Number Sequences with Known Parameters

Often, sequence statistics are checked at the output of a generator of uniformly distributed random or pseudorandom numbers in some range $[a, b]$ (such generators are most widely used for information security tasks as key entropy generators). Then the simplest analysis of the studied sequence allows us to determine the set of d.r.v. values at the generator output. If there is a sequence of numbers $\{x_i \in A\}$ from the alphabet A , then its cardinality $N = \max\{x_i\} - \min\{x_i\} + 1$, the range of d.r.v. values $X \in [\min\{x_i\}; \max\{x_i\}]$, its expectation $E(X) = (\min\{x_i\} + \max\{x_i\})/2$, and its variance $Var(X) = (N^2 - 1)/12$.

A similar estimate can be made if the size of the analyzed number sequence significantly exceeds the capacity of the alphabet. Otherwise there is a possibility to incorrectly define the lower or upper limit of d.r.v. values set. The probability that the minimum or the maximum value of N -symbol alphabet will not be present in a sequence of V symbols is equal to $P_{er} = 2((N-1)/N)^n$.

Thus, for the given values of alphabet capacity M and probability P_{er} , it is possible to calculate the required sample size $V \geq \log_{(N-1)/N} P_{er}/2$.

For example, for $N = 256$ and $P_{er} = 10^{-10}$ we get: $V \geq \log_{1-1/256} 10^{-10}/2 = 6061$.

For known parameters (expectation $E(X)$ and variance $Var(X)$) of a random process $X(t)$, the estimate of normalized ACF can be calculated by the expression

$$r_x'(\tau) = 1/n \cdot \sum_{i=0}^{n-1} [(x_{t+i} - E(X)) \cdot (x_{t+\tau+i} - E(X))] / Var(X). \quad (10)$$

In this case, for independent values x_i , as well as in accordance with the regularities stated in (1) and (2), and $n \rightarrow \infty$, $r_x'(\tau) \rightarrow N(0; 1/n)$, $\lim_{\substack{n \rightarrow \infty \\ T \rightarrow \infty}} \left(\sum_{\tau=1}^T (r_x'(\tau))^2 \right) = T/n$,

$$n \sum_{\tau=1}^T \left(r_x'(\tau) \right)^2 \rightarrow \chi_T^2. \quad (11)$$

For side lobes power $W'(T) = \sum_{\tau=1}^T \left(r_x'(\tau) \right)^2$ of ACF estimate (10), $\lim_{\substack{n \rightarrow \infty \\ T \rightarrow \infty}} (W'(T)) = T/n$ and the expression (11) can be rewritten as follows:

$$nW'(T) \rightarrow \chi_T^2. \quad (12)$$

In this case, the statistical criterion for corresponding an ACF estimate of uniformly distributed random number sequence to the white noise ACF provides calculating the normalized autocorrelation coefficients from (10), forming an estimate of the side lobes power $W'(T)$, and then estimating it with (12).

4 Description of the Criterion for Estimating Uniformly Distributed Random Number Sequences

The criterion for estimating uniformly distributed random number sequences is the following:

1. if the d.r.v. definition area is unknown, it is empirically determined;
2. the d.r.v. expectation and variance are calculated;
3. using the expression (10), the sequence of estimates $r_x'(\tau)$ of normalized autocorrelation coefficients is calculated for $\tau \in [1; T]$;
4. the side lobes power $W'(T) = \sum_{\tau=1}^T \left(r_x'(\tau) \right)^2$ of the ACF estimate is calculated;
5. if the value $nW'(T)$ for the selected level of significance does not exceed the quantile $\chi_{1-\alpha, T}^2$ of chi-square distribution with T degrees of freedom, the null hypothesis is accepted. It is that the numbers of the sequence under study are random. Otherwise, the null hypothesis is rejected.

5 Applying the Criterion for Estimating Uniformly Distributed Random Number Sequences

We implement the developed criterion for known PRNG, which pass all tests of the TestU01 test package [4]. For this purpose we conduct $N = 1000$ independent tests. We define the value $nW'(T)$. Then calculate the relative frequency of the event $A = \{nW'(T) \leq \chi_{1-\alpha, T}^2\}$.

We denote by Q the value that in each specific test takes a value of 1 if the event A is true and 0 if the event A is false. Then the relative frequency of the event $A = \{nW'(T) \leq \chi^2_{1-\alpha, T}\}$ in N independent tests is $p^* = \sum_{i=1}^N Q_i / N$.

The expectation of the relative frequency is $E(p^*) = 1 - \alpha$, its variance is $Var(p^*) = \alpha(1 - \alpha)/N$. Then the inequality $|p^* - (1 - \alpha)| < t_\gamma \sqrt{\alpha(1 - \alpha)/N}$ is satisfied with probability γ , where t_γ is the quantile of standard normal distribution with level γ . In other words, the calculated relative frequency p^* falls within the confidence interval $(1 - \alpha - t_\gamma \sqrt{\alpha(1 - \alpha)/N}; 1 - \alpha + t_\gamma \sqrt{\alpha(1 - \alpha)/N})$ with probability γ .

For testing we will choose $\alpha = \gamma = 0.05$. Then the confidence interval for p^* is $(0.9365; 0.9635)$. The test results are summarized in Table 1.

Table 1. The results of applying the criterion for estimating uniformly distributed PRNG

PRNG	TestU01 autocorrelation tests results (4 smallCrush + 4 bigCrush), passed/total tests	Test results of the developed criterion
LCG(2^{24} , 16598013, 12820163)	2/8	failed
LCG(2^{31} , 65539, 0)	6/8	failed
LCG(2^{32} , 69069, 1)	5/8	failed
LCG(2^{32} , 1099087573, 0)	5/8	failed
LCG(2^{46} , 5^{13} , 0)	7/8	failed
LCG(2^{48} , 25214903917, 11)	7/8	failed
LCG(2^{48} , 5^{19} , 0)	7/8	failed
LCG(2^{48} , 33952834046453, 0)	7/8	failed
LCG(2^{48} , 44485709377909, 0)	7/8	failed
LCG(2^{59} , 3^{13} , 0)	8/8	failed
LCG(2^{63} , 5^{19} , 1)	8/8	failed
LCG(2^{63} , 9219741426499971445, 1)	8/8	failed
LCG($2^{31} - 1$, $2^{31} - 2^{10}$, 0)	6/8	passed
LCG($2^{31} - 1$, 16807, 0)	7/8	passed
LCG($2^{61} - 1$, $2^{30} - 2^{19}$, 0)	8/8	passed
LCG($10^{12} - 11$, 427419669081, 0)	8/8	passed

The results indicate that some of the generators that successfully pass all the TestU01 autocorrelation tests do not meet the developed criterion.

6 Conclusions

The study has produced the following results:

- integral estimate of normalized autocorrelation coefficients (ACF side lobes) is theoretically obtained. This provides the basis for building statistical criteria for verifying correlation properties of random and pseudorandom number sequences by their empirical estimates;
- the first and second initial moments of estimates of normalized autocorrelation coefficients are presented for: PACF; ACF of fixed size sample calculated according to different approaches (for example, presented in [9, 10, 21]); "sliding window" ACF for long period sequences;
- the upper bound of the variance of estimates of normalized autocorrelation coefficients calculated by [9] or [10] is clarified. This allows increasing the accuracy of ACF side lobes integral estimate;
- the criterion for estimating autocorrelation of time series based on simultaneous analysis of several autocorrelation coefficients (similar to the Box-Pierce [14] and Ljung–Box [15] criteria) has been further developed by adapting it to uniformly distributed r.v. This made it possible to perform a complex estimate of ACF for sequences of uniformly distributed random and pseudorandom numbers;
- applying the criterion revealed statistical deviations for some PRNG that successfully pass all TestU01 autocorrelation tests.

References

1. Knuth, D. E.: The Art of Computer Programming: Seminumerical Algorithms, 3 ed., vol. 2. Boston, Addison-Wesley Longman Publishing Co., Inc. (1997)
2. Marsaglia, G.: DIEHARD Battery of Tests of Randomness. <http://www.stat.fsu.edu/pub/diehard>
3. Bassham, L. E. et al.: A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications. SP 800-22 Rev. 1a., Gaithersburg, MD, United States (2010)
4. L'Ecuyer, P., Simard, R.: TestU01: A C library for empirical testing of random number generators. ACM Transactions on Mathematical Software. **33**, no. 4, 22-es (2007). doi: 10.1145/1268776.1268777
5. Caelli, W., Dawson, E., Nielsen, L., Gustafson, H.: CRYPT-X Statistical Package Manual, Measuring the strength of Stream and Block Ciphers. Queensland University of Technology (1992)
6. Security requirements for cryptographic modules. US standard FIPS PUB 140-2 (2001)
7. Papoulis, A., Pillai, S. U.: Probability, random variables, and stochastic processes, 4th ed. Boston, McGraw-Hill (2002)
8. Ivanov, M.A., Chugunkov, I.V.: Theory, application and quality evaluation of pseudorandom sequence generators. Moscow, Kudits-Obraz (2003) (in Russian)
9. Smirnov, N.V., Dunin-Barkovsky, I.V.: Course in probability theory and mathematical statistics for technical applications. Moscow, Nauka (1969) (in Russian)
10. Kendall, M.G.: The advanced theory of statistics, vol. 2. London: C. Griffin & Company limited (1946)
11. Faure, E.V., Shcherba, A.I., Rudnytskyi, V.M.: The method and criterion for quality assessment of random number sequences. Cybernetics and Systems Analysis, **52**, no. 2, 277–284 (2016). doi: 10.1007/s10559-016-9824-3

12. Dirac, P.A.M.: The principles of quantum mechanics. London, OUP (1958). doi: 10.1063/1.3062610
13. Bolshev, L.N., Smirnov, N.V.: Tables of mathematical statistics. Moscow, Nauka (1983) (in Russian)
14. Box, G.E.P., Pierce, D.A.: Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American Statistical Association*, **65**, no. 332, 1509–1526 (1970). doi: 10.1080/01621459.1970.10481180
15. Ljung, G.M., Box, G.E.P.: On a measure of lack of fit in time series models. *Biometrika*, **65**, no. 2, 297 (1978). doi: 10.1093/biomet/65.2.297
16. Wentzel, Ye.S., Ovcharov, L.A.: Applied problems of probabilities theory. Moscow, Radio and Communication (1983) (in Russian)
17. Kuznetsov, V.M.: Generators of random and pseudorandom sequences on digital delay elements (basics of theory and methods of construction). Thesis for doctor of technical sciences degree. Kazan, Kazan State Technical University (2011) (in Russian)
18. Dixon, W.J.: Further contributions to the problem of serial correlation. *Ann. Math. Statist.*, **15**, no. 2, 119–144 (1944). doi: 10.1214/aoms/1177731279
19. Lemeshko, B.Yu., Komissarov, A.S., Shcheglov, A.Ye.: Application of tests for trend detection and checking for randomness. *Metrology*, **12**, 3–25 (2010) (in Russian)
20. Varakin, L.E.: Communication systems with noise-like signals. Moscow, Radio and Communication (1985) (in Russian)
21. Anderson, T.W., Walker, A.M.: On the asymptotic distribution of the autocorrelations of a sample from a linear stochastic process. *The Annals of Mathematical Statistics*, **35**, no. 3, 1296–1303 (1964). doi: 10.1214/aoms/1177703285
22. Moran, P.A.P.: Some theorems on time series: II the significance of the serial correlation coefficient. *Biometrika*, **35**, no. 3/4, 255–260 (1948). doi: 10.1093/biomet/35.3-4.255
23. Dufour, J.-M., Roy, R.: Some robust exact results on sample autocorrelations and tests of randomness. *Journal of Econometrics*, **29**, no. 3, 257–273 (1985). doi: 10.1016/0304-4076(85)90155-1
24. Faure, E.V.: Statistical characteristics of estimates of normalized autocorrelation coefficients of (pseudo) random number sequences. In: All-Ukrainian Scientific and Practical Internet Conference on Automation and Computer-Integrated Technologies in Production and Education: State, Achievements, Prospects of Development, pp. 46-47. Cherkasy (2015) (in Russian)
25. Orlov, A.I.: Applied statistics. Moscow, Examen (2004) (in Russian)
26. Lemeshko, B.Yu., Pomadin, S.S.: Correlation analysis of observations of many-dimensional random variables under violation of normality assumptions. *Siberian Journal of Industrial Mathematics*, **5**, no. 3 (11), 115-130 (2002) (in Russian)
27. Pomadin, S.S.: Investigation of statistics distributions of multidimensional data analysis in violation of normality assumptions, dissertation. Thesis for candidate of technical sciences degree. Novosibirsk, Novosibirsk State Technical University (2004) (in Russian)
28. Kendall, M.G., Stuart, A., Ord, J.K.: The advanced theory of statistics. 4 ed. V. 3: Design and analysis, and time-series. London: C. Griffin & Company limited (1983)
29. Moran P.A.P.: Testing for serial correlation with exponentially distributed variates. *Biometrika*, **54**, no. 3/4, 395–401 (1967). doi: 10.1093/biomet/54.3-4.395
30. Kramer, G.: Mathematical methods of statistics, 2 ed. Moscow, Mir (1975)
31. Fisher, R.A.: On the probable error of a coefficient of correlation deduced from a small sample. *Metron*, **1**, 3–32 (1921)

Analysis of Algorithms for Constructing Dense Sequencing of Digraphs Vertices

Valentina Turchyna¹[0000-0003-1051-9597], Kostiantyn Karavaiev²[0000-0001-6062-4577],

¹Oles Honchar Dnipro National University, D.Yavornitsky avenue, 35, Dnipro, 49000, Ukraine
vaturchina1949@gmail.com

²Oles Honchar Dnipro National University, D.Yavornitsky avenue, 35, Dnipro, 49000, Ukraine
karavaiev_k@fpm.dnulive.dp.ua

Abstract. In this work behavior of one exact polynomial algorithms for particular case of optimal sequencing problems when it's used for arbitrary graphs and sequencing widths is analysed. Based on this analysis several modifications of this algorithm are developed. Results of computational experiments has shown that proposed modifications are effective and substantially increase precision of finding optimal solution up to 98,5 percent. Additionally it is proved that an arbitrary optimal sequencing problem can be reduced to the problem with graph that has a dense sequencing for even sequencing width.

Keywords: discrete optimization, scheduling theory, optimal sequencing, approximate methods, level principle, maximum matching, dense sequencing

1 Introduction

When solving practical problems related to the optimal allocation of a finite set of jobs (tasks, projects, operations, etc.) between executors, two main classes are distinguished: the first - problems in which the order of execution is arbitrary, the second - problems in which technological constraints are imposed. This work is devoted to the latter. Since the technological constraints can be mathematically modeled with a directed acyclic graph, the problems under consideration are formulated as optimization problems on graphs. In general, these problems are NP-hard, so it is important to have effective approximate algorithms of polynomial complexity [1, 2].

2 Formulation of the problem

Consider one of the well-known problems of sequencing vertices of a directed graph. Suppose that a finite set of jobs is specified, on the order of execution of which the technological constraints are imposed [3-5]. Assuming that all jobs have the same execution time, the minimum time for which, without violating technological restrictions, all jobs can be executed by a given number of performers has to be determined.

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

It is natural to set technological constraints in a form of directed acyclic graph $G(V, U)$, $|V| = n$, where V is the set of vertices whose elements are associated with the jobs. Then the arrows correspond to the technological constraints. One of the optimization problems that arises in this case is the construction of optimal parallel sequencing, which is defined as such placement of the vertices of the digraph in line-arranged places for which the following conditions are satisfied:

- in each place there is no more than a given number of vertices, denoted h ;
- if a pair of vertices $(i, j) \in U$, then vertex i is located to the left of vertex j ;
- the number of non-empty places (called the sequencing length and denoted l) on which all vertices of the graph are placed is minimal.

In the general case, that is, for an arbitrary graph G and width h , this problem is NP-hard, so directional search schemes such as the branch and bound method are used to find the exact solution. Only for four partial cases the exact algorithms of polynomial complexity are found. This is the case when the graph G is directed forest [6] (this algorithm is called the level principle), when $h = 2$ (two algorithms are known, one based on lexicographic order [7], the other one is being under the scrutiny in this work), when the graph G is opposing forest [8] and when the graph G is graph of bounded height [9].

3 Algorithm based on maximal matching

Suppose that we have an arbitrary acyclic directed graph G and $h = 2$. In this case, in order to construct the desired optimal sequencing, one can use an algorithm that was historically first. This method of constructing S is called the algorithm based on the maximum matching.

Algorithm based on maximum matching.

1. For a graph $G(V, U)$, we construct an undirected graph $\bar{G}(V, E)$, where $(i, j) \in E$, if there is no path in the graph G neither from vertex i to j nor from j to i . Such a graph will be referred to as the reachability graph.
2. In the obtained graph $\bar{G}$, we find the maximal matching, which is denoted $M \subseteq E$, that is, a subset of the edges of the maximal cardinality with no vertices in common.
3. In the desired sequencing S^* , we consider all places empty and set $k = 1$.
4. If G is empty, then the algorithm is finished.
5. One of the following cases is possible:
 - (a) Among the open vertices of G , there exists one that does not belong to any of the edges in M , then we choose it to be placed.
 - (b) In the set M there is an edge (i, j) such that vertices i and j are open, then we choose them for the placement and remove (i, j) from M .

- (c) In the set M there is a pair of edges (i, p) and (j, q) such that vertices i and j are open, and between vertices p and q there is an edge in the graph $\overline{G}$, then we choose vertices i and j for placement, edges (i, p) and (j, q) are removed from M , and the edge (p, q) is added to M .
6. We place the selected vertices on the k^{th} place of the sequencing S^* and remove them from the graph G together with the arrows directed from them, if any. Set $G := G$, $k := k + 1$ and go to 4.

It is important to note that polynomial complexity algorithms, such as the blossom algorithm, are known for finding M for arbitrary undirected graphs.

It was proved in [10] that the sequencing obtained by this algorithm is optimal. It is not straightforward to apply this algorithm to arbitrary sequencing width h due to the specificity of operations in step 5 of the algorithm.

4 A class of graphs for which dense sequencing exists

Consider some of the statements and their corollaries, which will narrow down a set of parallel sequencing tasks that need consideration, namely, to the class of graphs for which dense sequencings for an even h exist. This will allow us to generalize the algorithm based on the maximum matching for the case of arbitrary input data.

To the class of graphs for which there are dense sequencings, D_h , belong such graphs that, for a given value of the width h , have sequencings in which at each place there are h vertices.

Theorem 1. If there is an exact algorithm A of polynomial complexity for some fixed sequencing width $\tilde{h}$, then there is a polynomial algorithm for any $h' < \tilde{h}$.

Proof. Suppose that there is some non-empty graph G with n vertices and some fixed value $h' < \tilde{h}$. It is clear that the length of its optimal sequencing l^* is in the range from 1 (if $n \leq h'$ and all vertices are isolated) to n (the graph is a chain) inclusive.

We construct a new graph G' , which is obtained from the graph G by adding $(\tilde{h} - h')$ chains of some length $l \in \overline{1, n}$ to it. We now construct the sequencing S of the graph G' for $h = \tilde{h}$ by Algorithm A . It is known that it will be optimal and can be obtained in polynomial time.

It is clear that if $l = l^*$, then the length of S will also be l^* . On the one hand, it cannot be greater than l^* , since we have $(\tilde{h} - h')$ chains occupying $(\tilde{h} - h')$ positions on each of the l^* places, and vertices, belonging to the graph G can be placed on l^* places in case of $h = h'$. On the other hand, it cannot be smaller than l^* , since the chains can only be placed in a line and their length is l^* . Thus, if we remove from S all vertices belonging to the chains, then we obtain the optimal sequencing of S^* for

the original problem in polynomial time, since the addition and removal of chains can be done in linear time.

Therefore, if we can find the length of chains $l = l^*$, which is generally unknown, in polynomial time, then we can use algorithm A to find the optimal sequencing for $h' < \tilde{h}$. This can be done through binary search.

Note that $l = l^*$ corresponds to the minimum value of l for which there will be $(\tilde{h} - h')$ vertices of the chains on all places in the resulting sequencing S obtained by the algorithm A for the graph G' and $h = \tilde{h}$. Indeed, for $l > l^*$, the length S will coincide with l , since it is necessary to place the chains, and the vertices of the graph G can be placed on a smaller number of places. For $l < l^*$, S will have places where there is less than $(\tilde{h} - h')$ vertices belonging to the chains, otherwise vertices of graph G could be placed at a smaller number of places than l^* for $h = h'$.

We apply binary search to this problem. We choose some integer value of $l_0 \in \overline{a_0, b_0} = \overline{1, n}$, which splits this interval approximately in half. We construct a graph G' , for that value of l . By applying algorithm A we find the optimal sequencing S for $h = \tilde{h}$. If there are $(\tilde{h} - h')$ vertices belonging to the chains on all the places of S , then we choose $l_1 \in \overline{a_1, b_1} = \overline{1, l_0}$, and continue our search using this interval. If S has places where there are less than $(\tilde{h} - h')$ vertices of the chains, then we choose $l_1 \in \overline{a_1, b_1} = \overline{l_0 + 1, n}$ and similarly continue the search. The algorithm convergence follows from the convergence of binary search for monotonic functions. The search requires logarithmic time, the check can be performed in linear time, and therefore we find the $l = l^*$ in polynomial time.

Since all the steps of the described algorithm can be performed in polynomial time, their number is finite, algorithm A has polynomial complexity under the condition of theorem, and thus described algorithm is exact and has polynomial complexity for any $h' < \tilde{h}$. ■

Theorem 2. If there is an exact algorithm A of polynomial complexity for some fixed sequencing width $\tilde{h}$ and graphs from $D_{\tilde{h}}$ (for which exists dense sequencing of width $\tilde{h}$), then there is a polynomial algorithm for an arbitrary graph for $h = \tilde{h}$.

Proof. Suppose there is some non-empty graph G having n vertices. It is clear that the length of its optimal sequencing l^* is in the range from 1 (if $n \leq h'$ and all vertices are isolated) to n (the graph is a chain) inclusive.

We construct a new graph G' , which we obtain from the graph G by adding $m = k * \tilde{h} + r$ isolated vertices to it, where $r = \tilde{h} * \left\lceil \frac{n}{\tilde{h}} \right\rceil - n$, k being some factor. This

factor is not less than 0 (if it is sufficient to add $r < \tilde{h}$ vertices to G to obtain dense sequencing, that is, for graphs for which there are optimal sequencings where there are r free positions). It reaches its greatest value in the case of a graph-chain when

we have $n * (\tilde{h} - 1) = k * \tilde{h} + r = k * \tilde{h} + \tilde{h} * \left\lceil \frac{n}{\tilde{h}} \right\rceil - n \Rightarrow k = n - \left\lceil \frac{n}{\tilde{h}} \right\rceil$. Therefore

$k \in \overline{0, n - \left\lceil \frac{n}{\tilde{h}} \right\rceil}$. Now let us construct a sequencing S of G' for $h = \tilde{h}$ using Algo-

rithm A . It is known that, if there is a dense sequencing for G' , it will be optimal and can be obtained in polynomial time.

It is clear that if $m = l^* * \tilde{h} - n$, then S will be dense and its length will be l^* . On the one hand, it cannot be less than l^* since at least l^* places are required to place the vertices of graph G . On the other hand, it cannot be larger than l^* , since in the optimal sequencing S^* for G there are $l^* * \tilde{h} - n$ positions that can be occupied by isolated vertices, and algorithm A finds the optimal dense sequencing, if any. Then if we remove from S all the added isolated vertices, then we obtain the optimal sequencing of S^* for the original problem in polynomial time, since the addition and removal of isolated vertices can be done in linear time.

Similarly to Proof of Theorem 1, if we can find the number of isolated vertices $m = l^* * \tilde{h} - n$ in polynomial time, or, what is the same, the value of the factor k for which the corresponding value is reached, then we can use algorithm A to find the optimal sequencing for any graph for $h = \tilde{h}$.

Note that $m = l^* * \tilde{h} - n = k * \tilde{h} + r$ corresponds to the minimum value of k , for which the obtained sequencing S by algorithm A for the graph G' and $h = \tilde{h}$ will be dense. Indeed, for smaller values of k , obtained sequencing won't be dense, since not enough isolated vertices have been added. For larger values of k , we also obtain dense sequencing, since we add the number of vertices that is a multiple of $\tilde{h}$.

We apply binary search to this problem. We choose some integer value of

$k_0 \in \overline{a_0, b_0} = \overline{0, n - \left\lceil \frac{n}{\tilde{h}} \right\rceil}$, which splits this interval approximately in half. We construct

a graph G' , for that value of k , then we find the sequencing of S using algorithm A for $h = \tilde{h}$. If S is dense, then we choose, similarly, $k_1 \in \overline{a_1, b_1} = \overline{0, k_0}$, and continue search within this interval. If S has free positions, then we choose

$k_1 \in \overline{a_1, b_1} = \overline{k_0 + 1, n - \left\lceil \frac{n}{\tilde{h}} \right\rceil}$ and similarly continue the search. The algorithm conver-

gence follows from the convergence of binary search for monotonic functions. The search requires logarithmic time, the check can be performed in linear time, so we find the required m in polynomial time.

Since all the steps of the described algorithm can be performed in polynomial time, algorithm A has polynomial complexity under the condition of the theorem, thus we obtained an algorithm that is exact and has polynomial complexity for any graph G and $h = \tilde{h}$. ■

Corollary 1. If there is an exact algorithm A of polynomial complexity for some fixed sequencing width $\tilde{h}$ and graphs from $D_{\tilde{h}}$ (for which exists dense sequencing of width $\tilde{h}$), then there is a polynomial algorithm for arbitrary graph and for any $h' < \tilde{h}$.

Proof. It follows directly from Theorems 1 and 2. ■

This corollary allows us to generalize the algorithms obtained for the class of graphs for which there are dense sequencings, for the case of all graphs and smaller h . In particular, it allows us to apply algorithms obtained for even values of h to smaller odd ones.

5 Generalization of the algorithm based on the maximum matching

This section will discuss the theorems that prove the feasibility of using and developing modifications of the algorithm based on maximal matching to graphs for which there are dense sequencings.

Further research will use the notion of "undirected graph clique" in the classical sense.

Definition. A clique is a subset of vertices of an undirected graph such that every two distinct vertices in the clique are adjacent [11].

Theorem 1. From the existence of dense sequencing for some graph G for sequencing width $h = \tilde{h}$ follows the existence of disjoint cliques of size $\tilde{h}$ in reachability graph, covering all its vertices.

Proof. Suppose that we have a graph G for which the optimal sequencing S^* of length l^* for $h = \tilde{h}$ is dense.

For any place $i \in \overline{1, l^*}$, since the vertices of $S^*[i]$ are in the same place, by the definition of parallel sequencing, it follows that for any two vertices $v_1 \in S^*[i], v_2 \in S^*[i]$ there is no directed path neither from v_1 to v_2 , nor from v_2 to v_1 .

It is known that in the reachability graph $\overline{G}$ corresponding to graph G , two vertices v_1 and v_2 are joined by an edge only when there is no directed path neither from v_1 to v_2 nor from v_2 to v_1 . From the above, it follows that there are edges in $\overline{G}$ between all pairs of vertices $v_1, v_2 \in S^*[i]$, and hence they form a clique of size $\tilde{h}$ in it.

Since the above reasoning is true for all places in the sequencing S^* and all of them has $\tilde{h}$ vertices of graph G , and that sequencing contains all vertices of graph G , this leads to the conclusion of the statement. ■

The above statement is a necessary condition for the existence of dense sequencing for the graph. Note that in the case of $h = 2$, it is sufficient. For $h > 2$ there are graphs for which there are cliques of size h covering all of its vertices, but dense sequencing doesn't exist.

It is known that pairs of vertices are considered in the algorithm based on the maximum matching, so it is more convenient to apply it to even values of $\tilde{h}$. The following statement proves the feasibility of using an algorithm based on maximal matching for this case.

Theorem 2. Among the maximal matchings M of the reachability graph $\overline{G}$ corresponding to the graph G , there are those that give the optimal sequencing when applying the algorithm based on the maximum matching, when the optimal sequencing for an even width $h = \tilde{h}$ is dense.

Proof. Suppose that a graph G is given for which the optimal sequencing S^* of length l^* for $h = \tilde{h}$ is dense.

We construct for it a reachability graph $\overline{G}$. It follows from Theorem 1 that there exists a cover of vertices of the graph $\overline{G}$ by cliques of size $\tilde{h}$. Let's look at an arbitrary clique from this cover and split the vertices that are belonged to it in pairs (we can do this since $\tilde{h}$ is even).

Let us now consider all such pairs obtained from all the cliques forming the cover. These pairs form the maximum matching for the graph $\overline{G}$, since each vertex belongs to only one pair (pairs contain vertices from one clique and cliques are disjoint), in each pair the vertices are adjacent (belong to the same clique), the pairs cover all vertices (pairs cover all vertices of the cliques, and cliques cover all vertices of the graph).

Consider now the algorithm based on the maximum matching, in which we will place as many pairs as possible on each place. If we apply that algorithm, where we use the resulting set of pairs as the maximum matching, then we can get the optimal sequencing by placing pairs from one clique in one place (vertices from one clique correspond to one place in S^*). ■

Note that the validity of the statement does not contradict the fact that for the reachability graph there may be maximal matchings, by using which we will not obtain the optimal sequencing. In fact, such graphs and corresponding matchings exist.

The above statements suggest that the algorithm based on maximal matching can, theoretically, find the optimal solution for problems where the sequencing width is even and the graphs have dense sequencing.

We also see that in its classical form, it is only approximate to this class of sequencing problems.

6 Computational experiment

To experimentally determine the accuracy of an algorithm based on the maximum matching, it was implemented according to algorithm discussed in section 3. This algorithm is further referred to as the classical algorithm based on the maximum matching. Preference was given to vertices with smaller labels, rather than random choice. This will not affect the generality of the results since the randomness of the selection is implemented by the randomness of the graph generation and the labeling

of their vertices. Also, to generalize the algorithm for the case of width $h > 2$ step 5 of the algorithm is repeated until free positions are exhausted or until we can't choose the next pair or single vertex.

During the experiment, only graphs were generated for which there are dense sequencings for even values of h . Their advantages include the fact that for such graphs the exact solution is known by generation. Transitive arrows are not removed from graphs because their presence does not affect the algorithm (they do not affect the reachability graph). The labeling of vertices in the graph is random to implement the randomness of the selection of pairs and vertices that do not belong to the pair.

In this and subsequent experiments, the accuracy of the algorithm based on the maximum matching was verified.

All of the experimental conditions are summarized in Table 1.

Table 1. - Values of parameters used in testing

n	h	Number of tests
[10,20]	{4,6,8,10}	10000
[21,40]	{4,6,8,10}	10000
[41,60]	{4,6,8,10}	10000
[61,100]	{4,6,8,10}	10000

The results of evaluating the accuracy of the classical algorithm based on the maximum matching are shown in Table 2. These results and the following contain three parameters: the number of cases when the algorithm based on the maximum matching found the exact solution; the average deviation of the length of the obtained sequencing from the length of the optimal one, and the number of cases in which the resulting sequencing is twice as long as the optimal sequencing. In this one and following tables AMM stands for algorithm based on maximal matching.

Table 2. - Experimental results for the classical algorithm

Classical algorithm		
AMM is accurate	Average deviation of AMM	Obtained sequencing two times longer than optimal
8688	1.160823	9
5718	1.562121	19
2564	2.165008	4
761	3.391384	0

Test results show that it has very low accuracy. The number of times it gives the optimal solution decreases rapidly. The average deviation of the obtained solutions also grows very fast and differs significantly from the one even for small graphs (21-40 vertices). In addition, there are cases where the sequencing obtained is twice as long as the optimal one, which exceeds the maximum accuracy estimate in [12]. Note that this does not contradict the statement, since the algorithm can sometimes leave

empty positions in sequencing when open vertices are still present. Such behavior is a case when they belong to pairs that cannot be placed in the current step.

From this, we can conclude that as the number of vertices in the graph increases, the number of maximal matchings that give suboptimal solutions increases faster than the number of matchings that give optimal ones.

It is known that the resulting matching is strongly depends on the labels of vertices in the graph. One can assume that if we can relabel the vertices so that algorithm is more likely to find matchings that give optimal sequencing, then we can improve its results. It is also known that the vast majority of optimal solutions for graphs are satisfying the level principle. Therefore, it is natural to expect that if we relabel the vertices of the graph according to the level principle, then we can increase the likelihood of getting a matching, which gives an optimal solution. The results of the corresponding experiment are shown in Table 3.

Table 3. - Experimental results for the classical algorithm with the relabeling of the vertices

Classical algorithm with relabeling		
AMM is accurate	Average deviation of AMM	Obtained sequencing two times longer than optimal
9290	1.083099	0
6485	1.317496	0
3078	1.684051	0
943	2.509551	0

From the experiment we observe that such relabeling slows down the decrease of accuracy of the algorithm and significantly slows the growth of the average deviation. There were no cases where sequencing that are twice as long as optimal were obtained, but the hypothetical possibility of their existence remains.

The sequencings that are twice as long as the optimal one are obtained, since the algorithm can leave empty positions because the open vertices are bound in pairs that cannot be placed. This behavior of the algorithm further aggravates the impact of the initial matching. An obvious way to reduce this effect is to allow splitting the pair if it cannot be placed in the current step, but at least one of its vertices is open. The resulting algorithm is later referred to as modified algorithm based on the maximum matching. The results of the accuracy evaluation of this algorithm are shown in Table 4.

Table 4. - Experimental results for the modified algorithm

Modified algorithm		
AMM is accurate	Average deviation of AMM	Obtained sequencing two times longer than optimal
9031	1.008256	0
8122	1.06869	0
7387	1.166475	0
6418	1.341709	0

From the results we see that such modification significantly improves the accuracy of the algorithm. The number of cases where the algorithm was accurate for graphs with 61-100 vertices increased almost 8.5 times, for other cases this number also increased significantly. The maximum average deviation from the previous experiment has halved and does not exceed 1.5. As expected there were no cases where the length of the resulting solution was twice the length of the exact one.

Note also that despite the smaller average deviation for the case of graphs with 10-20 vertices compared to the previous experiment, the number of cases where the algorithm is accurate is smaller. Therefore, such modification does not always effectively comply with the level principle. Based on the previous results, it is hypothesized that if we apply the modified algorithm to previously relabeled vertices according to the level principle, we will increase accuracy even further. The results obtained are shown in Table 5.

Table 5. - Results of the experiment for the modified algorithm with relabeling of the vertices

Modified algorithm with relabeling		
AMM is accurate	Average deviation of AMM	Obtained sequencing two times longer than optimal
9675	1.006154	0
8913	1.022079	0
8134	1.107181	0
7491	1.229175	0

We see that a significant improvement in performance of the algorithm is achieved. The number of cases where the algorithm was accurate for graphs with 61-100 vertices increased almost 10 times, compared to the classical algorithm. The average deviation for the first three cases is smaller than the average deviation for the first case for the classical algorithm and for the fourth case is comparable to it.

All this indicates the feasibility of introducing modifications to the algorithm. Therefore, the effect of the initial matching on the result obtained is much smaller. Note that the algorithm did not lose the ability to violate the level principle.

It can be assumed that if we apply, in addition to algorithms based on the level principle, such a modified algorithm, we can further improve the overall results, first of all due to cases in which to obtain optimal sequencing it is necessary to violate the level principle. To verify the latter assumption, an experiment was performed comparing the accuracy of a modified maximum matching algorithm with the best of results of lexicographic and double-label algorithms. The latter further along referred to as combined algorithm (CA). The same parameters were used as for the previous experiments.

The results consist of two parts: comparison of the algorithms with each other and comparison of the accuracy of the algorithms by the values of the objective function. A comparison of the algorithms with each other contains the number of test cases in which sequencing with a shorter length is obtained by the AMM; in which a sequencing with a shorter length is obtained by the combined algorithm, and in which sequencings of equal length are obtained by the algorithms. A comparison of the accu-

racy of the algorithms contains the number of cases when an exact solution was obtained by the algorithm based on the maximum matching; when an exact solution is obtained by the CA; the number of cases when one of the algorithms was accurate and the average deviation of the length of the solutions obtained by the algorithms from the length of the optimal one.

The results of the experiment are shown in Table 6.

Table 6. - Results of comparison of accuracy of the modified relabeling algorithm and the combined algorithm

Modified algorithm with relabeling and combined algorithm				
Sequencing obtained by AMM is shorter		Obtained sequencings have the same length		Sequencing obtained by CA is shorter
5		9673		322
30		8954		1016
50		8275		1675
73		7547		2380
AMM is accurate	CA is accurate	One of them is accurate	Average deviation of AMM	Average deviation of CA
9675	9992	9997	1.02508	1
8923	9906	9936	1.02507	1
8206	9815	9865	1.098105	1
7436	9674	9747	1.232059	1

The results obtained confirm that, even with all modifications, the algorithm based on maximal matching is significantly inferior to the algorithms based on the level principle on all indicators. However, we see that the number of cases where the algorithm based on maximal matching finds shorter sequencing increases. In addition, in all these cases, these sequencings were accurate (this can be seen by summing up the number of AMM wins and the number of cases where the CA was accurate and comparing it to the number of cases when one of the algorithms is accurate).

On the other hand, the number of cases when the combined algorithm gave shorter but not accurate sequencing increases with the number of vertices. This may indicate that there is some "barrier" that prevents the combined algorithm from finding the optimal solutions, and the more vertices there are in the graph the more noticeable it is. This barrier may be a level principle, more precisely the necessity to violate it.

Let's return to the idea that the labeling of vertices in the graph strongly influences the result of the algorithm. The experiments with relabeling, according to the level principle, showed the effectiveness of this approach. It can be assumed that, if we randomly relabel the vertices, we will get a new matching and a new resultant sequencing, which may be shorter than the original one. The more times we repeat this procedure, the more likely it is that one of the resulting sequencings will be optimal.

To test this hypothesis, an experiment was conducted in which the algorithm based on the maximum matching was applied to the generated graph, and if it gave a suboptimal solution, then the vertices of the graph were randomly relabeled and the algorithm was applied again. The described steps are repeated until the algorithm finds the

optimal sequencing or until the maximum number of repetitions is exhausted. We call this modification a random algorithm based on maximal matching. In graphs generation the same parameter values as in the previous experiments were used; the maximum number of repetitions for all cases is set to 10.

The results of checking the accuracy of a random classical algorithm based on the maximum matching are given in Table 7. These results contain the same columns as the previous experiments.

Table 7. - Experimental results for the random classical algorithm

Random classical algorithm		
AMM is accurate	Average deviation of AMM	Obtained sequencing two times longer than optimal
9946	1	0
9277	1.034578	0
7088	1.205357	0
3543	1.810129	0

From the results of the experiment we can see that the hypothesis finds its confirmation. Thus, the number of cases where the algorithm was accurate increased significantly, in particular for graphs with 10-20 vertices got almost perfect accuracy, and for the case of graphs with 61-100 vertices the accuracy increased almost 5 times compared to the classical algorithm. The average deviation significantly decreased. There were no cases where the length of the resulting solution exceeds twice the length of the exact one, which may be due to the fact that the matchings that give rise to them are very rare.

A similar experiment was performed for the modified algorithm. The results obtained are summarized in Table 8.

Table 8. - Experiment results for the random modified algorithm

Random modified algorithm		
AMM is accurate	Average deviation of AMM	Obtained sequencing two times longer than optimal
9950	1	0
9834	1	0
9600	1.005	0
9145	1.037427	0

In the experiment obtained results similar to the previous ones. The number of cases where the modified algorithm was accurate exceeds 91% for all graph sizes. For graphs with the number of vertices from 61 to 100 it is 12 times better than the classical algorithm. Similarly, the average deviation decreased.

The results obtained in the last experiment are comparable to those of the algorithms based on the level principle, which confirms the effectiveness of the proposed approach, especially considering that the chosen maximum number of repetitions was relatively small. Experiments with the relabeling of vertices, according to the level

principle, are not listed, because it does not qualitatively affect the operation of the algorithm and therefore its accuracy.

Considering the improvement obtained, the last algorithm was compared with the combined algorithm. The results can be seen in Table 9.

Table 9. - Accuracy comparison results for the random modified algorithm and the combined algorithm

Random modified algorithm and combined algorithm				
Sequencing obtained by AMM is shorter		Obtained sequencings have the same length		Sequencing obtained by CA is shorter
4		9957		39
59		9795		146
132		9518		350
195		9085		720
AMM is accurate	CA is accurate	One of them is accurate	Average deviation of AMM	Average deviation of CA
9960	9995	9999	1	1
9841	9928	9987	1	1
9599	9816	9948	1.014963	1
9160	9674	9867	1.041667	1.009202

From the results we observe that the scores of the random modified algorithm are much higher than the scores of the modified one. The number of cases when it and the combined one give the sequencing of same length exceeds 90%. The number of cases where the sequencing obtained by algorithm has a smaller length is also much larger and increases rapidly with the number of vertices in the graph. In contrast to the previous comparison, there are cases, when the algorithm found shorter sequencings, but they were not optimal. The number of cases in which at least one of the algorithms is accurate for graphs of all sizes exceeds 98.6%, which significantly improves the results when using only algorithms based on the level principle. From this we can conclude that the resulting algorithm not only effectively comply with the level principle, but also violates it if necessary.

All previous results further confirm the effectiveness of random algorithms based on maximal matching, as well as the importance and prospectivity of the modifications of the classical algorithm as such that may violate the level principle.

The disadvantages of the random algorithm compared to the algorithm with re-labeling, according to the level principle, include its non-determinism, which greatly complicates the analytical study of its properties.

The results obtained in this section have practically proved the feasibility of using the algorithm based on maximal matching and its modifications to the problems of optimal sequencing with graphs for which there are dense sequencings and even width.

7 Conclusion

The well-known class of discrete optimization problems (which are formulated as optimization problems on graphs), and requiring the development of new effective approximate polynomial algorithms, was investigated.

The theorems allowing to reduce the problem of optimal sequencing with arbitrary graph and sequencing width to the problem with graph, for which exists dense sequencing, and even width, have been proved. It was substantiated that to solve problems from this class it is suitable to apply an algorithm based on maximum matching.

For the algorithm based on maximum matching, several modifications have been developed. Experimental verification has established that they significantly increase its accuracy. In addition, the joint application of the proposed algorithms and known algorithms based on the level principle resulted in an accuracy exceeding 98%.

8 References

1. Garey, M., Johnson, D.: Computers and intractability. W.H. Freeman and Company, New York (2009).
2. Prot, D., Bellenguez-Morineau, O.: A survey on how the structure of precedence constraints may change the complexity class of scheduling problems. *Journal of Scheduling*. 21, 3-16 (2017).
3. Tanaev, V.S., Kovalyov, M.Y., Shafransky, Y.M.: Scheduling theory. Group Technologies, Minsk, Institute of Technical Cybernetics NAN of Belarus (1998). (in Russian)
4. Pinedo, M.: Scheduling. Theory, Algorithms, and Systems. (2016).
5. Brucker, P.: Scheduling algorithms. Springer, Berlin (2007).
6. Hu, T.: Parallel Sequencing and Assembly Line Problems. *Operations Research*. 9, 841-848 (1961).
7. Coffman, E., Graham, R.: Optimal scheduling for two-processor systems. *Acta Informatica*. 1, 200-213 (1972).
8. Garey, M., Johnson, D., Tarjan, R., Yannakakis, M.: Scheduling Opposing Forests. *SIAM Journal on Algebraic Discrete Methods*. 4, 72-93 (1983).
9. Dolev, D., Warmuth, M.: Scheduling precedence graphs of bounded height. *Journal of Algorithms*. 5, 48-59 (1984).
10. Fujii, M., Kasami, T., Ninomiya, K.: Optimal Sequencing of Two Equivalent Processors. *SIAM Journal on Applied Mathematics*. 17, 784-789 (1969).
11. Luce, R., Perry, A.: A method of matrix analysis of group structure. *Psychometrika*. 14, 95-116 (1949).
12. Bruno, J., Coffman, E.: Computer and job-shop scheduling theory. Wiley, New York (1976).

Algorithm for fixing singular defects of polygon meshes based on Half-Edge Data structure

Danylo Shovhelia^[0000-0003-0571-0899], Natalya Sokolova^{✉[0000-0002-2207-1884]}

Oles Honchar Dnipro National University, Gagarina av., 72, Dnipro, 49005, Ukraine
dshovhelia@gmail.com, n.olegowna@gmail.com

Abstract. The problem of the occurrence of singular defects in the creation of 3D models in polygonal meshes due to the peculiarities of the Half-Edge Data structure is considered. An algorithm for detecting defects in a model based on half-edge structures and mesh recovery is proposed. Software implementation of the method allows to create a model based on the source data without the formation of new defects. Appropriate tools have been developed for benchmarking, and results are presented.

Keywords. 3D printing, 3D models, polygonal mesh, singular defect, Half-Edge Data Structure, non-manifold elements.

1 Introduction

Additive Technologies (for example, three-dimensional or 3D printing) is a form of additive manufacturing technology where a three-dimensional object is created by imposing consecutive layers of material (printing, growing) according to the digital model. Digital 3D models are key components of this technology, and each specific program that deals with a 3D geometry has its quality requirements that limit the class of acceptable and supported models. In most cases, visualization is just one of many steps that make up the life cycle of a digital 3D model. 3D models need to be analyzed and processed using advanced algorithms, which usually have strict requirements for the quality and integrity of their input data. Adapting imperfect 3D models to these requirements is important.

Polygonal and triangular meshes are now a de facto standard in most 3D modeling areas. Polygonal meshes are directly supported by accelerated graphics equipment, which facilitated their expansion and use [1-4]. Most graphic formats commonly used for 3D model sharing (OFF, VRML, PLY) encode the surface mesh using indexed sets of faces. However, these file formats do not guarantee the proper simplicity complex, since they can easily encode non-manifolds and/or non-oriented polygons, isolated elements, and several other elements that are often the source of problems.

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

2 Formal problem statement

3D printing is one of the fastest growing technologies that captured many spheres of human life and has become their indispensable core, in particular, such as architecture, medicine and mechanical engineering. Models is a basis for 3D printing. Primitive models cannot display the whole variety of printed products, and modeling of more complex models based on a simplest one can cause a number of defects. The latter, may lead to models that could not exist in the real world. The urgent issue is the timely detection of such defects, development of algorithms for model fixation, methods restoration of polygonal meshes and their software implementation in existing tools.

3 Literature review

3.1 Defects in the representation of three-dimensional models

Specialized computer-aided design (CAD) systems are used to model the objects, which are based on a particular internal structure of the model representation. Move an object from one system to another or changing their representation can be completely unpredictable, which leads to the appearance of topological defects (defects of local connections) and consequently to geometric defects [1,5], without any possibility of their identification. Three-dimensional scanning of real-world objects can be accompanied by many defects, which are for the most part classified as geometric [6]. Such situations are unacceptable because they lead to a violation of the integrity of the printed object.

1. Local connectivity where a set of triangular polygons encoded in a file does not form a combinatorial manifold simplex complex. Correcting such problems depends largely on learning the available input geometry and most often involves creating new geometry. The following defects include [1,5]:

Isolated Vertices where a vertex that is not an element of any other simplex of a mesh. Such vertices can often be simply ignored and if necessary, be removed very easily. Many existing tools provide manual or automatic procedures for this.

Dangling Edges where edges that do not have incident triangles. A common strategy is to simply ignore or remove the edges. Edges may also be used by some approaches as a useful source of information about the assumed basic geometry to recover.

Singular (complex, non-manifold) Edges where more than two polygons have a common edge. Since the condition of the combinatorial set is required in several application contexts, there is a strong call for solutions that transform such a mesh into a combinatorial manifold. This problem is not as easy to solve as in a case of isolated vertices or edges because of the inherent ambiguity.

Singular Vertices where the vertex is not manifold in the topology of the abstract simplicial complex. Detecting such vertices is more difficult than detecting singular

edges. If it is very important to calculate the number of incident triangles for the edges, then it is necessary to calculate the number of connected components in the vicinity for the vertices. Solutions to this kind of problem are usually based on the duplication of one vertex, after which each component of the neighborhood should be re-assigned to one of the copies.

2. Global Topology (general topological characteristics of the surface, including the number connected components, the genus, the number of cavities and orientation) [1,5]:

Topological Noise [7] represents a situation when reconstructing a surface, starting with point clouds or when removing an isosurface from 3D images, tiny "tunnels" that were not present in the original object are introduced into the constructed digital model due to the effects of overlay or discrete noise basic data.

Inconsistent Orientation where the order of the sequences of vertex indices indicates the orientation of the polygon. To ensure full visibility across all imaging systems, it is necessary to provide a unique consistent orientation for all mesh polygons by selecting a seed surface and extending the orientation to adjacent faces. However, some configurations are essentially not oriented, which means that, the system must necessarily cut the surface for consistent orientation.

3. Geometric defects where the geometric realization of an abstract simplicial complex is considered. Geometric defects also relate to the characterized position of the vertices. This family is further complicated by the need to handle continuous coordinates and intersections, which (for efficiency reasons) must be expressed and processed with finite precision arithmetic, which usually leads to reliability problems. The geometric effects include the following [1,5]:

Surface Holes and Gaps. When scanning an object using standard scanners and designing a surface using standard CAD systems, adjacent patches are separated by unwanted holes due to the features of the equipment and displacement of tessellation patches, which must be compensated. Such steps are known as closing the gap (the area between two triangulated surface spots that must be permanently connected but not associated with the gap) and filling the holes (undesirably missing part of the surface in the triangulated patch). The main difference between the two cases is the connection of their boundaries. The boundary of the gap usually consists of two (or more) disconnected chains of edges. The border of the hole usually consists of one or more closed edge loops.

Degenerate Elements. Degenerate triangles are zero-area triangles, and they are a source of many problems for many applications, since many useful objects (ordinary vectors, circles, bar-centric coordinates) cannot be calculated on such triangles. Applications that specialize in calculating the above objects (for example, finite element analysis or Delaunay refinement implemented in the freely available Triangle package [5,8]) fail when the mesh contains degenerate triangles or becomes unstable when it contains nearly degenerate triangles.

Self-intersections. In several application contexts, it is assumed that the input mesh represents the boundary of some solid volume, and therefore cannot self-intersect. Although it is relatively easy to test the mesh for self-intersections, solving them is a difficult problem due to inherent ambiguities. Self-intersecting meshes are typically generated by multi-patch mosaics, mesh deformation, assembly of many parts without care, or by combining trays reconstructed from partial scanning of a 3D object. Due to ambiguity, there is no common strategy for solving this problem.

Sharp Feature Chamfering. Most restoration and contouring methods limit each sample or vertex to a specific line or curve, the position of which is completely determined by the pre-set pattern. In most cases, such a pattern cannot be adjusted so that it coincides with the sharp edges and angles of the model and, therefore, virtually none of the specimens rests on such sharp features. This leads to the imposition of artifacts, where sharp edges and angles of the original shape are removed by the sampling process and replaced by unevenly triangulated chamfers, which in turn leads to poor visualization and high distortion;

Data Noise. Each scan tool has a finite precision. Thus, the output of the sample model contains additive noise from different sources. The main problem is to remove noise while maintaining the morphology of the main sampling surface, with particular attention to high-frequency details such as angles, edges, or other sharp features.

3.2 Methods of mesh repairing

Non-manifold meshes, which contain singular edges and vertices, can be classified into two classes: those that bind so-called regular sets and those that do not have regular sets [5, 9]. The former are still well defined by the continuous volume, down to the singular contacts on the non-manifold edges and vertices. Singular edges and vertices are often deliberately created to avoid duplication of vertices and for the continued need to hold these duplicate vertices sequentially. In a general case (for example, meshes containing singular edges with an odd number of incident faces), which usually has ambiguities, specialized or global methods are required.

The two algorithms proposed by Guezic et al. [10, 11] transform meshes that restrict non-manifold regular sets into sets of combinatorially manifold meshes. Strictly speaking, a closed-mesh mesh can be a regular set only if it has no self-intersections. However, they are only bonded and therefore can also handle self-intersecting meshes. After identifying a singular edge having $2n$ incident faces, it splits into n multi-row edges having 2 faces each. Such representation of marginal diversity may still contain separate vertices that must be identified and duplicated.

Rossignac and Cardoze [12] proposed a strategy for performing the minimum number of such duplications. The approach introduces additional operations to merge the edge edges of the mesh section along singular edges or by simply joining the border (clamping) or stitching adjacent curves, and thus reducing the number of locking connected components.

Extensions of these works have been studied for higher dimensions by De Floriani et al. [13-16], who propose to retain some of the innocuous features as long as the model remains a initial quasi-manifold model, which is a weaker condition than a

manifold one. For the individual case of three dimensions, Attene [1] performed a detailed analysis of existing algorithms of restoring all types of defects for polygonal meshes, according to efficiency and proposed two algorithms for removing features from tetrahedral meshes. One algorithm offers combinatorial manifolds and the other prioritizes geometry. Any non-degenerate mesh that restricts regular recruitment can be tetrahedroned and processed.

Wang Zengbo [17] presents algorithms for the rapid recovery of triangular meshes imported from STL files.

Jixin Tan and Jianxun Chen [18] describe generalized algorithms for eliminating singular defects. The latter approach involves splitting models connected by singular vertices or complex edges into two independent objects.

However, it should be kept in mind that due to topology, a half-edge data structure, does not allow the creation of singular defects, which leads to the formation of defects of another type and the inability to determine the initial ones.

3.3 Typical structures for description three-dimensional models

Objects created using polygonal meshes should store different types of elements such as vertices, edges, faces, polygons and surfaces. Polygonal meshes can be represented in a variety of ways using different methods of storing vertices, edges, and faces. These include the following [5]:

- List of faces: description of faces is made by pointers to the list of vertices.
- Winged View: Each point of an edge points to two vertices, two faces, and four (clockwise and anti-clockwise) edges to which it belongs. Large memory is required for storing mesh description data.
- Half-edged meshes: the method is similar to a winged view, except that only half of the edge is traversed.
- Four-edged meshes that hold edges, half-edges, and vertices without any indication of polygons. Polygons are not explicitly expressed in the description, and can be calculated bypassing the structure. Memory requirements are similar to half-edge meshes.
- A table of angles that stores vertices in a predefined table, and the bypass of the table implicitly specifies polygons. Table of angels do not feed the mesh completely. Most meshes require multiple corner tables (triangles).
- Vertex description: Only vertices that point to other vertices are represented. The edge and edge information are not explicit in this view. However, the simplicity of the presentation allows to perform many effective operations over the mesh.

Each presentation has its advantages and disadvantages.

The choice of data structure is determined by the application, required performance, size of the data, and operations that to be performed. Compact, simple structures are required for hardware rendering. Low-end APIs such as DirectX and OpenGL usually include a table of angles (triangles).

Half-Edge Data structure is based on a principle of splitting an edge into two multidirectional halves. Each of these halves is a basic element and stores a topology in itself, which allows to manipulate a polygonal mesh. According to the definition of

this structure, a model should not contain singular vertices and/or complex edges, since creating a connection between such elements is impossible with a basic interpretation [5].

Singular defects often occur in the following situations:

1. The process of folding vertices.
2. Boolean operations on objects.
3. Errors of tessellation algorithms (faceting of a solid-state model).
4. Changing the structure of the presentation from a more primitive (vertex representation) to a more complex (half-edge data structure).
5. Duplicate elimination in trivial data structures.

One common representation of data structures in CAD systems is the spreadsheet, which has advantages in terms of memory savings and ease of rendering. Such an implementation is presented in OpenMesh [19]. However, due to the lack of connection in such a structure, it can contain defects of any type, and attempting to create such a model in a half-edge structure can lead to its destruction. Each polygon consists of half-edges and contains so-called pointers to half-edges from the general table structure.

4 An algorithm for creating a polygon model in OpenMesh

The development team of OpenMesh uses the following algorithm to create a polygon model:

1. For a set of triple coordinates:
 - 1.1. Add three coordinates to the structure.
 - 1.2. Return the unique triangle ID.
2. From a set of identifiers to create a polygon:
 - 2.1. For a pair of vertex ID, check that there is a half-edge associated with that vertex, if so proceed to 2.4.
 - 2.2. If a half-edge has not yet been created, then create one and add it to the face.
 - 2.3. Select the next pair of vertices, if any, and return to item 2.1, otherwise go to item 2.7.
 - 2.4. Check whether a given half-edge belongs to the face. If not, return to 2.2.
 - 2.5. Check for the opposite half-edge of the face. If not, return to item 2.2.
 - 2.6. Record an attempt to attach a new polygon to a pair of connected polygons or separate it completely.
 - 2.7. Add the triangle ID to the data structure.
3. Return the generated polygon ID.

The disadvantage of this algorithm is that it is capable of replacing one defect with another. Instead of creating a singular edge, the resulting model will contain a hole and an island polygon. Such polygons can be evaluated as garbage and removed along with isolated vertices and hanging edges. In the end, further processing of such a model may lead to unexpected results.

It is important to consider that the sequence of polygon creation plays a large role in this algorithm.

5 Modified algorithm for model polygon creation

One of the peculiarities of implementing a half-edge data structure in Open-Mesh is the presence of edges that allow iterating over pairs of half-edges, although the edges themselves do not exist. If the general method requires counting the falling edges of a selected edge, is it possible to create an edge that will not be a structural unit of the model but will retain a link between all contenders? In other words, it is a bridge that connects all contenders for a given edge. It is an edge-bridge that stores the pointers of each half-edge attached to it.

Suppose that there is an edge between two points along which two half-edges pass (Fig. 1). In that case, the edge knows about its pair of half-edges, or a contender container consists of 2 half-edges. In this case, when trying to add one more half-edge, ie, to combine more than two polygons in one edge, the outer half-edge of the attached polygon will be added to the claim container (Fig. 2).

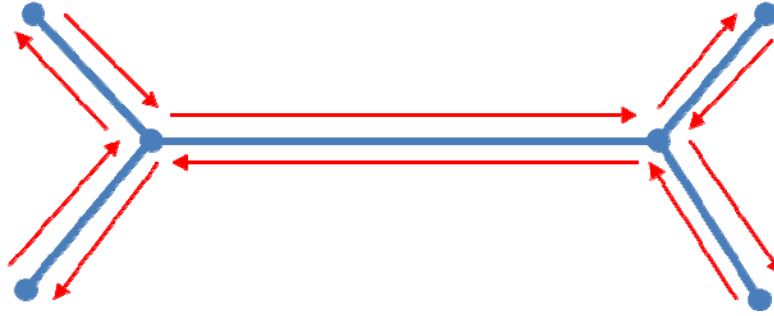


Fig. 1. Modified representation of edges

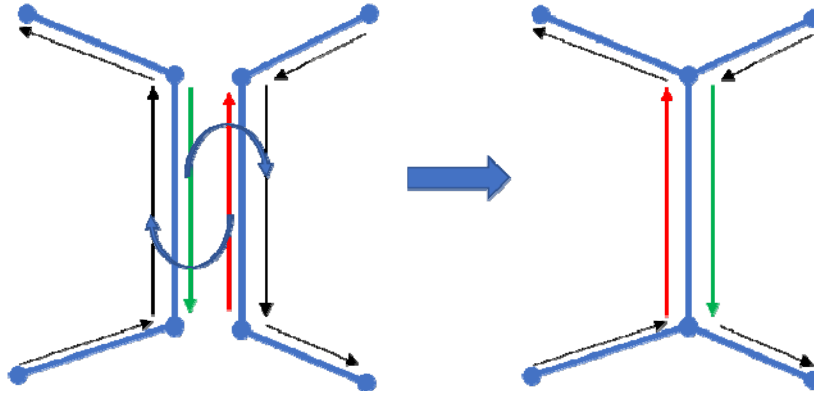


Fig. 2. Bonding of pairs

Given the fact that each half-edge can have only one partner, in that case, every $2n$ contender will form n pairs. From all of the above, the following changes can be made to the algorithm of creating a polygon:

1. For a set of triple coordinates:
 - 1.1. Add three coordinates to the structure.
 - 1.2. Return a unique vertex ID.
2. Create a polygon from a set of vertex identifiers:
 - 2.1. For a pair of vertex identifiers, to check if there is a half-edge associated with that vertex, if so proceed to 2.4.
 - 2.2. If a half-edge has not yet been created, then create one and add it to the face.
 - 2.3. Select the next pair of vertices, if any, and go to item 2.1, otherwise go to item 2.7.
 - 2.4. Check whether the given half-edge belongs to the face: if not, go to item 2.2.
 - 2.5. Check if the opposite half-edge contains the faces: if not, go to item 2.2.
 - 2.6. Invite a container of contenders for a given half-edge.
 - 2.7. If the number of contenders is not even, then find a pair by method of elimination of already created pairs; otherwise, add the ID to a container. Go back to item 2.3.
 - 2.8. Add the triangle ID to the data structure.
3. Return the generated polygon ID.

6 Implementation of the modified algorithm

Object-oriented bridge implementation based on modified algorithm and implementation of this bridge in OpenMesh allows to eliminate defects in models based on half-edge data structure [20]. Many algorithms for preliminary analysis of meshes, at the stage of determining the integrity of the model, as well as cleaning it from debris, can lead to holes.

Let's look at the example of building an OpenMesh model. The original model was created in Solid Works 2018 SP3 (Fig.3,a). As can be seen in a figure, the model has stiffeners. The model loaded as a Triangle Soup model does not carry any information other than visual shape and cannot be used for 3D printing (Fig.3,b.).

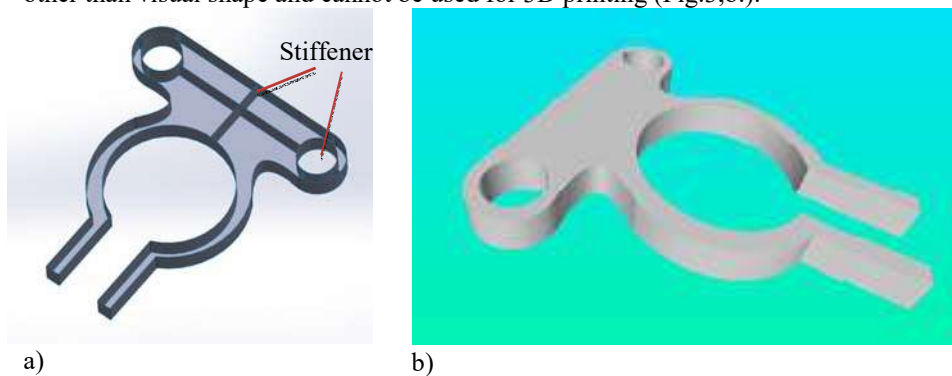


Fig. 3. Model # 1. a) Created in Solid Works 2018 SP3; b) 3D printing model loaded as Triangle Soup

When trying to create a model using OpenMesh, defects arise that are unacceptable with 3D printing (Fig.4).

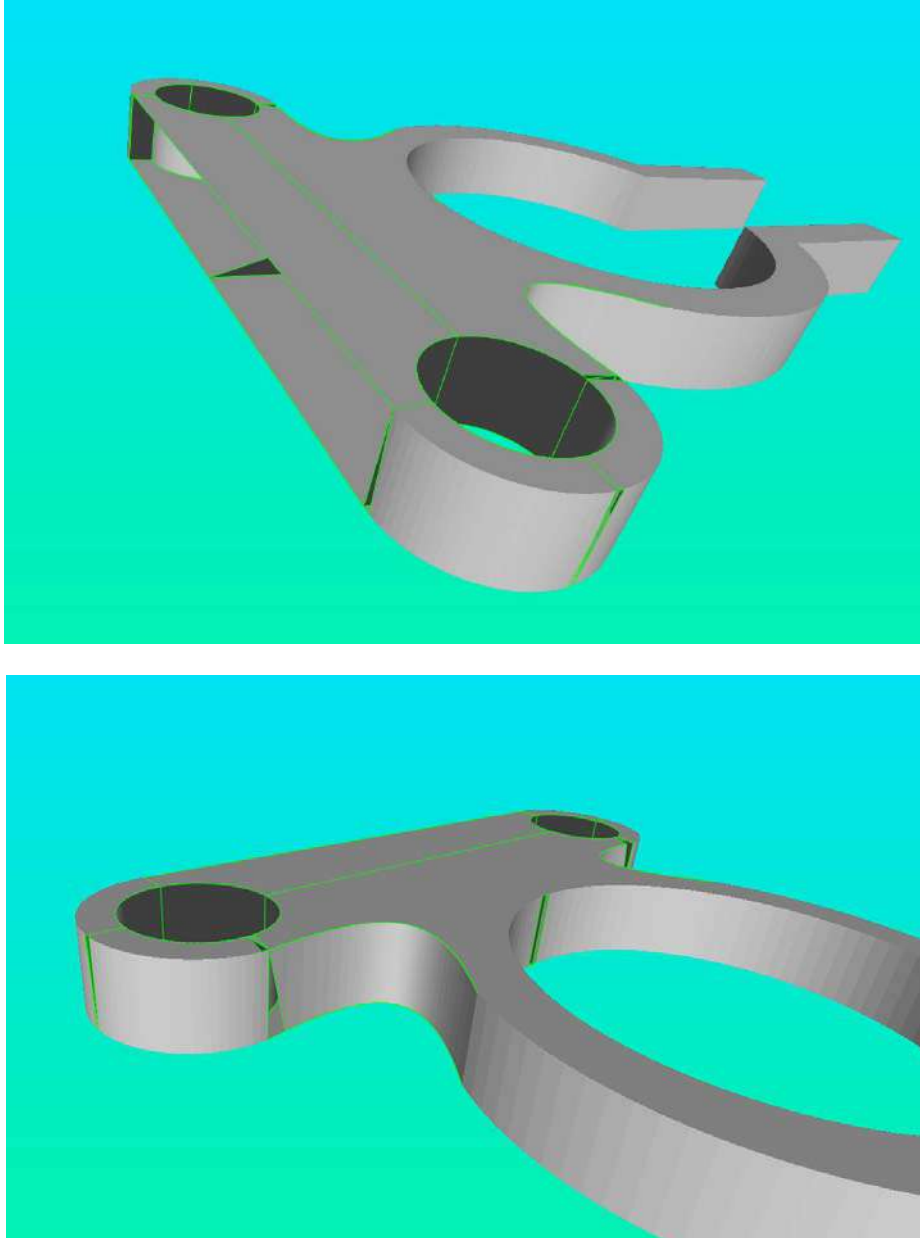


Fig. 4. A model that uses a basic Half-Edge Data Structure OpenMesh implementation.

This model is non-manifold and contains two types of defects (Singular Edges & Singular Vertices) created by Stiffener. The holes are highlighted in green (determined by the boundary edges). Because the base implementation does not support communication between more than two faces through one edge, all subsequent "contenders" for this edge will be disconnected with duplication of the corresponding points. The final analysis of a loaded model shows that the number of objects is greater than one, which is also inadmissible in some environments.

Using the modified algorithm based on the preservation of the connection between non-manifold elements gives the following result (Fig.5):

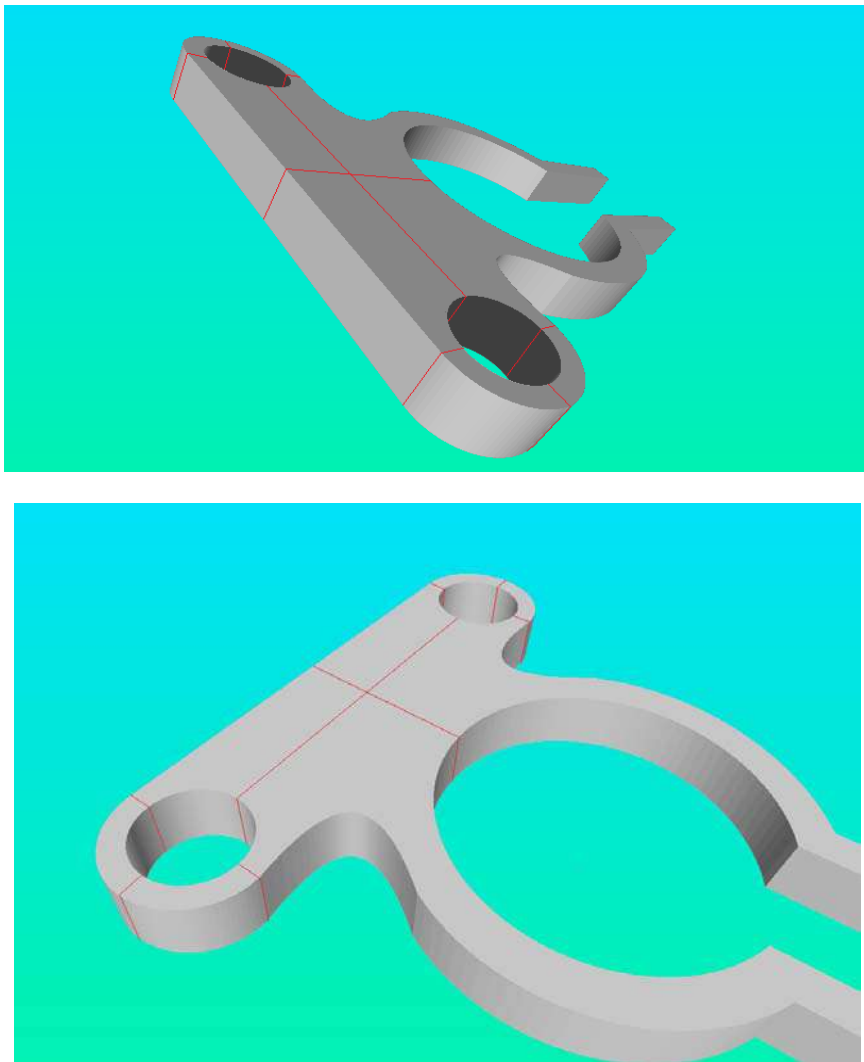


Fig. 5. Implementation of the approach with keeping the connection between non-manifold elements (Singular Edges are highlighted in red)

Another illustration compare the implementation of basic and modified algorithms.

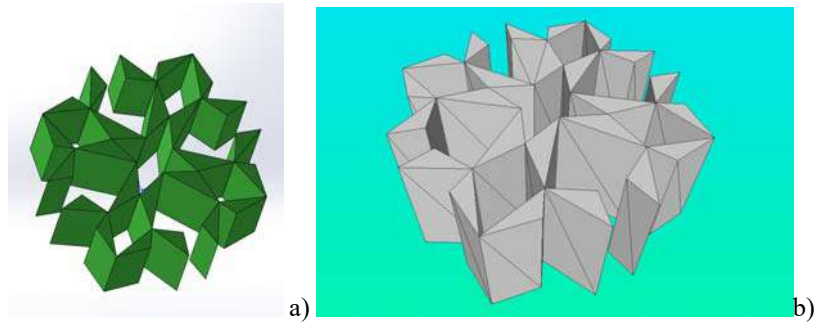


Fig. 6. Model # 2. a) Created in Solid Works 2018 SP3; b) Model for 3D printing, loaded as a Triangle Soup.

This model consists of 24 triangular prisms. In the absence of communication between all the elements, it can be exploded into 24 separate parts respectively (Fig.7).

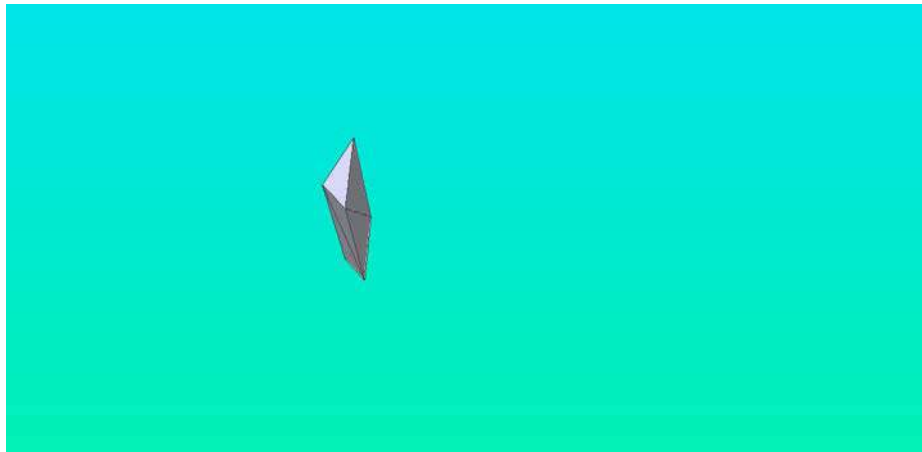


Fig. 7. Splitting the model into separate prisms in absence of communication between the elements

One solution to this problem is to replace singular defects with separate faces. After the orphaned triangles are removed, the result shown in Fig.8.

Based on the specifics of the model, about 1/8 faces will be removed. After loading this kind of object, it becomes impossible to trace the original source of data loss.

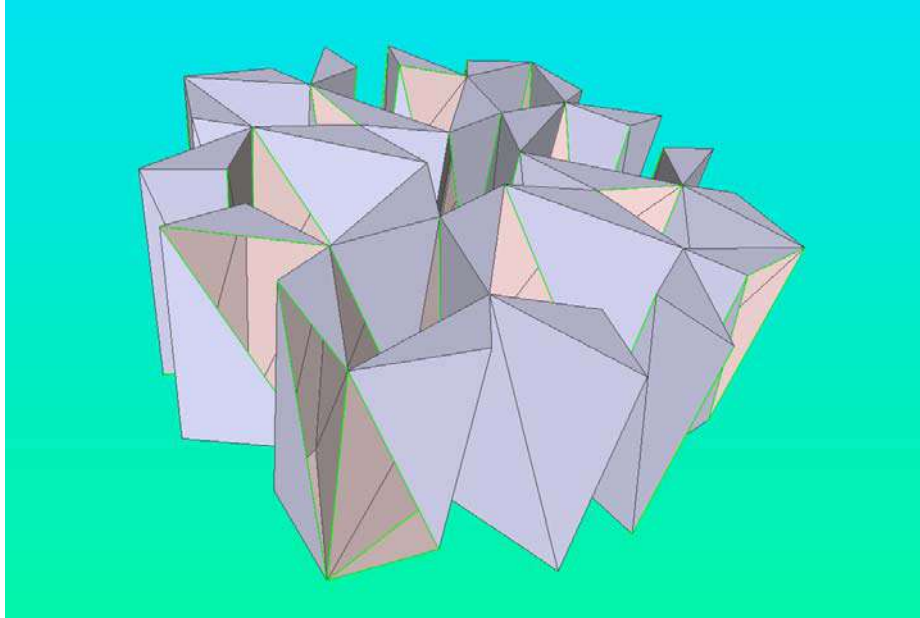


Fig. 8. Substitution of Singular Defects by Separate Edges (highlighted in green the boundary edges)

Using the algorithm of preserving the connection between the non-manifold faces and the software implementation of the bridge, the following results (Fig.9):

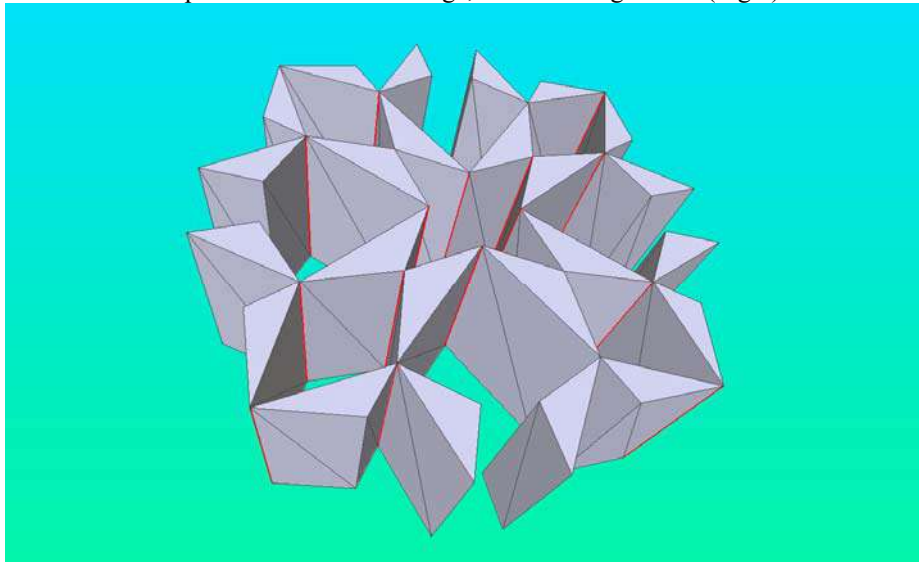


Fig. 9. Removal of singular defects by maintaining the connection between non-manifold faces (singular edges are red highlighted)

Keeping the connection between singular elements allows to load the original data without loss, as well as provide the necessary information to determine the defects of non-manifold edges and vertices.

7 Conclusion

A basic algorithm of the polygon creation implemented in OpenMesh leads to the destruction of the model if there is an attempt to create singular elements.

The newly proposed modified algorithm allows the connection between singular elements to be identified, and the overall structure of the model is preserved.

The advantage of the proposed algorithm is the simplicity and expansion of mesh recovery capabilities based on the earlier discussed algorithms.

The result of the proposed approach is the preservation of the connection in the topology during the formation of defects on polygonal meshes, which will enable detection and will make it possible to carry out various kinds of manipulations at the same time.

The disadvantage of this approach is in high memory requirements for storing links.

Further research will be devoted to improvement of the algorithm to ensure communication between all topological elements and its software implementation.

References

1. Atenne, M., Campen, M., Kobbelt, L.: Polygon Mesh Repairing: An Application Perspective : ACM Computing Surveys (scheduled to appear), 45(2):1-38. (2013). doi:10.1145/2431211.2431214
2. Lyon, M., Campen, M., Bommers, D., Kobbelt, L.: Parametrization Quantization with Free Boundaries for Trimmed Quad Meshing. ACM Transactions on Graphics (TOG), Vol. 38, Issue 4, Article No.: 51, pp 1–14 (2019). doi:10.1145/3306346.3323019 51:1-51:14.
3. Trettnerland, P., Kobbelt, L. Fast and Robust QEF Minimization using Probabilistic Quadrics. EUROGRAPHICS 2020 Volume 39 , Number 2 (2020).
4. Schmidt, P., Born, J., Campen, M., Kobbelt, L.: Distortion-Minimizing Injective Maps Between Surfaces ACM Trans. Graph., Vol. 38, No. 6, Art. No.: 156, pp 1–15 (2019). doi:10.1145/3355089.3356519
5. Botsch, M., Kobbelt, L. Pauly, M., Alliez, P., Lévy B.: Polygon Mesh A K Peters, Ltd. Natick, Massachusetts, (2010)
6. Ju, T.: Fixing Geometric Errors on Polygonal Models: A Survey. J. Comput. Sci. Technol. 24, 19–29 (2009). doi:10.1007/s11390-009-9206-7
7. Guskov, I., Wood, Z. 2001. Topological noise removal. GI '01: Proceedings of Graphics Interface 2001, pp.19–26, (2001). doi: 10.20380/GI2001.03
8. Shewchuk, J.R. Triangle: Engineering a 2D Quality Mesh Generator and Delaunay Triangulator. WACG 1996: Applied Computational Geometry Towards Geometric Engineering pp 203-222 (1996). doi:10.1007/BFb0014497
9. Ying, L., Zorin, D.N.: Nonmanifold subdivision. VIS '01: Proceedings of the conference on Visualization, pp.325–332, (2001)

10. Gueziec, A., Taubin, G., Lazarus, F., Horn, B.: Cutting and Stitching: Converting Sets of Polygons to Manifold Surfaces. IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS, Vol. 7, NO. 2, pp. 136-151 (2001)
11. Gueziec, A., Taubin, G., Lazarus, F., Horn, B.: Converting sets of polygons to manifold surfaces by cutting and stitching. SIGGRAPH '98: ACM SIGGRAPH 98 (1998). doi:10.1145/280953.281628
12. Rossignac, J., Cardoze, D.: Matchmaker: manifold BReps for non-manifold r-sets. SMA '99: Proceedings of the fifth ACM symposium on Solid modeling and applications, pp. 31–41 (1999). doi:10.1145/304012.304016
13. De Floriani, L., Hui, A.: A Scalable Data Structure for Three-dimensional Non-manifold Objects. In: Proc. of the Symp. on Geom. Proc., pp. 72-82 (2003)
14. De Floriani, L., Magillo, P., Puppo, E., Sobrero, D.: A Multi-resolution Topological Representation for Non-manifold Meshes. CAD Jour. 36(2), pp.141-159 (2004). doi:10.1145/566282.566307
15. De Floriani, L., Greenfieldboyce, D., Hui, A.: A data structure for non-manifold simplicial d-complexes. SGP '04: Proceedings of the 2004 Eurographics/ACM SIGGRAPH symposium on Geometry processing, pp. 83–92 (2004). doi:10.1145/1057432.1057444
16. De Floriani, L., Hui, A.: Data Structures for Simplicial Complexes: an Analysis and a Comparison. In: Proc. of the Symp. on Geom. Proc., pp. 119-128 (2005)
17. Zengbo, W.: Fast topological reconstruction algorithm for a STL file. Journal of Computer Applications, 34(9), pp.2720—2724 (2014).
18. Tan, J., Chen, J.: Research and Application on Model Repairing Algorithm of 3D Modeling Technology. 6th International Conference on Mechatronics, Computer and Education Informationization (MCEI 2016). doi: 10.2991/mcei-16.2016.48
19. OpenMesh <https://www.graphics.rwth-aachen.de/software/openmesh/>
20. Shovgelia, D.G., Sokolova, N.O.: Obnaruzhenie defektov v trekhmernykh geometricheskikh modelyakh, predstavlenykh na osnove Half-Edge Data Structure. Prykladni Pytannia Matematychnoho Modeliuvannia 1(2), pp.155-161 (2019). doi: <https://doi.org/10.32782/2618-0340-2019-3-14>

The Quality Indicators of Decision Tree and Forest Based Models

Sergey Subbotin¹[0000-0001-5814-8268]

¹National University "Zaporizhzhia Polytechnic", Zhukovsky str., 64,
Zaporizhzhia, 69063, Ukraine
subbotin@zntu.edu.ua

Abstract. The problem of quality model creation for models based on decision trees and forests is considered. The set of indicators characterizing properties of decision trees and forests is proposed. It allows to quantitatively evaluate such properties as diversity, equivalence, retraining, confidence in a decision-making, hierarchy, equifinality, generalization, nonlinearity, robustness, homogeneity, sensitivity to the input signals, plasticity, variability adaptability, symmetry, asymmetry, emergence (integrity), interpretability (logical transparency), learnability, and autonomy as for individually considered single tree model, as for an ensemble of tree models (forest).

Keywords: decision tree, forest, quality, property, model selection

1 Introduction

Automation of decision-making in applied tasks, as a rule, requires the construction of a decision-making model. To solve the problem of decision-making model constructing on the precedents, a wide class of computational intelligence methods has been proposed, including neural networks [1-5], neuro-fuzzy networks [6-9], decision and regression trees [10-17], forests of decision trees [18-22] and etc.

Usually, the quality of such models is characterized by the error function [1, 2]. As a result model is selected from several alternative obtained models, which has the smallest error. Note that for each class of methods, even for the same given training sample of observations, it is possible to obtain a wide range of different models with acceptable accuracy. At the same time, achieving the maximum accuracy (the smallest error) does not guarantee a high level of customer properties of the model.

Earlier in [23-26] author has proposed a set of indicators, applicable to models based on neural and neuro-fuzzy networks. However, most of these indicators are not applicable to the models based on decision trees and forests due to their paradigm difference from network models paradigm. Therefore, it is necessary to develop a quality model for decision trees and forests providing comparability of its indicators with the quality indicators of models based on neural and neuro-fuzzy networks proposed earlier in [23-25].

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

The properties of models can be affected not only by the structural parameters, but also by the properties of the training sample [27-31]. Therefore, it is necessary to take into account information about the properties of the sample when determining the quantitative indicators characterizing the properties of models.

The aim of this work was to create a quality model for models based on decision trees and forests as a set of quality indicators.

2 Formal problem statement

Let we have a training sample of observations in the form $\langle x, y \rangle$, where $x = \{x^s\}$, $x^s = \{x_j^s\}$, $y = \{y^s\}$, $s = 1, 2, \dots, S$, $j = 1, 2, \dots, N$, x^s is a set of input values (descriptive) features of s -th sample instance, x_j^s is a value of j -th input feature of s -th sample instance, y is an output feature value for the s -th instance in a sample, S is a number of instances in the training sample, N is a number of input (descriptive) features characterizing the instances of training sample.

Then the problem of model building for the dependence $y=f(w, x)$ on a sample of observations $\langle x, y \rangle$ based on the decision tree *tree* consists in identifying the model structure f and the values of its parameters w that provide an acceptable value of the given quality functional of the model $F(x, y, f, w)$ [17].

The problem of model building based on a forest of decision trees, in turn, can be represented as a problem of obtaining a set of models $forest = \{tree_t\}$, $tree_t = f_t(w^t, x)$, that provide an acceptable value of a given quality functional $F(x, y, \{f_t\}, \{w^t\})$, where t is a number of a tree in the forest, $t = 1, 2, \dots, T$, T is a number of trees in the forest, f is a model structure of a t -th tree, w^t is a set of model parameter values of a t -th tree [20].

It is obviously, that the problem of the quality functional creation of models based on decision trees and forests requires the determination of a set of indicators $\{I_i\}$ that quantitatively describes the properties of the models.

3 Primary model characteristics

Along with the sample parameters described above, we will use such notation for the characteristics of the samples: $\langle x_{test}, y_{test} \rangle$ is a test sample, S_{test} is a number of instances in the test sample; N, N' are, respectively, the number of signs in the original set and in the reduced set of features; f_{max}, f_{min} are, respectively, the maximum and minimum boundary values of a model output, y^{max}, y^{min} are, respectively, the maximum and minimum boundary values of output feature.

The basic properties characterizing the model structure are defined as: M is a number of levels in a tree, N_n is a number of nodes in the model, N_η is a number of nodes in the η -th layer of a tree, N_n^{max} is a maximum possible number of nodes in a model, f_i is an i -th node function, Δ_{min} is a smallest possible change of the real number, tak-

ing into account the bit grid of the computer, $\mathfrak{G}_{ao}(j)$ is a complexity of the j -th node, which can be defined similarly to [250] in units of elementary operations of addition and multiplication, $\mathfrak{G}_{aut}^{i*}$ is a characteristic of autonomy of a formation of i -th element of a model structure ($\mathfrak{G}_{aut}^{i*} = 0$, if the inclusion (or exclusion) of i -th node to the model determined only by the human; $\mathfrak{G}_{aut}^{i*} = 1$, if the inclusion (or exclusion) of i -th node to the model is automatically defined by the training method; $\mathfrak{G}_{aut}^{i*} = 0.5$, if the inclusion (or exclusion) of i -th node to the model can be determined by the human or training method), $\mathfrak{G}_{np}(i)$ is a characteristics of plasticity of i -th node of the tree, which is equal to the number of possible states of a node i (for leaf nodes containing singleton (not containing functions) $\mathfrak{G}_{np}(i) = 1$, for the nodes with functions the $\mathfrak{G}_{np}(i)$ should be taken equal to the number of different functions that may contained in the node, for the rest nodes of the tree we should take $\mathfrak{G}_{np}(i)$ as a number of branches of the node), w_{ij} is a connection of i -th and j -th nodes ($w_{ij} = 0$, if i -th and j -th nodes are not connected, and $w_{ij} = 1$, if i -th and j -th nodes are connected).

Let introduce the notation for the description of a model parameters: $w_{\max}, w_{\min}$ are, accordingly, the maximum and minimum possible values of a model parameters, N_w is a number of adjustable parameters of model node, $N_w^{\max}$ is a maximum possible number of adjustable parameters of a model nodes, $w_{i,j}^{\max}, w_{i,j}^{\min}$ are, respectively, the maximum and minimum possible values of j -th parameter in i -th node, Δw is a smallest possible change in weights taking into account the size of the computer bit grid, w_j is a j -th model parameter, $N_w(i)$ is a number of parameters of i -th model node, $\Delta w_{i,j}$ is a minimum possible change of j -th parameter of i -th node taking into account the bit grid size of a computer, $\mathfrak{G}_{sp}(i)$ is a characteristics of a plasticity of parameters of i -th node ($\mathfrak{G}_{sp}(i) = 0$ if the node has no adjustable parameters, otherwise set: $\mathfrak{G}_{sp}(i) = \sum_{j=1}^{N_w(i)} \text{round}((w_{i,j}^{\max} - w_{i,j}^{\min}) / \Delta w_{i,j})$),

We define the notation for describing the functioning of a model as: E_{tr}, E_{test} are, respectively, model errors for the training and test samples, $E(w)$ is a model error at a set of weights w .

The following notation will be used for the model training method parameters: N_{met} is a the number of training method parameters, N_{met}^{aut} is a number of parameters of a training method, which values of are determined automatically without a human intervention, $\mathfrak{G}_{aut}(w_i)$ is a characteristic of autonomy of i -th node parameter values setting ($\mathfrak{G}_{aut}(w_i) = 0$, if parameter values set only by a human; $\mathfrak{G}_{aut}(w_i) = 1$, if the values of the node parameters are determined automatically by the training method; $\mathfrak{G}_{aut}(w_i) = 0.5$, if the values of the node parameters can be determined by the human and method).

For tree models in a forest we denote: $w_{f,\max}, w_{f,\min}$, accordingly, maximum and minimum possible values of parameters of a forest model, N_{w_f} is a number of parameters of a forest model without taking into account the number of parameters of its trees, $T^{\max}$ is a maximum possible number of trees in the forest.

If necessary to distinguish the use of indicators for the decision tree we will use notations in the form of I^t or I^{tree} (here I is an indicator, t is a tree number, $tree$ is a tree symbol), and for the forest we will use notation of the form of I^{forest} (here I is an indicator, $forest$ is a forest symbol).

4 Model Quality Indicators

Diversity is defined by a number of different states of a system. In accordance with the law of "Requisite variety" of W.R. Ashby, creating of a system able to decide a problem, which has certain known diversity (complexity), it is necessary to provide for a system an even greater diversity (knowledge of solving methods) than the diversity of the problem being addressed, or to ensure the ability of the system to create this diversity within itself (it would have methodology, could have developed a methodic or proposed new methods for solving the problem) [32].

The absolute indicator of the limiting diversity of the synthesized model based on the decision tree *tree*, by analogy with [24], is defined as (1):

$$I_{div}(tree) = \text{round} \left(\frac{w_{\max} - w_{\min}}{\Delta w} \right)^{N_w N_a^{\max}} \prod_{i=1}^{N_w N_a^{\max}} (9_{np}(i)). \quad (1)$$

The absolute indicator of the limiting diversity of the synthesized model based on the forest of decision trees *forest* is defined as (2):

$$I_{div}(forest) = \text{round} \left(\frac{w_{f,\max} - w_{f,\min}}{\Delta w} \right)^{N_{w_f} T^{\max}} \prod_{t=1}^{N_{w_f} T^{\max}} (I_{div}(tree_t)). \quad (2)$$

The greater the value of limiting diversity, the wider the range of models we can obtained on its basis.

By analogy with [24], we define the diversity indicators of the tree model *tree* and forest model *forest*:

– in relation to the training sample as (3) and (4):

$$I_{div}(tree, < x, y >) = \frac{I_{div}(tree)}{I_{div}(x, y)}; \quad (3)$$

$$I_{div}(forest, < x, y >) = \frac{I_{div}(forest)}{I_{div}(x, y)}; \quad (4)$$

– in relation to the population universe as (5) and (6):

$$I_{div}(tree, X, Y) = \frac{I_{div}(tree)}{I_{div}(X, Y)}; \quad (5)$$

$$I_{div}(forest, X, Y) = \frac{I_{div}(forest)}{I_{div}(X, Y)}. \quad (6)$$

The more the value of $I_{div}(tree, \langle x, y \rangle)$ for a single tree, and the value of $I_{div}(forest, \langle x, y \rangle)$ for the forest, the more will be the model potential for approximating the relationship represented by the sample. The smaller the value of the corresponding indicator for the model at an acceptable level of error E , the better the approximation of the sample is.

The more the value of $I_{div}(tree, X, Y)$ for a single tree, and the value of $I_{div}(forest, X, Y)$ for a forest, the more the model will be able to solve the given problem. However, if the relevant indicator is greater than one, or is near to one, then the model is too excessive for the problem solving.

The equivalence of models is determined as follows: two models are equivalent if they have the same sets of answers (they respond equally to the same input stimuli) [33].

The equivalence coefficient of trained models based on decision trees t_1 and t_2 for the sample $\langle x, y \rangle$ we defined as (7):

$$I_{eq}(t_1, t_2) = \exp\left(-\frac{1}{S} \sum_{s=1}^S (f_{t_1}(x^s) - f_{t_2}(x^s))^2\right). \quad (7)$$

The equivalence coefficient of trained forests models can be determined similarly to the above, replacing the calculated tree outputs with the corresponding forest outputs.

The values of the equivalence indicator will be in the range from zero to one: the more similar the responses of the models with the same input influences, the greater the value of the equivalence coefficient.

Retraining of a model for the training sample x relatively to the test sample $\langle x_{test}, y_{test} \rangle \neq \langle x, y \rangle$ may be defined as (in the given form with substitutions) [24]:

– for classification problems as (8):

$$\delta(x, x_{test}) = \frac{1}{S} \sum_{x=1}^S \{1 | f(x^s) \neq y^s\} - \frac{1}{S_{test}} \sum_{x=1}^{S_{test}} \{1 | f(x_{test}^s) \neq y_{test}^s\}, \quad (8)$$

– for the evaluation problems as (9):

$$\delta(x, x_{test}) = \frac{1}{S} \sum_{x=1}^S \{1 | \delta \leq |f(x^s) - y^s|\} - \frac{1}{S_{test}} \sum_{x=1}^{S_{test}} \{1 | \delta \leq |f(x_{test}^s) - y_{test}^s|\}, \quad (9)$$

where δ is an error threshold.

Since the error threshold for an instance in practice cannot always be set, as well as for greater universality and uniformity in solving various problems we define it as (10):

$$\begin{aligned} \delta(x, x_{\text{test}}) = & -\frac{1}{S} \sum_{s=1}^S \exp \left(- \left(\frac{f(x^s) - y^s}{\max_{s=1,2,\dots,S} (y^s) - \min_{s=1,2,\dots,S} (y^s)} \right)^2 \right) + \\ & + \frac{1}{S_{\text{test}}} \sum_{s=1}^{S_{\text{test}}} \exp \left(- \left(\frac{f(x_{\text{test}}^s) - y_{\text{test}}^s}{\max_{s=1,2,\dots,S_{\text{test}}} (y_{\text{test}}^s) - \min_{s=1,2,\dots,S_{\text{test}}} (y_{\text{test}}^s)} \right)^2 \right). \end{aligned} \quad (10)$$

These indicators can be used not only for a model based on a single tree, but also for a forest of decision trees, using as f the output value determined by the ensemble of forest trees.

The higher the value of the retraining indicator, the worse the approximating properties of the model for data that did not used in the training.

Confidence in a decision-making is a subjective assessment by the model of the made decision [34].

Regarding the value at the model output for the instance x^s fed to its inputs, we determine the confidence indicator of the model in the made decision (11):

$$I_{\text{cert}}(x^s) = \exp \left(- (f(x^s) - y^s)^2 \right) \exp \left(- \sum_{j=1}^N (C_j^{u(x^s)} - x_j^s)^2 \right), \quad (11)$$

where C_j^q is a value of the coordinate on j -th feature of q -th cluster center corresponding to the node of the tree, to which the recognized instance x^s hits. For instances that are not included in the training set, instead of y^s it is possible to substitute the value of the output feature associated with the center of the corresponding cluster.

It is possible to estimate the coordinates of the cluster centers for leaf nodes on the basis of a training sample:

- for the decision tree constructed on the basis of the training sample, determine the belonging of the training sample instances to leaf nodes;
- for every q -th leaf node u_q form a cluster $C^q = \{C_j^q\}$ of instances fallen into this node, $q = 1, 2, \dots, Q$, where Q is a number of leaf nodes (clusters);
- for each j -th feature as the coordinate of the cluster center take the arithmetic average of the coordinates of the instances of the corresponding cluster according to the corresponding feature (12):

$$C_j^q = \frac{1}{S^q} \sum_{s=1}^S \{x_j^s \mid x^s \in u_q\}, j = 1, 2, \dots, N; q = 1, 2, \dots, Q, \quad (12)$$

where S^q is a number of instances of the training sample that fell into the q -th node (cluster).

– determine the function $u(x^s)$ that maps the recognized instance x^s to the node number of the tree into which it fell.

The average confidence of the decision tree for a sample x is defined as (13):

$$I_{cert}(x) = \frac{1}{S} \sum_{s=1}^S I_{cert}(x^s). \quad (13)$$

Indicators of subjective confidence of model based on a decision tree will take values in the range from zero to one: the higher the value, the closer the properties of a recognized instance to formed cluster center templates, and the more the model is confident in the made decision.

The confidence of the forest of decision trees for the instance x^s is defined as (14):

$$I_{cert}^{forest}(x^s) = \mathfrak{R}_k \left\{ I_{cert}^{forest}(x^s, k) \right\}, k = 1, 2, \dots, K, \quad (14)$$

where

$$I_{cert}^{forest}(x^s, k) = \mathfrak{N}_t \left\{ I_{cert}^t \mid f_t(x^s) = k \right\}, t = 1, 2, \dots, T, \quad (15)$$

f_t is an estimated value of the model output of t th tree, I_{cert}^{forest} is an indicator of confidence of forest models, $\mathfrak{R}_k$, $\mathfrak{N}_t$ are symbols of operators, defining the confidence of the forest in the decision for the k -th class and the t -th tree, respectively. As such operators it is possible to use the minimum, maximum, arithmetic mean of the set of arguments.

The indicator of averaged confidence of forest for sample x is defined as (16):

$$I_{cert}^{forest}(x) = \frac{1}{S} \sum_{s=1}^S I_{cert}^{forest}(x^s). \quad (16)$$

Indicators of subjective confidence of the forest will take values in the range from zero to one: the higher the indicator value, the more the trees ensemble confident in made decision.

The hierarchical organization of the structure, the integrity and crushability of elements allows to build models of complex objects from simpler ones; the work of the hierarchical structure requires that the information element in each hierarchical level behave as a whole, but when moving from level to level it must be fragmented, and when moving from the upper hierarchical level to the lower, this fragmentation corresponds to the allocation of its constituent elements, and when moving from the lower level to the top, it corresponds to the inclusion of a certain part of this element in a more complex object [33].

The hierarchy of the model based on the decision tree is defined by analogy with [25] as (17):

$$I_h = \frac{2 \sum_{\eta=1}^M \eta N_\eta}{M(M+1)N_n}, M \geq 1, N_n \geq 1. \quad (17)$$

The greater the I_h value, the greater the number of hierarchical levels in the model with respect to maximum possible number of levels for a given number of nodes N_n .

Estimate the maximum possible number of levels. Since the maximum number of levels in the tree will be at the minimum number of outcomes from nodes, then at each level there should be at least one node with two outcomes, and the rest of the nodes should be leafy. Moreover, the greatest number of levels will be achieved for the tree, where only one node at each level (except for the last) has two outcomes, and the rest are leafy and contain only two nodes at each level. Thus, for the deepest tree, the number of nodes of the highest level is 1 (root), for the lower level is 2 (leaves), for the remaining layers is 2, i.e. $1+2(M-1) = N_n$. From here we get: $M=0.5(N_n-1)+1$. Therefore, for the decision tree we get (18):

$$I_h = \frac{2 \sum_{\eta=1}^M \eta N_\eta}{0,25N_n^3 + N_n^2 + 0,75N_n}, M \geq 1, N_n \geq 1. \quad (18)$$

The hierarchy of the model based on the forest of decision trees is defined as (19):

$$I_h^{forest} = \max_{t=1,2,\dots,T} \{I_h^t\}. \quad (19)$$

The elasticity for a function $y(x)$ on the variable x_j in [35] is defined as (20):

$$El_{x_j}(y) = \lim_{\Delta x_j \rightarrow 0} \frac{\frac{\Delta y}{y}}{\frac{\Delta x_j}{x_j}} = \left(\lim_{\Delta x_j \rightarrow 0} \frac{\Delta y}{\Delta x_j} \right) \frac{x_j}{y}, \quad (20)$$

where $\Delta y = \frac{y(x_j + \Delta x_j) - y(x_j)}{y(x_j)}$, $x_j > 0, y > 0$.

The relative elasticity indicator on the variable x_j of approximating function $y=f(x)$ realized by the model at output y trained on a training sample $\langle x, y \rangle$ is defined similarly to [24] as (21):

$$El_{x_j}(y) = \frac{1}{2S} \sum_{s=1}^S \left(\frac{\tilde{x}_j^s (\tilde{f}_{(\tilde{x}_j^s + \Delta x_j)} - \tilde{f}_{(\tilde{x}_j^s)})}{\Delta x_j (\tilde{f}_{(\tilde{x}_j^s)})^2} \right), \quad (21)$$

where $\tilde{f}_{(\tilde{x}_j^s)}$ is a calculated value at the output of the model when applying normalized values of the features of s -th instance to its inputs; $\tilde{f}_{(\tilde{x}_j^s + \Delta x_j)}$ is a value of the model output when applying normalized values of features of s -th instance to its inputs, and corrected normalized by Δx_j value of j -th feature of s -th instance to j -th input.

The larger the value of elasticity indicator, the more elastic is the model. This indicator applies both to a single tree model and for a forest model.

Equifinality is a regularity of functioning and development of the system, characterizing its ultimate capabilities [36].

Relative equifinality of a model based on a decision tree *tree* defined similarly to [24] as (22):

$$I_{eqf}(tree, < x, y >) = \frac{N_w}{N_w^{\max}} \frac{N_n}{N_n^{\max}} \exp \left(-\frac{1}{S} \sum_{s=1}^S (f(w, x^s) - y^s)^2 \right). \quad (22)$$

Relative equifinality will receive the largest value (top-limited by one) for those models which have reached the maximum possible size and the number of parameters during synthesis as well as the smallest error (bottom limited by zero) in the learning process.

For the forest of decision trees, we determine the relative equifinality indicator as (23):

$$I_{eqf}(forest, < x, y >) = \frac{T}{T_{\max}} \prod_{t=1}^T I_{eqf}(tree_t, < x, y >). \quad (23)$$

Generalization is the model's ability to integrate partial data to determine patterns and prolongate results, that is, after training based on the training set, to give answers for test sample instances similar to the training sample but not included in it [37, 38].

The generalization indicator of the decision tree for the training $< x, y >$ and test $< x_{\text{test}}, y_{\text{test}} >$ samples is determined by analogy with [37, 38] as (24):

$$I_G = 1 - \exp \left(-\frac{1}{S_{\text{test}}} \sum_{p=1}^{S_{\text{test}}} \left(\frac{(f(x^{p*}) - y_{\text{test}}^p)^2 \sum_{j=1}^N (x_j^s - x_{j\text{test}}^p)^2}{N(y^s - y_{\text{test}}^p)^2} \right) \right), \quad (24)$$

$$s = \arg \min_{t=1, 2, \dots, S} \sum_{j=1}^N (x_j^t - x_{j\text{test}}^p)^2, \quad j = 1, 2, \dots, N, p = 1, 2, \dots, S_{\text{test}}.$$

In a similar way, the generalization indicator for the forest of decision trees will be determined.

Generalization indicator will take values in the range from zero to one, and will be the greater, the smaller the error of a model at instance recognition, and difference of recognized instance to nearest by features instance of a training sample is more.

The generalization indicator of the trained model is defined as (25):

$$I_{gen} = \frac{NS}{N_w N_n} \exp(-(E_{tr} - E_{test})^2). \quad (25)$$

If generalization indicator is significantly greater than one, then the model shows great ability to generalization, if the generalization indicator much smaller than one, then the model does not shows no generalizing properties.

Generalization indicator for a forest of decision trees is defined as (26):

$$I_{gen}^{forest} = \frac{NS}{\sum_{t=1}^T N_w^t N_n^t} \exp(-(E_{tr} - E_{test})^2). \quad (26)$$

The errors here are defined for the ensemble of trees.

Nonlinearity is a dependency that cannot be explained by a linear combination of variable inputs [39, 40].

The nonlinearity indicator for classification problems is defined similarly to [39] as (27):

$$I_{nl} = \frac{2}{S(S-1)} \sum_{s=1}^S \sum_{p=s+1}^S \left(\frac{\sum_{\ell=0}^S \left| f\left(\frac{\ell x^p}{S} + \left(1 - \frac{\ell}{S}\right) x^s\right) - f(x^p) \right|}{\sqrt{\sum_{j=1}^N (x_j^s - x_j^p)^2}} \right). \quad (27)$$

The nonlinearity indicator for estimation problems is defined as (28):

$$I_{nl} = \frac{2}{S(S-1)} \sum_{s=1}^S \sum_{p=s+1}^S \left(\frac{\sum_{\ell=0}^S \left| f(\ell x^p S^{-1} + (1 - \ell S^{-1}) x^s) - f(x^p) \right|}{\sqrt{\sum_{j=1}^N (x_j^s - x_j^p)^2}} \right). \quad (28)$$

The nonlinearity indicator of the classifier will take values in the range [0, 1]: the greater its value, the more nonlinear is a model. A disadvantage of this indicator is its applicability only for models with single output, and exclusively for classification problems.

The nonlinearity indicator for estimation problems is applicable also for classification problems.

In a similar way, the nonlinearity indicator for the forest of decision trees can be determined.

The indicator of compliance of nonlinearities of the sample and model is defined as (29):

$$\tilde{I}_{nl} = \frac{I_{nl}(< x, y >)}{I_{nl}}, \quad (29)$$

where $I_{nl}(< x, y >)$ is a nonlinearity indicator of the sample, determined according to [41, 42].

If the indicator $\tilde{I}_{nl}$ is equal to one, then we can conclude that the model corresponds to the sample in complexity. If the indicator $\tilde{I}_{nl}$ is less than one, then the smaller its value, the greater the effect of retraining will be, and it would show possible redundancy of a model. If the indicator $\tilde{I}_{nl}$ value exceeds one, then the model is not sufficient for good approximation (require additional training or change the model structure).

Robustness is a model property to reliably solve a problem when receiving incomplete and / or damaged data. In addition, the results must be consistent, even if some part of the model is damaged [43, 44].

The robustness of the model on the basis of a decision tree in relation to the input signals is defined as (30):

$$I_{Rb}^x = \min_{j=1, 2, \dots, N} \left\{ \frac{\min(\tau_1, \tau_2) - \min_{s=1, 2, \dots, S} \{x_j^s\}}{\max_{s=1, 2, \dots, S} \{x_j^s\} - \min_{s=1, 2, \dots, S} \{x_j^s\}} \right\}, \quad (30)$$

where

$$\begin{aligned} \tau_2 &= \min_{\frac{1}{S} \sum_{s=1}^S \left(f \left(x^s \left| \begin{matrix} x_b^s = x_b^s, \forall b \neq j, b=1, 2, \dots, N \\ x_j^s = x_j^s - \Delta x_j \end{matrix} \right. \right) - y_j^s \right)^2 > \varepsilon} \{ \Delta x_j \}, \\ \Delta x_j &= \min_{s=1, 2, \dots, S} \{x_j^s\} + \delta_x \left(\max_{s=1, 2, \dots, S} \{x_j^s\} - \min_{s=1, 2, \dots, S} \{x_j^s\} \right), \dots, \\ &\quad \max_{s=1, 2, \dots, S} \{x_j^s\} - \delta_x \left(\max_{s=1, 2, \dots, S} \{x_j^s\} - \min_{s=1, 2, \dots, S} \{x_j^s\} \right), \end{aligned}$$

$\delta_x \in (0, 1)$ is a constant that regulates the accuracy of determining a robustness indicator on model inputs.

The indicator is normalized smallest change in the input signal, resulting in a significant increase of model error.

The robustness of a model based on a decision tree with respect to weights (parameters) is defined as (31):

$$I_{Rb}^w = \min_{j=1, 2, \dots, N_w} \left\{ \frac{\min(\tau_1, \tau_2) - w_{\min}}{w_{\max} - w_{\min}} \right\}, \quad (31)$$

$$\tau_1 = \min_{\frac{1}{S} \sum_{s=1}^S \left(f(x^s | w_j = w_j + \Delta w) - y_i^s \right)^2 > \varepsilon} \{ \Delta w_j \},$$

$$\tau_2 = \min_{\frac{1}{S} \sum_{s=1}^S \left(f(x^s | w_j = w_j - \Delta w) - y_i^s \right)^2 > \varepsilon} \{ \Delta w_j \},$$

$$\Delta w_j = w_{\min} + \delta_w (w_{\max} - w_{\min}), \dots, w_{\max} - \delta_w (w_{\max} - w_{\min}),$$

where $\delta_w \in (0, 1)$ is a constant regulating accuracy of determination of the robustness indicator on model parameters.

The indicator is a normalized least change in weight values, leading to significantly increase in model error.

The integral robustness indicator for a decision tree can be defined as (32):

$$I_{Rb} = I_{Rb}^x \cdot I_{Rb}^w. \quad (32)$$

The indicator I_{Rb} will take values in the range $[0, 1]$. The closer its value to zero, the lower the robustness of the model, the more sensitive the model to a change in input signals or parameter values. The closer the value of the indicator I_{Rb} to one, the greater the robustness of the model, the less sensitive the model to changes in input signals or parameter values.

In this way, robustness indicators for the forest of decision trees can be determined.

The homogeneity of the elements lies in the fact that models are built from many simple unified standard elements that perform elementary actions and are interconnected by various connections [45].

The homogeneity of the functions of the nodes of a decision tree is defined by analogy with [25] as (33):

$$I_{hn} = \frac{2 \sum_{i=1}^{N_n} \sum_{j=i+1}^{N_n} \{1 | f_i \equiv f_j\}}{N_n (N_n - 1)}. \quad (33)$$

The homogeneity indicator will vary from zero to one: the more its value, the more uniform the corresponding elements of the model.

The homogeneity of the node functions of the of the forest trees is defined as (34):

$$I_{hn}^{forest} = \frac{\sum_{t=1}^T I_{hn}^t N_n^t (N_n^t - 1)}{\sum_{t=1}^T N_n^t (N_n^t - 1)}. \quad (34)$$

The sensitivity to the input signals is characterized by calculating the partial derivatives of the model error function [46–48]. However, this approach is computationally hard.

The averaged normalized indicator of the sensitivity of the output of the decision tree to a change in the input signal is defined as (35):

$$I_{tol} = \frac{1}{SN(y^{\max} - y^{\min})} \sum_{s=1}^S \sum_{j=1}^N \max(\tau_1, \tau_2), \quad (35)$$

$$\tau_1 = \left(f \left(x^{s*} \left| \begin{array}{l} x_b^{s*} = x_b^s, b \neq j, b = 1, 2, \dots, N, \\ x_j^{s*} = x_j^s + \Delta_{\min} \end{array} \right. \right) - y_i^s \right)^2,$$

$$\tau_2 = \left(f \left(x^{s*} \left| \begin{array}{l} x_b^{s*} = x_b^s, b \neq j, b = 1, 2, \dots, N, \\ x_j^{s*} = x_j^s - \Delta_{\min} \end{array} \right. \right) - y_i^s \right)^2.$$

The value of the sensitivity indicator will be in the range [0, 1]. The higher the value of the sensitivity indicator, the stronger the model reacts to changes in the input signal, the greater are its categorization capabilities. However, too high sensitivity may indicate a weak model resistance to noise and interference in the input signal.

The average indicator of the sensitivity of the forest to changes in the input signal is defined as (36):

$$I_{tol}^{forest} = \frac{1}{T} \sum_{t=1}^T I_{tol}(tree_t). \quad (36)$$

Plasticity determines the complexity of the model's behavior, which is considered as a result of the interaction of many elements, each of which limits the action of others and is limited by others on the way to the formation of global observable behavior [49, 50]. As an analogue of neural plasticity, where neuron nodes are considered as plastic elements for neural network models, with respect to decision trees, we will consider the plasticity of nodes. As an analogue of synaptic plasticity (modification of the strength of the synaptic connection between nodes, implemented by the scales in neural network models) as applied to the decision tree, we will consider the plasticity of tunable parameters of tree nodes.

The relative indicator of plasticity of the nodes of the model by analogy with [25] is defined as (37):

$$I_{np} = \frac{\sum_{i=1}^{N_n} \mathfrak{g}_{np}(i)}{N_n^{\max} \mathfrak{g}_{np}^{\max}}. \quad (37)$$

The indicator I_{np} will take values in the range from zero to one: the greater its value, the higher the level of plasticity of the model nodes.

The relative plasticity indicator of the adjustable model parameters, by analogy with [25], is defined as (38):

$$I_{sp} = \frac{\sum_{i=1}^{N_n} \mathfrak{G}_{sp}(i)}{N_n^2 \text{round}\left(\frac{w_{\max} - w_{\min}}{\Delta w}\right)}. \quad (38)$$

The coefficient I_{sp} will take values in the range from zero to one: the bigger its value, the higher the level of plasticity of the model parameters.

The relative indicator of plasticity of a model is defined as (39) [25]:

$$I_{pl} = I_{np} I_{sp}. \quad (39)$$

The relative indicator of plasticity will take values in the range from zero to one: the greater its value, the higher the level of plasticity of the model and, therefore, it has better adaptive abilities.

For the forest of decision trees, we define the plasticity indicators as (40)–(42):

$$I_{np}^{forest} = \frac{1}{T} \sum_{t=1}^T I_{np}^t, \quad (40)$$

$$I_{sp}^{forest} = \frac{1}{T} \sum_{t=1}^T I_{sp}^t, \quad (41)$$

$$I_{pl}^{forest} = \frac{1}{T} \sum_{t=1}^T I_{pl}^t. \quad (42)$$

Variability is an ability to obtain several different models for approximating dependencies from the same data sample using the same method [51-53]. As applied to decision trees, the variability of models is determined by the choice of a feature for the root node and the order in which features are added for checks at other nodes, the method of determining threshold values in nodes, etc.

The absolute indicator of the variability of the model we define similarly to [25] as (43):

$$I_v = \prod_{\eta=1}^M \prod_{i=1}^{N_\eta} \mathfrak{G}_v(n(\eta, i)) \mathfrak{G}_v(w(\eta, i)), \quad (43)$$

where $\mathfrak{G}_v(n(\eta, i))$ is a variability of the verification of the i -th node of η -th layer of the model: $\mathfrak{G}_v(n(\eta, i)) = 1$, if a non-random feature hit into the node; $\mathfrak{G}_v(n(\eta, i)) = N$, if a feature for checking in the node selected as random from all original feature set $\mathfrak{G}_v(n(\eta, i)) = N^*$, if a feature for checking in the node is randomly selected from the set of N^* not yet considered features (44):

$$N_{(\eta,i)}^* = \left(\sum_{\mu=1}^{\eta-1} \sum_{j=1}^{N_{\mu}} 1 \right) + i - 1, \quad (44)$$

where $\mathfrak{G}_v(w(\eta,i))$ is a variability of determining the values of the parameters of i -th node of η -th model layer: $\mathfrak{G}_v(w(\eta,i))$ is equal to the number of parameters in the node that can be configured non-deterministically, if all parameters in the node depend on previous nodes, then $\mathfrak{G}_v(w(\eta,i)) = 1$.

The more the I_v value, the more different models can be obtained based on the corresponding paradigm.

The absolute indicator of the variability of the forest of decision trees is defined as (45):

$$I_v^{forest} = \prod_{t=1}^T I_v^t. \quad (45)$$

Noise resistance is the property of the model to provide the correct response to an input signal containing noise [54].

The indicator of the resistance of the trained model to additive noise in the input signal at the j -th input is defined similarly to [24] as (46):

$$I_{tolj}^{\ell} = \exp\left(-\frac{1}{2} \sum_{s=1}^S (\tau_1 + \tau_2)\right), \quad (46)$$

$$\tau_1 = (f(x^s) - f({}_{(j^+)}x^s))^2, \tau_2 = (f(x_j^s) - f({}_{(j^-)}x^s))^2,$$

$${}_{(j^+)}x_g^s = \begin{cases} x_g^s, g \neq j, g = 1, 2, \dots, N; \\ x_g^s + \ell \left(\max_{p=1,2,\dots,S} \{x_g^p\} - \min_{p=1,2,\dots,S} \{x_g^p\} \right), g = j, \end{cases}$$

$${}_{(j^-)}x_g^s = \begin{cases} x_g^s, g \neq j, g = 1, 2, \dots, N; \\ x_g^s - \ell \left(\max_{p=1,2,\dots,S} \{x_g^p\} - \min_{p=1,2,\dots,S} \{x_g^p\} \right), g = j, \end{cases}$$

where ℓ is a given noise level, $0 < \ell < 1$. In order to automate the process of ℓ setting it is proposed to use such formula (47):

$$\ell = \min_{j=1,2,\dots,N} \left\{ \frac{\min_{\substack{s=1,2,\dots,S; \\ p=s+1,\dots,S}} \left\{ \|x_j^s - x_j^p\| \mid y^s \neq y^p \right\}}{\left| \max_{p=1,2,\dots,S} \{x_j^p\} - \min_{p=1,2,\dots,S} \{x_j^p\} \right|} \right\}. \quad (47)$$

The greater the value of the indicator of model resistance to noise in the input signal on j -th input, the less important this input to the decision making.

This indicator is applicable both to a model based on a single decision tree and to a forest-based model.

Adaptability is a property of structures to dynamically and independently change their behavior in response to an input stimulus [55]. In relation to a model based on decision trees, adaptability is determined, first of all, by plasticity, which determines the resources for adaptation: the greater the plasticity, the more adaptive the properties of the model.

Plasticity is a necessary but insufficient prerequisite for adaptability. Along with plasticity, the adaptive properties of the model are influenced by the sensitivity of the model, which determines the strength of the reaction of the model to the minimum change in the values of its parameters.

The adaptability indicator of a model is defined as (48):

$$I_{adapt} = I_{pl} \cdot I_{tol}. \quad (48)$$

The larger the adaptability indicator value, the greater the possibility has models to adapt to a given task.

The adaptability indicator in this way can be determined for the forest of decision trees.

Symmetry reflects the proportionality in the arrangement of the parts of the whole in a space, the complete correspondence (in location, size) of one half of the whole to the other half [56]. In relation to decision trees, it is possible to determine the indicators of symmetry and asymmetry.

The indicator of symmetry of the structure of the decision tree is defined as (49):

$$I_{sym}^n = \frac{2}{N_n^2 - N_n} \sum_{i=1}^{N_n} \sum_{j=i+1}^{N_n} \{1 | f_i \equiv f_j\} \quad (49)$$

The greater the I_{sym}^n value, the greater the symmetry of the structure of the decision tree.

The asymmetry indicator of the model structure based on the decision tree as (50):

$$I_{asym}^n = 1 - I_{sym}^n. \quad (50)$$

The greater the value of I_{asym}^n the greater the asymmetry of the model structure.

The symmetry indicator of the nodes of the decision tree is defined as (51):

$$I_{sym}^w = \frac{1}{N_n^2} \sum_{i=1}^{N_n} \sum_{j=1}^{N_n} \{1 | w_{ij} = w_{ji}\} \quad (51)$$

The higher the I_{sym}^w value the more the symmetry of model connections.

The asymmetry indicator of model connections based on the decision tree is defined as (52):

$$I_{asym}^w = 1 - I_{sym}^w. \quad (52)$$

The larger the I_{asym}^w value the greater the asymmetry of the model connections.

The general indicator of the symmetry of a model based on a decision tree is defined as (53):

$$I_{sym} = I_{sym}^n I_{sym}^w. \quad (53)$$

The higher the value of I_{sym} the bigger the symmetry of the decision tree.

The general indicator of the asymmetry of a model based on a decision tree is defined as:

$$I_{asym} = 1 - I_{sym}. \quad (54)$$

The larger the I_{asym} the greater the asymmetry of the model.

For the forest of decision trees, the symmetry and asymmetry indicators can be determined as the average values of the corresponding indicators of the individual trees included in the forest.

Emergence (integrity) is a regularity that manifests itself in the system in the appearance of new properties in it, which are absent in its elements. The integrity property is associated with the purpose for which the system is created. Let C_o is an intrinsic complexity, which is the total complexity (content) of system elements without interconnecting them (in the case of pragmatic information, the total complexity of the elements that affect the achievement of the goal), C_v is a mutual complexity characterizing the degree of interconnection of elements in the system (i.e. the complexity of its schema or structure). The degree of system integrity in accordance with [24, 57, 58] is defined as (55):

$$I_\alpha = C_v / C_o. \quad (55)$$

The emergence of a model based on a decision tree is defined similarly to [24] as (56):

$$I_\alpha = \frac{\sum_{i=1}^{N_n} \sum_{j=1}^{N_n} \vartheta_{av}(i, j)}{\sum_{j=1}^{N_n} \vartheta_{ao}(j)}. \quad (56)$$

The larger the I_α value, the more holistic is the model.

For a forest-based model, emergence is defined as (57):

$$I_\alpha^{forest} = \frac{T^2}{\sum_{t=1}^T I_\alpha^t}. \quad (57)$$

Interpretability (logical transparency) is a model property to be understandable for human perception and analysis [59, 60]. Obviously, a model is more interpretable if it

is hierarchical, and the average number of node connections does not exceed 5-7 (this number is caused by the peculiarities of the human psyche). Since each node in the decision trees has only one input, it is necessary to consider mainly the number of outcomes from the node.

Heuristically we define interpretability through hierarchy and the number of nodes [25] as (58):

$$I_{\text{interp.}} = \frac{I_h}{\sum_{i=1}^{N_n} \sum_{j=1}^{N_n} \{w_{ij} | j \neq i\}}. \quad (58)$$

The level of model interpretability increases with increasing of $I_{\text{interp.}}$ value. For the forest of decision trees, we define the interpretability as (59):

$$I_{\text{interp.}}^{\text{forest}} = \frac{1 + \max_{t=1,2,\dots,T} \{I_h^t\}}{2N_n^2}. \quad (59)$$

Learnability is the property of a model to improve its work (to learn or adapt), using examples to turn it to solve a particular problem [61].

The decision tree model learning indicator similarly to [25] is defined as (60):

$$I_{lr} = \frac{I_{pl} L(\text{tree})}{NSL}, \quad (60)$$

where L is the Lipschitz constant for the training sample [62] as (61):

$$L = \max_{\substack{s, p=1,2,\dots,S; \\ s \neq p}} \left\{ \|y^s - y^p\| / \|x^s - x^p\| \right\}, \quad (61)$$

$L(\text{tree})$ is a Lipschitz constant (complexity) of the model [63, 64], which, as applied to the binary decision tree, is estimated as (62):

$$L(\text{tree}) \leq 10^{\frac{N'}{K}} \sqrt{N' K} \prod_{\eta=0}^{N'/K} \left(\frac{N'}{2^\eta} \right). \quad (62)$$

The greater the value of the learnability indicator, the model has bigger potential for solving the problem of approximating dependence $y = f(x)$ given in tabular form.

For a forest of decision trees, we define the learnability indicator as (63):

$$I_{lr}^{\text{forest}} = \frac{\sum_{t=1}^T I_{pl}^t L(\text{tree}_t)}{NSL}. \quad (63)$$

Autonomy is an agent's ability to act without direct human intervention by control on its own actions and internal state. Autonomy also implies the possibility of learning based on experience [65].

Since the trained computational intelligence models, as a rule, in the process of their functioning in decision-making does not require human intervention, they are equally have a property of autonomous functioning. However, in the process of training the level of autonomy for different models and different training methods may vary considerably.

Therefore, we will consider further characteristics of model autonomy only in relation to the process of its learning. Since the ability to learn is determined by plasticity, the autonomy of learning (self-adaptivity) will be characterized by an indicator that depends on the plasticity characteristics of the model. On the other hand, the dependence of model learning from the human may be characterized by its influence (portion) on the formation of the structure and parameters of the model.

Combining these considerations, we obtain the indicator of autonomy of model training method (64):

$$I_{aut} = \frac{1 + N_{met}^{aut}}{1 + N_{met}} \cdot \frac{\sum_{i=1}^{N_n} (g_{aut}^{i*} g_{aut}(w_i))}{N_n}. \quad (64)$$

As I_{aut} increases, the level of model autonomy in the training process increases.

For a forest of decision trees, the indicator I_{aut} can be defined as a smallest of the indicators of forest trees (65):

$$I_{aut}^{forest} = \min_{t=1,2,\dots,T} \{I_{aut}^t\}. \quad (65)$$

5 Integral indicators of model quality

Information quality criteria is a family of integral indicators, depending on the model error E , the training sample volume S and the number of adjusted model parameters N_w . They include Hannan-Quinn Criterion [66], Bayesian Information Criterion [67], Akaike's Information Criterion [68], Corrected AIC [69], and Unbiased AIC [69]. A number of criteria in addition to the error, the sample size and the number of adjustable parameters also take into account the maximum possible number of adjustable parameters $N_w^{\max}$. They include a Minimum Description Length [70], Shortest Data Description [71], Consistent AIC [69] and Mallows Criterion [72].

At model constructing and comparing it is usually assumed to be identical the sample size. Therefore, it is advisable to exclude the sample size from the comparison criteria. At the same time, various synthesized models may not use all of the features of the sample. Therefore the number of features used in the models N^f should be seen as an important property of the models at their comparison.

On the basis of these considerations we can define integral information criterion of a model as (66):

$$IIC = \left(1 - \frac{N'}{NN_w^{\max}}\right) e^{-E}. \quad (66)$$

The IIC criterion will take values in the range from zero to one. The less its value the worse the model, and the bigger its value the better the model. Here, for different models, the error values E and the maximum number of adjustable parameters $N_w^{\max}$, as well as the number of used features N' , may differ. At the same time the number of features in the original set N will not differ, but is used in the formula (66) for normalizing N' .

This indicator is applicable for comparing models based on single decision trees and for ensembles (forests) of decision trees. For the case of ensembles, N will remain unchanged, and the error E , the number of selected features N' and the number of adjustable parameters $N_w^{\max}$ will be determined for the entire ensemble of trees.

The effectiveness (quality) of a problem solving by the model is determined by the accuracy (error) of problem solving for the training and test data, simplicity, logical transparency and speed of the resulting model, as well as by the cost of model building (by hardware requirements, iteration and time spent of training method).

The generalized indicator of the model effectiveness based on the indicators proposed above is determined by analogy with [25] as (67):

$$I_{ef} = I_{gen} I_{lr} \exp(-E). \quad (67)$$

The indicator I_{ef} can be used as for comparing the models and methods of their synthesis, as for optimization of the model building process.

Similarly, the I_{ef} indicator can be determined for a forest of decision trees.

In this case, the error E and indicators I_{gen}, I_{lr} will be determined for the entire ensemble of trees.

An alternative generalized indicator of model efficiency may be defined by analogy with [25] as (68):

$$I_{ef}' = \frac{2}{\pi} \operatorname{arcsec} \left(\frac{N}{N'} \frac{N_w^{\max}}{N_w} \frac{N_n^{\max}}{N_n} \frac{NS}{N_w} \right) \exp(-E), \quad (68)$$

$$N' \geq 1, N_n^{\max} \geq 1, N_n^{\max} \geq 1.$$

Having resulted similar, we receive (69):

$$I_{ef}' = \frac{2}{\pi} \operatorname{arcsec} \left(\frac{N^2 S N_w^{\max} N_n^{\max}}{N' N_w^2 N_n} \right) \exp(-E), \quad (69)$$

$$N' \geq 1, N_n^{\max} \geq 1, N_n^{\max} \geq 1.$$

An alternative generalized efficiency indicator I_{ef}' may be used as for comparing models and methods for their synthesis, as to optimize the process of model building.

I_{ef} indicator can be used also for models based on forest of decision-trees. For the case of ensembles, N and S will remain unchanged, and the error E , the number of selected features N' , and the number of adjustable parameters $N_w^{\max}$ will be determined for the entire ensemble of trees.

6 Results and Discussion

The set of indicators proposed above is extensive and for its application in practice it is advisable to analyze the proposed indicators.

The Fig. 1 presents the classification of a set of proposed indicators characterizing the properties of models based on decision trees and forests.

Horizontally at Fig. 1, indicators are divided into groups according to the complexity of data compilation: sample (indicators characterize the properties of the sample and are model independent), tree model (indicators are defined for a single decision tree model), forest (indicators are defined for a set of decision tree models of a forest).

The more to the right the indicator is located horizontally at Fig. 1, the higher the level of complexity of data generalization by the model is required to determine it.

Vertically at Fig. 1, indicators are divided into groups according to the level of computational complexity relative to primary characteristics: basic properties (easily identifiable characteristics of data and models), primary indicators (indicators determined on the basis of basic properties), secondary indicators (indicators determined on based on primary indicators), and integrative indicators (indicators determined on the basis of indicators of previous levels).

The higher the level of the indicator, the more difficult it is to calculate it with respect to the primary properties of the data and models.

7 Conclusion

The problem of creation of a quality model for models based on decision trees and forests is solved.

The set of indicators characterizing properties of decision trees and forests is proposed. It allows to quantitatively evaluate such model properties as diversity, equivalence, retraining, confidence in a decision-making, hierarchy, equifinality, generalization, nonlinearity, robustness, homogeneity, sensitivity to the input signals, plasticity, variability adaptability, symmetry, asymmetry, emergence (integrity), interpretability (logical transparency), learnability, and autonomy as for individually considered (single) tree model, as for an ensemble of tree models (forest).

The prospects of further study are to obtain estimates of the computational (time) and spatial (memory) complexity of calculating the proposed indicators, to conduct an experimental study of the set of the proposed indicators for assessing the properties of models in solving practical problems of diagnosis and automatic classification on features, to identify the relationships between different indicators of the properties of models based on decision trees and forests.

Integrative indicators		I_{ef}	I_{ef}'	IIC	
Secondary indicators		I_{Rb}	$\tilde{I}_{nl}$	I_{asym}	$I_{cert}^{forest}(x)$
		$I_{interp.}$	I_{gen}	I_{sym}	$I_{cert}^{forest}(x^s)$
		$I_{cert}(x)$	I_{lr}	I_{asym}^n	I_{pl}^{forest}
		$I_{div}(tree < x, y >)$	I_{adapt}	I_{asym}^w	$I_{div}(forest, < x, y >)$
		$I_{div}(tree, X, Y)$		I_{pl}	$I_{div}(forest, X, Y)$
Primary indicators	$I_{nl}(< x, y >)$	$I_{div}(tree)$	I_G	E	$I_{div}(forest)$
					$I_{cert}^{forest}(x^s, k)$
	$I_{div}(x, y)$	$I_{cert}(x^s)$	I_{nl}	I_{sym}^n	I_h^{forest}
		I_h	I_{aut}	I_{sym}^w	I_{gen}^{forest}
	$I_{div}(X, Y)$				I_{aut}^{forest}
		$I_{eq}(tree < x, y >)$	$L(tree)$	I_{np}	I_{lr}^{forest}
	L	I_{Rb}^x	I_α	I_{sp}	$I_{interp.}^{forest}$
					I_α^{forest}
	$I_{nl}(< x, y >)$	I_{Rb}^w	I_v	$El_{x_j}(y)$	I_v^{forest}
					I_{np}^{forest}
Basic properties	x_j^s	N	w	$\mathfrak{G}_v(n(\eta, i))$	$tree_t$
		M	w_{max}		
	y^s	N_n^{max}	w_{max}	$\mathfrak{G}_v(w(\eta, i))$	$w_{f, max}$
		N_η	Δw	N_{met}	
	S	f	N_w		$w_{f, min}$
Basic properties		f_{max}	N_w^{max}	N_{met}^{aut}	
	N	f_{min}	w_{ij}		N_{w_f}
		f_i	$w_{i,j}^{max}$	$\mathfrak{G}_{aut}(w_i)$	
	y^{max}	Δ_{min}	$w_{i,j}^{min}$		
		$\mathfrak{G}_{ao}(j)$	w_j		T^{max}
		$\mathfrak{G}_{aut}^{i*}(i)$	$N_w(i)$		
	y^{min}	$\mathfrak{G}_{np}(i)$	$\Delta w_{i,j}$		
		$\mathfrak{G}_{sp}(i)$			
Level	Sample	Tree model		Forest model	

Fig. 1. Analysis of decision tree and forest indicators

References

1. Haykin, S.: Neural Networks and Learning Machines. Pearson, London (2008).
2. Bishop, C.: Neural networks for pattern recognition. Oxford University Press, New York (1995).
3. Subbotin, S.: The special deep neural network for stationary signal spectra classification. In: Proceedings of 14th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering (TCSET 2018), Slavske, pp. 123-128. IEEE, Los Alamitos (2018). doi: 10.1109/tcset.2018.8336170
4. Subbotin, S.: Neural network modeling of medications impact on the pressure of a patient with arterial hypertension. In: Proceedings of the International Conference on Information and Digital Technologies (IDT 2016), Zilina, pp. 249-260. IEEE, Los Alamitos (2016). doi: 10.1109/dt.2016.7557182
5. Subbotin, S.A.: The neural network model synthesis based on the fractal analysis. Optical Memory and Neural Networks (Information Optics), 26(4): 257-273 (2017). doi: 10.3103/s1060992x17040099
6. Kumar, K. V.: Neural networks and fuzzy logic. S. K. Kataria & Sons, New Delhi (2009).
7. Rutkowska, D.: Neuro-Fuzzy Architectures and Hybrid Learning. Studies in Fuzziness and Soft Computing, Springer, Berlin (2002).
8. Oliinyk, A.O., Zayko, T.A., Subbotin, S.O.: Synthesis of Neuro-Fuzzy Networks on the Basis of Association Rules. Cybernetics and Systems Analysis, 50(3): 348-357 (2014). doi: 10.1007/s10559-014-9623-7
9. Subbotin, S.: The neuro-fuzzy network synthesis and simplification on precedents in problems of diagnosis and pattern recognition. Optical Memory and Neural Networks (Information Optics), 22(2): 97-103 (2013). doi:10.3103/s1060992x13020082
10. Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J.: Classification and regression trees. Chapman and Hall, Wadsworth, New York (1984).
11. Rabcan, J., Rusnak, P., Subbotin, S.: Classification by fuzzy decision trees inducted based on Cumulative Mutual Information. In: Proceedings of 14th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering (TCSET 2018), pp. 208-212. IEEE, Los Alamitos (2018). doi: 10.1109/tcset.2018.8336188
12. Geurts, P., Irtuthum, A., Wehenkel, L.: Supervised learning with decision tree-based methods in computational and systems biology. Molecular Biosystems, 5(12):1593-1605 (2009).
13. Rabcan, J., Levashenko, V., Zaitseva, E., Kvassay, M., Subbotin, S.: Application of Fuzzy Decision Tree for Signal Classification. IEEE Transactions on Industrial Informatics, 15(10): 5425-5434 (2019). doi: 10.1109/tii.2019.2904845
14. Kamiński, B., Jakubczyk, M., Szufel, P.: A framework for sensitivity analysis of decision trees. Central European Journal of Operations Research, 26 (1): 135-159 (2017). doi:10.1007/s10100-017-0479-6
15. Rabcan, J., Levashenko, V., Zaitseva, E., Kvassay, M., Subbotin, S.: Non-destructive diagnostic of aircraft engine blades by Fuzzy Decision Tree. Engineering Structures, 197 (2019). doi: 10.1016/j.engstruct.2019.109396
16. Quinlan, J. R.: Induction of decision trees. Machine learning, 1(1):81-106 (1986).
17. Subbotin, S., Kirsanova, E.: The regression tree model building based on a cluster-regression approximation for data-driven medicine. CEUR Workshop Proceedings, 2255: 155-169 (2018).
18. Breiman, L.: Random Forests. Machine Learning, 45 (1): 5-32 (2001). doi:10.1023/A:1010933404324

19. Denisko, D., Hoffman, M.: Classification and interaction in random forests. *Proceedings of the National Academy of Sciences of the United States of America*, 115 (8): 1690–1692 (2018). doi:10.1073/pnas.1800256115.
20. Subbotin, S.: A random forest model building using a priori information for diagnosis. *CEUR Workshop Proceedings*, 2353: 962-973 (2019).
21. Boulesteix, A.-L., Janitza, S., Kruppa, J., König, I. R.: Overview of random forest methodology and practical guidance with emphasis on computational biology and bioinformatics. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 2(6):493–507 (2012).
22. Lin, Y., Yongho, J. Random forests and adaptive nearest neighbors. *Journal of the American Statistical Association*. 101 (474): 578–590 (2006). doi:10.1198/016214505000001230
23. Subbotin, S.: Methods of data sample metrics evaluation based on fractal dimension for computational intelligence model building. In: *Proceedings of 4th International Scientific-Practical Conference Problems of Infocommunications Science and Technology (PICS and T 2017)*, pp. 1-6 (2017). doi: 10.1109/infocommst.2017.8246136
24. Subbotin, S. A.: Models of criteria of comparison of neural networks and neuro-fuzzy networks in the problems of diagnosis and pattern classification. *Scientific Reports of Donetsk National Technical University. Serie "Informatics, Cybernetics and Computers"*, 12 (165): 148-151 (2010).
25. Subbotin, S. A.: Analysis of properties and criteria of comparison of neural network models for solving diagnostics and pattern recognition problems. *Data Registration, Saving and Processing*, 11 (3): 42–52 (2009).
26. Subbotin, S. A.: Methodics and criteria of comparison of models and algorithms of artificial neural network synthesis. *Radio Electronics, Computer Science, Control*, 2: 109–114 (2003).
27. Subbotin, S.A., Oliynyk, A.A.: The dimensionality reduction methods based on computational intelligence in problems of object classification and diagnosis. *Advances in Intelligent Systems and Computing*, 543: 11-19 (2017). doi: 10.1007/978-3-319-48923-0_2
28. Subbotin, S.: Quasi-relief method of informative features selection for classification. In: *Proceedings of IEEE 13th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT)*, pp. 318-321. IEEE, Los Alamitos (2018). doi: 10.1109/stc-csit.2018.8526627
29. Subbotin, S.A.: Methods of sampling based on exhaustive and evolutionary search. *Automatic Control and Computer Sciences*, 47(3): 113-121 (2013). doi: 10.3103/s0146411613030073
30. Subbotin, S., Oliynyk, A.: The sample and instance selection for data dimensionality reduction. *Advances in Intelligent Systems and Computing*, 543: 97-103 (2017). doi: 10.1007/978-3-319-48923-0_13
31. Subbotin, S.: The instance and feature selection for neural network based diagnosis of chronic obstructive bronchitis. *Studies in Computational Intelligence*, 606: 215-228 (2015). doi: 10.1007/978-3-319-19147-8_13
32. Ashby, W. R.: *An Introduction to Cybernetics*. Martino Fine Books, Eastford (2015).
33. Dopico, J. R., de la Calle, J. D., Sierra, A. P.: *Encyclopedia of artificial intelligence*. Information Science Reference, New York (2009).
34. Luger, G. F.: *Artificial Intelligence. Structures and Strategies for Complex Problem Solving*. Pearson Education, London (2011).
35. Nievergelt, Y.: The Concept of Elasticity in Economics. *SIAM Review*, 25 (2): 261–265 (1983). doi:10.1137/1025049.

36. Beven, K.J., Freer, J.: Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems. *Journal of Hydrology*, 249: 11–29 (2001).
37. Pedrycz, W.: *Fuzzy modelling: paradigms and practice*. Springer, Berlin (1996).
38. Hoekstra, A.: Generalisation in feed forward neural classifiers. Technische Universiteit Delft, Delft (1998).
39. Hoekstra A., Duin, R.: On the nonlinearity of pattern classifiers. In: *Proceedings of 13 International conference on Pattern recognition*, Vienna, 25-29 August 1996, vol. 4, pp. 271–275. IEEE, Los Alamitos (1996)
40. Grossberg, S.: Nonlinear neural networks: principles, mechanisms, and architectures. *Neural Networks*, 1(1): 17-61 (1988).
41. Subbotin, S. A.: The training set quality measures for neural network learning. *Optical Memory and Neural Networks (Information Optics)*, 19 (2): 126–139 (2010). doi: 10.3103/s1060992x10020037
42. Subbotin, S.A. The sample properties evaluation for pattern recognition and intelligent diagnosis. In: *10th International Conference on Digital Technologies (DT 2014)*, Zilina, pp. 321-332. IEEE, Los Alamitos (2014). doi: 10.1109/dt.2014.6868734
43. Weng T.-W., Zhang, H., Chen, P.-Yu, Yi, J., Su, D., Gao, Y., Hsieh C.-J., Daniel, L.: Evaluating the Robustness of Neural Networks: An Extreme Value Theory Approach. <https://arxiv.org/pdf/1801.10578>
44. Carlini, N., Wagner, D.: Towards Evaluating the Robustness of Neural Networks. In: *2017 IEEE Symposium on Security and Privacy (SP)*, vol. 1, pp. 39-57. IEEE, Los Alamitos (2017). DOI:10.1109/SP.2017.49
45. Krus, D.J., Blackman, H.S., Test reliability and homogeneity from perspective of the ordinal test theory. *Applied Measurement in Education*, 1: 79–88 (1988).
46. Alippi, C., Piuri, V., Sami, M.: Sensitivity to errors in artificial neural networks: a behavioral approach. *IEEE transactions on circuits and systems – I: Fundamental theory and applications*, 42 (6): 358–361 (1995).
47. Hashem, S.: Sensitivity analysis for feedforward artificial neural networks with differentiable activation functions. In: *Proceedings of International Joint Conference on Neural Networks*, Baltimore, 7-11 June 1992, vol. I., pp. 419–424. IEEE, Los Alamitos (1992).
48. Tao C.-W., Nguyen H.T., Yao J.T., Kreinovich V.: Sensitivity analysis of neural control. In: *Proceedings of Fourth International Conference on Intelligent Technologies*, Chiang Mai, 17-19 December 2003, pp. 478–482. Chiang Mai University, Chiang Mai (2003).
49. Gerrow, K. Synaptic stability and plasticity in a floating world. *Current Opinion in Neurobiology*, 20 (5): 631–639 (2010). doi:10.1016/j.conb.2010.06.010. PMID 20655734
50. Meyer, D., Bonhoeffer, T., Scheuss, V.: Balance and Stability of Synaptic Structures during Synaptic Plasticity. *Neuron*, 82 (2): 430–443 (2014).
51. Kick, D.R., Schulz, D.J.: Variability in neural networks. *eLife*, 7: e341532018 (2018). doi: 10.7554/eLife.34153
52. Norris, B.J., Wenning, A., Wright, T.M., Calabrese, R.L.: Constancy and variability in the output of a central pattern generator. *Journal of Neuroscience*, 31:4663–4674 (2011). doi: 10.1523/JNEUROSCI.5072-10.2011
53. Masquelier, T.: Neural variability, or lack thereof. *Front. Comput. Neurosci.*, 7: 7 (2013). doi: 10.3389/fncom.2013.00007
54. Rusiecki, A., Kordos, M., Kamiński, T., Greń, K.: Training Neural Networks on Noisy Data. In: *International Conference on Artificial Intelligence and Soft Computing (ICAISC 2014)* pp. 131-142. Springer, Cham (2014).

55. Martín J.A., de Lope, J., Maravall, D.: Adaptation, Anticipation and Rationality in Natural and Artificial Systems: Computational Paradigms Mimicking Nature. *Natural Computing*, 8(4): 757-775 (2009).
56. Mainzer, K.: *Symmetry And Complexity: The Spirit and Beauty of Nonlinear Science*. World Scientific, Singapore (2005).
57. Bunge, M. A.: *Emergence and Convergence: Qualitative Novelty and the Unity of Knowledge*, University of Toronto Press, Toronto (2003).
58. Wan, P.: Emergence à la Systems Theory: Epistemological Totalausschluss or Ontological Novelty? *Philosophy of the Social Sciences*, 41 (2): 178–210 (2011). doi:10.1177/0048393109350751
59. Japaridze, G., De Jongh, D. The logic of provability. in Buss, S., ed., *Handbook of Proof Theory*, pp. 476-546 North-Holland, Amsterdam (1998).
60. Molnar, Ch.: *Interpretable Machine Learning. A Guide for Making Black Box Models Explainable*. <https://christophm.github.io/interpretable-ml-book/>
61. Valiant, L.: A theory of the learnable. *Communications of the ACM*, 27 (11): 1134–1142 (1984). doi:10.1145/1968.1972.
62. Iuliano, E.: A Comparative Evaluation of Surrogate Models for Transonic Wing Shape Optimization. In: Andrés-Pérez, E., González, L.M., Periaux, J., Gauger, N.R., Quagliarella, D., Giannakoglou, K.C., eds., *Evolutionary and Deterministic Methods for Design Optimization and Control With Applications to Industrial and Societal Problems*, pp. 161-180. Springer, Cham (2019).
63. Scaman, K., Virmaux, A.: Lipschitz regularity of deep neural networks: analysis and efficient estimation. In: 32nd Conference on Neural Information Processing Systems (NeurIPS 2018), Montréal, Canada. <https://papers.nips.cc/paper/7640-lipschitz-regularity-of-deep-neural-networks-analysis-and-efficient-estimation.pdf>
64. Lipschitz constant. *Encyclopedia of Mathematics*. http://www.encyclopediaofmath.org/index.php?title=Lipschitz_constant&oldid=30687
65. Russell, S., Norvig, P.: *Artificial intelligence: a modern approach*. Prentice Hall, Upper Saddle River (2009).
66. Hannan, E. J., Quinn, B. G.: The determination of the order of an autoregression. *Journal of the Royal Statistical Society, Ser. B*, 41(2): 190–195 (1979).
67. Schwarz, G. E.: Estimating the dimension of a model. *Annals of Statistics*, 6 (2): 461–464 (1978).
68. Akaike, H.: A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19 (6): 716–723 (1974).
69. Gheissari, N., Bab-Hadiashar, A.: Model selection criteria in computer vision: are they different? In: *Proceedings of Digital Image Computing: Techniques and Applications*, Sydney, 10-12 December 2003, pp. 185-194. CSIRO, Collingwood (2003).
70. Grünwald, P., Myung J., Pitt, M.: *Advances in Minimum Description Length: Theory and Applications*. MIT Press, Cambridge (2005).
71. Rissanen, J.: Modeling by shortest data description. *Automatica*, 14 (5): 465-471 (1978).
72. Mallows, C. L.: Some Comments on CP. *Technometrics*, 15 (4): 661–675 (1973). doi:10.2307/1267380

Online Credibilistic Fuzzy Clustering of Data Using Membership Functions of Special Type

Alina Shafronenko¹[0000-0002-8040-0279], Yevgeniy Bodyanskiy¹[0000-0001-5418-2143], Iryna Klymova¹[0000-0003-0455-6180], Olexii Holovin²[0000-0003-4662-4559]

¹Kharkiv National University of Radio Electronics, Nauky Ave., 14, Kharkiv, Ukraine
alina.shafronenko@nure.ua, yevgeniy.bodyanskiy@nure.ua,
iryna.klymova@nure.ua

²Central Research Institute of Weapons and Military Equipment of the Armed Forces
of Ukraine, Povitroflotsky Ave., 28, Kyiv, Ukraine
a_a_golovin@ukr.net

Abstract. In the paper new online method of credibilistic fuzzy clustering of data was proposed. This algorithm is based on credibilistic approaches using on batch and online modes of information processing. Using proposed approach it's possible to solve clustering task in on-line mode when data are fed to processing sequentially, possible in real time.

Keywords: fuzzy clustering, learning rule, possibilistic fuzzy clustering, probabilistic fuzzy clustering, credibilistic fuzzy clustering, measure of similarity.

1 Introduction

The task of clustering (classification in the self-learning mode) of multidimensional data is an important part of Data Mining, within which a number of directions and approaches have developed [1, 2]. One of these areas is formed by fuzzy clustering methods, that are based on the assumption that the generated clusters - classes mutually overlap so that each vector - observation with different levels of membership-probability-possibility can belong to several or to all classes.

Here, the algorithms of probabilistic fuzzy clustering and, first of all, the fuzzy c-means method (FCM) are most widely used [3, 4]. The possibilities of this approach are limited by probabilistic restrictions on membership levels so that observations "contaminated" with disturbances and outliers can be assigned to different classes with almost identical membership levels.

In this regard, in [5] was proposed the possibilistic fuzzy clustering method (PCM) that is more resistant to noise and disturbances. At the same time, PCM algorithms suffer from the so-called coincident problem, when during the processing of information some clusters begin to merge with each other, that leads to an incorrect estimate of the number of formed clusters.

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

Algorithms of credibilistic fuzzy clustering [6-8], based on the apparatus of the theory of credibility [9], are devoid of these shortcomings. As part of this approach in the calculate process to evaluate not only the fuzzy membership levels, but also credibility levels based on a membership measure of a special type [10]. The experimental results have shown [7, 8] that the credibilistic methods provide a higher quality of clustering comparatively with probabilistic and possibilistic methods.

The initial information for solving the problem of fuzzy clustering is an array of n - dimensional observations - vectors $X = \{x_1, x_2, \dots, x_N\} \subset R^n$, $x(k) \in X$, $k = 1, 2, \dots, N$, that should be divided into m classes-clusters with a certain level of membership – probability – possibility $U_q(k)$ of k^{th} vector x_k to q^{th} cluster ($1 < m < N, 1 \leq q \leq m$). It should also be noted that the initial data are pre-processed so that $-1 \leq x_{ki} \leq 1$ ($1 \leq i \leq n$) where x_{ki} – i^{th} component of vector x_k .

Thus, the clustering problem is solved in batch mode, when the entire data array is processed multiple times based on alternating cluster estimation [8]. If the data arrive for processing in the form of a stream or form big data, the batch mode does not allow to solve the problem under consideration effectively. In this situation, the most effective are the recursive fuzzy clustering procedures that allow to solve the problem online and refine the desired solution as each new observation that arrives. Thus, in the [11, 12] recurrent variants of FCM have been proposed that are essentially gradient optimization procedures adopted by the goal function, and in [13, 14] recurrent PCM modifications have been introduced designed for sequential data processing.

In this regard, it seems appropriate to develop a recurrent modification of the method of credibilistic fuzzy clustering, that allows clarifying the desired characteristics of the clusters as each new observation arrives.

2 Recurrent Method Of Credibilistic Fuzzy Clustering (RCCM)

The most popular method of probabilistic fuzzy clustering is associated with minimizing the goal function (1) [4]

$$E(U_q(k), w_q) = \sum_{k=1}^N \sum_{q=1}^m U_q^\beta(k) D^2(x_k, w_q) \quad (1)$$

with constrains $\sum_{q=1}^m U_q(k) = 1$, $0 < \sum_{q=1}^m U_q(k) < N$. Solving the nonlinear programming problem using the method of indefinite Lagrange multipliers, we arrive to the known result (2) and (3)

$$U_q^{(\tau+1)}(k) = \left(D^2(x_k, w_q^{(\tau)}) \right)^{\frac{1}{1-\beta}} \left(\sum_{l=1}^m \left(D^2(x_k, w_l^{(\tau)}) \right)^{\frac{1}{1-\beta}} \right)^{-1}, \quad (2)$$

$$w_q^{(\tau+1)} = \sum_{k=1}^N \left(U_q^{(\tau+1)}(k) \right)^\beta x_k \left(\sum_{k=1}^N \left(U_q^{(\tau+1)}(k) \right)^\beta \right)^{-1}, \quad (3)$$

where $U_q(k)$ – membership levels of vector observation x_k to q^{th} cluster $Cl_q (1 \leq q \leq m)$, w_q – prototype – centroid of q^{th} cluster, $\beta > 1$ fuzzifier, that defines the “blurring” of boundaries between classes, $D(x_k, w_q)$ – the distance between x_k and w_q in adopted metric, $\tau = 0, 1, 2, \dots$ – index of the epoch of information processing in the alternate estimation mode. In this case, the calculation process continues until the conditions

$$w_q^{(\tau+1)} - w_q^{(\tau)} \leq \varepsilon \forall 1 \leq q \leq m$$

is satisfied, where ε – reassigned threshold calculation accuracy.

In the case, when $\beta = 2$ and Euclidean metric $D^2(x_k, w_q) = \|x_k - w_q\|_2^2$ is used, we obtain to the popular fuzzy c-means algorithm (FCM) [15] in the form (4), (5)

$$U_q^{(\tau+1)}(k) = \frac{\|x_k - w_q^{(\tau)}\|^{-2}}{\sum_{l=1}^m \|x_k - w_l^{(\tau)}\|^{-2}}, \quad (4)$$

$$w_q^{(\tau+1)} = \sum_{k=1}^N \left(U_q^{(\tau+1)}(k) \right)^2 x_k \left(\sum_{k=1}^N \left(U_q^{(\tau+1)}(k) \right)^2 \right)^{-1}. \quad (5)$$

If the data are processed sequentially online, the nonlinear programming task can be solved using the Arrow-Hurwitz-Uzawa algorithm, which is essentially a gradient procedure for finding the saddle point of the Lagrange function based on criterion (1) with constraints on the sum of membership levels.

In this case relations (2), (3) can be rewritten in the form (6)

$$\begin{cases} U_q(k+1) = \left(D^2(x_{k+1}, w_q^{(k)}) \right)^{\frac{1}{1-\beta}} \left(\sum_{l=1}^m (x_{k+1}, w_l(k))^{\frac{1}{1-\beta}} \right)^{-1}, \\ w_q(k+1) = w_q(k) + \eta(k+1) U_q^\beta(k+1) (x_{k+1} - w_q(k)) \end{cases}, \quad (6)$$

where $\eta(k)$ – learning rate parameter, and (4), (5) can be rewritten in the form (7)

$$\begin{cases} U_q(k+1) = \frac{\|x_k - w_q(k)\|^{-2}}{\sum_{l=1}^m \|x_k - w_l(k)\|^{-2}}, \\ w_q(k+1) = w_q(k) + \eta(k+1) U_q^2(k+1) (x_{k+1} - w_q(k)) \end{cases}, \quad (7)$$

which are a generalization of the recurrent procedures of Park-Dagger [11] and Chung-Lee [12].

Possibilistic algorithms for fuzzy clustering are based on minimizing the goal function (8) [5]

$$E(U_q(k), w_q, \mu_q) = \sum_{k=1}^N \sum_{q=1}^m U_q^\beta(K) D^2(x_k, w_q) + \sum_{q=1}^m \mu_q \sum_{k=1}^N (1 - U_q(k))^\beta, \quad (8)$$

where $\mu_q \geq 0$ determines the distance at which the membership level takes the value 0.5, i.e. $U_q(k) = 0$ if $D^2(x_k, w_q) = \mu$.

Minimization of criterion (8) allows us to obtain analytical solution in the form (9) – (11)

$$U_q^{(\tau+1)}(k) = \left(1 + \left(\frac{D^2(x_k, w_q^{(\tau)})}{\mu_q^{(\tau)}} \right)^{\frac{1}{\beta-1}} \right)^{-1}, \quad (9)$$

$$w_q^{(\tau+1)} = \sum_{k=1}^N (U_q^{(\tau+1)}(k))^\beta x_k \left(\sum_{k=1}^N (U_q^{(\tau+1)}(k))^\beta \right)^{-1}, \quad (10)$$

$$\mu_q^{(\tau+1)} = \sum_{k=1}^N (U_q^{(\tau+1)}(k))^\beta D^2(x_k, w_q^{(\tau+1)}) \left(\sum_{k=1}^N (U_q^{(\tau+1)}(k))^\beta \right)^{-1}, \quad (11)$$

which in the quadratic case takes the form (12) – (14)

$$U_q^{(\tau+1)}(k) = \left(1 + \frac{\|x_k - w_q^{(\tau)}\|^2}{\mu_q^{(\tau)}} \right)^{-1}, \quad (12)$$

$$w_q^{(\tau+1)} = \sum_{k=1}^N (U_q^{(\tau+1)}(k))^2 x_k \left(\sum_{k=1}^N (U_q^{(\tau+1)}(k))^2 \right)^{-1}, \quad (13)$$

$$\mu_q^{(\tau+1)} = \sum_{k=1}^N (U_q^{(\tau+1)}(k))^2 \|x_k - w_q^{(\tau+1)}\|^2 \left(\sum_{k=1}^N (U_q^{(\tau+1)}(k))^2 \right)^{-1}. \quad (14)$$

Online versions (9) – (14) at the same time have the form (15) [13, 14]

$$\begin{cases} U_q(k+1) = \left(1 + \left(\frac{D^2(x_{k+1}, w_q(k))}{\mu_q(k)} \right)^{\frac{1}{\beta-1}} \right)^{-1}, \\ w_q(k+1) = w_q(k) + \eta(k+1) U_q^\beta(k+1) (x_{k+1} - w_q(k)), \\ \mu_q(k+1) = \sum_{p=1}^{k+1} U_q^\beta(p) D^2(x_p, w_q(k+1)) \left(\sum_{p=1}^{k+1} U_q^\beta(p) \right)^{-1} \end{cases} \quad (15)$$

and in the case $\beta = 2$ (16)

$$\begin{cases} U_q(k+1) = \left(1 + \frac{\|x_{k+1} - w_q(k)\|^2}{\mu_q(k)} \right)^{-1}, \\ w_q(k+1) = w_q(k) + \eta(k+1)U_q^2(k+1)(x_{k+1} - w_q(k)), \\ \mu_q(k+1) = \sum_{p=1}^{k+1} U_q^2(p) \|x_p - w_q(k+1)\|^2 \left(\sum_{p=1}^{k+1} U_q^2(p) \right)^{-1}. \end{cases} \quad (16)$$

Credibilistic fuzzy clustering is associated with minimizing the goal function (17)

$$E(Cr_q(k), w_q) = \sum_{k=1}^N \sum_{q=1}^m Cr_q^\beta(k) D^2(x_k, w_q) \quad (17)$$

with constraints $0 \leq Cr_q(k) \leq 1 \forall q, k$; $\sup Cr_q(k) \geq 0, 5 \forall k$; $Cr_q(k) + \sup Cr_l(k) = 1$ for any q and k for which $Cr_q(k) \geq 0.5$. Here $Cr_q(k)$ – credibility that observation x_k belongs to a cluster Cl_q . In this case, the membership level is calculated based on the membership function (18) [15]

$$U_q(k) = \varphi_q(D(x_k, w_q)) \quad (18)$$

where: $\varphi_q(\cdot)$ – monotonically decreases in the interval $[0, \infty]$, $\varphi_q(0) = 1$, $\varphi_q(\infty) \rightarrow 0$.

It is easy to see that function (18) is essentially measure of similarity based on distance [16]. As such a function, it was proposed in [15] to use the expression (19)

$$U_q(k) = \left(1 + D^2(x_k, w_q) \right)^{-1}. \quad (19)$$

It is interesting to note that expression (2) can be rewritten in the form (20)

$$\begin{aligned} U_q(k) &= \left(D^2(x_k, w_q(k)) \right)^{\frac{1}{1-\beta}} \left(\sum_{l=1}^m \left(D^2(x_k, w_l(k)) \right)^{\frac{1}{1-\beta}} \right)^{-1} = \\ &= \left(D^2(x_k, w_q(k)) \right)^{\frac{1}{1-\beta}} \left(D^2(x_k, w_q(k)) \right)^{\frac{1}{1-\beta}} + \sum_{\substack{l=1 \\ l \neq q}}^m \left(D^2(x_k, w_l(k)) \right)^{\frac{1}{1-\beta}} \right)^{-1} = \\ &= \left(1 + \left(D^2(x_k, w_q(k)) \right)^{\frac{1}{1-\beta}} \sum_{\substack{l=1 \\ l \neq q}}^m \left(D^2(x_k, w_l(k)) \right)^{\frac{1}{1-\beta}} \right)^{-1}, \end{aligned} \quad (20)$$

and for the Euclidean metric and $\beta = 2$ takes the form of a Cauchy distribution density function with a width parameter σ_q^2 (21), (22), [17]

$$U_q(k) = \left(1 + \frac{\|x_k - w_q(k)\|^2}{\sigma_q^2} \right)^{-1}, \quad (21)$$

$$\sigma_q^2 = \left(\sum_{\substack{l=1 \\ l \neq q}}^m \|x_k - w_l(k)\|^{-2} \right)^{-1}. \quad (22)$$

It is easy to see that the membership function (19) is a special case of (21) for $\sigma_q^2 = 1$.

Finally, a batch algorithm of credibilistic fuzzy clustering can be written in the form (23) – (26) [7, 8]:

$$U_q^{(\tau+1)}(k) = \left(1 + D^2(x_k, w_q^{(\tau)}) \right)^{-1}, \quad (23)$$

$$U_q^{*(\tau+1)}(k) = U_q^{(\tau+1)}(k) \left(\sup_l U_l^{(\tau+1)}(k) \right)^{-1}, \quad (24)$$

$$Cr_q^{(\tau+1)}(k) = \frac{1}{2} \left(U_q^{*(\tau+1)}(k) + 1 - \sup_{l \neq q} U_l^*(k) \right), \quad (25)$$

$$w_q^{(\tau+1)} = \sum_{k=1}^N \left(Cr_q^{(\tau+1)}(k) \right)^\beta x_k \left(\sum_{k=1}^N \left(Cr_q^{(\tau+1)}(k) \right)^\beta \right)^{-1}. \quad (26)$$

Based on (17), (21) – (26), we can write online version of the credibilistic fuzzy clustering method in the form (27)

$$\left\{ \begin{array}{l} \sigma_q^2(k+1) = \left(\sum_{\substack{l=1 \\ l \neq q}}^m \|x_{k+1} - w_l(k)\|^{-2} \right)^{-1}, \\ U_q(k+1) = \left(1 + \frac{\|x_{k+1} - w_q(k)\|^2}{\sigma_q^2(k+1)} \right)^{-1}, \\ U_{(k+1)}^* = U_q(k+1) \left(\sup_l U_l(k+1) \right)^{-1}, \\ Cr_q(k+1) = \frac{1}{2} \left(U_q^*(k+1) + 1 - \sup_{l \neq q} U_l^*(k+1) \right), \\ w_q(k+1) = w_q(k) + \eta(k+1) Cr_q^\beta(k+1) (x_{k+1} - w_q(k)). \end{array} \right. \quad (27)$$

Therefore, from a computational point of view, the online algorithm for credibilistic fuzzy clustering is no more complicated than the recurrent versions of FCM and PCM, while retaining the advantages of a credibility approach.

3 Experiments

To check the performance efficiency of the developed methods as well as to prove their benefits over the analogs, experimental research was conducted with the help on two different databases. Also conducted a comparative analysis of the quality of clustering data on the main characteristics quality ratings, such as: Partition Coefficient (PC) defines "overlapping" between groups of points, Partition Index (SC) quantifies the ratio sum of compactness and separation of the clusters, Xie and Beni's Index (XB) gauges a ratio of the total variability inside clusters and their separation, of existing clustering methods and proposed method.

The experimental results are given in Table 1 and Table 2.

Such clustering tools as fuzzy c-means (FCM) and the algorithm by Gustafson-Kessel (GK), Gath-Geva (GG), Adaptive probabilistic fuzzy clustering, Adaptive fuzzy possibilistic data clustering and Adaptive fuzzy credibilistic data clustering were chosen for comparison with the developed procedure.

Table 1. Evaluation of the quality of fuzzy clustering methods using first data set

Data clustering methods	First Data Set		
	PC	SC	XB
Fuzzy C-means (FCM)	0.50	1.62	0.19
Gustafson-Kessel	0.27	1.66	1.62
Gath-Geva	0.25	1.54	1.35
Adaptive probabilistic fuzzy clustering	0.25	1.44	0.01
Adaptive fuzzy possibilistic data clustering	0.26	1.22	0.01
Adaptive fuzzy credibilistic data clustering	0.21	1.13	0.01

Table 2. Evaluation of the quality of fuzzy clustering methods using second data set

Data clustering methods	Second Data Set		
	PC	SC	XB
Fuzzy C-means (FCM)	0.48	1.60	0.19
Gustafson-Kessel	0.26	1.64	1.62
Gath-Geva	0.26	1.50	1.35
Adaptive probabilistic fuzzy clustering	0.25	1.42	0.01
Adaptive fuzzy possibilistic data clustering	0.37	1.11	0.18
Adaptive fuzzy credibilistic data clustering	0.23	1.22	0.01

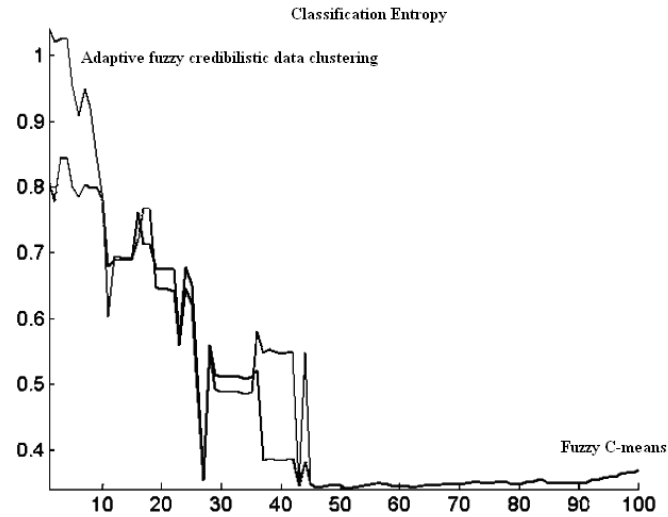


Fig. 1. Comparison CE of Adaptive fuzzy credibilistic data clustering and Fuzzy C-means (FCM) methods for First Data Set

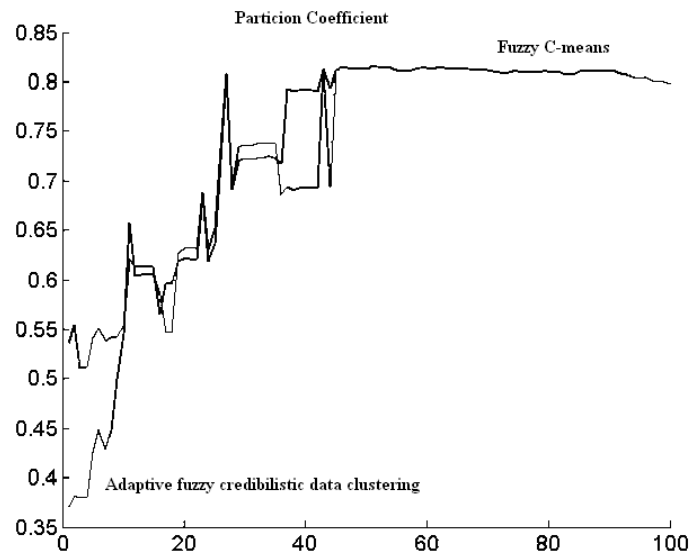


Fig. 2. Comparison PC of Adaptive fuzzy credibilistic data clustering and Fuzzy C-means (FCM) methods for First Data Set

In comparison with the well-known methods, this approach to data clustering demonstrates somewhat reliable results (Fig. 1, Fig. 2, Fig. 3, Fig. 4).

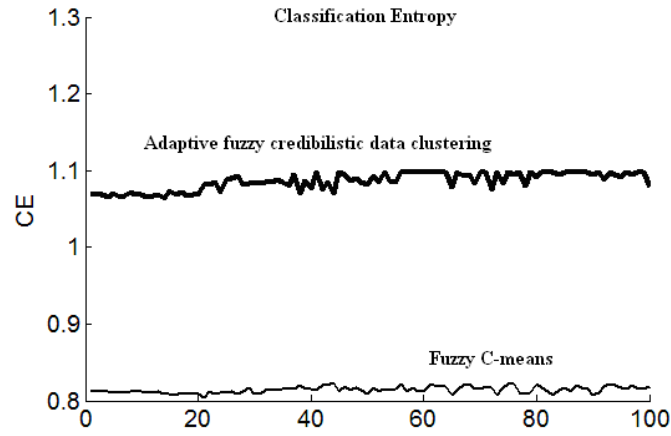


Fig. 3. Comparison CE of Adaptive fuzzy credibilistic data clustering and Fuzzy C-means (FCM) methods for Second Data Set

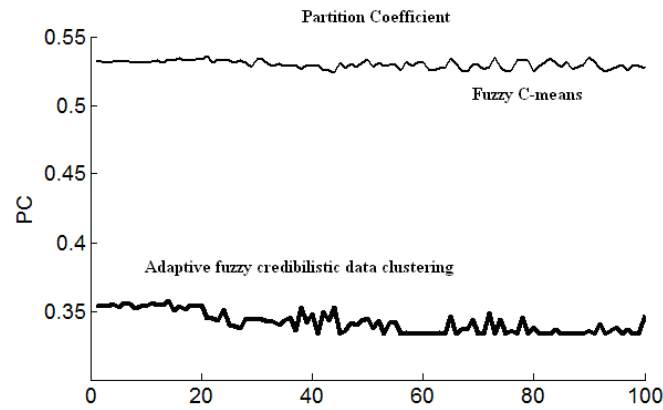


Fig. 4. Comparison PC of Adaptive fuzzy credibilistic data clustering and Fuzzy C-means (FCM) methods for Second Data Set

4 Conclusion

The problem of fuzzy clustering based on probabilistic, possibilistic and credibilistic approaches based online mode of information processing was considered. A recurrent version of credibilistic algorithm is introduced, which is essentially a gradient optimization procedure for the accepted criterion for fuzzy credibilistic clustering. A modification of the membership function is introduced, which is essentially a measure of similarity and a generalization of previously known functions. The considered

recursive procedures are simple in numerical implementation and are intended to solve problems arising in the framework of Data Stream Mining.

References

1. Xu, R., Wunsch, D.C.: Clustering. Hoboken, N.J. John Wiley & Sons, Inc. (2009)
2. Aggarwal, C.C.: Data Mining: Text Book. Springer (2015).
3. Bezdek, J.C.: Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum Press, New York (1981).
4. Höppner, F., Klawonn, F., Kruse, R., Runkler, T.: Fuzzy Clustering Analysis: Methods for Classification, Data Analysis and Image Recognition. Chichester, John Wiley & Sons (1999)
5. Krishnapuram, R., Keller, J.M.: A possibilistic approach to clustering. Fuzzy Systems, №2, pp.98-110 (1993)
6. Chintalapudi, K. K., Kam, M.: A noise resistant fuzzy c-means algorithm for clustering. In: IEEE Conference on Fuzzy Systems Proceedings, vol. 2, pp. 1458-1463 (1998)
7. Zhou, J., Wang, Q., Hung, C.-C., Yi, X.: Credibilistic clustering: the model and algorithms. Int.J. of Uncertainty, Fuzziness and Knowledge-Based Systems, №4, pp.545-564 (2015).
8. Zhou, J., Wang, Q., Hung, C. C.: Credibilistic clustering algorithms via alternating cluster estimation. J. Intell. Manuf., pp.727-738 (2017).
9. Liu, B., & Liu, Y. Expected value of fuzzy variable and fuzzy expected value models. In: "IEEE Transactions on Fuzzy Systems", № 4, pp. 445–450 (2002).
10. Liu, B.: A survey of credibility theory. Fuzzy Optimization and Decision Making, №4, pp. 387–408 (2006).
11. Park, D.C., Dagher, I.: Gradient based fuzzy c-means (GBFCM) algorithm. In: Proc. IEEE Int. Conf. on Neural Networks, pp.1626-1631(1984).
12. Chung, F.L., Lee,T.: Fuzzy competitive clustering. Neural Networks, 7(3), pp.539-552. (1994).
13. Bodyanskiy, Ye, Kolodyazhniy, V., Stephan, A.: Recursive fuzzy clustering algorithms. In : Proc 10th East West Fuzzy Coll. Zittau-Görlitz, HS, pp.276-283 (2002).
14. Bodyanskiy, Ye.: Computational intelligence techniques for data analysis. In: Lecture Notes in Informatics. Bonn: V.P. – 72, GI, pp.15-36 (2005).
15. Zhou, J., & Hung, C.-C.: A generalized approach to possibilistic clustering algorithms. Int. J. of Uncertainty, Fuzziness and Knowledge-Based Systems. 15. pp. 117–138. (2007).
16. Young, F.W., Hamer, R.M.: Theory and Applications of Multidimensional Scaling-Hillsdale, N.J.: Erlbaum (1994).
17. Hu, Zh., Bodyanskiy, Ye, Tyshchenko, O., Shafronenko, A.: Fuzzy clustering of incomplete data by means of similarity measures. In: 2019 IEEE 2nd Ukraine Conference on Electrical and Computer Engineering UKRCON - 2019, Conference Proceedings, July 2-6, 2019, Lviv, Ukraine, pp.149-152 (2019) doi: 10.1109/UKRCON.2019.8879844.

Intelligent Data Analysis in HR Process Management

Mykola Ivanov ^{1[0000-0002-1908-0763]}, Nataliia Maksyshko ^{1[0000-0002-0473-7195]},
Sergey Ivanov ^{1[0000-0003-1086-0701]} and Nataliia Terentieva ^{1[0000-0001-6930-879X]}

¹Zaporizhzhya National University, Zhukovsky str., 66, Zaporizhzhia, 69063, Ukraine
nn_iva@ukr.net, maxishko@ukr.net, flydaiver@gmail.com,
terenteva_nataliya@ukr.net

Abstract. This article discusses the problem of constructing a method of data mining in HR process management and modeling the assessment of the degree of personnel efficiency. A method of data mining in the management of HR processes in the form of an algorithm with a fuzzy inference procedure implemented in the MATLAB R2017a system is proposed, which made it possible to assess the degree of personnel work efficiency.

This method and procedures of data mining allows you to simulate multidimensional HR processes.

The application of the method and procedures of data mining in HR-process management allows us to develop problem solving approaches.

Keywords: intelligent data analysis, HR management systems, fuzzy sets, fuzzy inference procedure, fuzzy modeling, procedure

1 Introduction

Today, business is forced to solve a whole range of complex and unique tasks. So, to solve the problems of increasing and stabilizing the management of economic facilities in modern conditions, new approaches and solutions are required, which determined the emergence of a new concept of the German economist Klaus Schwab, President of the World Economic Forum in Davos [1]. According to this concept, it is stated that we live in the era of the fourth industrial revolution, where the virtual world is combined with the physical world using information technology. The fourth industrial revolution is characterized by a change in economic relations and the widespread use of intelligent technologies (Cloud technologies, Big Data, artificial neural networks and fuzzy sets, data mining and others), which is the basis of the digital economy.

It should be noted that there are various approaches to the determination and measurement of the dynamics of the digital economy, and it is also difficult to assess its volume. According to estimates [2], the share of the digital segment of the global economy in 2020 could be 23% (\$ 17 trillion). This leads to the expansion and rapid development of the Internet market using Big Data, followed by modeling by fuzzy logic approaches.

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

2 Formal problem statement

Therefore for successful management of economic facilities in the digital economy an important role is given to HR management systems (Human resources) which are able to offer a radically new mechanism that allows companies to remain competitive in the market.

3 Literature review

The first direction is determined by the level of application of intelligent systems. High information technologies embrace the world and replace the classical methods of HR-process management. Robot programs are now being applied, offering employees of the enterprise to undergo an express interview or interview using yet expert systems.

The next direction is the analysis and evaluation of young professionals who are able to learn and solve modern problems. Now, new specialists up to 25 years of age who have completely different knowledge, interests and fundamentally different ideas about modern work are entering the labor market. Young specialists are able to quickly make managerial decisions and push new projects forward, and thereby increase the level of work in the enterprise.

In addition to these areas, it should be noted that for a long time HR management was aimed at standardization and versatility. However, this approach is gradually becoming obsolete today. This approach is being replaced by HR management methods that are focused on the maximum use of the intellectual capital of employees. This is stimulated with the simultaneous satisfaction of individual needs, desires, employee capabilities and their synchronization with the tasks of the enterprise. Modern HR specialists are beginning to more closely monitor the development of employees within the enterprise, which allows flexible management of career growth, which can be adjusted taking into account the proposals of the employees themselves.

HR strategy is part of the overall strategy of enterprises and long-term planning of their business activities. An important role in these plans is played by assessments of the degree of personnel efficiency as a factor in updating and increasing production efficiency in the overall economic strategy of an enterprise. The creation of modern catalogs of employee data and their management requires the processing of a large amount of information. This is due to a wide range of organizational, economic and technical and technological tasks that are performed by personnel. Therefore, data analysis in the management of HR processes is an urgent task.

The influence of information technology in the management of economic systems was presented in the work of M. Ivanov [3]. However, the problems of rating management have not been resolved. The solution to the problem of rating management was the work of Yu. Lysenko, V. Petrenko L. [4]. However, the level of staff development and their assessment were not considered. This work was devoted to the solution of this problem, in which the theoretical aspects of personnel development are studied, in particular, the concept, main tasks and directions of personnel develop-

ment at the enterprise. V. Helman [5] considers the development of enterprise personnel as a change in its qualitative characteristics, in which indicators in the form of degree of activity were proposed.

Human resource management as a strategic human resource management today is seen as going beyond the boundaries of management tasks, such as motivation, the level of remuneration. Instead, managers need to consider HR management as a process that contributes to the success of the enterprise. Therefore, the work of Brian E. Becker [6] considers approaches where all managers should be involved in the management process, where the role of employees is important for the competitive advantage of the enterprise. In addition, organizations that value their skilled employees are more profitable than those that do not. The results of the works of scientists Mark A. Huselid [7], Jeffrey Pfeffer and John F. Veiga [8] show that successful enterprises have several common characteristics: stable job security, the use of self-government methods, perfect pay and access to information. These tasks were considered in the work of Brian E. Becker, Mark A. Huselid [9], when an enterprise develops and motivates the development of human capital. The most successful enterprises manage HR as a strategic asset, and evaluate the effectiveness of labor resources in terms of its impact. Dennis R. Briscoe [10] in his work write about situation when each employee of the enterprise effectively fulfills his duties and creates a high-performance work system in which the employee has the maximum involvement and responsibility.

In modern enterprises, an important task is to ensure a balance between the need for coordination and synchronization of units located both in different cities and around the world, which was reviewed by Randall S. Schuler, Pawan S. Budhwar, and Gary W. Florkowski [11]. Achieving this balance is becoming increasingly difficult as the level of functional diversity to which enterprises are exposed increases.

Today, new intelligent technologies are emerging to solve these problems. Such intelligent technologies include research and project management recommendations. Such works include the works of S. Gottwald [12] and N.Vedula, P. Nicholson [13].

These projects on improvement of resource processes are the ones of the most important tasks of production in enterprise management. Currently, the number of works in this direction is increasing by the authors W. Siler, J. Buckley [14], J. Waitelonis, C. Exeler, H. Sack [15] and A. Abraham [16]. The obvious advantage of the proposed methods is the use of mathematical tools, the theory of fuzzy sets, which allows implementing the process of selecting conditions, the introduction of the method in the managerial processes of the enterprise. It should be noted that this works did not solve the problem of data mining in HR-process management with the possibility of modeling the degree of personnel efficiency, which determined the topic and relevance of this article.

4 The purpose of the article

The article is devoted to development of a sequence of procedures for data mining in human resource management using fuzzy logic approaches.

5 Data Mining method in HR process management

The task of data mining in the management of HR processes is the effective extraction and analysis of the existing data array of employees with subsequent personnel management using cloud solutions. This will allow the rapid implementation of a new personnel management system, obtaining a new level of accessibility and increasing its mobility.

The resulting performance indicators of personnel at the enterprise can be represented in the form of multidimensional structures, where each measurement is represented by the corresponding indicators of the enterprise management system.

The following sequence diagram (the method) of data analysis in HR process management is proposed, which is presented in Fig. 1.

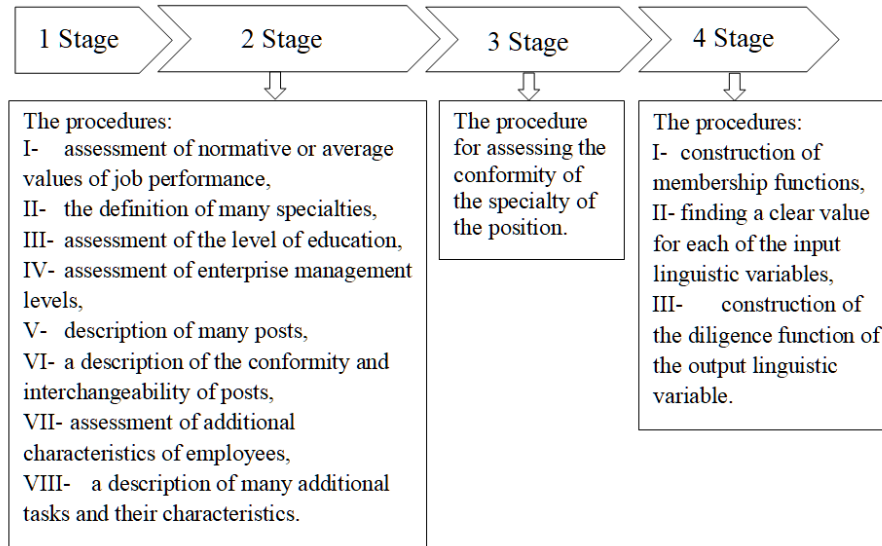


Fig. 1. Sequential data mining diagram in HR process management.

The proposed method includes four stages.

5.1 Stage 1 Procedures

The first stage solves the problem of choosing the analyzed indicators. For this, a lot of ratings are determined:

$$P = \{P_i = (op1, op2, ep1, ep2, ep3)\}, i = \overline{1, N}, \quad (1)$$

where *op1* is a generalized indicator of job compliance, characterizing the degree of conformity of qualifications and work experience of the post, level of responsibility, as well as the quality of the performance of current work and duties,

$op2$ - a generalized indicator of diligence, characterizing the effectiveness of the tasks (complexity, quality, timeliness),
 $ep1$ - ambitiousness, a single indicator of personality characteristics,
 $ep2$ - the quality of a leader, an indicator of personality characteristics,
 $ep3$ - the level of attitude in the team, a single indicator of personality characteristics.

5.2 Stage 2 Procedures

At the second stage, the initial information is determined, which is necessary for calculating indicators based on expert assessments, analytical indicators (for example, work experience, quality of work performed, and others).

To describe the formalized set of sets of source information, we introduce the rules, namely, if the set $P = \{P_i = (op1, op2, op3)\}$, is defined, then to use the value of the component $op02$ of unit level 0, we will use the notation $op02$ (join operator).

At the second stage, procedures are applied that allow:

The first (I) procedure allows you to evaluate the regulatory or average value of the performance of official duties by employees – P :

$$P = \{P_i = (op01, op02, op03)\}, \quad (2)$$

where $op01$ - normative or average value of the job performance of the i -th employee,

$op02$ - normative or average value of the level of assessment of the performance of tasks of the i -th employee.

The second (II) procedure is aimed at identifying many specialties (economist, programmer, builder and others) – SP :

$$SP = \{sp_r\}, \quad r = \overline{1, L^{SP}}, \quad (3)$$

where sp_r is the r - th specialty,

L^{SP} - is the number of specialties.

The third procedure allows you to assess the level of education (secondary, bachelor, master and others) - UO :

$$UO = \{u_{ov} = (name, \mu)\}, \quad r = \overline{1, L^{OV}}, \quad (4)$$

where u_{ov} - is the vector of characteristics of the v -th category,

$name$ - category name,

μ - assessment of the level of education for the category in points,

L^{OV} - is the number of categories.

The fourth (IV) procedure is aimed at assessing the level of enterprise management (higher, middle and lower level) - UD :

$$UD = \{ud_V = (name, \gamma)\}, \quad r = \overline{1, L^{UD}}, \quad (5)$$

where ud_V - is the vector of characteristics of the V -th level,

$name$ - level name,

γ - is an estimate of the level in points,

L^{UD} - the number of levels, which is determined by the scale of the enterprise.

The fifth (V) procedure solves the problem of describing many positions in the enterprise - D :

$$D = \{d_j = (name, ud, uo, kl, SPD) : ud \in UD, uo \in UO\}, \quad j = \overline{1, L^d}, \quad (6)$$

$$SPD = \{spd_r = (sp, \beta) : sp \in SP, 0 \leq \beta \leq 1\}, \quad r = \overline{1, L_j^{SPD}}, \quad (7)$$

where d_j is the j -th position,

$name$ - job title,

ud - position level in the organizational and staff structure of the enterprise,

uo - the level of education required for the j -th position,

kl - required work experience (minimum number of years) in a given position for an optimal qualification level,

SPD - many specialties related to this position,

spd_r - is the vector of the correspondence characteristics of the r -th specialty of the j -th position β ,

β - is the correspondence coefficient of the specialty sp of the j -th position,

L^d - is the number of posts,

L_j^{SPD} - number of specialties in the j -th position.

The sixth (VI) procedure solves the tasks of describing correspondence and job interchangeability - SD :

$$SD = \{sd_f = (d1, d2, \alpha) : d1 \in D, d2 \in D, 0 \leq \alpha \leq 1, (d1 = d2) \Rightarrow \alpha = 1\}, \quad f = \overline{1, L^{SD}}, \quad (8)$$

$$L^{SD} = (L^d)^2, \quad (9)$$

where sd_f - is the f -th vector of job matching characteristics $d1$ and $d2$,

α - is the compliance coefficient.

The seventh (VII) procedure is aimed at assessing additional characteristics of employees - A :

$$A = \{a_i = (ds, ST, OB) : ds \in D, uo \in UO\}, \quad i = \overline{1, N}, \quad (10)$$

$$ST = \{st_j = (d, kl) : d \in D, d \in UO\}, j = \overline{1, L_i^{ST}}, \quad (11)$$

$$OB = \{ob_w = (sp, uo, god) : sp \in SP, uo \in UO\}, w = \overline{1, L_i^{OB}}, \quad (12)$$

where a_i - is the vector of characteristics of the i -th employee,

ds - the position held by the employee,

ST - many posts in which the employee previously worked and experience in them,

st_j - vector of characteristics of work experience in previous positions,

kl - length of service (number of years) in the position d ,

OB - value, reflects the education received by the i -th employee; sp - specialty;

uo - level of education,

god - year of receipt of the qualification document (certificate, certificate, diploma and others),

L_i^{ST} - the number of posts previously held by the i -th employee,

L_i^{OB} - the number of specialties in which the employee was educated by the i -th employee.

The eight (VIII) procedure allows you to describe many additional tasks (determined by orders) and their characteristics in the enterprise:

$$Z = \{z_k = (t0, tk, tk', usz), 0 \leq \beta \leq 1\}, k = \overline{1, M}, \quad (13)$$

where z_k - is the vector of characteristics of the k -th task,

$t0$ & tk - the value of the beginning and end of tasks, determines the term for completing the task in units of measurement of working time (for example, working days, hours and others),

tk' - the value of time, determines the critical deadline for completing the task, after which the task is either canceled or transferred to another performer,

usz - task difficulty level,

M - is the number of tasks.

The set of completing additional IZ tasks by employees can be written as follows:

$$IZ = \{iz_k = (a, z, uv^p, uv) : a \in A, z \in Z, 0 \leq uv^p \leq 200\}, k = \overline{1, M}, \quad (14)$$

where iz_k - is the characteristic of the k -th job,

a - an employee who performs additional tasks,

iz - the task

uv^p - is the percentage of the task according to the plan at the current time t ($uv^p = 0$ at time $t0$, $uv^p = 200$ at time k),

uv - is the percentage of the task at the current time t .

In case of failure to perform additional tasks, the value of $IZ = 0$.

5.3 Stage 3 Procedures

At the third stage, the procedure for assessing the conformity of the specialty of the position is performed. The function $f\beta$ returns the value of the correspondence of the specialty xsp to the position xd :

$$f\beta(xsp, xd) = \begin{cases} d_{j0}.spd_{i0}.\beta, \exists j_0 r_0 : (d_{j0} = xd) \wedge (d_{j0}.spd_{r0}.sp) \\ 0, & \neg \exists j_0 r_0 : (d_{j0} = xd) \wedge (d_{j0}.spd_{r0}.sp) \end{cases} . \quad (15)$$

The function $f\alpha$ returns the value of the coefficient of correspondence and interchangeability of the post xsp post xd :

$$f\alpha(xsp, xd) = \begin{cases} d_i.spd_g.\alpha, \exists g_0 : (sd_{g0}.d1) \wedge (sd_{g0}.d2 = xd) \\ 0, & \neg \exists g_0 : (sd_{g0}.d1) \wedge (sd_{g0}.d2 = xd) \end{cases} . \quad (16)$$

To determine job conformity is the level of education of the position held in conjunction with work experience in similar or related positions:

$$p_i.op1 = \delta(op11.op12).op13, \quad (17)$$

$$op11 = \sum_{i=1}^{L_i^{ob}} f\beta(a_i.ob_w.sp.a_i.ds) \cdot \frac{a_i.ob_w.god}{godT}, \quad (18)$$

$$op12 = \sum_{i=1}^{L_i^{st}} f\delta(a_i.ds.f.a_i.st/d) \cdot \frac{a_i.st_jkl}{a_i.ds.kl}, \quad (19)$$

where $godT$ - is the value of the current year,

$op11$ - qualification level of education received,

$op12$ - qualification level, which is determined by work experience,

$op13$ - quality of job performance, determined by an expert.

When solving the problem of data mining in the management of HR processes, fuzzy logic methods are used to display the result on the interval $[0; 1]$.

5.4 Stage 4 Procedures

Therefore, at the fourth stage, the procedure for constructing membership functions based on the theory of fuzzy sets is performed.

The following “position”, “level”, “education” can be attributed to numerical linguistic variables of employees, and “conflict”, “level of substitution” to linguistic variables. Numerical linguistic variables and their meanings serve for a qualitative description of a quantitative quantity. The values of linguistic variables are determined by experts.

It should be noted that a linguistic variable, like its original term set, is associated with a specific dimensional scale on which all arithmetic operations are defined.

To assess the characteristics of employees in Table 1, linguistic variables and their dimensions are proposed.

Table 1. Linguistic variables of employee characteristics

Term set	The metric and type of exposure	$X_{\min}^1$	$X_{\max}^1$	The term designation
T_1^1	Performance of duties <position> - stimulation	0	1	Not performed Partially completed Performed
T_2^1	Job Interchangeability <Interchangeability Level>, discouragement	1	3,0	Low Average High
T_3^1	Level of education <education>, stimulation	1	3,0	Secondary education Bachelor Master
T_4^1	Conflict <conflict>, stimulation	0	3,0	Low Average High
T_5^1	The importance level of the specialty < SpecialtyLevel >, discouragement	0	0,5	Low Average High

The use of the concept of stimulation and destimulation is applied taking into account the influence on the degree of personnel efficiency, namely, stimulation - the effect on the increase and destimulation - on the reduction of the factor.

Therefore, the term set $T_i^n = \{T_i^{n'}\}$ is associated with the set $T_i^{n'}$, where $T_i^{n'} = \langle x, \mu_{T_i^{n'}}(x) | x \in [x_{\min}, x_{\max}] \rangle$ - is a fuzzy number, $i = \overline{1, m}$, m - is the number of term sets, n - is the number of employees.

To eliminate the influence of changes in the input variables of the metrics and, as a consequence, the correction of term sets, a transition to a normalized function is proposed. Let the previously defined term set T_i be the original one.

The normalized linguistic variable is a mapping on the interval [0; 1]:

$$D_i^n = \{D_i^{n'}\}, D_i^{n'} = \langle z, \mu_{D_i^{n'}}(z) | z \in [0, 1] \rangle, \quad (20)$$

where z is a fuzzy number corresponding to the term set D_i' on the interval $[0; 1]$, n - is the number of employees.

These functions allow you to display heterogeneous input variables in a single normalized interval $[0;1]$, which allows you to reduce errors associated with different quantities and their dimensions. This provides a convenient representation of the values, as well as their interpretation.

6 The method of constructing a model of data mining in HR process management

The structural model of data mining in HR process management is presented in Fig. 2.

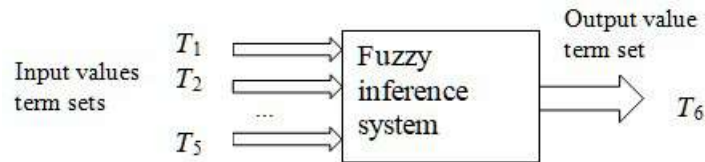


Fig. 2. The structural model of the model of data mining in the management of HR-process.

In the structural model, $T = \{T_i\}$ is a term set, where $i = \overline{1, n}$, n - is the number of sets, each of which is represented by a fuzzy variable with a domain of definition x .

The process of modeling fuzzy values is based on a fuzzy inference system, which allows you to convert expert estimates into fuzzy values.

In the fuzzy inference system, the procedure for finding a clear value for each of the input linguistic variables based on defuzzification is applied. Defuzzification in a fuzzy inference system is the process of finding a value for each of the output linguistic variables of the set $W = \{x_1, x_2, \dots, x_n\}$. The task of defuzzification is to use the results of accumulation of all output linguistic variables. It is necessary to obtain a quantitative value of each of the output variables. Output variables can be used in a fuzzy inference system relative to the input linguistic variable.

Accumulation of fuzzy inference is the process of finding the membership function for each of the output linguistic variables of the set.

The transformation of a fuzzy set into list of values of variables is named as defuzzification.

The defuzzification procedure is performed by a sequence that considers each of the output linguistic variables β and the fuzzy set $T_i = \{T_i^j\}$ related to it. The result of defuzzification for the output linguistic variable is defined as a quantitative value.

The defuzzification process is considered complete when quantitative values are determined for each of the output linguistic variables. For the fuzzy inference system, the Mamdani algorithm was applied [17].

The Mamdani algorithm includes the following steps [18, 19]:

- the formation of a rule base for fuzzy inference systems;
- fuzzification of input variables;
- aggregation of conditions in fuzzy rules to find the degree of truth of the conditions of each of the rules of fuzzy logic;
- accumulation of conclusions of fuzzy production rules;
- defuzzification of output variables based on the center of gravity method.

An example of one of 234 rules has the following form:

If (Position is Npfd) and (InterchangeabilityLevel is low) and (Education is secondary) and (Conflict is low) and (SpecialtyLevel is low) then (StaffEfficiency is NotHardworking).

Depending on the nature of the domain X, numerical linguistic variables can be defined.

The values of linguistic variables are determined on an ordinal scale. It should be noted that a linguistic variable, like its original term set, is associated with a specific scale on which all arithmetic operations are defined. Therefore, the term set $T_i = \{T_i^j\}$ is associated with the set T_i^j , where $T_i^j = \langle x, \mu_{T_i^j}(x) \mid x \in [x_{\min}, x_{\max}] \rangle$; $i = \overline{1, n}$; $j = \overline{1, m}$; n - is the number of term sets, m is the number of terms.

A model that satisfies these fuzzy sets is their union:

$$\mu T_i = \sup \{ \mu_{T_i^j}(x) \} \quad T_i = \cup T_i^j. \quad (21)$$

We construct membership functions for the linguistic variable characteristics of employees, presented in Table 1.

The process of converting experts' qualitative assessments into fuzzy quantities consists in mapping the elements of the original term set in the form of constructing membership functions of fuzzy quantities $T_i^j \in T_i$.

The description of linguistic variables is as follows:

- $\langle \text{position, \{not fulfilled, partially fulfilled, fulfilled\}, [0; 1]} \rangle$,
- $\langle \text{level of interchangeability, \{low, medium, high\}, [1; 3]} \rangle$,
- $\langle \text{education, \{secondary, bachelor, master\}, [1; 3]} \rangle$,
- $\langle \text{conflict, \{low, medium, high\}, [1; 3]} \rangle$.

Moreover, the values of the sets are in the range $[0; 1]$ & $[1; 3]$.

7 Experiments and results

The use of trapezoidal membership functions is due to the fact that the calculation uses the definition of fuzzy numbers. This form of analytical approximation use (LR) - functions, which include the trapezoid functions. From this it follows that fuzzy numbers are determined to perform operations with similar procedures.

The constructed membership functions of the input linguistic variables are presented in Fig. 3.

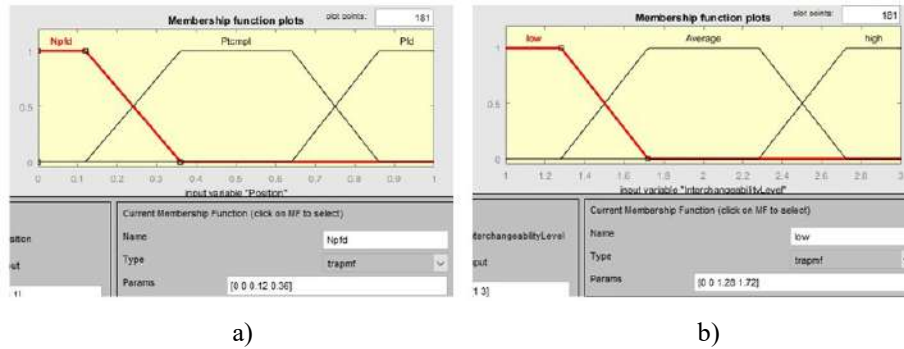


Fig. 3. The membership function of the input linguistic variables: a) “Fulfillment of duties”, b) “Interchangeability of posts”.

To assess the degree of personnel efficiency, developed knowledge bases are applied, which are presented in the form of a rule base.

In the fuzzy inference procedure for managing HR processes, it is necessary to consider the work of employees at all levels of work. The fuzzy inference procedure is implemented in the MATLAB R2017a system, which allowed obtaining the following results of assessing the degree of personnel efficiency.

To perform the procedure, we built the diligence function of the output linguistic variable “assessment of the degree of personnel efficiency”, which is presented in Fig. 4.

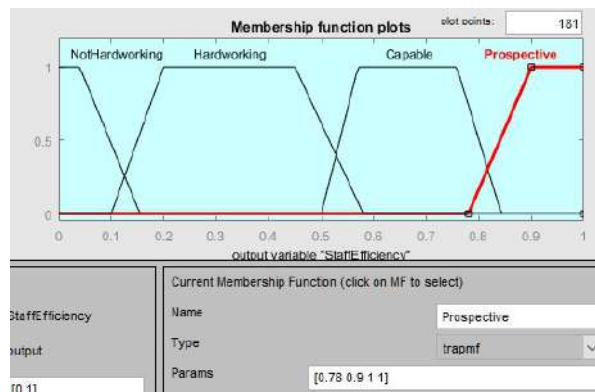
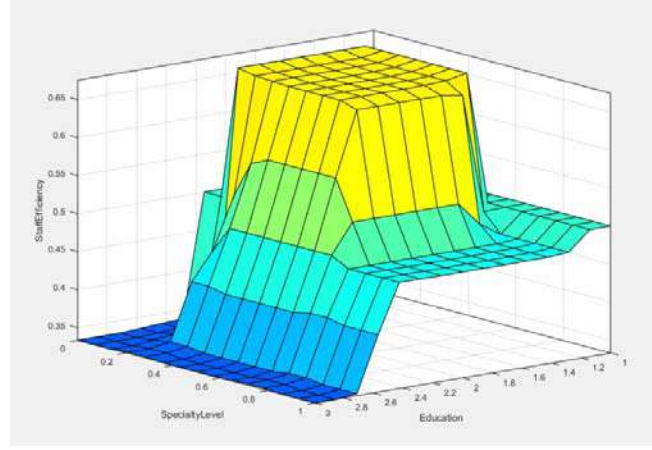
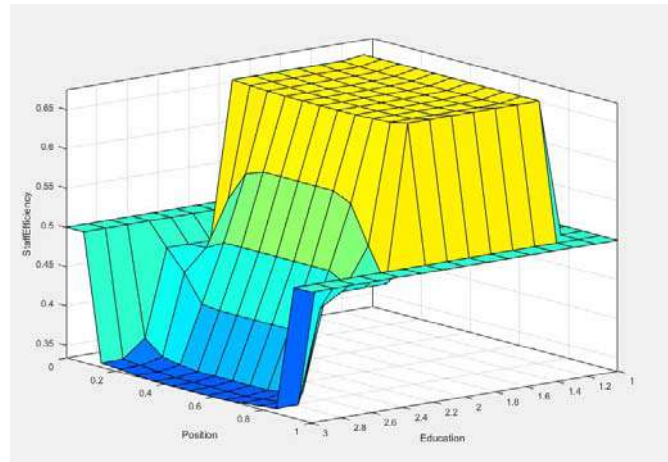


Fig. 4. The membership function of the output linguistic variable "assessment of the degree of personnel efficiency".

The simulation results of assessing the degree of personnel efficiency, which is presented in Fig. 5.



a)



b)

Fig. 5. Modeling the assessment of the degree of personnel efficiency: a) level of importance of the specialty(SpecialtyLevel) and education, b) performance of duties (Position) and education.

The results of modeling the potential effectiveness of the staff showed that five sets: responsibilities, job experience, duration of continuous work per day, job interchangeability, conflict, the level of importance of the specialty and the level of education have an intersection ($\bigcap_{i=1}^5 T_i^1 \neq \emptyset$). The value of the potential efficiency of the employee's staff is equal to the intersection of sets and it is 0.65, which corresponds to a high score (Fig. 5).

The result can be used by business entities to improve the efficiency of staff and in the selection of personnel.

8 Conclusion

The authors of the article propose a solution for constructing a sequence of data mining of human resources and modeling the degree of personnel efficiency based on fuzzy sets.

The proposed data analysis method involves four steps. At the first stage, the problem of choosing the analyzed indicators is solved. At the second stage, eight procedures are performed to solve the following problems: to assess the normative or average values of labor productivity, to determine many specialties, to assess the level of education, to evaluate the levels of enterprise management, to describe a lot of posts, to describe correspondence and job interchangeability, to evaluate additional characteristics of employees, for descriptions of many additional tasks and their characteristics. At the third stage, the procedure for assessing the conformity of the specialty of the post is carried out. At the fourth stage, the procedure for constructing membership functions based on the theory of fuzzy sets is performed. The fuzzy inference procedure is implemented in the MATLAB R2017a system, which makes it possible to assess the degree of personnel efficiency.

Prospects for the application of the proposed method and procedures for data mining of human resources lie in the possibility of using large data warehouses and cloud computing.

References

1. Schwab, K.I.: The Fourth Industrial Revolution. Currency, Geneva (2016).
2. Ryan, J.: Engaging Professional Audiences on LinkedIn, <https://www.emarketer.com/topics/topic/b2b>.
3. Ivanov, M.: Cloud-based Digital Marketing. CEUR Workshop Proceedings. Vol. 2422, pp. 395-404 (2019).
4. Lysenko, Yu.G., Petrenko, V.L., Bogatov, O.I., Skobelev, V.G.: Rating management of economic systems. Yugo-Vostok, Donetsk (1999).
5. Helman, V.M., Makazan, Ye.V., Buriak, A.M.: Personnel development as pledge of the success of enterprise. Bulletin Zaporizhzhia national university. Economic sciences. Vol. 43(3), pp. 112 - 115 (2019). doi: 10.26661/2414-0287-2019-3-43-18
6. Becker, Br. E., Huselid, M. A.: Strategic Human Resources Management: Where Do We Go from Here?. Journal of Management. Vol. 32(6), pp. 898–925 (2006). doi: 10.1177/0149206306293668
7. Huselid, M. A.: The Impact of Human Resource Management Practices on Turnover, Productivity, and Corporate Financial Performance. Academy of Management Journal. Vol. 38(3), pp. 635–672 (1995). Available at SSRN: <https://ssrn.com/abstract=1803666>
8. Pfeffer, J., Veiga, J. F.: Putting People First for Organizational Success. Academy of Management Executive. Vol. 13(2), pp. 37–48 (1999). doi: 10.5465/ame.1999.1899547
9. Becker, Br. E., Huselid, M. A., Ulrich, D.: Six Key Principles for Measuring Human Capital Performance in Your Organization. Working paper, School of Management and Labor Relations (2002).
10. Briscoe, D. R., Schuler, R. S., Claus, L.: International Human Resource Management, 3rd ed.. Chartered Institute of Personnel and Development, New York (2009)

11. Schuler, R. S., Budhwar, P. S., Florkowski, G. W.: International human resource management: review and critique. *International Journal of Management Reviews*. Vol. 4(1), pp. 356–414 (2002). doi: 10.1111/1468-2370.00076
12. Gottwald S.: Universes of Fuzzy Sets and Axiomatizations of Fuzzy Set Theory. Part I: Model-Based and Axiomatic Approaches. *Studia Logica*. Vol. 82(2), pp. 211–244. Springer (2006). doi: 10.1007/s11225-006-7197-8
13. Vedula, N., Nicholson, P.K., Ajwani, D., Dutta, S., Sala, A., Parthasarathy, S.: Enriching taxonomies with functional domain knowledge. *SIGIR '18*, pp. 745–754 (2018). doi: 10.1145/3209978.3210000
14. Siler, W., Buckley, J.: *Fuzzy Expert Systems and Fuzzy Reasoning*. Wiley Interscience, Birmingham (2005).
15. Waitelonis, J., Exeler, C., Sack, H.: Linked data enabled generalized vector space model to improve document retrieval. *International Semantic Web Conference (ISWC)*. Vol. 1581(4), (2015).
16. Abraham, A.: *Intelligent Systems: Architectures and Perspectives, Recent Advances in Intelligent Paradigms and Applications*. Springer, Verlag (2003).
17. Mamdani, E.H., Assilian, S.: An experiment in linguistic synthesis with a fuzzy logic controller. *Int. J. Man-Mach*. Vol. 7, pp. 1–13 (1975).
18. Gacto, M., Alcalá, R., Herrera, F.: Interpretability of linguistic fuzzy rule - based systems: An overview of interpretability measures. *Information Sciences*. Vol. 181(20), pp. 4340–4360 (2011).
19. Gang F.: *Analysis and synthesis of fuzzy control systems: a model-based approach*. CRC Press (2010).

Bagging-based instance selection for instance-based classification

Dmytro Kavrin¹[0000-0002-8952-4067], Sergey Subbotin²[0000-0001-5814-8268],

^{1,2}National University "Zaporizhzhia polytechnic", Zhukovsky str., 64,
Zaporizhzhia, 69063, Ukraine

¹kavrin@zntu.edu.ua, ²subbotin@zntu.edu.ua

Abstract. The task of reducing marked large size samples for building diagnostic and recognizing models by precedents is considered. The method allowing to reduce essentially the size of a training sample increasing at the same time its efficiency, through removal of irrelevant and redundant instances is proposed. The given method provides an opportunity to estimate each instance of a training sample by synthesis of an ensemble of weak classifiers using a bagging model, and to create a reduced sample of the most significant instances by estimations. Software is developed to implement the proposed method. This software has been experimentally investigated in solving the task of reducing synthetic and real world data. The results of the conducted experiments allow recommending the use of the developed method and its software realization for solving the task in the sphere of technical diagnostics.

Keywords: base classifier, class, classification, ensemble of classifiers, instance, meta-estimator, metric, sampling, training sample.

1 Introduction

The constantly increasing volume of available information requires significant computing resources for successful data processing. So the task of reducing dimension data for further building models based on them is urgent [1-4]. It is especially necessary when solving practical tasks of industrial diagnostics, when the diagnostic system should immediately respond to any deviations in the operation of equipment. In the conditions of continuous production, diagnostic systems of nondestructive testing, which allow making diagnostics in real time, are especially important. The main component of nondestructive testing diagnostic systems is the model of pattern classification by precedents (classifier) [5-8]. The classifier is a set of rules, which allow determining whether new observations (instances) belong to one of the existing classes. To generate a model of recognition by precedents it is necessary to have a set of instances (precedents) with known values of classes (training sample), and a classification method, which will form the rules of recognition (training) using the training sample [9-10].

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

There are many classification methods with different principles and approaches [11-12]. The family of metric classification methods based on precedents is quite effectively used in building diagnostic models. Metric methods of training refer to the geometrical paradigm of machine learning, which assumes that instances have a geometric structure, where each of them is described by numerical features and is considered as a point in a multidimensional feature space. These methods are based on the assumption of local compactness of classes, from which it follows that the similarity of two instances on N independent features also assumes their similarity on one dependent feature $N+1$. Thus, only instances of the same class can exist in the neighborhood of one class instances. And the closer a control instance is to the neighborhood of a class, the more likely it will belong to this class [13-14]. Metric methods have such advantages as simplicity of implementation, clear logic of methods' work, geometrical nature, simplicity of model results interpretation, developed theoretical base, adaptation to the necessary task by metric selection. The disadvantage of the metric methods recognition is the necessity to store the whole training sample in the computer memory. Thus, the use of large size training samples may require significant computing resources and classification time. In technical diagnostics systems, where the speed of decision making is a priority, models based on large training samples may be ineffective. Reducing the size of training samples will help reduce both the classification time and the computational complexity of diagnostic models.

The most widely used approach to reducing the data dimension is the selection of informative features [15-17], which implies the selection from the initial set of features a smaller subset of features, sufficient to solve the problem with the required accuracy or meeting some criterion. However, if the size of the feature space is small, or features individually are low informative, but together they contain enough information for building a model, selection of informative features does not allow to reduce the data dimension effectively.

Another approach to reducing data dimensional is the selection of instances [1, 18-22]. To date, the instance selection has been considered a necessary preliminary data processing procedure [1, 23]. Successful use of instance selection methods allows for the selection of a small sample size, independently of the model in which it will be used in a future and without performance loss [2]. In the process of selecting instances, irrelevant and redundant instances are removed from the sample, so in some cases the performance of models that are built on processed samples may be higher than on the initial data.

Many instance selection methods have been proposed in the past few decades, each with weaknesses and advantages [23-26]. However, there is no universal method which can achieve equally high results with various data samples. In general, the task of instance selection is to select the most relevant instances of a training sample. Virtually, for the selection of instances must be solved the problem of binary classification where each instance of the training sample can be classified as selected or unselected. Therefore, it can be assumed that the approaches and methods used to solve the classification problem can also be applied to the instance selection task.

One of the successful directions for increase of models productivity in the tasks of classification is the use ensembles of classifiers [27-29].

In this paper we study the possibility of using ensembles of metric classifiers to select the most relevant instances in solving the problem of reducing training samples and increasing their representativeness.

2 Formal problem statement

The task of instance selection for classification model building based on precedents is to generate such a training subsample of minimum size from the most relevant instances of the initial training sample, which will allow to classify new unmarked data with an accuracy at least as accurate as using the initial training sample [2, 30]. Formally, the task can be written in the following way.

Let the initial training sample be presented as a set of S precedents of dependence $y(x)$ and be defined by the expression $X = \langle x, y \rangle$ is a sign matrix of input features x and output features y . The set of input attributes x is defined by the standard object – feature matrix:

$$x = \|x_{ij}\|_{S \times N}, \quad (1)$$

where S is a number of instances, N is a number of input features, x_{ij} is a value of the j -th feature on the i -th instance. The set of output features is defined by the vector:

$$y = (y_1, \dots, y_S)^T, y_i \in \{1, 2, \dots, K\}, \quad (2)$$

where K is a number of classes in the sample ($K > 1$). Each i -th instance is represented as $X_i = \langle x_i, y_i \rangle$. Then the task of selecting instances is to select from the initial sample $X = \langle x, y \rangle$ such a subsample $X' = \langle x', y' \rangle$ that the following conditions are executed:

$$x' \in x, y' = \{y_i \mid x_i \in x'\}, S' = |y'|, S' < S, f(\langle x', y' \rangle, \langle x, y \rangle) \rightarrow opt, \quad (3)$$

where S' is a number of instances of the resulting subsample, x' is a set of input features of the resulting subsample, y' is a set of output features of the resulting subsample.

3 Review of the literature

At solving the complex practical problems of pattern recognition by precedents, there may be occasions when no classifier provides the necessary accuracy of pattern recognition (classes). Increase in accuracy allows creating a model composed of a set of classifiers (ensemble). The strategy of ensembles is that a set of independent models

of classification (base classifiers) is created, and the results of their work are combined. Thus, the productivity of base classifiers increases due to compensation of errors of some classifiers by correct work of other ones [27-29]. Generally, the ensemble of classifiers can be described as follows:

$$y(x) = \Omega(b_1(x), \dots, b_r(x)), \quad (4)$$

where b_i is a base classifier, Ω is a meta-estimator that creates a decisive rule that the recognized instance belongs to a certain class $y(x)$.

The basic properties of classifiers are the ability of each base classifier to independently solve the initial classification task and the possibility to use existing classification training standard methods.

The one of the most important conditions for the efficiency of the classifier ensembles in pattern recognition problems is the requirement for a sufficient variety of base classifiers [27-28]. Thus there is a compensation of classification errors of some base classifiers by work of other ones. Therefore, it is necessary to combine the results of the base classifiers, so as to increase the influence of true decisions and minimize the influence of wrong decisions on the response of the ensemble. The basic strategies for building ensembles are the synthesis of independent classifiers and decision-making on a bagging basis [31-33], special coding of target values and reduction of the task solution to solving several tasks (error-correcting output code) [34], building of meta-signatures on the basis of responses of base classifiers on subsets of samples and training meta-functions on them (stacking) [28, 35], sequential summarization of several classifiers, with each next classifier being trained taking into account the errors of the previous ones (boosting) [27-29, 32-33, 36-37], heuristic methods of combining the answers of base classifiers by training in special subspaces and visualizations (mixture-of-experts) [28, 38-39], recursive synthesis of homogeneous ensembles (neural networks) [40].

Bagging-based ensembles are the most common when it comes to solving real world tasks due to their simplicity of implementation and high generalization ability. The main advantages of bagging are the ability to perform parallel computations at high classification accuracy.

During ensemble forming by the bagging method, each base classifier is trained on a random subset of the training sample. With this approach, the variety of methods is achieved even with the use of one classification method for all base classifiers [27, 33, 38]. In the traditional bagging model, the bootstrap technique [27, 29, 31, 33, 41] is used to select random subsets of a training sample, which implies the formation of subsets by random selection with return. The basic bagging method works well on small samples, when exclusion of a small number of instances leads to a significant distribution transformation in the sample. For larger training samples it is possible to use other sampling methods. In this case, the task of selecting the optimal size of the selected subsamples S .

There are many different ways to extract subsets (resampling) of a training sample for the synthesis of bagging ensembles [11, 31, 41-46]. The most known methods are

bootstrap [31, 41-46], pasting [47], random subspaces [28-29], random patches [48] and cross-validation [41].

The base classifiers forming an ensemble using a bagging strategy do not need a separate test subsample to evaluate the accuracy of the generated model. It is possible because each base classifier is trained on a subsample containing only part of the training sample. Thus, it allows to estimate the accuracy of each base classifier with instances not included in the selected subsample. Provided that the number of base classifiers is quite numerous, the evaluation will be performed for almost every instance of the training sample. Moreover, this evaluation will be independent, because the accuracy of each base classifier will be evaluated by a subsample of instances with the dependent variable values unknown to it. Further training and test samples of base classifiers will be called local for certainty.

By constructing ensembles of classifiers, it is important to take into account that there are classification methods stable with respect to the selection of random subsets (for example, the SVM (support-vector machine) method, or the kNN (k-nearest neighbors) method at $k > 3$) [27, 49]. Application of such methods in ensembles based on bagging is ineffective because diversity of methods is not achieved.

The result of ensemble work depends on the choice of meta-estimator. In the most common case, the meta-estimator is a majority vote function for the classification task:

$$y(x) = \Omega(b_1(x), \dots, b_T(x)) = \arg \max_{k=1}^K \sum_{t=1}^T b_t^k. \quad (5)$$

In more complex cases, weighted voting can be used, where each base classifier has a weighting characteristic:

$$y(x) = \Omega(b_1(x), \dots, b_T(x)) = \arg \max_{k=1}^K \sum_{t=1}^T w_t b_t^k, \quad (6)$$

where w_t is a weight characteristic of the base classifier. The weighting characteristic primarily depends on the accuracy parameter of new instances recognition by the classifier. Such characteristic can be a relative number of correctly recognized instances of the test sample (relative accuracy):

$$E = \frac{1}{S^b} \sum_{i=1}^{S^b} [1 | b(x_i) = y_i], \quad (7)$$

where E is a relative accuracy of the classifier, S^b is a number of instances in the local test sample of the current classifier, b is an approximating function of the classifier.

Another important parameter depends on the number of instances in the selected training sample. According to the size minimization task, in order to calculate the weighting characteristics of each base classifier it is possible to combine the charac-

teristics of classification accuracy within the local test sample and the share of instances reduced from the initial training sample:

$$w = \gamma E + (1 - \gamma)r, \quad (8)$$

where $\gamma \in [0...1]$ is a coefficient indicating the degree to which factors affect the value of the overall score, r is a share of reduced instances $r = S^h/S$, since the local test sample is formed from instances not included in the local training sample.

The relative accuracy of classification (7) gives an objective assessment of the classifier provided it is sufficiently balanced by classes the test sample. If the test sample has an imbalance of classes, for example, a minority class is 1% of the sample, it is possible that a classifier that incorrectly classified all minority instances and correctly classified the majority class will have abnormally high relative accuracy ($E = 99\%$). Selection of instances, when constructing a bagging-ensemble, is carried out randomly, so it is impossible to guarantee the balancing of training and test local samples. Stratified instances selection for the local training sample can be one solution to the problem, but if the initial sample is imbalanced by classes, the local test sample will also have class imbalances. Therefore, for imbalanced samples, an assessment based on a confusion matrix may be more appropriate [50]. The confusion matrix is a way of grouping the instances depending on the combination of the true answer and the classifier's answer and allows to get a set of different metrics. In case of binary classification, instances can be divided into four categories (Table 1).

Table 1. – Confusion matrix

	$y = 1$	$y = 0$
$b(x) = 1$	True Positive (TP)	False Positive (FP)
$b(x) = 0$	False negative (FN)	True Negative (TN)

The instances of the class of greater interest are called positive instances and another class is called negative. When dealing with imbalanced data, the minority class is usually presented as positive. Using the confusion matrix, it is possible to obtain precision and recall metrics. The precision:

$$P = TP/(TP + FP), \quad (9)$$

where TP is the correctly classified positive instances, FP is a incorrectly classified positive instances, shows the share of correctly predicted positive instances. The recall:

$$R = TP/(TP + FN), \quad (10)$$

where FN is the incorrectly classified negative instances, shows the share of correctly predicted positive instances of all predicted instances as positive instances. Obviously, the higher the values of these metrics, then the classifier is better. However, it is impossible in practice to reach the maximum values of precision and recall simulta-

neously, so it is necessary to choose which characteristic is more important for a particular task, or to search for balance between these values. The harmonic mean of the precision and recall (F-measure) allows combining these parameters [51]:

$$F = 2PR/(P + R). \quad (11)$$

4 Materials and methods

To select the most relevant instances of a training sample using an ensemble of classifiers based on a bagging model, you need to solve the problem of binary classification for each instance of the training sample. In addition, each instance will be assigned to the selected class of instances, or to the class of instances that are not meeting the selection condition.

The classification of instances from the initial sample is based on the voting results of base classifiers. The base classifier is a model trained on the marked sample, in which the value of the output feature $y(x)$ for each instance is known. To synthesize the base classifier of a bagging ensemble, a random subset of instances is selected from the initial sample by the bootstrap method. The resulting local sample is used to train the base classifier. The nearest neighbor kNN method with one nearest neighbor and Manhattan distance metric was chosen as training method [52]. This choice is conditioned by the necessity of obtaining less stable classifiers and increasing the rate of synthesis of a large number of base classifiers. This model uses a passive learning strategy, in which there is no learning phase of the classifier, instead, the learning sample is stored in memory, which is used to classify new data. The main advantage of this model is the ability to use new data without retraining, simply adding new significant instances to the sample. However, in such a model, large size training samples will require significant memory resources for storage. Using a basic bootstrap method for building an ensemble of classifiers implies retrieving subsamples of the same size as the initial sample. Thus, each ensemble classifier must store almost the entire training sample in memory. Since this research is aimed at reducing the size of large training samples, extraction of random subsamples was performed by random selection with return, but in contrast to bootstrap, the length of the subsamples was randomly determined in a predefined range. When creating each classifier, the unselected instances were used as a local test sample for the particular classifier estimate. The weighting parameter w of each base classifier was obtained using the following equation (8).

The primary aim of the study was to select the most representative data from the training sample, so the task of the selection method is to investigate each instance of the sample and assess its relevance. Random selection methods do not guarantee the examination of each instance, so at the preliminary stage of creating an ensemble it is proposed to divide the initial sample into some number approximately equal in size subsamples and then to classify each subsample using ensembles built on the remaining training subset. With this approach, it is possible to ensure that every instance of the initial training sample is examined. The number of subsamples can take the value

$M > 1$, increasing the number of subsamples will lead to more stable and accurate results, but on the other hand will increase the computational complexity of the model and the processing time.

Formally, the proposed instance selection method can be presented as follows:

1. Set the initial training sample $X = \langle x, y \rangle$, initialize the resulting sample $X' = \langle x', y' \rangle = \emptyset$. Set the number of subsamples $M > 1$. Set the number of base classifiers T . Set the value of the coefficient $\gamma \in [0...1]$. Set the threshold $\delta \in [0...1]$, instance selection in a new training sample.
2. Split the initial sample X into M subsamples of approximately equal size:

$$X = \{X^m = \langle x^m, y^m \rangle, m = 1...M\}. \quad (12)$$

3. Set the number of subsamples $m = 1$.
4. Set the number of base classifiers $t = 1$.
5. Using a method of simple random selection with return, take a local training sample $X^\wedge$ from the subsample $X \setminus X^m$. To define as a local test sample $X^\top$ the set of instances not selected in the sample $X^\wedge$.
6. Train the base classifier using the local training sample $b = fit(X^\wedge)$.
7. Calculate the harmonic mean value F (11) for the current base classifier. If $F < 0.5$, go to step 5.
8. Calculate the weight characteristic of the current base classifier w (8).
9. Classify the subsample X^m and calculate the weight of each instance, taking into account the weight characteristic of the base classifier :

$$\chi_t = \{w[b(x_i^m) = y_i^m]\}_{i=1}^{S^m}. \quad (13)$$

10. Set $t = t + 1$. If $t \leq T$, go to step 5.
11. For each instance of a subsample, calculate the value of the meta-estimator:

$$\Omega_i^m = \left\{ \sum_{t=1}^T \chi_t \right\}_{i=1}^{S^m}. \quad (14)$$

12. Set $m = m + 1$. If $m \leq M$, go to step 4.
13. Merge M vectors of meta-estimators and normalize the values to a unit segment:

$$\left\{ \Omega_i = \frac{\Omega_i}{\arg \max \{\Omega_i\}_{i=1}^S} \right\}_{i=1}^S. \quad (15)$$

14. Form a new training sample:

$$X' = \langle x_i, y_i \rangle | \Omega_i > \delta \}_{i=1}^S. \quad (16)$$

5 Experiments and Results

To obtain a summary evaluation of the method, the experiments were conducted on two different samples, which differed in the number of instances, features and classes (Table 2). To evaluate the obtained training samples, at the first stage of the experiment, the initial data set X^0 was divided by the stratification method [53] into training X and validation X^v samples in a ratio of 75/25. The training sample obtained by stratification method was later considered as the initial sample. Classifiers were built on the basis of the initial and the resulting samples and tested with the validation sample. Then values of relative accuracy of classification and number of instances of samples were compared. The nearest neighbor method with the Euclidean distance metric was used as a method of classifying the recognition model. The resulting sample was formed using an ensemble of base classifiers. All base classifiers used the nearest neighbor method with one neighbor and Manhattan distance metric. Using the Manhattan distance metric reduced the stability of the base classifiers and computational complexity. A variety of base classifiers was achieved through the use of simple random selection with return when forming a local training sample of each base classifier.

Table 2. - Experimental data sets

Dataset	S^0	S	S^v	N	K
Pulsars Recognition (Pulsar) [54]	17 898	13 424	4 474	8	2
Sample Classification (Banana) [55]	5 300	3 975	1 325	2	2

To guarantee the evaluation of each instance, the training sample was randomly divided into two subsamples with approximately equal numbers of instances. An ensemble of base classifiers was built upon each such subsample. Figure 1 shows the scheme of work of the instance selection method based on an ensemble of classifiers.

The number of basic classifiers varied in the interval $T \in [1...200]$. The instance selection for the local training sample of each base classifier was performed by simple random selection with return. The local sample size was determined randomly in the range of $[1\%...100\%]$ of the initial subsample size. For the base classifier the F-measure value F was calculated using the local test sample, consisting of instances not included in the local training sample. Moreover, if the F-measure value F of the base classifier was less than 50%, the synthesis procedure for this classifier was repeated. Then the weight of each base classifier was calculated according to the equation (8) with parameter $\gamma = 0,75$. The second subsample was classified by each base classifier according to its weight. Thus, a vector of weights corresponding to all the base classifiers was formed for each instance of the initial sample. The resulting training sample X' was formed from the instances having the maximum total weight.

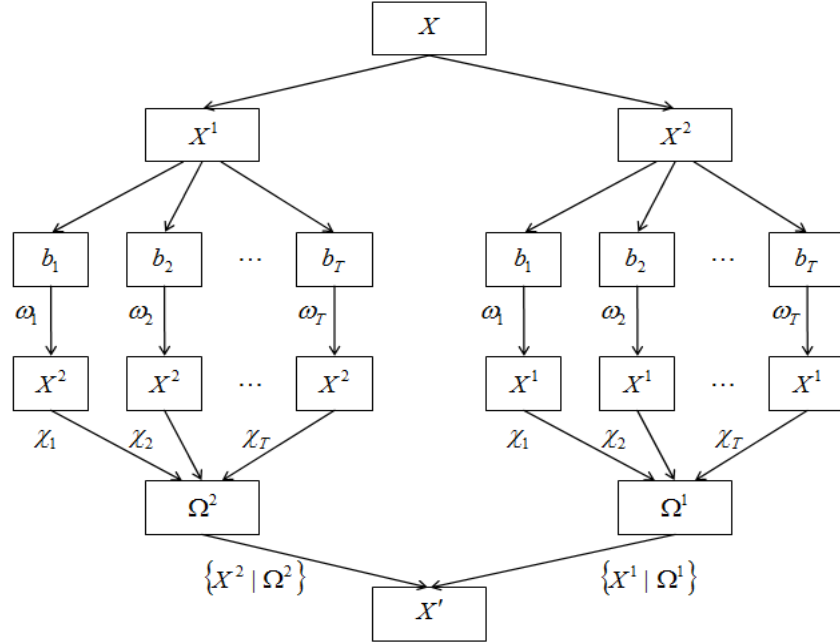


Fig. 1. – Scheme of work of the instance selection method based on an ensemble of classifiers

At the next stage of the study, classification of the validation sample X^v by the method of the nearest neighbor with one nearest neighbor and the Euclidean distance metric was performed. The initial sample X and the resulting sample X' were used as training samples. The relative accuracy of the models was calculated according to the equation (7). Using the obtained data, the dependencies of relative accuracy and sample length on the number of base classifiers were plotted (Fig. 2-5).

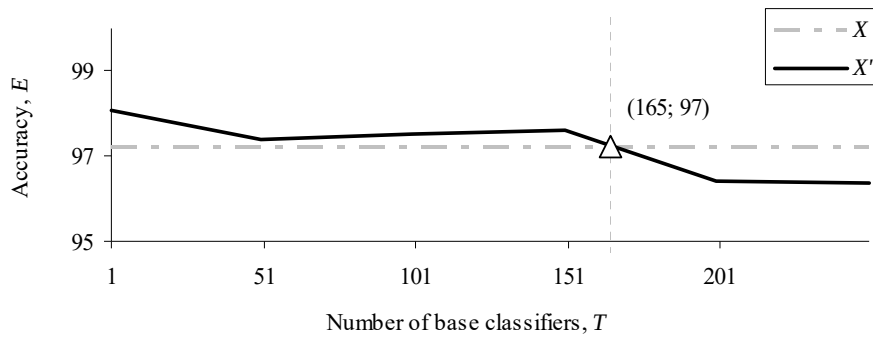


Fig. 2. – Dependence of model accuracy on the number of base classifiers for the Pulsar dataset

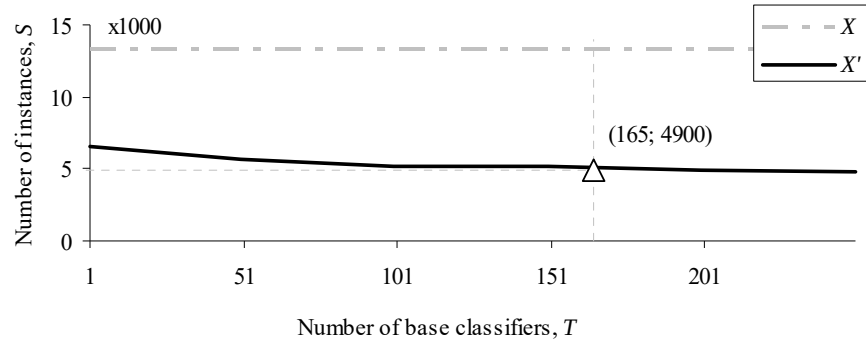


Fig. 3. – Dependence the number of instances on the number of base classifiers for the Pulsar dataset

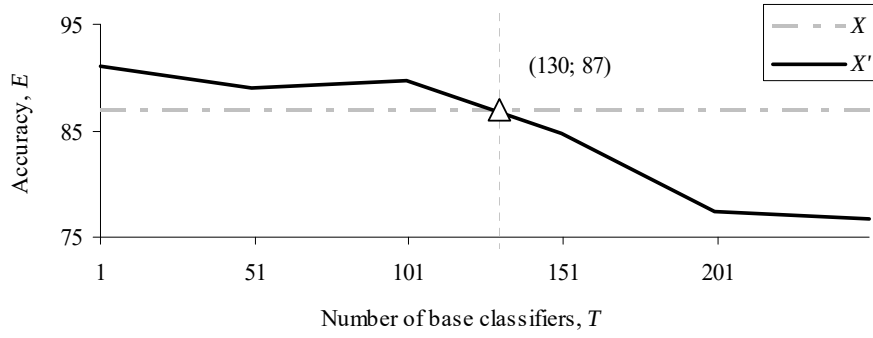


Fig. 4. – Dependence of model accuracy on the number of base classifiers for the Banana dataset

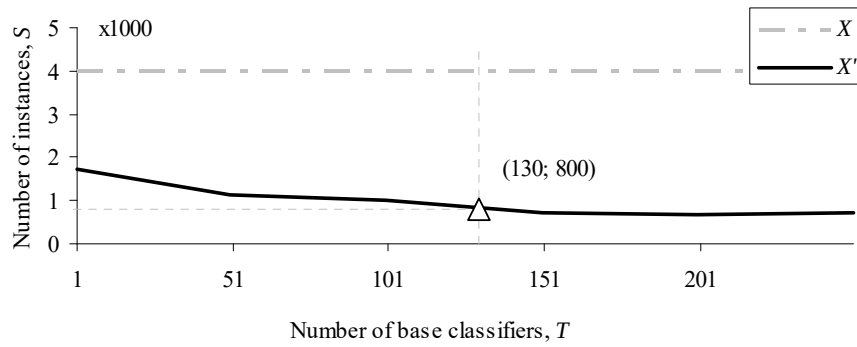


Fig. 5. – Dependence the number of instances on the number of base classifiers for the Banana dataset

For clarity, the figures showed the local areas within the critical values of model accuracy, at which the accuracy of classification of the resulting sample became less than the accuracy of the initial training sample. Thus, it was possible to estimate the critical number of instances of the resulting sample, below which the relative accuracy of the model based on the resulting sample became lower than the relative accuracy of the initial sample model.

6 Discussion

The proposed method showed high efficiency on all investigated datasets. All models on the basis of obtained training samples had relative accuracy higher than models with initial samples. At the same time, the size of the obtained samples was less than two times than the initial samples, even with a minimum number of classifiers. The increase in the number of base classifiers led to a decrease in the size of the resulting sample and a decrease in the relative accuracy of the model. Such results are due to the fact that with the increase in the number of base classifiers decreased the number of instances, the total weight of which reached the value of a given threshold, and probably removed significant instances, which reduced the effectiveness of the model. Despite the disadvantages of the proposed method, in practical application there is a range of the number of classifiers, in which the relative accuracy of the method will be higher, the classifier based on the initial sample. Also, the proposed method of instance selection requires a quite large number of initial parameters. Therefore, there is a need to create a method capable of independently estimating the initial parameters of the model. For this purpose, it is necessary to develop mechanisms of model evaluation and determination of the initial parameters of the method.

7 Conclusions

The task of reducing marked data samples of large size for building diagnostic and recognizing models by precedents is considered. The results of the experiments have shown the efficiency of the proposed method on all investigated samples.

The scientific novelty of the obtained results is that a new method has been created, reducing the size of the marked samples, saving the most significant instances and removing the less informative ones. Thus, the proposed method allows to solve the data reduction problem, increasing the efficiency of the training sample, by removing irrelevant and redundant instances.

The practical significance of the results obtained is that the software implementing the proposed method has been developed. The given software has been experimentally investigated at the decision of problems reduction of synthetic and real world data. The conducted experiments have confirmed working capacity of the developed software. The results of the performed experiments allow recommending the use of the developed method and its software for solving the problems of technical diagnostics.

Further research in the field of reducing training samples by building ensembles of classifiers can be conducted in the following directions:

- development of adaptive ensembles of classifiers with a minimum number of initial parameters;
- using ensembles of classifiers different types;
- the use of different approaches to the formation of local samples for base classifiers, to create balanced data sets;
- search for optimal classification methods for the synthesis a base classifiers;
- implementation of the proposed method for multiprocessor systems operating in parallel modes.

8 References

1. García, S., Luengo, J., Herrera, F.: Data Preprocessing in Data Mining. Springer, Switzerland (2015)
2. García, S., Ramírez-Gallego, S., Luengo, J. et al.: Big data preprocessing: methods and prospects. In: Big Data Anal, vol. 1(9). Springer (2016). doi: 10.1186/s41044-016-0014-0
3. Chen, M., Mao, S., & Liu, Y.: Big Data: A Survey. In: Mobile Networks and Applications, vol. 19(2), pp. 171–209. Springer (2014). doi:10.1007/s11036-013-0489-0
4. Han, J., Kamber, M., Pei, J.: Data Mining: Concepts and Techniques. Morgan Kaufmann, USA (2011)
5. Subbotin, S.: The sample properties evaluation for pattern recognition and intelligent diagnosis. In: The 10th International Conference on Digital Technologies 2014, Zilina, 9–11 July 2014, pp. 332–343. IEEE, Los Alamitos (2014)
6. Subbotin, S., Oliinyk, A., Levashenko, V., Zaitseva, E.: Diagnostic rule mining based on artificial immune system for a case of uneven distribution of classes in sample. In: Communications - Scientific Letters of the University of Zilina, vol. 18(3), pp. 3-11 (2016)
7. Boguslayev, A. V., Oleynik, Al. A., Oleynik, An. A. et al.: Progressivnyye tekhnologii modelirovaniya, optimizatsii i intellektual'noy avtomatizatsii etapov zhiznennogo tsikla aviatsionnykh dvigateley: monografiya [Progressive technologies for modeling, optimization, and intelligent automation of aircraft engine life cycle stages: monograph]. OAO «Motor Sich», Zaporozh'ye (2009)
8. Harley, J. B., Sparkman, D.: Machine learning and NDE: Past, present, and future. In: AIP Conference Proceedings, vol. 2102(1) (2019). doi:10.1063/1.5099819
9. Murphy, K.: Machine Learning: A Probabilistics Perspective. The MIT Press, Cambridge, Massachusetts (2012)
10. Bishop, C. M.: Pattern recognition and machine learning. Springer, New York (2014)
11. Lavrakas, P.J.: Encyclopedia of survey research methods. Sage Publications, Thousand Oaks (2008)
12. Fernández-Delgado, M., Cernadas, E., Barro, S. et al.: Do we need hundreds of classifiers to solve real world classification problems? In: Journal of Machine Learning Research, vol. 15(1), pp. 3133-3181 (2014)
13. Samarev, R., Vasnetsov, A., Smelkova, E.: Generalization of metric classification algorithms for sequences classification and labelling. In: ArXiv: 1610.04718 (2016)
14. Abu Alfeilat, H. A., Hassanat, A. B. A., Lasassmeh, O. et al.: Effects of Distance Measure Choice on K-Nearest Neighbor Classifier Performance: A Review. In: Big Data, vol. 7(4), pp. 221-248. Mary Ann Liebert, Inc. (2019). doi:10.1089/big.2018.0175

15. Subbotin, S.: The neuro-fuzzy network synthesis and simplification on precedents in problems of diagnosis and pattern recognition. In: Optical Memory and Neural Networks (Information Optics), vol. 22(2), pp. 97-103. Springer (2013)
16. Subbotin, S.: Quasi-relief method of informative features selection for classification. In: 2018 IEEE 13th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT 2018), Lviv, 11-14 September 2018, pp. 318-321. Vezha i Ko, Lviv (2018)
17. Subbotin, S.: Methods of data sample metrics evaluation based on fractal dimension for computational intelligence model building. In 4th International Scientific-Practical Conference. Problems of Infocommunications. Science and Technology (PICS&T), Kharkov, 10-13 Oct. 2017, pp. 1-6. IEEE, Los Alamitos (2017)
18. Haro-García, A., Cerruela-García, G., García-Pedrajas, N.: Instance selection based on boosting for instance-based learners. In: Pattern Recognition, vol. 96. Elsevier (2019). doi: 10.1016/j.patcog.2019.07.004
19. Hamidzadeh, J., Monsefi, R., Sadoghi Yazdi, H.: IRAHC: Instance Reduction Algorithm using Hyperrectangle Clustering. In: Pattern Recognition, vol. 48(5), pp. 1878-1889. Elsevier (2015). doi: 10.1016/j.patcog.2014.11.005
20. Subbotin, S., Oliynyk, A.: The sample and instance selection for data dimensionality reduction. In: Szewczyk, R., Kaliczynska, M. (eds.) Advances in Intelligent Systems and Computing, vol. 543, pp. 97-103. Springer, Cham (2017)
21. Subbotin, S.: The instance and feature selection for neural network based diagnosis of chronic obstructive bronchitis. In: Bris, R., Majernik, J., Pancerz, K., Zaitseva, E. (eds.) Studies in Computational Intelligence, vol 606, pp. 215-228. Springer, Cham (2015)
22. Subbotin, S.: Methods of sampling based on exhaustive and evolutionary search. In: Automatic Control and Computer Sciences, vol. 47(3), pp. 113-121. Springer (2013)
23. Liu, H., Motoda, H.: On issues of instance selection. In: Data Mining and Knowledge Discovery, vol. 6, pp. 115-130. Springer (2002)
24. Olvera-López, J. A., Carrasco-Ochoa, J. A., Martínez-Trinidad, J. F. et. al.: A review of instance selection methods. In: Artificial Intelligence Review, vol. 34(2), pp. 133-143. Springer (2010). doi:10.1007/s10462-010-9165-y
25. Garcia, S., Derrac, J., Cano, J. R. et. al.: Prototype Selection for Nearest Neighbor Classification: Taxonomy and Empirical Study. In: IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 34(3), pp. 417-435. IEEE (2012). doi: 10.1109/tpami.2011.142
26. Kavrin, D., Subbotin, S.: The sampling method preserving interclass boundaries. In: CEUR Workshop Proceedings of the Second International Workshop on Computer Modeling and Intelligent Systems (CMIS-2019), vol. 2353, pp. 664-673. CEUR-WS (2019)
27. Kuncheva, L. I.: Combining pattern classifiers: methods and algorithms. John Wiley & Sons, Inc., Hoboken, New Jersey (2014)
28. Polikar, R.: Ensemble based systems in decision making. In: IEEE Circuits and Systems Magazine, 6(3), pp. 21-45. IEEE (2006). doi:10.1109/mcas.2006.1688199
29. Valentini, G., Re, M.: Ensemble methods: a review. In: Advances in Machine Learning and Data Mining for Astronomy, pp. 563-594. Chapman & Hall (2012)
30. Sammut, C., Webb, G.: Encyclopedia of Machine Learning. Springer (2017)
31. Breiman, L.: Bagging predictors. In: Machine Learning, vol. 24, pp. 123-140. Springer (1996)
32. Kotsiantis, S., B.: Bagging and boosting variants for handling classifications problems: a survey. In: The Knowledge Engineering Review, 29(1), pp. 78-100. Cambridge University Press (2013). doi:10.1017/s0269888913000313
33. Zhou, Z., H.: Ensemble Methods Foundations and Algorithms. Chapman & Hall (2012)

34. Dietterich, T.G., Bakiri, G.: Solving multiclass learning problems via error-correcting output codes. In: *Journal of Artificial Intelligence Research*, vol. 2, pp. 263–286. AI Access Foundation, USA (1995)
35. Wolpert, D., H.: Stacked generalization. In: *Neural Networks*, vol. 5(2), pp. 241–259. Elsevier (1992). doi: 10.1016/S0893-6080(05)80023-1
36. Yoav, F., Schapire, R., Abe, N.: A Short Introduction to Boosting. In: *Journal of Japanese Society For Artificial Intelligence*, vol. 14(5), pp. 771–780. (1999)
37. Blachnik, M.: Instance Selection for Classifier Performance Estimation in Meta Learning. In: *Entropy*, vol. 19(11) (2017). doi:10.3390/e19110583
38. Rokach, L.: Ensemble-based classifiers. In *Artificial Intelligence Review*, vol. 33, pp. 1–39. Springer (2010). doi:10.1007/s10462-009-9124-7
39. Jordan, M., I., Jacobs, R., A.: Hierarchical mixtures of experts and the EM algorithm. In: *Neural computation*, vol. 6(2), pp. 181–214. IEEE (1994). doi: 10.1162/neco.1994.6.2.181
40. Haykin, S., O.: *Neural Networks and Learning Machines*. Pearson (2008)
41. Kuhn, M., Johnson, K.: *Applied Predictive Modeling*. Springer (2018)
42. Thompson, S.K.: *Sampling*. John Wiley & Sons, Hoboken (2012)
43. Cochran, W.G.: *Sampling Techniques*. John Wiley & Sons, New York (1977)
44. Chaudhuri, A., Stenger, H.: *Survey sampling theory and method*. Chapman & Hall, New York (2005)
45. Good, P., I.: *Resampling methods: a practical guide to data analysis*. Birkhäuser (2005)
46. Scheaffer, L., Mendenhall, W., Lyman Ott, R. at. al.: *Elementary Survey Sampling*. Cengage Learning (2011)
47. Breiman, L.: Pasting Small Votes for Classification in Large Databases and On-Line. In: *Machine Learning*, vol. 36, pp. 85–103. Springer (1999)
48. Louppe, G., Geurts, P.: Ensembles on Random Patches. In: *Lecture Notes in Computer Science*, pp. 346–361. Springer (2012). doi:10.1007/978-3-642-33460-3_28
49. Wu, X., Kumar, V., Quinlan, J. at. al.: Top 10 algorithms in data mining. In: *Knowledge and Information Systems*, vol. 14, pp. 1–37. Springer (2008). doi: 10.1007/s10115-007-0114-2
50. Elkan, C.: The foundations of cost-sensitive learning. In: *17th international joint conference on Artificial intelligence 2001*, vol. 2, pp. 973–978. Morgan Kaufmann Publishers Inc. (2001)
51. Fawcett T.: An Introduction to ROC Analysis. In: *Pattern Recognition Letters*, vol. 27(8), pp. 861–874. Elsevier (2005). doi: 10.1016/j.patrec.2005.10.010
52. Zhang, S., Cheng, D., Deng, Z. at. al.: A novel KNN algorithm with data-driven k parameter computation. In: *Pattern Recognition Letters*, vol. 109, pp. 44–54. Elsevier (2018). doi: 10.1016/j.patrec.2017.09.036
53. Parsons, V., L.: Stratified Sampling. In: *Wiley StatsRef: Statistics Reference Online*. John Wiley & Sons (2017). doi: 10.1002/9781118445112.stat05999.pub2
54. Lyon, R., J., Stappers, B., W., Cooper, S. at. al.: Fifty Years of Pulsar Candidate Selection: From simple filters to a new principled real-time classification approach. In: *Monthly Notices of the Royal Astronomical Society*, vol. 459(1), pp. 1104–1123. Oxford (2016). doi: 10.1093/mnras/stw656
55. Alcalá-Fdez J., Fernandez A., Luengo J. at. al.: KEEL Data-Mining Software Tool: Data Set Repository, Integration of Algorithms and Experimental Analysis Framework. In: *Journal of multiple-valued logic and soft computing*, vol. 17(4), pp. 255–287. Old city publishing (2010)

Optical-Electronic System of Automatic Detection and High-Precise Tracking of Aerial Objects in Real-Time

Igor Shostko^[0000-0002-5612-3080], Andriy Tevyashev^[0000-0001-5261-9874],
Yuliia Kulia^[0000-0001-6541-7913], Anton Koliadin^[0000-0001-5552-5080]

Kharkiv National University of Radio Electronics, Kharkiv, 14 Nauky Ave., UKRAINE
ihor.shostko@nure.ua, tad45ua@gmail.com,
yuliia.kulia@nure.ua, anton.koliadin@nure.ua

Abstract. The article describes the results of the development of digital video processing technology in the visible and infrared frequency bands for the automatic detection and precision tracking of aerial objects in real-time. Algorithm and software based on automatic detection and precision tracking of aerial objects in real-time developed. The algorithm is tested and its performance is evaluated. The algorithm performance was evaluated by measuring the time spent processing each frame in sequence. As a result of testing, it was found that when executing the Field-Programmable Gate Array (FPGA) algorithm, the time spent processing the frame does not depend on the object configuration, frame-filling, and background characteristics. The speed of the algorithm when it is executed on the FPGA at frame size 1920x1080 exceeds the speed of execution on the Personal computer (PC) more than 20 times.

Keywords: optical-electronic system, digital video processing, detection and tracking of moving objects.

1 Introduction

Optical-electronic systems (OES) with automatic detection and tracking of moving objects are used in solving various problems in systems:

- machine vision;
- video surveillance;
- weapons management.

The development of high-tech specimens of equipment using OES is increasingly demanding the detection, tracking and delivery of measurement information about the parameters of the movement of objects of observation. At this time, the most demanding of the delay in processing the image of the video stream are air defense systems, weapons management. Therefore, there is a need to develop OES capable of auto-detecting and precision real-time tracking of aerial objects.

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

2 Relevance of research

The automatic optical-electronic system in the process of detection and tracking of the target solves the following tasks:

- improving image quality (reducing visible noise on the image, improving image contrast, improving clarity);
- adaptation of the camera to poor working conditions (low light, fog and other atmospheric influences).
- detection of all moving objects in the video stream;
- search for a moving object based on a priori information about its type, shape, speed and nature of movement;
- measurement of geometric parameters of the objects of observation;
- automatic identification of the object type by comparison with the standard in shape, color, surface texture, speed of movement;
- formation of special points of the accompanying object - the most informative or vulnerable points, the center of mass, etc.
- accurate support of the object on the selected points on its surface;
- predicting the trajectories of objects accompanied;
- re-capture of the object in case of failure of its accompaniment.

Despite the progress achieved in solving these problems separately [1 - 9], there is currently no comprehensive solution to the problem of automatic detection in the optical range of moving objects and their high-precision real-time tracking. Therefore, the task of developing a technology for detecting and tracking moving objects in the air environment in the visible and infrared frequencies in real-time is urgent.

Object of study - the process of digital processing of video stream in real-time for detection and high-precision tracking of moving objects in the air environment.

The purpose of the work is the development of digital video processing technology in the visible and infrared frequency bands for automatic detection and precision tracking of aerial objects in real-time.

Research Methods - Mathematical, computer-aided modeling and experimental study of digital video processing for real-time high-precision detection of aerial objects.

Detection and precision tracking of aerial objects consists of an orderly sequence of procedures for obtaining, processing, analyzing video information, making and implementing decisions for precision escorting an object in the context of high a priori uncertainty about the behavior of the surveillance object and the environment. The technology is implemented in the form of mathematical, algorithmic and hierarchically ordered software complex for digital processing of video streams coming from television and thermal imager channels. The software package provides the following functions: image acquisition; pre-filtering the image; detecting moving objects in the video stream; optical flow calculation; formation of object support point, high-precision support of moving objects in the air environment; Search for missing objects recapture of lost objects that were accompanied.

Video streams from CCTV cameras are processed in real-time using a Xilinx Zynq UltraScale + MPSoC ZCU104 boards and a personal computer.

3 Literature review

The development of high-tech weapons and military equipment is increasingly demanding the detection, tracking and delivery of measurement information on target movement parameters. Currently there are both high-precision stationary measuring complexes and fully automated optical-electronic and television measuring complexes (USA - Raptor system manufactured by IEC Infrared Systems; CZECH REPUBLIC - Sirius Sea system; NORWAY - Multi-Sensor DC system; FRANCE - Thales and its Margot 8000 system, SAFRAN and its TEOS; UK - Instro's Advanced Stabilized Sensor Platform (ASSP) system, SelexEs Nerio-URL system; RUSSIA - Frigate Station; AUSTRALIA - ETS Imaging's ATS 3000 system; TURKEY - Seaeye-CUPRA station). Today, with its pivoting platform Mantis, Instro is one of the leaders in the world. High-precision measurement in these systems is ensured by state-of-the-art technology and an elemental base, combined with the high dynamic performance of electromechanical tracking systems with the processing of measurement results on a computer.

Analytical review of optical-electronic systems can conclude that there are too many systems for object tracking and detection, but they do not have the hardware for trajectory measurements of object motion in real-time and mode. Therefore, it is currently urgent to develop a new automated optical-electronic measuring system and special algorithms for its operation, aimed at working in automatic mode and in real-time.

4 Materials and methods

OES is a complex of software and hardware modules:

- optoelectronic module;
- support-rotary device;
- digital image processing module;
- software for real-time detection and precision tracking of aerial objects.

Optoelectronic module includes digital cameras of the tracking system in the visible wavelength range and in the infrared range. Each digital camera is an optical module with a lens, a light-sensitive sensor, an exposure and aperture control unit, as well as a control interface.

The operation of the OES is possible with the receipt of a preliminary target from external means or offline. Targeting automatically captures the target for auto-tracking. The target is maintained in the visible wavelength range or in the infrared range, providing round-the-clock use of the OES. Visual accompaniment is recorded on a digital information storage.

A structural diagram of the OES is shown in Fig. 1. The block diagram of the algorithm of functioning of the OES is shown in Fig. 2.

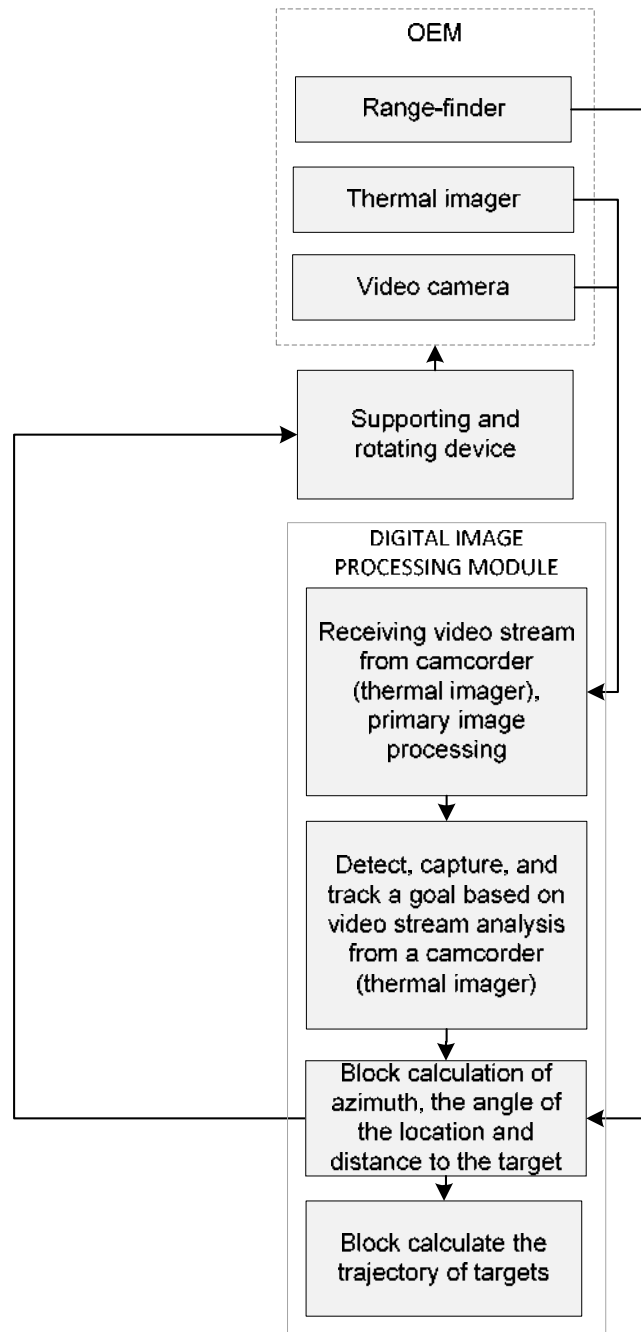
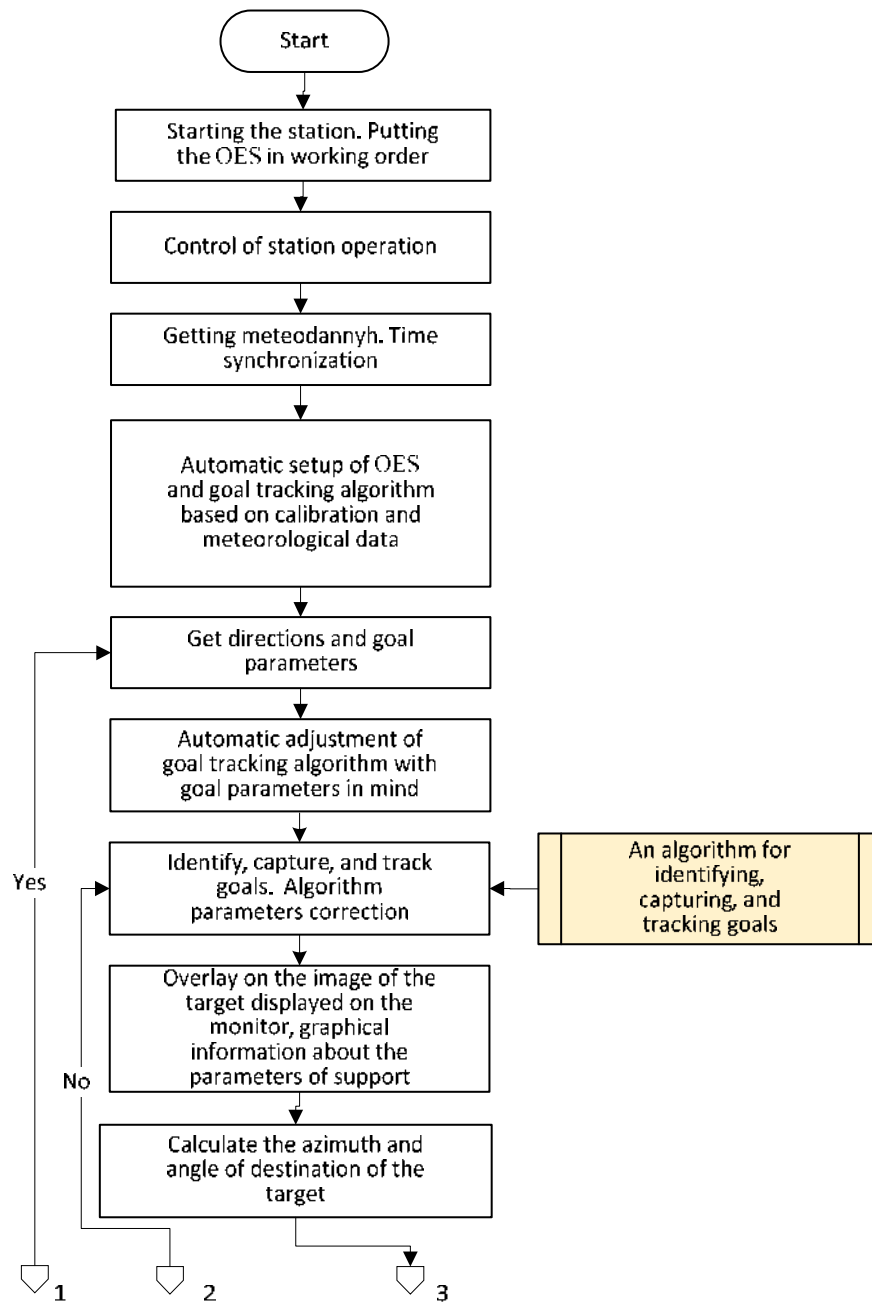


Fig. 1 Block diagram OES



continuation of the flowchart

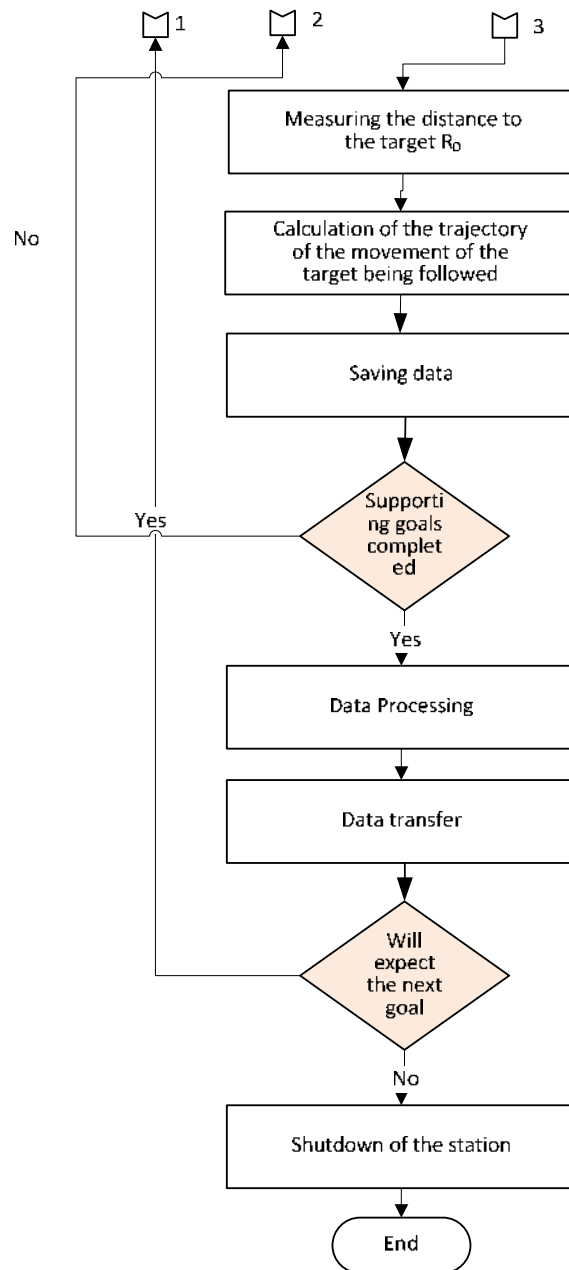


Fig. 2. Flowchart of the OES functioning algorithm

5 Experiment.

5.1 Development of prototype OES

Development of a prototype OES detecting and tracking moving objects in the air environment was carried out using the licensed software package SolidWorks - Computer-aided design system for automation of the stages of design and technological preparation of production in the Microsoft Windows environment.

SolidWorks's three-dimensional solid-state and surface parametric design principle was used in the development of the OEC reference-rotation device, which made it possible to create volumetric structural parts and assemble assemblies in the form of three-dimensional electronic models, which created two-dimensional drawings and specifications in accordance with the design requirements documentation.

Three-dimensional modeling of the Support-Rotary Device has given a number of significant advantages over the traditional two-dimensional design: elimination of errors of assembly of the product at the design stage; creation by electronic model of the part of the control program for processing on the machine with the Numerical Software Control of the fork of the Support and Rotary Device. Three-dimensional parts of the Support-Rotary Device were obtained as a result of a combination of three-dimensional primitives. The consistent extension of 3D objects allowed us to eventually obtain a Supporting Rotary Device that meets all technical and technological requirements of the OES. To carry out strength studies of the design of the Support-Rotary Device, the SOLIDWORKS Simulation package was used, which allowed to accelerate the design process of the Rotary-Rotary Device with the guaranteed properties.

On the basis of the developed design documentation, a prototype of the OES was made to detect and tracking moving objects in the air environment at the research plant Kharkov National University Research of Fig. 3.

The OES prototype consists of: support-rotary device; optical-electronic module in the composition of the television camera, thermal imager and laser rangefinder; digital image processing module that is designed to identify and support air objects based PC and board Xilinx Zynq UltraScale + MPSoC ZCU104.

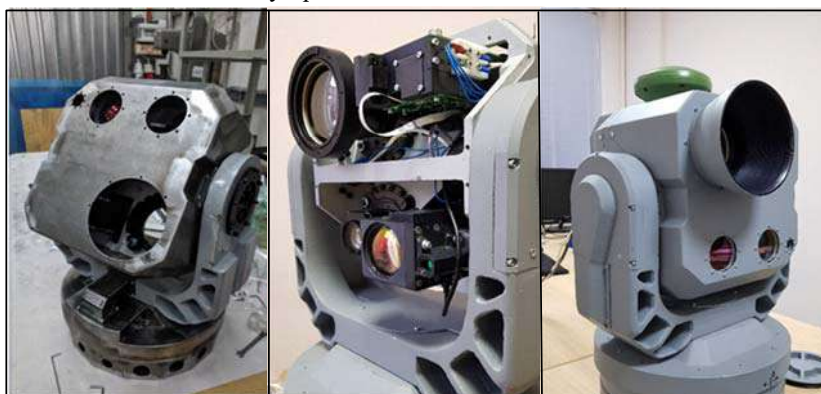


Fig. 3. Production of optical-electronic system

5.2 Digital image processing module

Digital image processing module (DIPM) provides:

- receive video from one of the cameras, primary image processing and real-time detection, capture and tracking of goals based on video stream analysis;
- the overlay on the image of the purpose of graphic information and parameters of support;
- the calculation of the trajectory of the movement of the target being accompanied;
- formation of control commands of the supporting-turning device.

The following basic features and limitations related to the nature of the tasks being solved have been taken into account in the development of image analysis methods and algorithms for processing these images in the DIPM module:

- 1) a priori information about the characteristics of the observed objects and background is often missing or includes only approximate object sizes;
- 2) image processing and analysis must be performed in real-time;
- 3) the capture and tracking of the target must take place offline or with minimal involvement of the operator.

Basic approaches to improving vision in complex surveillance conditions are based on the application of different methods of linear and nonlinear spatial-temporal filtering, algorithms for estimating the parameters of geometric image transformations.

Among the approaches to the detection and measurement of coordinates of objects that have proven, there are four main classes of methods.

1. The benchmark-based methods can be used to measure the coordinates of moving and fixed objects observed on a homogeneous and heterogeneous background, with small signal-to-noise ratios. To use this method, it is necessary to further develop a database of portraits (spatial-frequency characteristics) of typical goals.

2. Statistical segmentation methods are intended to highlight moving and fixed objects that are observed on a relatively homogeneous background. They are based on the use of a priori information about the difference between the statistical properties of the object and the background. When using a thermal imaging camera, if the temperature of the target or its fragments, such as an engine, significantly exceeds the temperature of the surrounding background elements, then adjusting the contrast and brightness of the image can get rid of the background. This method is chosen as the main one working at night, when accompanying an object with a thermal imaging camera.

3. Methods for object selection using spatial filtering. These methods are based on the use of linear and non-linear spatial filtering of the object in the image by color, design feature, by the presence of graphic or text characters applied to the housing. This class of methods is most effective at highlighting moving and fixed objects against a clear or cloudy sky when the camera is moving.

4. Methods based on the allocation of dynamic changes are focused on solving the problem of selection of moving objects observed on a homogeneous or heterogeneous background. The principle of operation of such algorithms is based on the detection of changes that occur over time in the observed sequence of images. In the case of a

moving camera, the offset of the picture is initially offset by the movement of the camera. This method is effective when the accompanying object moves against a sky or static background surface. Therefore, this method is chosen as the main one when working in the daytime on moving targets.

Among the promising concepts for the development of information technologies for image processing and analysis in optical-electronic systems can be distinguished:

- 1) application of complex information received by different image registration channels: television, thermal imaging, radar; it allows to increase reliability of detection, allocation and estimation of parameters of objects at work at long distances, in conditions of low visibility and at use of various masking means;

- 2) the use of structural methods for detecting and evaluating the parameters of objects based on the detection and analysis of visual primitives (nodal points, segments, arcs) corresponding to the observed object; this class of methods is most effective for detecting and measuring the coordinates of objects observed under conditions of image deformation;

- 3) analysis of the background on which the objects are observed, and automatic selection of the algorithm for detecting and measuring the coordinates of the objects, which is most effective in the current observation conditions; this makes it possible to increase the degree of autonomy of the optical-electronic system, which eliminates in many cases the need for human operator intervention.

A significant feature of the presented concept is the presence of a stage analysis of the target environment, which results in the decision to use an algorithm for tracking and measuring coordinates of objects.

At the heart of the DIPM is the FPGA processor and matrix. The processor manages and the FPGA performs the calculations. The combination of these devices is due to the fact that modern processors lack performance. System performance is determined not only by the speed of image compression of the video frame, but also by the speed of pre- and post-processing (scaling, sweep conversion, filtering, color conversion, etc.). These procedures take longer than the actual compression process.

The project uses Xilinx Zynq UltraScale + MPSoC ZCU104 board. The board provides camera image capture and hardware processing, which minimizes delays and increases the speed of the DIPM. Xilinx Zynq UltraScale + MPSoC ZCU104 Central Plug-in Module is a FPGA-integrated, quad-core 64-bit ARM® Cortex®-A53 processor (up to 1.5 GHz) with hardware virtualization, AMP and Trust technology. The co-processor of this system is a two-core Cortex-R5 real-time processor (up to 600 MHz). For hardware acceleration of video processing, the system has a graphical accelerator ARM Mali-400 (clock speed up to 667 MHz), as well as video codec H.264-H.265, support for DisplayPort, MIPI and HDMI. The board has high-speed USB3.0, SATA3.1, GigabitEthernet, SD / SDIO peripherals. Power management is carried out in a separate subsystem PMU, which performs the functions of control and management of power throughout the system. The reVISION platform for the ZCU104 provides software development in the SDSoc programming environment with OpenCV libraries, allowing you to develop effective real-time streaming video algorithms.

The effectiveness of combining the processor and the FPGA matrix on the same platform can be demonstrated by the example of the solution of the problem of image stabilization. To reconcile the two frames, a parameter called the sum of the absolute

difference (SAD) is calculated. If the core of the system contains only the processor, then the SAD calculation can take up to 65% of the processor time. Considering that this is not the only task that the processor performs, it turns out that it is fully loaded. FPGA application makes it easier for DSPs - the matrix will count SAD 10 times faster. This will free up computing capabilities for other tasks. However, when transferring tasks from the processor to the FPGA, they will exchange data. Therefore, the choice between the processor and the FPGA is made on the principle: if the calculations require the execution of hundreds of millions of accumulations per second, they are carried out on the processor; if such operations are more than a billion per second, then the FPGA. On FPGAs it is advisable to perform tasks such as median filtering or feature selection of an accompanying object.

5.3 Description of algorithms for identifying, capturing and tracking targets

The algorithm is based on modern methods of finding and tracking moving objects in a video stream obtained from a camcorder or infrared camera.

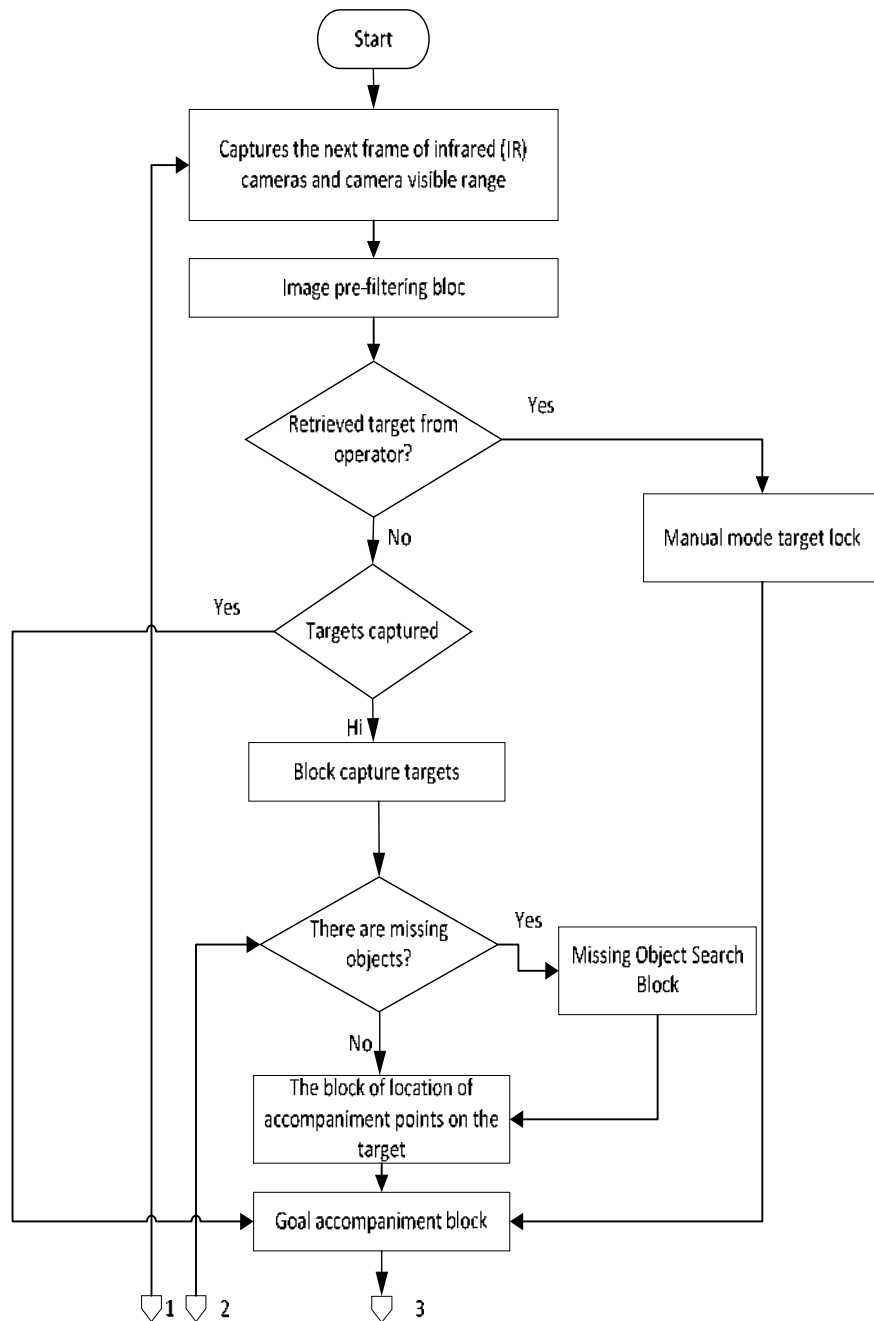
The algorithm can operate in full automatic mode of capture and tracking of targets, in semi-automatic mode when the target is specified from an external targeting system or by the operator, and tracking is automatic, and in full manual mode, when the operator directly controls the actuators with the joystick.

The algorithm consists of individual blocks.

The main blocks of the algorithm:

- 1) The block of receiving the image.
- 2) Image pre-filtering unit.
- 3) The block capture targets.
- 4) Missing object search block.
- 5) The block of location of support points on the target.
- 6) Goal support block.
- 7) Block of filtering of the monitored points.
- 8) Trajectory prediction unit.
- 9) Block analysis of support errors.

A block diagram of the algorithms for identifying, capturing and tracking targets is shown in Fig. 4.



continuation of the flowchart

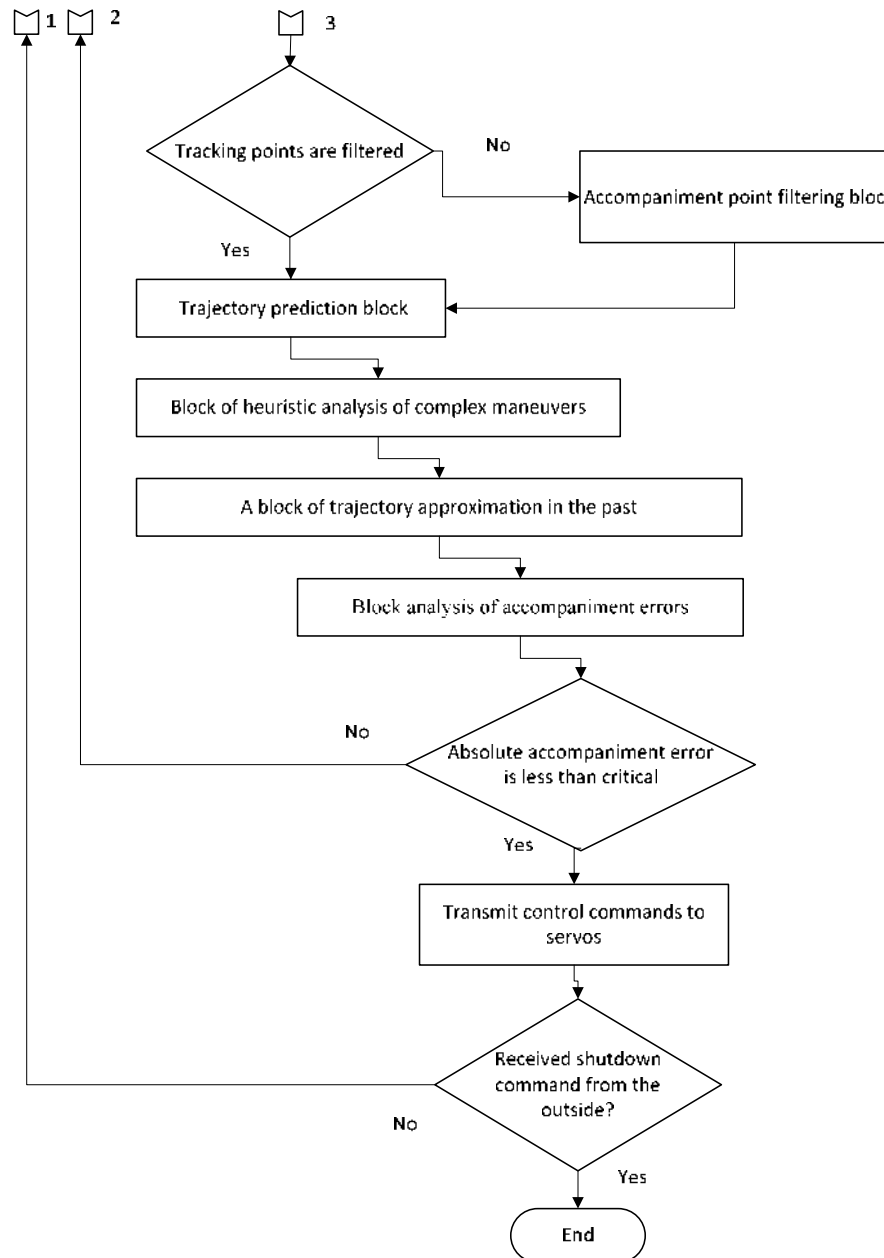


Fig. 4. Flowchart of goal detection, capture and tracking algorithms

1. Getting started with the algorithm. The block of receiving the image.

The algorithm starts by choosing the source of the video stream - an infrared camera or a visible range camera, or both cameras together.

The image receiving unit performs the tasks of managing the cameras and converting the image to an easy-to-process format. The camera controls depend on the particular camera model. The main adjustable parameters are optical zoom, focus, frame exposure time, matrix sensitivity (ISO), resolution, frame rate. Some of the settings can be adjusted automatically by the camera without external commands.

An important parameter for the algorithm operation is the resolution of the frame. It depends on the speed of the algorithm and the quality of the image. Most pre-processing, object search, and frame analysis operations have complexity that is directly proportional to the number of pixels in the frame. According to preliminary tests of the software prototype, image processing at 640x480 resolution was in the range of 150-300 frames per second, at 1280x720 resolution - 40-100 frames per second.

Increasing the resolution allows distinguishing more details on the target, which gives further information on facilities for algorithms using analysis of form and texture of the object. Using a color camera adds more information about objects and allows you to use algorithms that are sensitive to the target's color histogram. But due to physical limitations, increasing the resolution of the matrix or using a color camera will reduce the physical size of pixels on the matrix, which increases the noise and reduces the quality of the image (especially in low light conditions). Therefore, it is advisable to use the main camera in monochrome, with a large matrix size and a small resolution (640x480) with a frame rate of 30-60 per second.

2. Pre-filtering unit of the image.

Standard image filtering operations such as linear pixel averaging, median filtration, Gaussian blur, mathematical morphology, anisotropic pixel diffusion, Wiener spatial filter, and others are used during image pre-processing. Filters are selected according to the following methods of capture and tracking purposes.

3. The block capture targets.

A. Manual capture (semi-automatic mode). The goal is set manually by the operator. In this case, the auto-capture detector is applied to the operator-selected area of the so-called area of interest (ROI).

B. Fully automatic mode. The choice of capture method in automatic mode depends on the time of day, weather conditions, and the parameters of the object being accompanied.

Moving object detector (Highlighting dynamic changes). For the primary capture of moving targets, a cascade motion detector consisting of sequential image processing operations is used, characterized by a high speed of operation, versatility in the appearance of objects, and the ability to add or remove operations from an image processing sequence during operation. At this point, it is preferable that the platform does not move so as not to interfere with the capture stage, but if this is not possible, the offset of the image shift is pre-combined by maximizing the correlation between the two frames using fast Fourier transform and phase correlation.

Detection of moving objects is performed as a result of the following operations:

- 1) calculation of inter-frame difference;
- 2) morphological operation of erosion;

- 3) morphological expansion operation;
- 4) morphological operation of selection of object boundaries;
- 5) iterative approach of connected domains.

The detector contrast objects. Instead of a motion detector, statistical object segmentation is used to capture still objects based on the contrast of the object and background, or an object search is performed for certain features, such as color and shape (spatial filtering). That is, instead of comparing the inter-frame difference with the threshold, the values of the colors or the intensity of the pixels with the given thresholds are compared. Subsequent operations of the morphology and iterative traversal of the areas remain the same.

Detector objects by color (in case of color camera use). Object Detector Structure by Color

- 1) highlight areas of a particular color;
- 2) morphological operation of erosion;
- 3) morphological expansion operation;
- 4) morphological operation of selection of object boundaries;
- 5) iterative bypass connected regions.

Areas of a particular color are highlighted by converting the image to a HSV color space, in which each pixel is described by color and color saturation. Thus, we select areas whose color lies within the specified range and the intensity is above the threshold.

Instead of selecting objects of a particular color, you can select all objects that are different from the color of a homogeneous background.

Items 2 through 5 remain the same as in the contrast object detector.

Detector of objects of a given contour shape. To search for a form object, the form template is first taken and vectorized by the calculation of the statistical moments.

As a pre-processing of the initial image, a Gaussian filter is used to reduce the gradient at the boundaries.

An object detector based on a template base. To capture objects of a certain type, a method of analyzing histograms of directional gradients is used, where the number of gradient directions and the cascade Viola-Jones cascade classifier that uses Haar traits as descriptors of singular points are given special points. These methods require its own database of objects of all types of interest to us, who sought on the image. For grading purposes, pixel sizes should be larger than 10x10 pixels, otherwise there will be a high probability of errors, so it is not always possible to determine the type or look for objects of a particular type for long-range purposes.

Capture objects at long range. Only the contrast object search method, the color object search method, and partially the moving object search method, can be used to detect long-range aerial targets that appear to be a point or a few pixels. The object should have a high contrast with the background and the background should have a high uniformity.

Additional features can be used to extend the range of capture conditions for a 1 by 1 pixel target. One such feature is the inversion trace, which has much larger linear dimensions, and its shape and direction unmistakably points to a moving object and enables it to detect its type. Search the inversion trace is done by finding objects of a given linear form.

4. Block of missing objects search.

If an accompanying object is lost, then the object is searched in the predicted areas, using known information about the object's previous appearance and image comparison algorithms based on special point descriptors (SIFT, SURF, ORB), boundary shape (ShapeMatch), or cross-correlations, depending on what the most characteristic features an object had before losing it.

5. Object support (block of detection and tracking of characteristic points on the target).

A bloc tracking designed to identify and track feature points in a series of video frames. This bloc is implemented on Xilinx Zynq UltraScale + MPSoC ZCU104. A Harris angle detector is used to detect characteristic points.

Angles are areas of the image with a large difference in pixel intensity in all directions. To find the angle, it is necessary to calculate the difference $E(u,v)$ when shifting in all directions (1):

$$E(u,v) = \sum_{x,y} w(x,y) [I(x+u,y+v) - I(x,y)]^2 \quad (1)$$

where $I(x,y)$ – intensity at the point;

$I(x+u,y+v)$ – intensity in the shifted point;

$w(x,y)$ – window function.

The window function is either a rectangular window or a Gaussian window that gives weight to the difference when displaced.

To identify the angles need to maximize function $E(u,v)$.

The result of image processing by the Harris function is an image in grayscale where the angles are indicated by dots.

A modified version of the Lucas Kanade optical stream is used to track the goal. The main part of the algorithm uses the current and subsequent frames as input and displays a list of characteristic points to be tracked. The current image is the first frame in the set in which the algorithm will detect and track feature points. The number of frames in the set in which you want to track the characteristic points is specified as input.

In block detection and tracking characteristic points on goals using five hardware functions of the library xfOpenCV, which are combined into one new feature tracking:

```
xf :: cornerHarris,
xf :: cornersImgToList,
xf :: cornerUpdate,
xf :: pyrDown,
xf :: densePyrOpticalFlow.
```

The interconnection of hardware functions is shown in fig. 5.

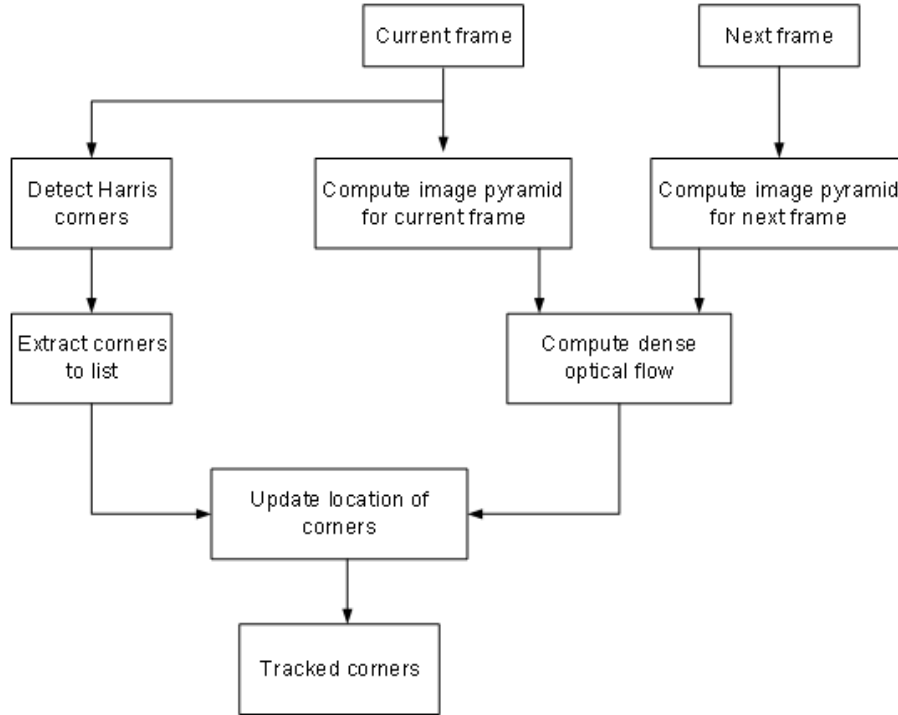


Fig. 5. Tracking feature points using a rarefied optical stream

The tracking function takes the characteristic points from the Harris angle detector, and the dense optical flux vectors from the dense pyramidal optical flux function, and outputs updated coordinates of the characteristic points, tracking the input angles, using the dense flux vectors, thus simulating the behavior of the sparse optical flux. This hardware function is clocked at 300 MHz for 10,000 characteristic points of the image at 720p resolution, adding a minimum delay for the conveyor.

6. Block filtering points that are tracked.

The algorithm uses a screening out operation to remove unnecessary and erroneous points. Unnecessary and erroneous dots appear on the image as a result of background motion. The n extreme points farthest from the center of mass (center of the object) are eliminated in the algorithm, and points whose motion is significantly different from the movement of most points of the object are also rejected. The resulting image can be considered as a binary mask where we are interested only in

To find the bounding frame around the points in the image, we check for each non-zero pixel that its position is outside the current rectangle. If so, we update the location of the rectangle. After the overlapping of the detected characteristic points and the bounding frame on the original image (Fig. 6), we obtain in each frame of the video sequence a frame that moves behind the object (Fig. 7).



Fig. 6. Overlay of detected feature points and bounding box on the original frame image



Fig. 7. Supporting the air target (on the video)

7. Analysis of objects at intersections with other objects and object detection points disappearance from view. (The block of approximation of the trajectory in the past moments and the block of distinction of intersections of goals with other goals and forms of relief (analysis of complex maneuvers)).

The obtained image of moving objects is compared with the previous ones, and in case of loss of any object from the field of view, it is followed in the predicted trajectory, and in the following frames the objects in the forecast areas are searched. It uses all the known information about object previews and image comparison algorithms based on special point descriptors (SIFT, SURF, ORB), boundary shape (ShapeMatch), or cross-correlation, depending on what the most characteristic features were have object before its loss.

8. Block trajectory prediction.

To predict the trajectory of the object using a Kalman filter in which each frame the coordinates of the targets that are not marked as lost.

9. Block analysis of maintenance errors.

The analysis of the reasons for exceeding the error of tracking a given threshold is performed on the basis of the method of hierarchy analysis and allows to establish one of the possible reasons for this exceedance:

- changing the nature of the object's movement (changing speed, maneuvering, hovering, etc.);
- change of video surveillance conditions (change of brightness, contrast, partial or full overlapping of moving object, etc.).

After establishing the causes, one proceeds to one of the previous stages of the algorithm. This process continues until the escort error is less than the specified threshold.

6 Results

Testing the algorithm of tracking the aerial objects and assessing its performance was carried out in the laboratory on static and dynamic scenes using a simulated moving air object, placed on the background, which allows to simulate different conditions of visibility (contrast) of the accompanying object. In the experiment, the movement of the object was recorded using a video camera, real-time video stream was processed using FPGA, video image after processing was displayed on the monitor screen. As a result of the video stream processing, the coordinates of the object are calculated and the object is maintained and held in the frame.

The algorithm performance was evaluated by measuring the time spent processing each frame in sequence. As a result of testing, it was found that when running the FPGA algorithm, the time spent processing the frame does not depend on the object configuration, frame-filling, and background characteristics.

The results of a study of the performance of the algorithm developed and a comparison of the performance of the calculations to implement the Harris function on a personal computer (PC) (CPU: i7 4500u, memory: 8GDDR3) and FPGA are shown in table 1.

Table 1. The results of the comparison

Video resolution	frame	Processing time using a PC, ms	Processing time using FPGA, ms
640x480		17,5	0,98
1280x480		68,2	4,3
1920x1080		272,3	13,6
1920x1920		613,5	46,7

Table 1 shows, that the FPGA-based approach shows much better performance than the CPU. For example, the speed of an algorithm when executed on an FPGA at a frame size of 1920x1080 exceeds the speed of execution on a PC more than 20 times.

Zynq Series Xilinx boards are ideal for video processing. Benefits of the Zynq platform:

- the bandwidth is 12 times higher than the alternative SoCs currently on the market;
- allows you to get 6x better images / sec / watt compared to embedded GPUs and typical SoCs;
- allows you to use video sensor combinations.

The software implementation of the algorithm discussed in this article can be implemented on the Xilinx Zynq-7000 and ZU + MPSoC platforms, including ZCU102, ZCU104.

7 Conclusions

1. The prototype of the OES for the detection and maintenance of moving targets in the air environment is developed. The development of a new automated OES¹ and special algorithms for its operation, aimed at working in automatic mode and in real-time, so the detection and tracking of characteristic points on the target performed on FPGA. We recommend the software implementation of the target point detection and tracking unit using on the Xilinx Zynq-7000 and ZU + MPSoC platforms, including ZCU102, ZCU104 boards.

2. Digital video processing technology in the visible and infrared frequency bands has been developed for the automatic detection and precision tracking of aerial objects in real-time. The main components of the flowchart of the algorithms for identifying, capturing and tracking targets are identified. Software implementation of all digital video processing features on FPGA is not possible, so at this time a compromise solution was proposed using a personal computer (CPU: i7 4500u, memory: 8GDDR3) and ZCU104.

3. Was a study of the performance of the algorithm developed and a comparison of the performance of the calculations to implement the Harris function on a personal computer and FPGA. The speed of the algorithm when it is executed on the FPGA at frame size 1920x1080 exceeds the speed of execution on the PC more than 20 times.

References

1. Yane, B.: *Cifrovaya obrabotka izobrazhenij* (Digital image processing). Tehnosfera. 584 p. (2007) ISBN 978-5-94836-122-2

¹ The work was carried out with the assistance of the Department of Infocommunication Engineering V.V. Popovsky, Department of Applied Mathematics and the test plant of Kharkiv National University of Radio Electronics.

2. Lukyanica, A.A., Shishkin, A.G.: Cifrovaya obrabotka izobrazhenij (Digital image processing). Aj-Es-Es Press. 518 p (2009)
3. Murphy, K.P.: Models for Generic Visual Object Detection. Technical report, Department of Computer Science. University of British Columbia. Vancouver. Canada, p 8 (2005)
4. Viola, P., Jones, M.: Robust Real-Time Object Detection. Intl. J. Computer Vision. 57(2):137-154. Vol. 57(2), p. 137–154 (2004)
5. Bulychev, Yu.G., Manin, A.P.: (2000) Matematicheskie aspekty opredeleniya dvizheniya letatelnyh apparatov (Mathematical aspects of aircraft motion determination). Mashinostroenie. 256 p (2000)
6. Bay, H., Ess, A., Tuytelaars, T., Van Gool, L.: SURF: Speeded Up Robust Features. Computer Vision and Image Understanding (CVIU). Vol. 110, N 3, p. 346–359 (2008)
7. Titov, I.O., Emelyanov, G.M.: Sistema kompyuternogo zreniya dvizhushegosya vozdušnogo obekta (Computer vision system of moving air object). Kompyuternaya optika. T. 35, № 4. pp. 491–494 (2011)
8. Danelljan, M., Khan, F.S., Felsberg, M., van de Weijer, J.: Adaptive Color Attributes for Real-Time Visual Tracking. Conference on computer vision and pattern recognition, pp. 1090–1097. (2014) doi:10.1109/CVPR.2014.143
9. Henriques, J.F., Caseiro, R., Martins, P., Batista, J.: High-Speed Tracking with Kernelized Correlation Filters. IEEE transactions on pattern analysis and machine intelligence. Mar;37(3):583-96. (2015) doi: 10.1109/TPAMI.2014.2345390

Neural model of conveyor type transport system

Pihnastyi Oleh¹[0000-0002-5424-9843] Khodusov Valery²[0000-0003-1129-3462]

¹ National Technical University "KPI" Kharkiv, Ukraine
pihnastyi@gmail.com

² Kharkiv National University named after V.N. Karazin
vdkhodusov@karazin.ua

Abstract. In this paper, a model of a transport conveyor system using a neural network is demonstrated. The analysis of the main parameters of modern conveyor systems is presented. The main models of the conveyor section, which are used for the design of control systems for flow parameters, are considered. The necessity of using neural networks in the design of conveyor transport control systems is substantiated. A review of conveyor models using a neural network is performed. The conditions of applicability of models using neural networks to describe conveyor systems are determined. A comparative analysis of the analytical model of the conveyor section and the model using the neural network is performed. The technique of forming a set of test data for the process of training a neural network is presented. The foundation for the formation of test data for learning neural network is an analytical model of the conveyor section. Using an analytical model allowed us to form a set of test data for transient dynamic modes of functioning of the transport system. The transport system is presented in the form of a directed graph without cycles. Analysis of the model using a neural network showed a high-quality relationship between the output flow for different conveyor sections of the transport system.

Keywords: conveyor, PDE– model, distributed system, transport delay.

1 Introduction

The transport conveyor is a complex dynamic stochastic distributed system. The transport conveyor is an integral part of the technological process at enterprises with the flow method of organizing production [1]. Conveyor transport is widely used in the mining industry [2–6]. Table 1 shows a number of basic characteristics of conveyor-type transport systems. One way to save energy, which is necessary for the functioning of such systems, is to increase the level of congestion of the conveyor line [7–9]. To reduce the energy costs required to move one ton of material along the transportation route, systems are used to control the speed of the belt or the intensity of the material at the entrance of the conveyor section from the input bunker [10–12]. The effectiveness of the conveyor control system is largely determined by the model of the transport system. This fact acquires special significance when designing control systems for a transport system consisting of a large number of sections.

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

Table 1. Characteristics of long-ranged conveyor transport system

Conveyor name	Length (km)	sections	Power (kW)	Speed (m/sec)	Capacity (t/h)
From the Bu Craa mine to the coast at El Aaiún, Western Sahara, [2]	128.7	11			2000
Sasol's Impumelelo project in South Africa (2015), [2]	27.5	1		6.5	2400
The Henderson Coarse Ore Conveying System, the North American Continental Divide (2000), [3,4]	24.0	3	12700	4.5	2270
Çöllolar Lignite Open Pit Mine, Turkey (2011), [5]	17.4	26	46300		9350
From a mine in India to a cement plant in Bangladesh (2005), [2]	16.5	1		6.5	
Neyveli Lignite Corp., India (2007), [5]	14.0	8	2520	5.4	
Open Cast Mine Reichwalde, Germany (2010), [5]	13.5	6	19350	5.5	6000
Coarse ore conveyor system Minera Los Pelambres, Chile (1998), [5]	12.7	3	25000		8700
Tianjin China Port Authority, China (2005) [6]	8.98	1	4x 1500	5.6	6000
Baumgartner Tunnel, the Metropolitan St. Louis (Missouri) , USA [6]	6.18			2.54	200
Barcelona Tunnel (the Metro (Train) Extension Project), Spain (2005) [6]	4.71	1	3.5		1500

2 Literature review

To build the models on which the systems for controlling the speed of the belt or the intensity of material input at the entrance of the conveyor section from the input bunker are based, use the finite element method [13–18]; finite difference method [18,19]; Lagrange method [19]; a method using the aggregated equation of state [20]; system dynamics method [10]; multiple regression method [26–28]. Most often used in models for calculating flow parameters a finite element method. This method allows you to determine the value of the flow parameters of the conveyor section for dynamic transient conditions, taking into account the distribution of material along the transportation route. The finite element method, before the advent of the analytical model (PiKh–model) of the conveyor-type transport system [12], was perhaps the main method used by researchers to construct the conveyor model. The use of neural network methods and multiple regression methods to describe flow parameters was less promising than the finite element method. One of the reasons is that the researchers focused on modelling a single section of the conveyor. Another, no less important reason is the lack of test data in the right amount for training a neural network or for building a regression model. When considering a model of a transport system, which consists of a large number of separate conveyor sections, the use of the finite element method is unreasonable even when modelling a transport system consisting of several dozen separate sections. A good tool, in this case, is the PiKh–

veyor system model. In this case, a separate model is built for each separate section. Combining sections into a common system leads to a system of equations [29, 30]. In the event that a separate section does not include an accumulating bunker, the number of equations of the system is equal to the number of sections. In [29], a model of a conveyor system consisting of 2 sections is presented. In [30], the principles of constructing a model of the main conveyor are considered.

3 Formal problem statement

If the transport system is a conveyor [31], which consists of tens or even hundreds of separate sections, and each section has a system for controlling the rate of material input from the input bunker and a belt speed control system, then using analytical models can be associated with significant difficulties. In this case, the application of methods using the neural network and multiple regression methods is of scientific and practical interest for solving the problem. The more the number of sections in the transport system, the stronger the interest of researchers in applying methods using the neural network and multiple regression methods. In this regard, in this work, we will pay attention to constructing a model of an assembly line using a neural network.

4 Conveyor section model

To describe the conveyor section (Fig. 1) let us use the classic dynamic distributed model of the conveyor in a dimensionless form (PiKh-model) [12]:

$$\frac{\partial \theta_0(\tau, \xi)}{\partial \tau} + g(\tau) \frac{\partial \theta_0(\tau, \xi)}{\partial \xi} = \delta(\xi) \gamma(\tau) \quad (1)$$

$$\theta_0(0, \xi) = H(\xi) \psi(\xi). \quad (2)$$

The state of the flow parameters of the conveyor line at a point in time t at the point of the transport route with the coordinate S is described by dimensionless variables:

$$\tau = t / T_d, \quad \xi = S / S_d, \quad (3)$$

$$\theta_0(\tau, \xi) = [\chi]_0(t, S) / \Theta, \quad \psi(\xi) = \Psi(S) / \Theta, \quad (4)$$

$$\gamma(\tau) = \lambda(t) \frac{T_d}{S_d \Theta}, \quad \gamma_b(\tau) = \lambda_b(t) \frac{T_d}{S_d \Theta}, \quad \gamma_{\max} = \lambda_{\max} \frac{T_d}{S_d \Theta}, \quad (5)$$

$$g(\tau) = a(t) T_d / S_d, \quad \Theta = \max\{\Psi(S), \lambda(t) / a(t)\}, \quad \delta(\xi) = S_d \delta(S), \quad H(\xi) = H(S), \quad (6)$$

$$[\chi]_1(t, S) = a(t) [\chi]_0(t, S), \quad \mathcal{G}(\tau) = \sigma(t) \frac{T_d}{S_d \Theta}, \quad 0 \leq \lambda(t) \leq \lambda_{\max}, \quad (7)$$

where S_d is length of the conveyor line; T_d is the characteristic time of the passage of the material along the transport route; $[\chi]_0(t, S)$, $[\chi]_1(t, S)$ is the linear density of material distribution and material flow at a point in time t at the point of the transport route with the coordinate $S \in [0, S_d]$; Θ is the limit value of the linear density of the material for the analyzed conveyor section; $\Psi(S)$ is the initial distribution of material along the technological route; $\lambda_b(t)$ is the intensity of the flow of material into the bunker; $\lambda(t)$ is the output flow of material from the bunker to the input of the conveyor section, limited by $\lambda_{\max}$; $a(t)$ is conveyor belt speed; $\sigma(t)$ is the predicted output flow of the material from the conveyor section; $\delta(S)$ is delta function; $H(S)$ is Heaviside function.

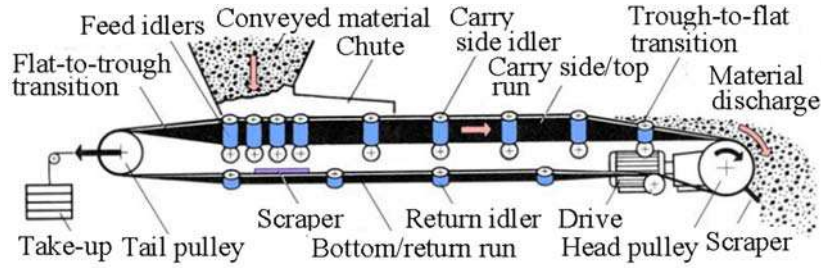


Fig. 1. Schematic diagram of the conveyor line [2]

Equation (1) with initial conditions (2) corresponds to the solution [12]:

$$\theta_0(\tau, \xi) = (H(\xi) - H(\xi - G(\tau))) \frac{\gamma(G^{-1}(G(\tau) - \xi))}{g(G^{-1}(G(\tau) - \xi))} + H(\xi - G(\tau)) \psi(\xi - G(\tau)) \quad (8)$$

$$\theta_1(\tau, \xi) = g_0 \theta_0(\tau, \xi) = \gamma(\tau - \xi / g_0), \quad G(\tau) = \int_0^\tau g(\alpha) d\alpha. \quad (9)$$

The system of equations (8), (9) determines the behaviour of the flow parameters of the conveyor. The linear density of the material along the transport route $\theta_0(\tau, \xi)$ at an arbitrary point in time τ can be determined if the intensity $\gamma(\tau)$ of the rock entering the conveyor line entrance and the speed of the conveyor belt $g(\tau)$ are known. The linear density of the material $\theta_0(\tau, 1)$ and the material flow $\theta_1(\tau, 1)$ at the output from the transport conveyor system $\xi = 1$ is determined by the expressions

$$\theta_0(\tau,1) = \begin{cases} \frac{\gamma(G^{-1}(G(\tau)-1))}{g(G^{-1}(G(\tau)-1))}, & G(\tau)-1 \geq 0; \\ \psi(1-G(\tau)), & G(\tau)-1 < 0; \end{cases} \quad \theta_1(\tau,1) = \theta_0(\tau,1)g(\tau). \quad (10)$$

Equation solution

$$G(\tau_{tr}) - 1 = 0 \quad (11)$$

allows you to calculate the duration of the transition period $\Delta\tau_{tr} = \tau_{tr} - \tau_0$, during which the material flow at the exit from the transport system is determined by the type of expression of the linear density of the material $\psi(\xi)$ at the initial time τ_0 . The linear density of the material $\theta_0(\tau, \xi)$ at an arbitrary point ξ at the time $\tau \geq \tau_{tr}$ is related to the linear density of the material $\theta_0(\tau_\xi, 0)$ at the input of the transport system at the time $\tau = \tau_\xi$

$$\theta_0(\tau, \xi) = \gamma(\tau_\xi) / g(\tau_\xi) = \theta_0(\tau_\xi, 0), \quad \tau \geq \tau_{tr}, \quad \tau_\xi = G^{-1}(G(\tau) - \xi). \quad (12)$$

If we introduce a definition for the delay time $\Delta\tau_\xi = \tau - \tau_\xi$, then expression (13) can be represented as follows

$$\theta_0(\tau, \xi) = \frac{\gamma(\tau - \Delta\tau_\xi)}{g(\tau - \Delta\tau_\xi)} = \theta_0(\tau - \Delta\tau_\xi, 0). \quad (13)$$

The delay time $\Delta\tau_\xi$ sets the period of time during which the element of material received at the entrance of the transport system at a time τ_ξ passes the path along the transportation route equal to ξ . When $\xi = 1$, the expression

$$\theta_0(\tau, 1) = \frac{\gamma(\tau - \Delta\tau_1)}{g(\tau - \Delta\tau_1)} = \theta_0(\tau - \Delta\tau_1, 0) = \theta_0(\tau_1, 0), \quad \tau \geq \tau_{tr}, \quad (14)$$

determines the relationship between the linear density of the material at the input and output. The value of the linear density at the output is equal to the value of the linear density at the input with a delay $\Delta\tau_1$.

5 Conveyor section model using a neural network.

The system of equations (8), (9) determines the linear density of the material along the transportation route and allows you to calculate the material flow at an arbitrary location of the transport path of a separate conveyor section. At a constant speed of movement of the conveyor belt, the expression determining the linear density of the

material $\theta_0(\tau, \xi_m)$ and the material flow $\theta_1(\tau, \xi_m)$ at the time τ at the output of the transportation route $\xi = \xi_m$ takes a simple form with a constant time value of the delay time $\Delta\tau_{\xi m} = \tau - \tau_{\xi m}$. If the speed of the belt is not constant in time, then to calculate the flow parameters of the conveyor transport system, it is necessary to determine the value of the delay time $\Delta\tau_{\xi m}$ for each m-th section from the equation

$$\tau_{\xi m} = G_m^{-1}(G_m(\tau) - \xi_m), \quad G_m(\tau) = \int_{\tau_{0m}}^{\tau} g_m(\alpha) d\alpha. \quad (15)$$

If the transport system consists of a large number M of individual sections, then it is required to solve the M-equations (8), (9). Additional restrictions are imposed due to the complexity of constructing an analytical system of equations that determines the flow of material from the place of production to the place of processing [32, Fig.1 and Fig.2]. Therefore, with a large number M of individual sections, it is advisable to build aggregated models of transport systems. One of the approaches to designing aggregated models of conveyor transport systems is the use of neural networks [32–37]. To describe the functioning of a separate conveyor section of the transport system, we use dimensionless variables (4) - (7) of the model (1), (2), which allow us to determine the state of the flow parameters of the individual conveyor section at a time τ : $\gamma_m(\tau)$ is the intensity of the input flow of material; $g_m(\tau)$ is conveyor belt speed; ξ_m is section transport route length. Let's move on to the construction of a neural network using the example of a branched transport system. As an option for analysis, we will use the structure of the transport conveyor shown in Fig. 2, which consists of 8 separate sections (M = 8). It should be noted that the state of the flow parameters at the output sections (section m = 7,8) is determined by the parameters of the 4 input sections (section m = 1,2,4,5). The transport system has nodes where the material flows converge (Fig. 5.a) and nodes where the material flows diverge (Fig. 3). When considering, let's assume that there is no bunker control. The amount of material flow through the bunker remains unchanged. This situation is common, it represents the case when the parameters of the bunker are not controlled. In this case, the bunker at the entrance of a separate section does not contain material. For nodes in which the material flows converge, the intensity of the input material flow is determined through the parameters of the converging sections. For the case when the node contains two incoming flows and one outgoing (Fig. 3), the balance relation holds:

$$\gamma_3(\tau) = \gamma_1(\tau - \Delta\tau_{\xi 1}) \frac{g_1(\tau)}{g_1(\tau - \Delta\tau_{\xi 1})} + \gamma_2(\tau - \Delta\tau_{\xi 2}) \frac{g_2(\tau)}{g_2(\tau - \Delta\tau_{\xi 2})}, \quad (16)$$

$$\Delta\tau_{\xi m} = \tau - \tau_{\xi m} = \tau - G_m^{-1}(G_m(\tau) - \xi_m). \quad (17)$$

For nodes in which the material flows diverge, the intensity of the input material flow is also determined through the parameters of the converging sections. For the case

when the node contains an incoming flow and two outgoing flows (Fig. 3), the balance ratio has the form:

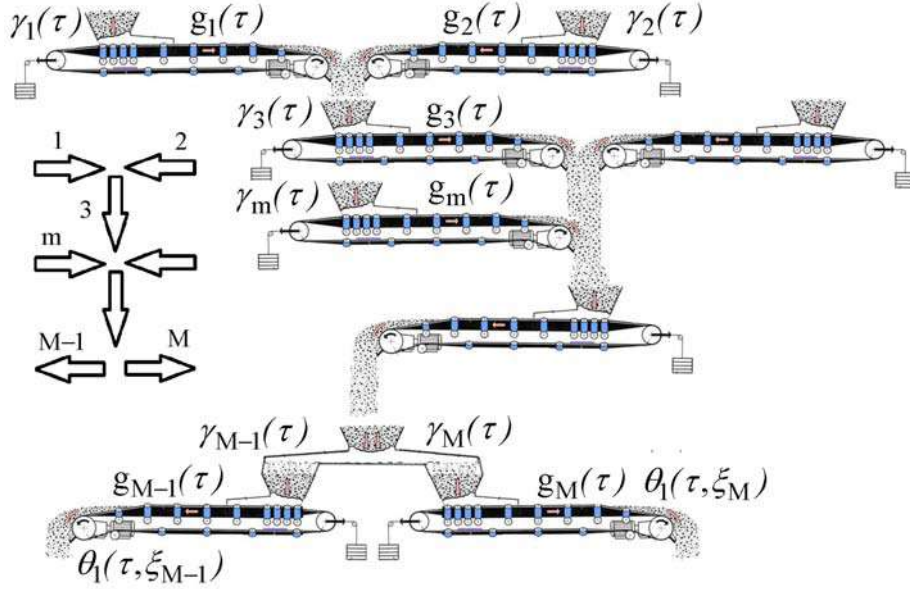


Fig. 2. Diagram of a branched conveyor transport route

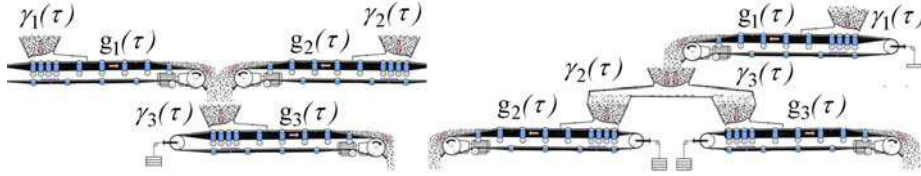


Fig. 3. Schemes for calculating the balance of flows in the nodes of the conveyor transport route: a) converging node; b) diverging node

$$\frac{\gamma_2(\tau)}{\gamma_3(\tau)} = \gamma_{23} = \text{const}, \gamma_3(\tau) \neq 0. \quad (18)$$

Let us assume that the state of the transport system is determined at a moment in time τ , if at that moment in time the parameters of each individual conveyor section are determined: $\gamma_m(\tau)$, $g_m(\tau)$. When constructing an aggregated model of a conveyor transport system in the absence of control, we exclude from consideration the parameters $\gamma_m(\tau)$ of the internal nodes, which can be determined through the flow parameters of the mated sections.

The architecture of the neural network to build an aggregated model (Fig. 4.) Let us introduce the notation for the parameters of the input layer of the neural network

$$x_{3m-2} = \gamma_m(\tau), x_{3m-1} = g_m(\tau), x_{3m} = \xi_m, m=1..M, \quad (190)$$

where m is the number of the conveyor section (Fig. 2). For the transport system model in Fig. 2, the input parameters (20) $x_7 = \gamma_3(\tau)$, $x_{16} = \gamma_6(\tau)$, $x_{19} = \gamma_7(\tau)$, $x_{22} = \gamma_8(\tau)$ are excluded. Similarly, let us exclude the velocities for sections $m=3,6$ from the input layer. We introduce the notation for the parameters of the output layer

$$y_1 = \theta_{17}(\tau, \xi_7), y_2 = \theta_{18}(\tau, \xi_8). \quad (20)$$

The output parameters y_1 and y_2 correspond to the output material flow for $m=7,8$ sections of the transport system Fig. 2.

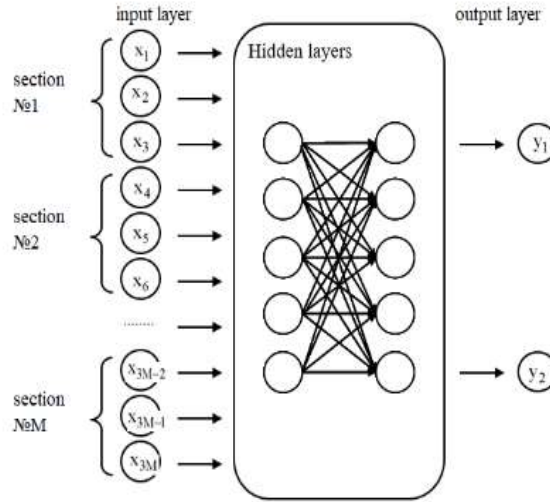


Fig. 4. Neural network architecture

The topology of the hidden layer of the neural network for models of the conveyor section using part of the parameters (20) was considered in [34]. For forecasting, one hidden layer with six nodes was used. As an activation function, the Logistic function was selected:

$$f(x) = \frac{a}{1 + \exp(-bx)} \quad (21)$$

Weights are initialized with random values. In [33], a 4-20-1 conveyor system model was considered to study the dependence of the output material flow on 4 input parameters, among which an important parameter is $g_m(\tau)$. The inner layer contains 20 nodes. In [35], the topology of a neural network of the form $(m_1 - m_2 - 14) =$

(4-9-14) was considered, where $m_2 = 2m_1 + 1$ is the number of hidden layers; $m_1 = 4$ is the number of nodes in the input layer. In this paper, let's focus on the topology 9-3-2. This architecture corresponds to the transport system model of 4 sections with parameters (20) $x_{3m-2} = \gamma_m(\tau)$, $x_{3m-1} = g_m(\tau)$ and one node whose value is one. The hidden layer contains 3 nodes. The output layer contains 2 nodes (21). The activation function has the form (22). The length of the conveyor is different.

6 Preparation of test data

As noted above, for existing transport systems it is almost impossible to obtain complete experimental data for training a neural network for transient modes. For training of the neural network, test data is required that contain a wide range of values. However, the functioning of the transport system in such a range of flow parameters is associated with high energy costs. Additionally, the lengths of the sections of the existing transport system are defined and cannot be changed. In this regard, let's use the PiKh – model (1), (2) [12] to prepare the test data, which allows us to construct an exact solution that determines the state of the flow parameters of the transport system. Let us believe that the intensity of the material flow $\lambda_m(\tau)$ to the input of the m -th non-node section of the conveyor and the belt speed $a_m(t)$ m -th section is known:

$$a_m(t) = a_{0m} + a_{1m} \sin(\omega_a m t + \varphi_{am}), \quad (22)$$

$$\omega_a m = \frac{m\pi}{T_a}, \quad \varphi_{am} = \frac{m\pi}{4}, \quad a_{0m} = a_{1m} = a_0 \frac{3+m}{8}, \quad (23)$$

$$\lambda_m(t) = \lambda_{0m} + \lambda_{1m} \sin(\omega_\lambda m t + \varphi_{\lambda m}), \quad (24)$$

$$\omega_\lambda m = \frac{m\pi}{T_\lambda}, \quad \varphi_{\lambda m} = -\frac{m\pi}{4}, \quad \lambda_{0m} = \lambda_{1m} = \lambda_0 \frac{3+m}{8}, \quad (25)$$

at the initial linear density along the route of the conveyor section

$$\Psi_m(t) = \Psi_{0m} + \Psi_{1m} \sin(k_m S + \varphi_{\Psi m}), \quad (26)$$

$$\varphi_{\Psi m} = \frac{m\pi}{4}, \quad \Psi_{0m} = \Psi_{1m} = \Psi_0 \frac{3+m}{8}. \quad (27)$$

To go to dimensionless coordinates, we choose the characteristic size S_d , the characteristic process time T_d for the transport system

$$S_d = S_6, \quad T_d = T_6. \quad (28)$$

The choice of characteristic quantities is arbitrary and is used to select the scale of the scale for measuring system parameters for conducting a numerical experiment. We assume that the 6th section will be one of the most loaded elements of the system, or at least for a functioning transport system this section will be in operation for maximum time. To simplify the dependencies $\lambda_m(\tau), a_m(t), \Psi_m(t)$, used to form test data, we assume

$$T_a = T_d, T_\lambda = T_d, S_\lambda = S_d. \quad (29)$$

Let us also introduce the characteristic flow of material in the network $\lambda_{kh} = 3\lambda_0$. The value of the characteristic flow is equal to the sum of the average values λ_{0m} of the non-nodal sections. It should be noted that the choice of characteristic values is arbitrary and also determines the scale of the variables λ_{0m} of the problem under consideration. Taking into account (30), should write

$$\tau = t / T_d, \xi_{dm} = S_{dm} / S_d, \quad (30)$$

$$g_{0m} = g_{1m} = a_{0m} \frac{T_d}{S_d}, \gamma_{0m} = \gamma_{1m} = \frac{\lambda_{0m}}{\lambda_{kh}}, \psi_{0m} = \psi_{1m} = \frac{\Psi_{0m}}{\lambda_{kh}} \frac{S_d}{T_d}, \quad (31)$$

$$g_m(\tau) = a_m(t) \frac{T_d}{S_d} = g_{0m} + g_{1m} \sin\left(m\pi\tau + \frac{m\pi}{4}\right), \quad (32)$$

$$\gamma_m(\tau) = \lambda_m(t) \frac{T_d}{S_d \Theta} = \gamma_{0m} + \gamma_{1m} \sin\left(m\pi\tau - \frac{m\pi}{4}\right), \quad (33)$$

$$\psi_m(t) = \frac{\Psi_m(t)}{\Theta} = \psi_{0m} + \psi_{1m} \sin\left(m\pi\xi + \frac{m\pi}{4}\right). \quad (34)$$

Since the choice T_d is arbitrary, the parameter T_d is defined in such a way that equality

$$1 = a_0 T_d / S_d, \quad (35)$$

then dimensionless coefficients can be written as

$$g_{0m} = \frac{3+m}{8}, \gamma_{0m} = \frac{3+m}{24}, \psi_{0m} = a_0 \frac{\Psi_0}{\lambda_0} \frac{3+m}{24}. \quad (36)$$

For training the neural network, test data were used [38]. Test data was generated on the foundation of the model (1) - (7) [12] in accordance with the scheme of the transport system of Fig. 3 and the architecture of the neural network shown in Fig. 4. Test data is based on an analytical model (PiKh-model). The parameters of the analytical model correspond to conditions (23)–(37).

7 Analysis of the results

Fig.5 shows the output flow of section 8 of the transport system (Fig. 2, Fig. 4) for the analytical model and the neural network model. The calculation of the parameters of the output flow for the neural network model is performed for the number of epochs equal to 300,000 and the learning coefficient $\alpha = 10^{-5}$

$$W_{j,k,n+1} = W_{j,k,n} - \alpha \nabla E, \quad E = \frac{1}{2} \sum_{m=1}^{N_m} (z_m - y_m)^2, \quad (37)$$

where the updated weight value $W_{j,k,n+1}$ (for an epoch $n+1$) is calculated on the basis of its old value and the error $E = E(z_m, y_m)$, determined by the N_m parameters of the output layer between the test data z_m and the values y_m of the neural network model (Fig. 4). To analyze the process of training a neural network, data from a test run comes in a strictly specified order. This allows for multiple repetitions of training

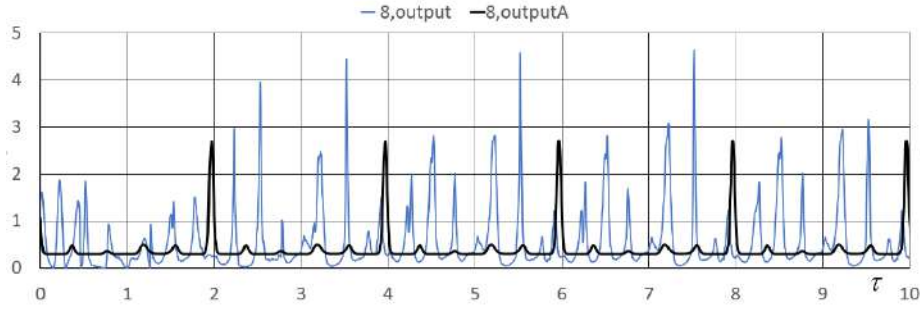


Fig. 5. The value of the output flow of section 8, calculated using the analytical model (8, output) and the neural network model (8, outputA)

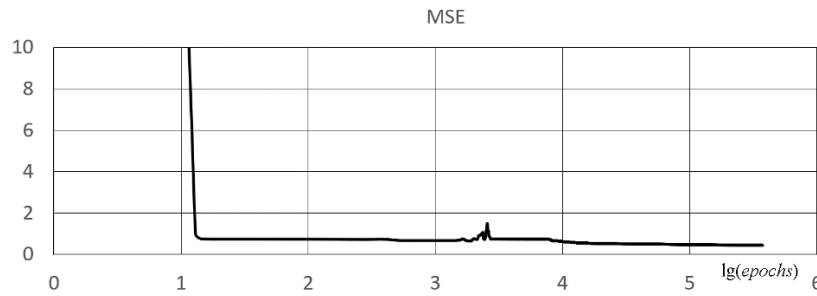


Fig. 6. The magnitude of the MSE error depending on the number of training epochs

with various parameters of the network architecture and compares the effect of changing parameters. The value of MSE (Mean squared error) depending on the number of training epoch is shown in Fig. 6. For training with 375000 epochs and a test sample

of 9000 lines, the MSE is 0.445 [38]. The learning results show that the predicted values of the output flow of section 8 obtained using the neural model are different from the values of the flow parameters of the test sample (analytical model). This result is explained by a rather large MSE value. Further comparative qualitative

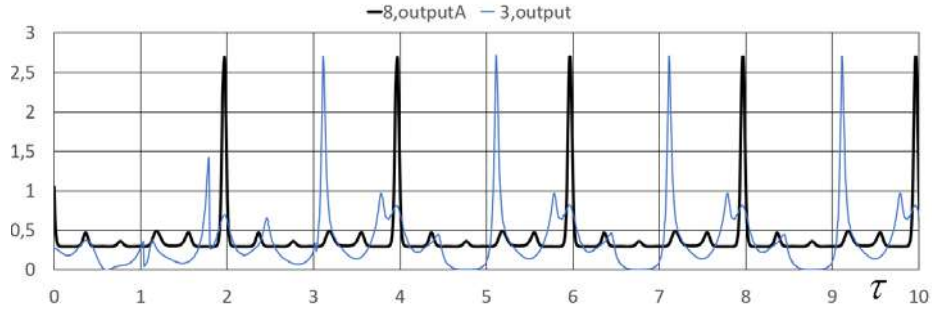


Fig. 7. The value of the output stream of section 3 calculated using the analytical model (3, output) and the output flow of section 8 calculated using the neural network model (8, outputA)

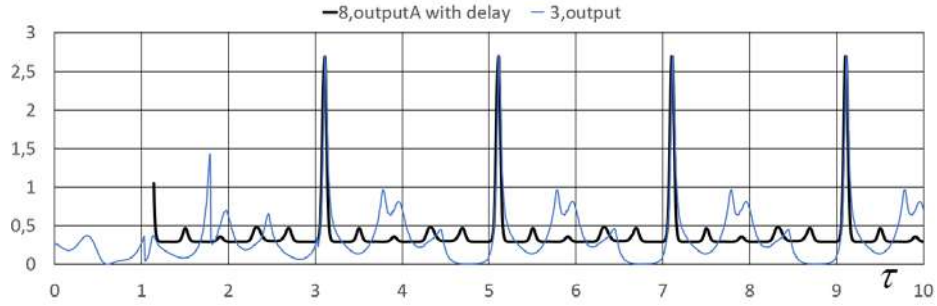


Fig. 8. The value of the output flow of section 8 with an offset by the delay $\Delta\tau_{68} = 1.15$

analysis of the output streams of material from two different sections gave a rather interesting result. The value of the parameter of the output flow of section 8 (neural model) qualitatively repeats the values of the parameter of the output flow of section 3 (neural network model). Repetition is carried out with some delay (Fig. 7). The delay value is the value $\Delta\tau_{68} = 1.15$ (Fig.7) [38]. This value can also be calculated using the test data table [38]. On the other hand, using test data, let's calculate the average delay for section 6 and section 8. Material that leaves section 3 should go through section 6 and section 8 until it reaches the output of section 8 of the transport system. Thus, it should be assumed that the delay between the value of the material flow in section 6 and in section 8 is $\Delta\tau_{68} \approx \Delta\tau_6 + \Delta\tau_8$. Test data allow us to determine the average value of the delay for sections 6 and 8: $\langle\Delta\tau_6\rangle = 0.8754$, $\langle\Delta\tau_8\rangle = 0.4287$, $\langle\Delta\tau_{68}\rangle = 1.3041$, which is pretty close to the value synthesized by

the neural model ($\Delta\tau_{68} = 1.15$). The offset of the output flow of section 8 is shown in Fig. 8.

The developed model of a transport system using the neural network opens up new prospects for the design of control systems for a multi-section conveyor. Also, one of the differences between this work and works [21–25] is that a new method of data preparation for training a neural network is proposed. This allows you to significantly expand the field of study of the behaviour of the parameters of the transport system for various established operating modes of individual sections. A separate area of further research is the definition of similarity criteria for transport systems. Such an approach will make it possible to determine the basic models for various operating modes of the transport conveyor and to study in detail the characteristics of the flow of the material for individual sections.

8 Conclusion

A model of a conveyor transport system using a neural network is one of the tools for research a transport system with a large number of separate sections. The advent of the analytical PiKh-model [12] made it possible to generate test data that are necessary for training a neural network. The lack of a test data set is one of the problems that impede the process of constructing neural models. In many cases, the formation of a set of test data in the required range of parameter changes is an insurmountable obstacle. In this regard, an important result of this work is the development of a method for generating a set of test data for training a neural network that simulates conveyor-type transport systems. A distinctive feature of the modelling of transport systems is that they are complex dynamic distributed systems in which the signal propagates with a delay. And this article presents the first results of constructing a neural model on the foundation of an analytical model. Using the simple architecture of a neural network (9-3-2) as an example, a qualitative relationship between the output flows of the material of different sections is shown. The qualitative relationship between the output flow parameters of section 3 and section 8 is shown. A quantitative assessment of the estimated time of movement of the material through the sections of the transport system is given.

References

1. Pihnastyi, O.: Statistical theory of control systems of the flow production. Lambert Academic Publishing, Beau Bassin (2018)
2. Conveyorbeltguide Engineering: Conveyor components.
<http://conveyorbeltguide.com/examples-of-use.html>
3. Kung, W.: The Henderson Coarse Ore Conveying System. A Review of Commissioning, Start-up, and Operation, Bulk Material Handling by Belt Conveyor, Society for Mining, Metallurgy and Exploration (2004)
4. Alsbaugh, M.: Latest developments in belt conveyor technology. In: MINExpo 2004, New York, Las Vegas (2004)

5. Siemens. Innovative solutions for the mining industry. <http://www.siemens.com/mining>
6. Alspaugh, M.: Longer Overland Conveyors with Distributed Power. In: Overland Conveyor Company, Lakewood. (2005)
7. DIN 22101:2002-08. Continuous conveyors. Belt conveyors for loose bulk materials. Basics for calculation and dimensioning. [Normenausschuss Bergbau (FABERG), Deutsches Institut für Normung e.v. Normenausschuss Maschinenbau (NAM)]. (2002).
8. Pihnastyi O.M.: Control of the belt speed at unbalanced loading of the conveyor. Scientific bulletin of National Mining University. 6, 122–129 (2019). doi:10.29202/nvngu/2019-6/18
9. Semenchko, A., Stadnik, M., Belitsky, P., Semenchko, D., Stepanenko, O.: The impact of an uneven loading of a belt conveyor on the loading of drive motors and energy consumption in transportation. Eastern European Journal of Enterprise Technologies, 4/1(82), 42–51 (2016). <https://doi.org/10.15587/1729-4061.2016.75936>
10. Wolstenholm, E.: Designing and assessing the benefits of control policies for conveyor belt systems in underground mines. Dynamica. 6(2), 25-35 (1980)
11. Lauhoff, H.: Speed Control on Belt Conveyors – Does it Really Save Energy? Bulk Solids Handling Publ. 25(6), 368-377 (2005)
12. Pihnastyi, O., Khodusov, V.: Model of conveyor with the regulable speed. Bulletin of the South Ural State University. Ser. Mathematical Modelling, Programming and Computer Software 10, 64–77 (2017). <https://doi.org/10.14529/mmp170407>.
13. Nordell, L.K., Ciozda, Z.P.: Transient belt stresses during starting and stopping: Elastic response simulated by finite element methods. Bulk Solids Handling 4(1), 99–104 (1984). <http://www.ckit.co.za/secure/conveyor/papers/troughed/transient/transient-belt-stresses.htm>
14. He, D., Pang, Y., Lodewijks, G., Liu, X.: Determination of Acceleration for Belt Conveyor Speed Control in Transient Operation. International Journal of Engineering and Technology 1.8(3), 206–211 (2016). <http://dx.doi.org/10.7763/IJET.2016.V8.886>
15. Alspaugh, M.A.: Latest Developments in Belt Conveyor Technology. Overland Conveyor Co., In: MINExpo 2004, pp.1–11. Las Vegas, (2004). [http://www.overlandconveyor.com/pdf/Latest Developments in Belt Conveyor Technology.pdf](http://www.overlandconveyor.com/pdf/Latest%20Developments%20in%20Belt%20Conveyor%20Technology.pdf)
16. Karolewski, B., Ligocki, P.: Modelling of long belt conveyors. Maintenance and reliability. 16 (2), 179–187 (2014).
17. Pascual, R., Meruane, V., Barrientos, G.: Analysis of transient loads on cable-reinforced conveyor belts with damping consideration. In: the XXVI Iberian Latin-American Congress on Computational Methods in Engineering, pp.1–15, Santo, Brazil (2005). <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.494.34&rep=rep1&type=pdf>
18. Wheeler, C.A.: Predicting the main resistance of belt conveyors. In International Materials Handling Conference (Beltcon) 12, Johannesburg, South Africa (2003). <http://www.saimh.co.za/beltcon/beltcon12/paper1208.htm>
19. Mathaba and Xia, 2015 T. Mathaba, X. Xia, A parametric energy model for energy management of long belt conveyors. Energies 8(12), 13590–13608 (2015). <https://doi.org/10.3390/en81212375>
20. Reutov, A.: Simulation of load traffic and steeped speed control of conveyor. In: IOP Conference Series: Earth and Environmental, 87, pp.1–4 (2017). <https://doi.org/10.1088/1755-1315/87/8/082041>.
21. Kirjanów, A.: The possibility for adopting an artificial neural network model in the diagnostics of conveyor belt splices. Interdisciplinary issues in mining and geology 6, 1–11 (2016)

22. Więcek, D., Burduk, A., Kuric, I.: The use of ANN in improving efficiency and ensuring the stability of the copper ore mining process. *Acta Montanistica Slovaca* 24(1), 1–14 (2019). <https://actamont.tuke.sk/pdf/2019/n1/1wiecek.pdf>
23. Li, W., Wang, Z., Zhu, Zh., Zhou, G., Design of Online Monitoring and Fault Diagnosis System for Belt Conveyors Based on Wavelet Packet Decomposition and Support Vector Machine. *Advances in Mechanical Engineering* 2013, 1–10 (2013) <http://dx.doi.org/10.1155/2013/797183>
24. Xinglei L., Hongbin, Yu.: The Design and Application of Control System Based on the BP Neural Network. In the 3rd International Conference on Mechanical Engineering and Intelligent Systems, pp. 789–793 (2015). <https://doi.org/10.2991/icmeis-15.2015.148>
25. Pingyuan, Xi, Yandong, Song : Application Research on BP Neural Network PID Control of the Belt Conveyor. *JDIM* 9(6), 266–270 (2011).
26. Andrejiova, M, Marasova, D.: Using the classical linear regression model in analysis of the dependences of conveyor belt life. *Acta Montanistica Slovaca* 18(2), 77–84 (2013). <https://actamont.tuke.sk/pdf/2013/n2/2andrejiova.pdf>
27. Grincova, A., Li, Q.: A regression model for prediction of idler rotational resistance on belt conveyor. *Measurement and Control* 52(5), 441–448 (2019). <https://doi.org/10.1177/0020294019840723>
28. Karolewski, B., Marasova, D.: Experimental research and mathematical modelling as an effective tool of assessing failure of conveyor belts. *Maintenance and reliability*. 16 (2), 229–235 (2014). <http://www.ein.org.pl/sites/default/files/2014-02-09.pdf>
29. Pihnastyi , O., Khodusov, V.: Calculation of the parameters of the composite conveyor line with a constant speed of movement of subjects of labour. *Scientific bulletin of National Mining University*. 4 (166), 138–146 (2018). <https://doi.org/10.29202/nvngu/2018-4/18>
30. Pihnastyi , O., Khodusov, V.: Model of a composite magistral conveyor line. In *IEEE International Conference on System analysis & Intelligent computing*, pp.68–72 (2018). <https://doi.org/10.1109/saic.2018.8516739>
31. Bastian Solutions Conveyor System Design Services. <https://www.bastiansolutions.com/solutions/technology/conveyor-systems/design-services>
32. Zimroz, R., Krol, R.: Failure analysis of belt conveyor systems for condition monitoring purposes. *Mining Science* 128(36), 255–270 (2009).
33. Abdollahpor, S., Mahmoudi, A., Mirzazadeh, A.: Artificial neural network prediction model for material threshing in combine harvester. *Elixir Agriculture* 52, 11621–11626 (2012). [https://www.elixirpublishers.com/articles/1353477517_52%20\(2012\)%2011621-11626.pdf](https://www.elixirpublishers.com/articles/1353477517_52%20(2012)%2011621-11626.pdf)
34. Xinglei, L., Hongbin, Yu.: The Design and Application of Control System Based on the BP Neural Network. In the 3rd International Conference on Mechanical Engineering and Intelligent Systems, pp.789–793 (2015). <https://doi.org/10.2991/icmeis-15.2015.148>
35. Yuan, Y., Meng, W., Sun, X. Research of fault diagnosis of belt conveyor based on fuzzy neural network. *The Open Mechanical Engineering Journal* 8, 916–921 (2014). <https://doi.org/10.2174/1874155X01408010916>
36. Selcuk, N., Birbir, Y.: Application of artificial neural network for harmonic estimation in different produced induction motors. *Int. J. of Circuits, Systems and Signal Processing* 4(1), 334–339 (2007). <https://pdfs.semanticscholar.org/ba2f/6d5ea4c91720be4044e0f3544efb60fd6bb4.pdf>
37. Chen, W., Li, X. : Model predictive control based on reduced order models applied to belt conveyor system. *ISA Transactions* 65, 350–360 (2016). doi:10.1016/j.isatra.2016.09.007
38. Pihnastyi, O.: Test data set for the conveyor transport system, Mendeley Data, V3, (2020). <http://dx.doi.org/10.17632/4vc843t76.3>

Multicomponent Analyzer of Volatile Compounds Characterization Based on Artificial Neural Networks

Taras Chaikivskyi¹[0000-0002-1166-8749], Bogdan Sus²[0000-0002-2566-5530],
Oleksandr Bauzha²[0000-0002-4920-0631] and Sergiy Zagorodnyuk²[0000-0003-3415-7746]

¹Lviv Polytechnic National University, Bandera str, 12, Lviv, 79013, Ukraine
taras.v.chaikivskyi@lpnu.ua

²Taras Shevchenko National University of Kyiv, Volodymirskaya str., 64, Kyiv 01033, Ukraine
bnsuse@gmail.com, asb@mail.univ.kiev.ua, kola@univ.net.ua

Abstract. An automated multicomponent system for measuring the content of monatomic phenolic compounds in alcohol solutions has been developed. To assure the quality control of widespread alcohol solutions, including medications and strong alcoholic beverages, a new multi-sensor automated quality monitoring system has been proposed. The system is based on a multi-layer artificial neural network for the digital processing of sensor signals. Taking into account the cross-sensitivity of the sensors and the application of selective sensors reduce the error in determining the concentration of light phenolic compounds. The sensor signal processing process is complemented by parallel channels for calculating the ethyl alcohol concentration as well as the illumination and temperature values by linearly converting the output signals. Simulation models using sensors based on the electronic theory of impurity sorption on the surface of semiconductors were acquired.

Keywords: semiconductor sensor, neural network, volatile solutions, microcontroller.

1 Introduction

The application of semiconductor sensors as working elements in multi-sensor systems makes it possible to effectively solve typical problems of modern analytical chemistry - to analyze the compositions of liquids and gases. Such analyzing systems, also known as "electronic tongues" and "electronic noses," work with chemical sensors of different operating principles and biosensors in combination with multi-step mathematical algorithms for data processing. Therefore, the development and implementation of new analytical systems is an urgent technological challenge and is fully in line with global trends. The practical principle of the electronic nose is best explained through its natural counterpart. Volatile compounds reach the olfactory epithelial tissues, where they reach the olfactory receptors. This generates a corresponding neuron, which transmits information about this fact to the brain. In it,

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

the information of neurons is consolidated and structured according to existing patterns that enable the subject to recognize a specific odor [1].

Preferably, there is a need to make extensive application of artificial electronic sensor systems during manufacturing products on an industrial scale. In particular, the control and monitoring of the processes of growing, maturing and spoiling organic raw materials are extremely important and valuable for ensuring the safety and quality of food products [2]. In the production of wine and spirits for quality control and identification of the aroma mandatory state certification for alcoholic beverages is used [3]. Today, the quality of the wine is mainly evaluated by physical and chemical indicators, such as alcohol content and subjective evaluations, as expert tasters. Obviously, this procedure, which is performed over a long period of time and requires expert evaluation, has a fairly high cost. Analysis of light volatile compounds released by alcoholic beverages can help to classify its specific characteristics, to identify in a timely manner the features and complexities of processing and long-term storage. Chemical analytical methods also require long and complex investigations, but in the result of the accumulation and structuring of the results of the data, these methods give higher reliability and accuracy of measurements than the subjective evaluation of a strong drink by a specialist taster [4]. Most of the chemical components that provide the consumable quality of the beverage can alter the signals of the sensors, and therefore provide the beverage categorization according to qualitative and quantitative criteria.

The presence and concentration of light volatile compounds in alcoholic beverages are also monitored by means of multi-component semiconductor gas analyzers. Semiconductor sensor systems provide the ability to control the concentrations of several lightweight volatile compounds. But concentration measurement gives a relatively high error, which limits their scope. This error is related to the cross-sensitivity of the gas analyzers. One method of reducing this error is to design and use highly selective sensor systems. They are an important component of the concept of "lab-on-a-chip" or "micro-total-analysis". This is the concept of making miniature devices that enable the sequencing of chemical processes or chain chemical reactions on a single chip with a small area. In addition, state-of-the-art lab-on-a-chip devices provide real-time chemical or quantitative chemical analysis [5].

Semiconductor gas-sensitive sensors have a low response time, but they are also sensitive to other physical parameters, including light, temperature, humidity, and background concentrations in the air of other organic volatile compounds. The effect of these factors has a significant effect on the measuring signal. Currently, several technological approaches are being used to minimize the impact of these factors. These approaches include the method of temperature compensation for gas-sensitive sensors and the pulse mode of sensor heating. By applying these methods, the effect of evaporation of water and ethanol on the sensitivity of the sensors was significantly reduced [6]. In this work, the measuring cell was purged for 10 additional minutes with pure helium to completely evaporate and dry the sensors before taking the measurements. It should be noted that such a procedure is technologically complicated and costly. Such solutions do not always provide the required level of error and may have a low reaction time for gas analyzers. Some studies have shown

that under conditions of light irradiation, the photocurrent of a semiconductor strongly depends on the recombination characteristics and the charge state of the surface. The approach of using solar cells as sensory structures is interesting. Induced light flux (LBIC) measurements reflect the spatial distribution of the photocurrent of a solar cell. Depending on the diameter of the excitation beam and the distance, a lateral resolution of several 10^{-4} m to 10^{-6} m can be achieved. The concept of the LBIC sensor system is described in [7]. The conductivity of the sensitive layer of a semiconductor sensor depends on the concentration of free electrons, which is proportional to the fraction of the surface directly covered by the molecules of the adsorbed light compound.

However, even such sensors also have low selectivity and significant measurement error. The responses of each low-selective sensor are slightly dependent on the type of adsorbent. One option to increase selectivity is to increase the number of sensors and create their array. The set of multiple responses to the sensory system forms a unique imprint for each substance composition. The artificial neural network apparatus can be used to detect the composition of volatile compounds in sensor systems. A computer program can evaluate the signal pattern and can compare the flavors of different samples. The measurement results are comparative, not quantitative, and therefore presented as an "imprint".

Currently, artificial neural networks are used in many fields of science and technology. They are powerful tools for solving complex problems whose answers are not obvious and cannot be determined by classical mathematical algorithms. Neural networks have been increasingly used since the introduction of the backpropagation algorithm [8]. The study of fingerprint recognition using multisensory data is an area where rapid technological development of the data collection system is required, which are quantitative indicators of human quality of life. In particular, multiple sensor data collected from a smart home using several deep neural networks were investigated in [9]. Similarly, the application of machine learning methods, including linear regression, neural network, and vector machine support, to determine the dependence of wine quality on the aromatic content and to predict wine quality were explored in [10]. The paper demonstrates that the value of a dependent variable can be more accurately predicted if only the important features of the wine are taken into account in the forecasting but not all the features are taken into account. In this case, a linear regression was implemented to determine the dependence of wine quality on the various 11 physicochemical characteristics. Wine quality assessment is one of the key elements in this work and this assessment can be used for product certification. This type of quality certification helps to ensure long-lasting quality of wine and its competitiveness. The wine has different characteristics, including density, pH, strength, acidity, etc. Wine quality can be assessed by two types of tests; the first is a physicochemical test and the second is a sensory test [11].

Low-accuracy electronic sensor systems were used for pattern recognition and selectivity. Artificial neural networks can also be effectively used to increase the selectivity of semiconductor sensor systems. Therefore, the creation of multi-component gas analyzers based on neural networks that provide simultaneous quantitative analysis to determine the concentrations of components, increase

selectivity and reduce sensitivity to environmental factors are and remain relevant technical problems of analytical chemistry.

2 Hardware Setup

A block diagram of a multicomponent analyzer measuring complex for volatile compounds is shown in Fig. 1.

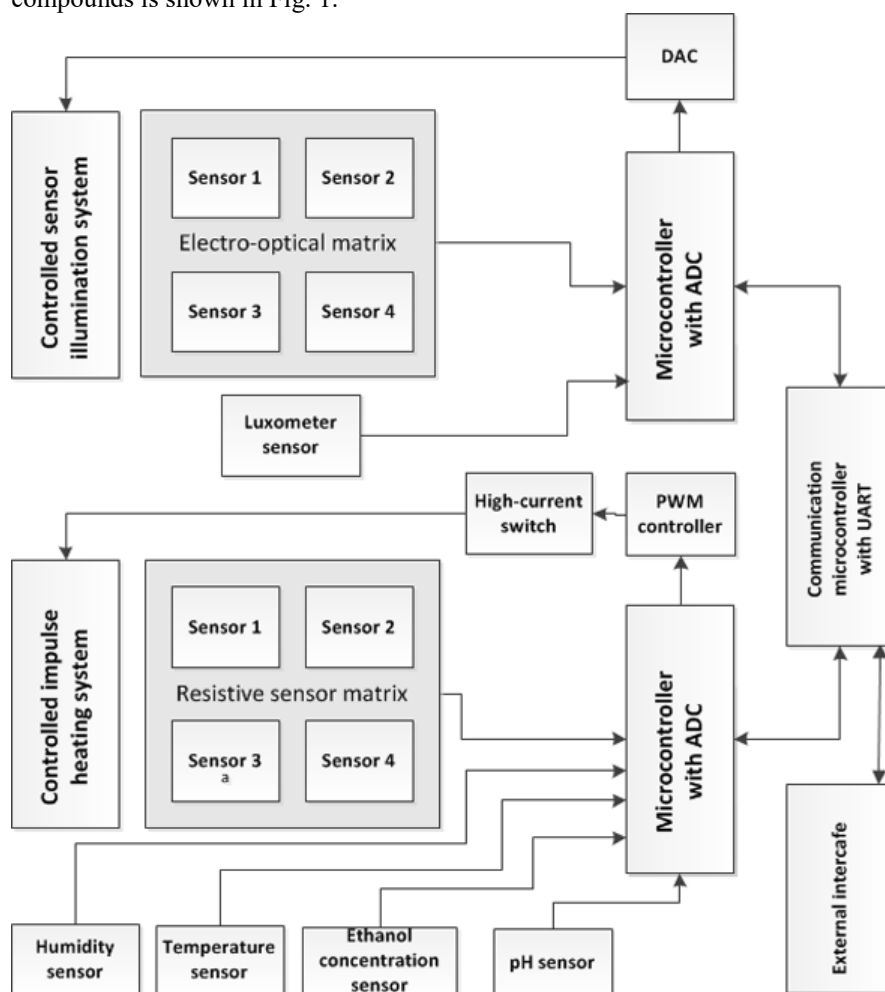


Fig. 1. Structural diagram of the measuring complex.

The measuring chamber is shown in fig. 2.

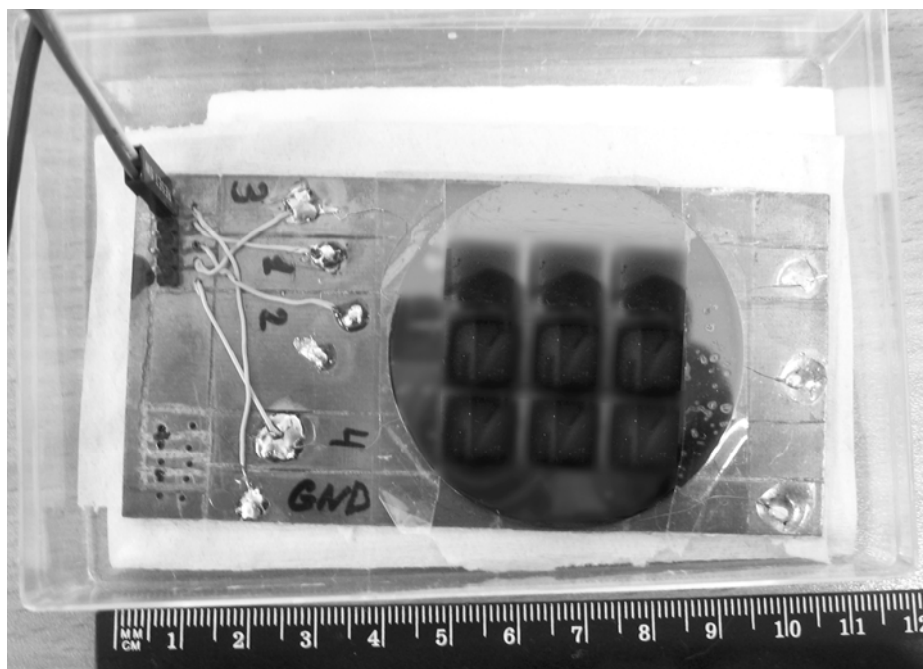


Fig. 2. Measuring chamber with Optical Conversion Sensor.

Nowadays, the most promising and convenient are sensor systems based on compact digital devices - microcontrollers equipped with built-in ADCs. The complex consists of sensors, sensitive to light volatile compounds and sensors of environmental parameters, microcontrollers of control and preprocessing of signals, and also the communication microcontroller of communication. The microcontroller (1) checks the illumination brightness of the sensor plate in an assigned intensity range. Brightness values are obtained from the light sensor. Brightness ratios and corresponding photoresponses of different samples are stored by the neural network. ADC microcontrollers convert analog sensor signals into digital signals, perform signal correction and normalization, control sensor backlight modes with optical recreation and change of temperature measurement modes. The communication microcontroller receives data from ADC microcontrollers and processes the data according to algorithms implemented by the artificial neural network. Additionally, software and hardware encryption of the accumulated data may be implemented [12]. The developed setup can also operate with wireless data modules and be practiced for laboratory work during the learning process [13-15].

When using a 4-channel barrier sensor, a bias voltage should be applied to the p-n junctions. The photocurrent is determined by the rate of superficial recombination of charge carriers. Each channel has a distinct sensitivity to analytes. The equivalent circuit of the sensor is shown in fig. 3.

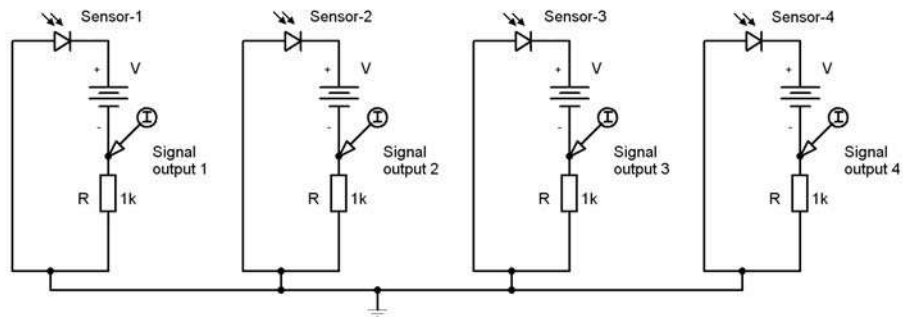


Fig. 3. Equivalent circuit of 4-channel sensor with p-n junction

Since the sensor measurement data also depend on the illumination brightness of the p-n junctions, the illumination intensity varies linearly with the DAC converter in the specified range during the measurements. There may be several photocurrent extremes in the range of illumination intensity, which could be used as additional informative parameters.

3 Experiment

Experiments with a set of substances of light alcohol compounds were carried out. The evaporation of the substances was above the sensors of the electronic nose. The sensors were illuminated by the incident light of a given intensity. The results of photocurrent measurements from 4 electronic nose sensors are shown in Fig. 4.

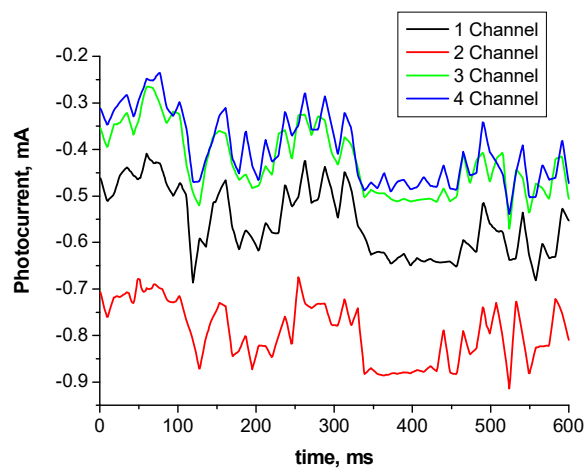


Fig. 4. Results of photocurrent measurements (sample # 1).

The dependencies of the sensor data from different samples are shown in Fig. 4.. The normalized values of the photocurrent are shown on the X-axis.

The currents density passed into the measurement sampling is shown on the Y-axis, Fig.5. Photocurrents from different channels are shown in different colors. As can be seen in the Fig. 5., the values on the sensors from different channels overlap.

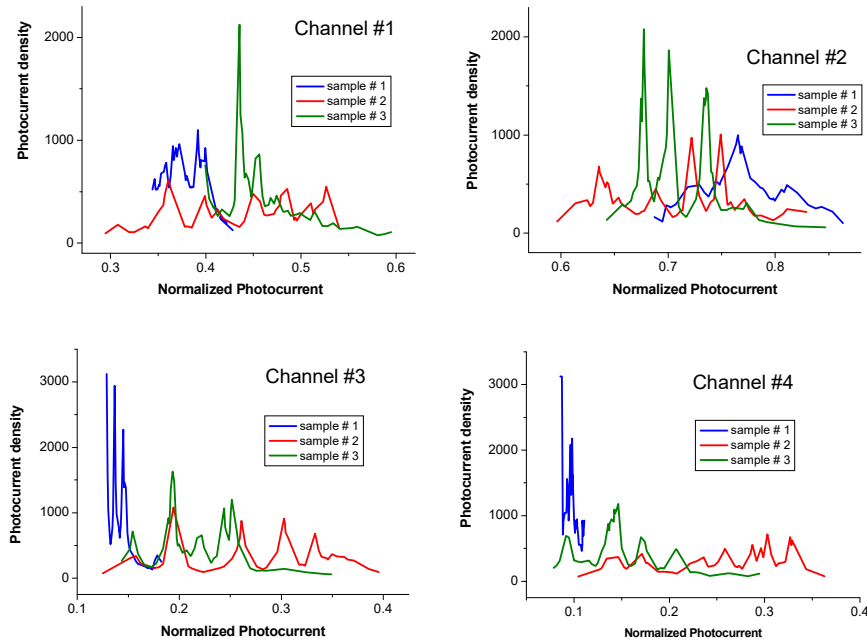


Fig. 5. Sampling of the sensor density measurement results.

4 Software implementation

During the measurements of different samples of alcoholic beverages, it turned out that the results of the sensors for the related samples have similar values. The expected correlation of the sensors with the specimens of the test substances (cognacs) was not observed. Obviously, under such conditions, it is advisable to use an artificial neural network to classify substances. The multilayered, fully connected feedforward neural network architecture has been selected. The artificial neural network receives sets of values of variables, specifically the values of equivalent concentrations of the mixture components of volatile compounds from optical and resistive sensor matrices. The normalized values of the concentrations of the components of the volatile compounds are calculated for each set of input variables. Values depend on the normalized output signals of the sensors. The input layer of the

neural network consists of 8 input neurons. Input signals were normalized to 1 and reflected 4 channels of photocurrents from electronic nose sensors. The neural network contains three hidden layers, each of which has 10 neurons. The output layer has 6 neurons. The structure of the neuron network is presented in Fig. 6. This artificial network allows increasing the required number of neurons in layers.

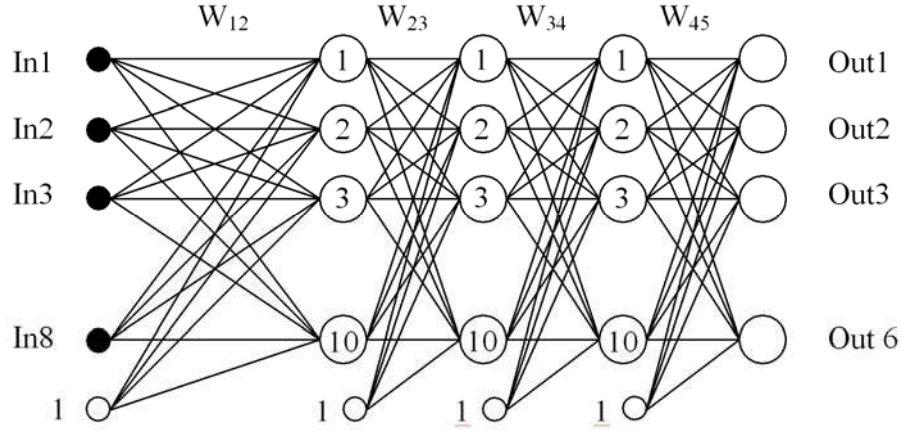


Fig. 6. Multilayer feedforward artificial neural network.

The sigmoid curve was chosen as the activation function.

$$f(x) = \frac{1}{1 + e^{-x}}$$

Artificial neural network training was carried out by the method of training with the teacher. To determine the minimum of the error function, the backpropagation method in stochastic gradient descent algorithm was used. The algorithm of artificial neural network training is shown in Fig. 7. During training, the signals In, Out - are the input and output vectors of the neural network, respectively. The W array is a matrix of neural network weights.

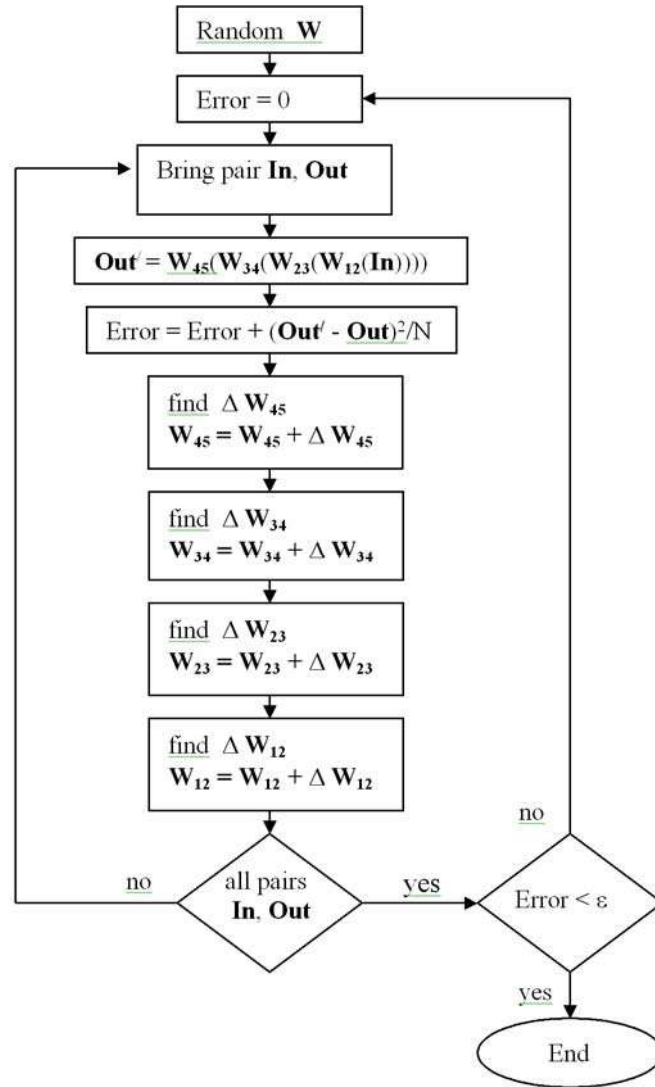


Fig. 7. Backpropagation algorithm.

The user interface of the sensor software with the neural network implementation is shown in Fig. 8.

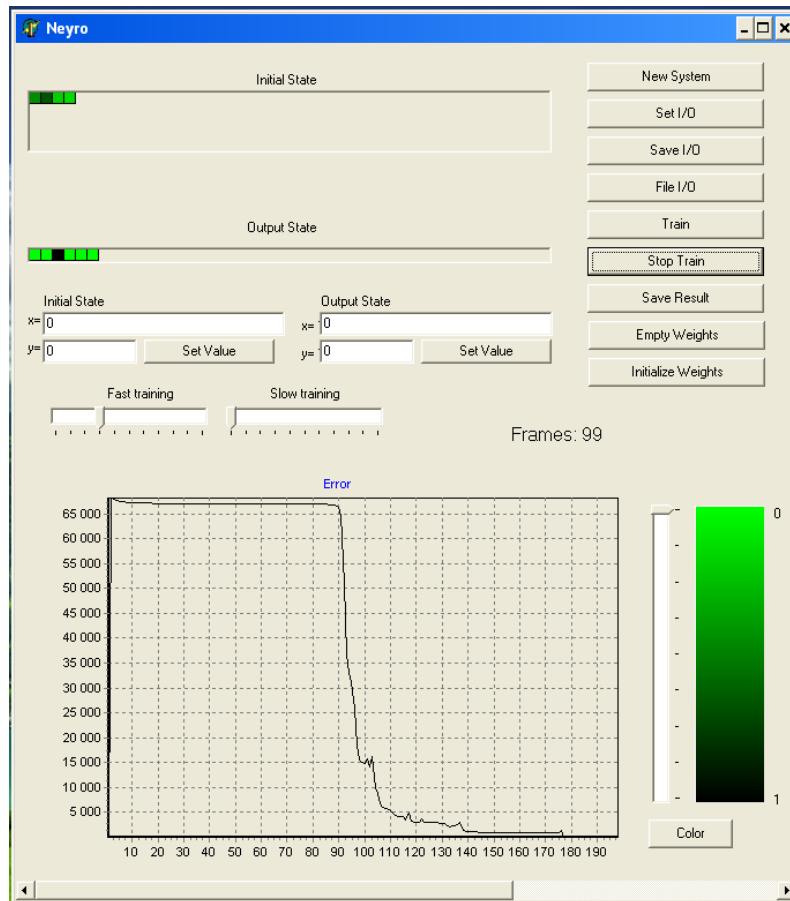


Fig. 8. Program interface

The “New System” button allows the user to configure the neural network. During configuration, it is possible to specify the number of input, output neurons and the number of neurons in each of the 3 intermediate layers. The “Set I / O”, “Save I / O”, “Set Value”, and “File I / O” buttons define the input vectors of the input and output neurons (frames). The process of training the neural network begins with the “Train” button. The Empty Weights button sets the weight of the W_{ij} training to its initial, random position. The “Initialize Weights” button changes the learning scales and exits the system from the point of local minimum error.

The values of the input and output neurons in the frame are represented at the top left in the form of a matrix of rectangles. The color of the rectangles is determined by the value of the corresponding neurons. The scale for matching (and changing) the color and meaning of the neurons are shown at the bottom right.

In Fig. 8. it is shown that the neural network has finished the training and the schedule of the error change can be observed below.

5 Experimental data processing

The experiment was carried on three sample substances. For each substance, 51 measurements of photocurrent were executed, and therefore 153 measurements of photocurrent from three samples were executed, in the future, these measurements are called frames. The influence of time and the influence of the measurement history from the sensors on the current measurement result were not taken into account. A sample of 99 frames was randomly selected from 153 frames, 33 frames from each sample for neural network training. The remaining 54 frames, 16 frames from each sample are left as controls to test the operation of the already trained neural network. The neural network was trained for about 200 epochs, after which the total error was less than $\varepsilon = 0.01$, which was admitted a positive result. After training, the system was tested on control frames. The test gave 98% correct results. In 2% of cases, the non-authentic odor (with indicators of 50-92%) was the main one, while the authentic odor (with indicators of 8-50%) took second place. The result demonstrates the possibility of a system error (due to a lack of training frames) and demonstrates the need to consider not only the winners (first place) but also those source neurons that took second place.

In turn, among 98% of the correct results, 2% of cases were the authentic odor (82-93%), while the non-authentic odor (7-18%) ranked second. In the rest (96%) of control shots, the authentic smell was the undisputed leader (more than 99% probability).

6 Conclusions

During creating automated multicomponent systems for measuring the concentration of mixtures of volatile compounds, the processes occurring on the surface of the semiconductor sensitive layer as the temperature, light, and humidity of the environment are taken into account.

The use of multicomponent low-selection sensors gives the possibility to significantly reduce measurement errors in determining the concentrations of light mixtures, including when exposed to external disturbances.

For the training of neural networks, it is necessary to create and develop specialized mathematical models that describe the transformation processes in the used sensors. With the help of a neural network, sensor parameters can be normalized and calibrated to increase selectivity and to predict complex integral features in alcohol solution samples.

When considering a component that is recognized by a neural network, not only the winner neurons from the set of source neurons but also the source neurons that received the second result should be considered.

An increase in the number of measurement samples leads to an increase in neural network complexity, as well as an increase in the number of sensors needed to accurately classify the test substances.

Many substances have a specific range of odors that can confuse the neural network and generate incorrect neuron at the output, and therefore the correct classification of substances by constituents is an important technical task.

In the multi-component analytical system implemented, optimal conditions for recognizing the concentration of monatomic phenolic compounds in alcohol solutions were determined and formulated.

The main areas of further research are the development of methods for determining the age of strong drinks, including the justification for the use of markers and their linear combinations to improve the accuracy and reliability of the result, as well as the transfer of the measuring complex to the platform of programmable digital integrated circuits FPGA.

References

1. Lvova, L., Kirsanov, D.: Editorial: Multisensor Systems for Analysis of Liquids and Gases: Trends and Developments. *Front. Chem.* 6, 591 (2018). <https://doi.org/10.3389/fchem.2018.00591>.
2. Cuypers, W., Lieberzeit, P.A.: Combining Two Selection Principles: Sensor Arrays Based on Both Biomimetic Recognition and Chemometrics. *Front. Chem.* 6, 268 (2018). <https://doi.org/10.3389/fchem.2018.00268>.
3. Macías, M., Manso, A., Orellana, C., Velasco, H., Caballero, R., Chamizo, J.: Acetic Acid Detection Threshold in Synthetic Wine Samples of a Portable Electronic Nose. *Sensors*. 13, 208–220 (2012). <https://doi.org/10.3390/s130100208>.
4. Waterhouse, A.L., Ebeler, S.E. eds: Chemistry of wine flavor. American Chemical Society ; Distributed by Oxford University Press, Washington, DC : [New York] (1998).
5. Ghafar-Zadeh, E., Sawan, M.: CMOS capacitive sensors for lab-on-chip applications: a multidisciplinary approach. Springer, Dordrecht ; New York (2010).
6. Guadarrama, A., Fernández, J.A., Íñiguez, M., Souto, J., de Saja, J.A.: Array of conducting polymer sensors for the characterisation of wines. *Analytica Chimica Acta*. 411, 193–200 (2000). [https://doi.org/10.1016/S0003-2670\(00\)00769-8](https://doi.org/10.1016/S0003-2670(00)00769-8).
7. Litvinenko, S.V., Kozinetz, A.V., Skryshevsky, V.A.: Concept of photovoltaic transducer on a base of modified p–n junction solar cell. *Sensors and Actuators A: Physical*. 224, 30–35 (2015). <https://doi.org/10.1016/j.sna.2015.01.014>.
8. Rumelhart, D.E., McClelland, J.L., University of California, S.D., PDP Research Group: Parallel distributed processing. explorations in the microstructure of cognition. (1987).
9. Park, J., Jang, K., Yang, S.-B.: Deep neural networks for activity recognition with multi-sensor data in a smart home. In: 2018 IEEE 4th World Forum on Internet of Things (WF-IoT). pp. 155–160. IEEE, Singapore (2018). <https://doi.org/10.1109/WF-IoT.2018.8355147>.
10. Gupta, Y.: Selection of important features and predicting wine quality using machine learning techniques. *Procedia Computer Science*. 125, 305–312 (2018). <https://doi.org/10.1016/j.procs.2017.12.041>.
11. Ebeler, S.E.: Linking Flavor Chemistry to Sensory Analysis of Wine. In: Teranishi, R., Wick, E.L., and Hornstein, I. (eds.) *Flavor Chemistry*. pp. 409–421. Springer US, Boston, MA (1999). https://doi.org/10.1007/978-1-4615-4693-1_35.

12. Tmienova, N., Sus, B.: Hardware data encryption complex based on programmable microcontrollers. In: CEUR Workshop Proceedings, pp. 199–208 (2018). <http://ceur-ws.org/Vol-2318/paper17.pdf>
13. Chaikivskyi, T., Bauzha, O., Sus, B. B., Tmienova, N., Zagorodnyuk, S.: 3D simulation of virtual laboratory on electron microscopy. In: CEUR Workshop Proceedings 2533, pp. 282–291. (2019). <http://ceur-ws.org/Vol-2533/paper26.pdf>
14. Sus, B., Tmienova, N., Revenchuk, I., Vialkova, V.: Development of Virtual Laboratory Works for Technical and Computer Sciences. In: Damaševičius, R. and Vasiljevičienė, G. (eds.) Information and Software Technologies. pp. 383–394. Springer International Publishing, Cham (2019). https://doi.org/10.1007/978-3-030-30275-7_29.
15. Bauzha, O., Sus, B., Zagorodnyuk, S., Stuchynska, N.: Electrocardiogram Measurement Complex Based on Microcontrollers and Wireless Networks. In: International Scientific-Practical Conference on Problems of Infocommunications Science and Technology, PIC S and T, pp. 345–349. (2019).

Adaptive Fuzzy Clustering Approach Based on Evolutionary Cat Swarm Optimization

Alina Shafronenko¹[0000-0002-8040-0279], Yevgeniy Bodyanskiy¹[0000-0001-5418-2143]

¹Kharkiv National University of Radio Electronics, Nauky Ave., 14, Kharkiv, Ukraine
yevgeniy.bodyanskiy@nure.ua, alina.shafronenko@nure.ua

Abstract. Computational intelligence methods are widely used to solve many complex problems, including, of course, traditional: Data Mining and such new directions as Dynamic Data Mining, Data Stream Mining, Big Data Mining, Web Mining, Text Mining, etc. One of the main areas of computational intelligence are evolutionary algorithms that essentially represent certain mathematical models of biological organisms evolution. In the paper adaptive methods of fuzzy clustering using on evolutionary cat swarm optimization were proposed. Using proposed approach it's possible to solve clustering task in on-line mode.

Keywords: fuzzy clustering, learning rule, online probabilistic fuzzy clustering, online possibilistic fuzzy clustering, online credibilistic fuzzy clustering, cat swarm.

1 Introduction

The task of classification in the self-learning mode (clustering) of multidimensional data is an important part of traditional intellectual analysis such as Data Mining, Dynamic Data Mining, Data Stream Mining, Big Data Mining, Web Mining [1, 2].

One of the main areas of computational intelligence are the so-called evolutionary algorithms, which are mathematical models of biological organisms evolution. The problem connected with by vector-observations, clustering often occurs in many applications of data mining, and first of all in data fuzzy clustering when processing vector-observation with different levels of probability possibility, credibility etc. can belong to more than one class.

Very effective are Kohonen self-organizing maps [3] and the evolutionary algorithms that can improve data clusterization in case, when the data are processed sequentially in online mode.

The problem of fuzzy clustering of data arrays is considered in the conditions when the formed clusters arbitrarily overlap in the space of features. The source information for solving the problem is an array of multidimensional data vectors, formed by a set of vector-observations $X = (x(1), x(2), \dots, x(k), \dots, x(N)) \subset R^n$ where k in the general case the observation number in the initial array, $x(k) = (x_1(k), \dots, x_i(k), \dots, x_n(k))^T$. The result of clustering is the partition of this array on m overlapped classes with

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

prototypes-centroids $Cl_j \in R^n$, $j=1,2,...,m$, and computing of membership levels $0 \leq U_j(k) \leq 1$ of each vector-observation $x(k)$ to every cluster Cl_j .

2 Adaptive algorithm for probabilistic fuzzy clustering (APrFC)

The goal function of popular probabilistic clustering has the form [4]

$$E(U_j(k), Cl_j) = \sum_{k=1}^N \sum_{j=1}^m U_j^\beta(k) D^2(\tilde{x}_k, Cl_j) \quad (1)$$

where $\sum_{j=1}^m U_j(k) = 1$, $0 < \sum_{k=1}^N U_j(k) < N$. Solving the nonlinear programming problem, we obtain the probabilistic fuzzy clustering algorithm

$$\begin{cases} U_j^{(\tau+1)}(k) = D^2(\tilde{x}_k, Cl_j^{(\tau)})^{\frac{1}{1-\beta}} * \left(\sum_{l=1}^m (D^2(\tilde{x}_k, Cl_l^{(\tau)}))^{\frac{1}{1-\beta}} \right)^{-1}, \\ Cl_j^{(\tau+1)} = \sum_{k=1}^N (U_j^{(\tau+1)})^\beta \tilde{x}_k * \left(\sum_{k=1}^N (U_j^{(\tau+1)}(k))^\beta \right)^{-1}, \end{cases} \quad (2)$$

in case when the $\beta = 2$ - parameter that is called fuzzyfier and defines "blurring" the boundaries between classes we come to a decision, that has the form

$$\begin{cases} U_j^{(\tau+1)}(k) = \|\tilde{x}_k - Cl_j^{(\tau)}\|^{-2} * \left(\sum_{l=1}^m \|\tilde{x}_k - Cl_l^{(\tau)}\|^{-2} \right)^{-1}, \\ Cl_q^{(\tau+1)} = \sum_{k=1}^N (U_j^{(\tau+1)}(k))^2 \tilde{x}_k * \left(\sum_{k=1}^N (U_j^{(\tau+1)}(k))^2 \right)^{-1}. \end{cases} \quad (3)$$

The process of fuzzy clustering can be organized in on-line adaptive mode

$$\begin{cases} U_j(k+1) = (D^2(\tilde{x}_{k+1}, Cl_j^{(k)}))^{\frac{1}{1-\beta}} * \left(\sum_{l=1}^m (D^2(\tilde{x}_{k+1}, Cl_l^{(k)}))^{\frac{1}{1-\beta}} \right)^{-1}, \\ Cl_j(k+1) = Cl_j(k) + \eta(k+1) U_j^\beta(k+1) (\tilde{x}_{k+1} - Cl_j(k)). \end{cases} \quad (4)$$

3 Adaptive algorithm for possibilistic fuzzy clustering (APosFC)

The goal function of possibilistic clustering has the form

$$E(U_j(k), Cl_j, \mu_j) = \sum_{k=1}^N \sum_{j=1}^m U_j^\beta(k) D^2(\tilde{x}_k, Cl_j) + \sum_{j=1}^m \mu_j \sum_{k=1}^N (1 - U_j(k))^\beta \quad (5)$$

where $\mu \geq 0$ - the scalar parameter that specifies the distance at which the level of membership equals 0.5.

Minimizing the equation (5) relatively, $U_j(k)$, Cl_j and μ_j we obtain the system of equations (6)

$$\begin{cases} U_j^{(\tau+1)}(k) = \left(1 + \left(\frac{D^2(\tilde{x}(k), Cl_j^{(\tau)})}{\mu_j^{(\tau)}} \right)^{\frac{1}{\beta-1}} \right)^{-1}, \\ Cl_j^{(\tau+1)} = \sum_{k=1}^N (U_j^{(\tau+1)}(k))^\beta \tilde{x}(k) * \left(\sum_{k=1}^N (U_j^{(\tau+1)}(k))^\beta \right)^{-1}, \\ \mu_j^{(\tau+1)} = \sum_{k=1}^N (U_j^{(\tau+1)}(k))^\beta D^2(\tilde{x}(k), Cl_j^{(\tau+1)}) * \left(\sum_{k=1}^N (U_j^{(\tau+1)}(k))^\beta \right)^{-1}, \end{cases} \quad (6)$$

in case when the $\beta = 2$ we come to a decision, that has the form (7):

$$\begin{cases} U_j^{(\tau+1)}(k) = \left(1 + \frac{\|\tilde{x}(k) - Cl_j^{(\tau)}\|^2}{\mu_j^{(\tau)}} \right)^{-1}, \\ Cl_j^{(\tau+1)} = \sum_{k=1}^N (U_j^{(\tau)}(k))^2 \tilde{x}(k) * \left(\sum_{k=1}^N (U_j^{(\tau)}(k))^2 \right)^{-1}, \\ \mu_j^{(\tau+1)} = \sum_{k=1}^N (U_j^{(\tau)}(k))^2 \|\tilde{x}(k) - Cl_j^{(\tau+1)}\|^2 * \left(\sum_{k=1}^N (U_j^{(\tau)}(k))^2 \right)^{-1}. \end{cases} \quad (7)$$

In the online mode algorithms (6), (7) can be written in recurrent form (8)

$$\begin{cases} U_j(k+1) = \left(1 + \left(\frac{D^2(\tilde{x}(k+1), Cl_j(k))}{\mu_j(k)} \right)^{\frac{1}{\beta-1}} \right)^{-1}, \\ Cl_j(k+1) = Cl_j(k) + \eta(k+1) U_j^\beta(k+1) (\tilde{x}(k+1) - Cl_j(k)), \\ \mu_j(k+1) = \sum_{p=1}^{k+1} U_j^\beta(p) D^2(\tilde{x}(p), Cl_j(k+1)) * \left(\sum_{p=1}^{k+1} U_j^\beta(p) \right)^{-1} \end{cases} \quad (8)$$

and

$$\begin{cases} U_j(k+1) = \left(1 + \frac{\|\tilde{x}(k) - c_j(k)\|^2}{\mu_j(k)} \right)^{-1}, \\ c_j(k+1) = c_j(k) + \eta(k+1)U_j^2(k+1)(\tilde{x}(k+1) - c_j(k)), \\ \mu_j(k+1) = \sum_{p=1}^{k+1} U_j^2(p) \|\tilde{x}(p) - c_j(k+1)\|^2 * \left(\sum_{p=1}^k U_j^2(p) \right)^{-1}, \end{cases} \quad (9)$$

that permits to solve the fuzzy clustering task in sequential mode.

4 Adaptive Credibilistic Fuzzy Clustering (ACrFC)

Credibilistic fuzzy clustering is associated with minimizing the goal function (10)

$$E\left(Cr_q(k), w_q\right) = \sum_{k=1}^N \sum_{q=1}^m Cr_q^\beta(k) D^2(x_k, w_q) \quad (10)$$

with constraints $0 \leq Cr_q(k) \leq 1 \forall q, k$; $\sup Cr_q(k) \geq 0,5 \forall k$; $Cr_q(k) + \sup Cr_l(k) = 1$ for any q and k for which $Cr_q(k) \geq 0,5$. Here $Cr_q(k)$ – credibility that observation x_k belongs to a cluster Cl_q . In this case, the membership level is calculated using on the membership function [5, 6]

$$U_q(k) = \varphi_q\left(D(x_k, Cl_q)\right) \quad (11)$$

where: $\varphi_q(\cdot)$ – monotonically decreases in the interval $[0, \infty]$, $\varphi_q(0) = 1$, $\varphi_q(\infty) \rightarrow 0$.

It is easy to see that function (11) is essentially measure of similarity based on distance [7]. As such a function, it was proposed in [8] to use the expression

$$U_q(k) = \left(1 + D^2(x_k, Cl_q) \right)^{-1}. \quad (12)$$

It is interesting to note that expression (12) can be rewritten in the form

$$\begin{aligned}
U_q(k) &= \left(D^2(x_k, Cl_q(k)) \right)^{\frac{1}{1-\beta}} \cdot \left(\sum_{l=1}^m \left(D^2(x_k, Cl_l(k)) \right)^{\frac{1}{1-\beta}} \right)^{-1} = \\
&= \left(D^2(x_k, Cl_q(k)) \right)^{\frac{1}{1-\beta}} \left(D^2(x_k, Cl_q(k)) \right)^{\frac{1}{1-\beta}} + \sum_{\substack{l=1 \\ l \neq q}}^m \left(D^2(x_k, Cl_l(k)) \right)^{\frac{1}{1-\beta}} \right)^{-1} = \quad (13) \\
&= \left(1 + \left(D^2(x_k, Cl_q(k)) \right)^{\frac{1}{1-\beta}} \sum_{\substack{l=1 \\ l \neq q}}^m \left(D^2(x_k, Cl_l(k)) \right)^{\frac{1}{1-\beta}} \right)^{-1}
\end{aligned}$$

that for the Euclidean metric and $\beta=2$ takes the form of a Cauchy distribution density function with a width parameter σ_q^2 [9]

$$U_q(k) = \left(1 + \frac{\|x_k - Cl_q(k)\|^2}{\sigma_q^2} \right)^{-1}, \quad (14)$$

$$\sigma_q^2 = \left(\sum_{\substack{l=1 \\ l \neq q}}^m \|x_k - Cl_l(k)\|^2 \right)^{-1}. \quad (15)$$

It is easy to see that the membership function (13) is a special case of (14) for $\sigma_q^2 = 1$.

Finally, a batch algorithm of credibilistic fuzzy clustering can be written in the form [5, 6]:

$$U_q^{(\tau+1)}(k) = \left(1 + D^2(x_k, Cl_q^{(\tau)}) \right)^{-1}, \quad (16)$$

$$U_q^{*(\tau+1)}(k) = U_q^{(\tau+1)}(k) \left(\sup_l U_l^{(\tau+1)}(k) \right)^{-1}, \quad (17)$$

$$Cr_q^{(\tau+1)}(k) = \frac{1}{2} \left(U_q^{*(\tau+1)}(k) + 1 - \sup_{l \neq q} U_l^*(k) \right), \quad (18)$$

$$Cl_q^{(\tau+1)} = \sum_{k=1}^N \left(Cr_q^{(\tau+1)}(k) \right)^\beta x_k \left(\sum_{k=1}^N \left(Cr_q^{(\tau+1)}(k) \right)^\beta \right)^{-1}. \quad (19)$$

Based on this formulas, we can introduce into consideration online version of the credibilistic fuzzy clustering method in the form

$$\left\{ \begin{array}{l} \sigma_q^2(k+1) = \left(\sum_{\substack{l=1 \\ l \neq q}}^m \|x_{k+1} - Cl_l(k)\|^2 \right)^{-1}, \\ U_q(k+1) = \left(1 + \frac{\|x_{k+1} - Cl_q(k)\|^2}{\sigma_q^2(k+1)} \right)^{-1}, \\ U_{(k+1)}^* = U_q(k+1) \left(\sup_{l \neq q} U_l(k+1) \right)^{-1}, \\ Cr_q(k+1) = \frac{1}{2} \left(U_q^*(k+1) + 1 - \sup_{l \neq q} U_l^*(k+1) \right), \\ Cl_q(k+1) = Cl_q(k) + \eta(k+1) Cr_q^\beta(k+1) (x_{k+1} - Cl_q(k)). \end{array} \right. \quad (20)$$

Therefore, from a computational point of view, the online algorithm for credibilistic fuzzy clustering is no more complicated than the recurrent versions of FCM and PCM, while retaining the advantages of a credibility approach.

5 Evolutionary Cat Swarm Optimization

To search global extremum of a function (1) (5) and (10) it is expedient to use the bio-inspired evolutionary particles swarm optimization algorithms [10]. Among the swarm algorithms, one of the fastest ones are the so-called algorithms of the cats swarm [11, 12], that proved to be effective in solving a wide range of Data Mining tasks.

Cat swarm-CS model of behavior, assumes that each cat cat_p of swarm consisting of Q individuals ($p=1,2,...,Q$), can be in one of two states: Seeking Mode (SM) and Tracing Mode (TM). In this case, the seeking mode is associated with slow movements with a slight amplitude near the initial position (space scanning in the vicinity of the current position), and the tracing mode that is determined by fast jumps with a large amplitude and allows the cat cat_p go put from local extremum, if she got there. The combination of local scanning and abrupt changes in the current state makes it more likely to find a global extremum compared to traditional multi-extrema optimization methods.

In the general case, both of these modes for each of the cats swarms can be described by the recurrent optimization procedure [13]

$$cat_p(\tau+1) = cat_p(\tau) - \alpha(cat_p(\tau) - cat_p(\tau-1)) - \eta \hat{\nabla} E_M(cat_p(\tau)) + \eta_\xi \Xi(\tau), \quad (21)$$

where $cat_p(\tau+1)$ - state of p -th cat of swarm on τ -th iteration of the search, α - parameter that determines the inertia properties of the tracing mode. In a case when $\alpha = 0$ process optimization approaches to the standard gradient search. η - seeking mode step, $\hat{\nabla}E(cat_p(\tau))$ - gradient estimate of the goal function (1), (5) and (10) in the neighborhood of the point $cat_p(\tau)$, $\Xi(\tau)$ - a random component that introduces additional stochastic motions into the tracing process, η_ξ - parameter that specifies the amplitude of these movements.

In this algorithm, each cat can have two parallel states: search mode and tracking mode. This approach provides a search for a global extremum in the case when the number of cats in the swarm is sufficient.

6 Experimental research

Fuzzy clustering based on evolutionary cat swarm optimization (CSO) was performed on four different data samples: Iris, Cancer, Wine and Glass. Each of the data sets has a number of parameters, presented in Table 1.

Table 1. Characteristic Parameters of the Samples

Data Set	Number of clusters	Number of attributes	Number of observations.
Iris	3	4	150
Cancer	2	9	683
Wine	3	13	178
Glass	6	8	214

Table 2. Parameters of cat swarm optimization algorithm (CSO)

Parameters	Value
SRD	Random [0,1]
Seeking memory Pool (SMP)	5
Population size	Number of clusters
r_1	Random in [0,1]
c_1	Const
SPC	Random in [0,1]
Number of iteration	Manually

Table 3. Comparative results of time processing of clustering algorithms such as Fuzzy c-means algorithm (FCM), Particle Swarm Optimization (PSO), Gauss – Seidel algorithm (GSA), CSO, APrFCCSO, APosFCCSO and ACrFCCSO

Data Set	FCM	PSO	GSA	CSO	APrFC CSO	APosFC CSO	ACrFC CSO
Cancer	0.009	0.138	0.204	0.026	0.007	0.010	0.012
Glass	0.010	0.431	0.431	0.021	0.020	0.025	0.024
Iris	0.008	0.020	0.022	0.043	0.012	0.017	0.015
Wine	0.009	0.282	0.098	0.076	0.013	0.015	0.013

Also conducted a comparative analysis of the quality of clustering data on the main characteristics quality ratings, such as: Partition Coefficient (PC), Partition Index (SC), Xie and Beni's Index (XB) of existing clustering methods and proposed method.

Table 4. Evaluation of the quality of fuzzy clustering methods data

Data clustering methods	PC	SC	XB
Fuzzy C-means	0.50	1.62	0.19
Adaptive probabilistic fuzzy clustering	0.25	1.44	0.01
Adaptive possibilistic fuzzy data clustering	0.26	1.22	0.01
Adaptive credibilistic fuzzy data clustering	0.24	1.25	0.01
Adaptive probabilistic fuzzy clustering CSO (APrFCCSO)	0.22	1.37	0.25
Adaptive possibilistic fuzzy data clustering CSO (APosFCCSO)	0.23	0.95	0.01
Adaptive credibilistic fuzzy data clustering CSO (ACrFCCSO)	0.22	0.70	0.15

Table 5. Results of clustering CSO, APrFCCSO, APosFCCSO and APosFCCSO with a different number of iterations (average error in%)

Data Set	Number of iterations CSO			Number of iterations APrFCCSO			Number of iterations APosFCCSO			Number of iterations APosFCCSO		
	50	100	150	50	100	150	50	100	150	50	100	150

Iris	23.34	20.84	21.67	17.55	14.78	16.46	17.65	14.88	16.55	17.45	14.78	16.45
Cancer	40,23	40,55	41,47	38.89	39.22	39.15	37.49	38.42	39.65	37.45	38.44	39.65
Wine	24,55	21,44	22,20	18.43	17.37	16.32	18.95	18.37	19.32	18.85	18.47	19.22
Glass	56.34	56.48	55.67	51.63	51.7	49.79	52.03	51.88	49.39	52.13	51.78	49.49

Table 6 – Comparative results of performance indicators of algorithms such as PSO, CSO, APrFCCSO, APosFCCSO and APosFCCSO for 1st database

Data Set	MSE	PSO	CSO	APrFCCSO	APosFCCSO	APosFCCSO
Iris	Best	4×10^{-7}	9×10^{-10}	8.4×10^{-11}	8.8×10^{-11}	8.56×10^{-11}
	Median	7×10^{-6}	7.3×10^{-9}	5.2×10^{-10}	5.4×10^{-10}	5.5×10^{-10}
	Worst	1.2×10^{-5}	9.3×10^{-9}	9.6×10^{-10}	9.4×10^{-10}	9.4×10^{-10}
Cance	Best	1.3×10^{-7}	8×10^{-10}	8.5×10^{-11}	8.4×10^{-11}	8.35×10^{-11}
	Median	7×10^{-6}	4.4×10^{-9}	7.6×10^{-11}	6.6×10^{-11}	6.55×10^{-11}
	Worst	2.02×10^{-5}	6.8×10^{-9}	7.8×10^{-10}	7.78×10^{-10}	7.8×10^{-10}
Wine	Best	1.4×10^{-6}	7×10^{-10}	9.5×10^{-11}	9.0×10^{-11}	8.95×10^{-11}
	Median	5×10^{-5}	4×10^{-9}	6.6×10^{-11}	6×10^{-11}	5.95×10^{-11}
	Worst	2×10^{-5}	6×10^{-9}	6.8×10^{-10}	6.7×10^{-10}	6.69×10^{-10}
Glass	Best	2.5×10^{-7}	7.9×10^{-10}	8.7×10^{-11}	8.4×10^{-11}	8.35×10^{-11}
	Median	6.9×10^{-6}	5×10^{-9}	7.7×10^{-11}	7.8×10^{-11}	7.77×10^{-11}
	Worst	3×10^{-5}	5.9×10^{-9}	7.7×10^{-10}	7.5×10^{-10}	7.49×10^{-10}

7 Conclusion

The problem of fuzzy clustering based on probabilistic, possibilistic and credibilistic online approaches was considered. The recurrent modifications of well-known batch procedures designed to solve problems of Data Stream Mining allow to process information in online mode as sequential additions to the system under consideration. Since the goal functions of fuzzy clustering in the general case are multi-extremal, was is proposed to refine the solutions using swarm evolutionary optimization

algorithms. The modification introduced on the basis of the optimization procedure cat swarms with improved properties through the use of a stochastic gradient estimate was proposed.

The experiments confirmed the effectiveness of the developed approach, which is characterized by simplicity of numerical implementation and a sufficiently high convergency rate.

References

1. Xu, R., Wunsch, D.C.: Clustering. Hoboken, N.J. John Wiley & Sons, Inc. (2009)
2. Aggarwal, C.C.: Data Mining: Text Book. Springer (2015)
3. Kohonen, T.: Self-Organizing Maps. Berlin: Springer-Verlag (1995)
4. Bezdek, J.C.: Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum Press, New York (1981)
5. Zhou, J., Wang, Q., Hung, C.-C., Yi, X.: Credibilistic clustering: the model and algorithms. *Int.J. of Uncertainty, Fuzziness and Knowledge-Based Systems*, №4, pp.545-564 (2015).
6. Zhou, J., Wang, Q., Hung, C. C.: Credibilistic clustering algorithms via alternating cluster estimation. *J. Intell. Manuf.*, pp.727-738 (2017).
7. Young, F.W., Hamer, R.M.: Theory and Applications of Multidimensional Scaling-Hillsdale, N.J.: Erlbaum (1994).
8. Zhou, J., & Hung, C.-C.: A generalized approach to possibilistic clustering algorithms. *Int. J. of Uncertainty, Fuzziness and Knowledge-Based Systems*. 15. pp.117–138. (2007).
9. Hu, Zh., Bodyanskiy, Ye, Tyshchenko, O., Shafronenko, A.: Fuzzy clustering of incomplete data by means of similarity measures. *Proc.2019 IEEE 2nd Ukr. Conf. on Electrical and Computer Engineering (UKRCON)*, Track 6. Lviv, Ukraine, pp.149-152 (2019).
10. Grosan, C., Abraham, A., Chis, M.: Swarm intelligence in Data Mining - Studies in Computational Intelligence (2006)
11. Chu, S.-C., Tsai, P.-W., Pan, J.S.: Cat swarm optimization. In: *Lecture Notes in Artificial Intelligence*. Berlin Heidelberg: Springer-Verlag (2006).
12. Chu, S.-C., Tsai, P.-W.: Computational Intelligence based on the behavior of cats. In: “*Int. J. of Innovative Computing, Information, and Control*”, №1, pp.163 – 173 (2007)
13. Shafronenko, A., Bodyanskiy, Ye. Pliss, I., Patlan, K.: Fuzzy Clusterization of Distorted by Missing Observations Data Sets Using Evolutionary Optimization. In: *Proceedings “Advanced Computer Information Technologies (ACIT’2019)”*, Česke Budejovice, Czech Republic, June 5-7, 2019, pp. 217-220 (2019). doi: 10.1109/ACITT.2019.8779888
14. Hu, Zh., Bodyanskiy, Ye, Tyshchenko, O., Shafronenko, A.: Fuzzy clustering of incomplete data by means of similarity measures. In: *2019 IEEE 2nd Ukraine*

Conference on Electrical and Computer Engineering UKRCON - 2019, Conference Proceedings, July 2-6, 2019, Lviv, Ukraine, pp.149-152 (2019) doi: 10.1109/UKRCON.2019.8879844.

15. Shafronenko, A. Yu, Bodyanskiy, Ye. V., Pliss I.P.: The Fast Modification of Evolutionary Bioinspired Cat Swarm Optimization Method. In: 2019 IEEE 8th International Conference on Advanced Optoelectronics and Lasers (CAOL), Sozopol, Bulgaria, 6-8 Sept. 2019, pp. 548-552. (2019) doi: 10.1109/CAOL46282.2019.9019583

A Constrained Method of Constructing the Logic Classification Trees on the Basis of Elementary Attribute Selection

Igor Povkhan¹[0000-0002-1681-3466],

¹ Uzhgorod National University, 89 B, Zankovetska str., Uzhgorod, 88000, Ukraine
Igor.povkhan@uzhnu.edu.ua

Abstract. A problem of constructing the logic classification tree model on the basis of a constrained elementary attribute selection method for the geologic data array has been considered. The method of the real data array approximation by a set of elementary attributes with a fixed criterion of the branching procedure stopping at the stage of constructing the classification tree has been suggested. This method allows the desired model accuracy to be provided, its structural complexity to be reduced and the necessary efficiency indices to be achieved. Based on the suggested elementary attribute selection method modification, the software has been developed enabling a set of different-type applied problems to be used.

Keywords: logic classification tree, image recognition, classification, attribute, branching criterion.

1 Introduction

The classification tree methods, the regressive trees used actively both for the artificial intellect theory problems, being the means of decision adoption support and big data array analysis, and for the related practical field of economy, management etc [1] still remain an important segment of the decision tree concept application. The principal available methods of training selection processing at the recognition function constructing do not allow the required accuracy level of the recognition system to be achieved and their complexity in the data system construction process to be regulated [2, 3]. This shortcoming is not peculiar for the recognition system construction methods based on the logic classification tree methods.

Classification trees are one of the basic methods of automatic data analysis. The first comprehensive studies and concepts of the idea of using the decision/ classification trees date back to the works of S. Hawland and E. Hunt [4]. Note that the logic classification tree (LCT) flexibility, i.e. the ability to take into account and investigate sequentially the effect of the influence of certain variables and structure attributes, is their important specific feature. Respectively, there are a series of reasons providing the LCT structures with higher flexibility as compared to the traditional data analysis methods and tools. For instance, the ability to perform one-dimensional branching to

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

analyze the influence (importance, quality) of certain variables allows one to work with different-type variables in a form of predicates [5].

The problem of selecting the branching stopping rule (the LCT construction stopping criterion) is a basic issue in the classification tree construction. Analyzing the available logic tree construction methods and algorithms, one may distinguish two basic approaches in this direction, i.e. the use of the statistical methods of estimation of the necessity of further manifold partitioning/branching (the so called early stopping or pre-pruning) and application of the logic tree depth (number of layers) limitation scheme.

Thus, let us note that the LCT construction method scheme described, e.g., in [6, 7], has one principal disadvantage related to the fact that the number of elementary attributes in the tree increases considerably with the number of vertices in the logic tree structure (the LCT layer increase). Certainly, such complication of the resulting LCT affects negatively the apparatus possibilities of the recognition system (i.e. memory, processor time). The LCT method modification (the constrained method) on the basis of a stage-by-stage selection of elementary attributes shall be suggested to overcome the above negative issues.

In practice, the LCT construction algorithms and methods quite often provide at the output the structurally complicated logic trees (in the meaning of the number of vertices, number of branches, belonging to the class of non-regular trees) that are not uniformly filled with the data and have different number of branches. Such complicated tree-like structures are quite difficult to be used in the external analysis due to a large number of nodes (vertices) and step-by step partitions of the initial training selection (TS) that contain minimal number of objects (in the worst case, possibly, even single objects). Obviously, it is better to have a certain LCT with a minimal number of vertices (nodes) that correspond to the absolute majority of the initial TS objects. Just at this stage a principal problem of the LCT theory appears, i.e. the problem of possible construction of all the logic tree variants that correspond to the initial TS and the minimal by depth (number of layers) logic tree selection [8–12].

2 Formal problem statement

Let the TS be defined in the following form:

$$(x_1, f_R(x_1)), \dots, (x_m, f_R(x_m)). \quad (1)$$

Note that here $x_i \in G$, $f_R(x_i) \in \{0, 1, \dots, k-1\}$, $(i = 1, 2, \dots, m)$, m is a number of objects with TS; $f_R(x_i)$ is a certain finitely significant function that defines the partition R of the manifold G into the classes (images) $H_0, H_1, \dots, H_{k-1}$. Relation $f_R(x_i) = l$, $(l = 0, 1, \dots, k-1)$ means that $x_i \in H_l$, where $x_i = \{x_{i_1}, x_{i_2}, \dots, x_{i_n}\}$, x_{i_j} are the values of the j -th attribute for the object x_i , $(j = 1, 2, \dots, n)$, and n is a number of attributes in the TS.

Thus, the TS is the array (more strictly, the sequence) of certain sets, and each set is the array of values of certain attributes and functions of the above set. One may say that the array of the attribute values is a certain pattern, whereas the value of the function relates the above pattern to the relevant image [13]. A problem is set to construct $LCT - L$ on the basis of the initial TS (1) array and determine the values of its structural parameters p (i.e. $F(L(p, x_i), f_R(x_i)) \rightarrow opt$).

3 Literature review

Present study continues a cycle of works dedicated to the problem of the tree-like recognition schemes and the discrete object logic classification methods and algorithms [3, 7–13]. They deal with the issues of constructing, analyzing and optimizing the logic classification trees. For instance, it is known [14] that the resulting classification rule (scheme) constructed by the arbitrary method or by the branched attribute selection algorithm has a tree-like logic structure. The logic tree consists of the vertices (attributes) grouped in layers that are obtained at a certain step (stage) of the recognition tree construction [15–17]. The problem of synthesizing the recognition trees that will be represented by the algorithm tree (graph) is an important one, according to Ref. [18–21]. Unlike the available methods, the main peculiarity of the tree-like recognition systems is that the importance of certain attributes (attribute groups or algorithms) is determined with respect to the function that defines the object partition into the classes [22]. For instance, in Ref. [23], the case of the decision tree construction for a set of low-informative attributes is considered.

The ability of the LCT to execute one-dimensional branching to analyze the influence (importance, quality) of certain variables gives a possibility to work with the different-type variables in a form of predicates (while in case of the algorithmic classification trees (ACT) the relevant autonomous classification/recognition algorithms shall be applied) [24, 25]. Such a logic tree concept is being used actively in the intellectual data analysis, where the target goal is to synthesize a model that shall predict the value of the target variable on the basis of a set of initial data at the system input [26]. Since the principal idea of the branched attribute selection methods and algorithms could be defined as the optimal approximation of a certain initial TS by a set of elementary attributes (the object attributes), then their central problem, i.e. the issue of selecting an efficient branching criterion (vertices, attributes and discrete object attributes) comes on line [27–29]. The work [17] is devoted just to these principal problems by raising the questions of qualitative estimation of certain discrete attributes, sets and their fixed combinations. This allows an efficient mechanism of branching realization to be implemented.

Note that the absolute majority of known decision tree construction algorithms belong to the ‘greedy algorithm’ class. In other words, if at a certain stage some vertex (attribute, node) was selected and used for the initial selection partitioning into the initial selection subsets, the algorithm fails to return at the next stage to selecting the other vertex (node) with higher-quality partition indices. This means, in fact, that at the stage of the logic tree construction it is impossible to determine whether the se-

lected branching vertex will perform some optimal partition [30, 31] at the final stage. For instance, at the decision tree synthesis, the issues of choosing the criterion of the attribute (the LCT vertex), after which the initial TS will begin, the training (the LCT structure construction) stopping criterion and the logic tree (the LCT sub-tree) branch withdrawal criterion will remain the central ones [32].

The logic classification tree structures obtained on the basis of the branched attribute selection methods are characterized, on the one hand, by the compactedness, and, on the other hand, by the non-uniformity of the layer filling (rarity) as compared to the regular trees (the algorithm with a single estimation of the attribute importance [10]. Note that the issues of convergence of the LCT construction process according to the elementary attribute selection methods and those of selecting the criterion of the logic tree synthesis process stopping (e.g. limitation by the logic tree depth or complexity, the accuracy or the structure error number) still remain important [33]. The solution of these issues from the viewpoint of a constrained (modified) LCT construction method shall be presented below.

Thus, notwithstanding a certain range of problems that arise when constructing and using the decision trees (the LCT concept in the general sense), one has to fix their following advantages:

- i) the tree-like models are distinguished by a principally fast stage of training the recognition system;
- ii) there exists a possibility to synthesize a manifold of decision rules within the area, where it is difficult even for a qualified expert to form a set of recommendations;
- iii) synthesis of rules (classification rules, decision rules) in the natural language;
- iv) the resulting tree-like model obtained is intuitively understandable;
- v) the prediction constructed at the final stage of model operation is characterized by a high accuracy even as compared to the statistical and neural network models;
- vi) a possibility to construct non-parametric model.

Hence, taking into account the aforementioned, one may conclude that the decision tree models together with the neural network approach [22, 34] are an important and relevant tool for the data structure extensive analysis and presentation.

4 A concept of training selection data approximation by a set of ranked elementary attributes

Consider some defined general-form (1) training selection (HS) and a certain system (set) of elementary attributes for the initial selection array $\varphi_1(x), \varphi_2(x), \dots, \varphi_n(x)$. Let us introduce the following manifolds:

$$G_{r_1, \dots, r_n} = \{x \in G / a_1(x) = r_1, \dots, a_n(x) = r_n\}. \quad (2)$$

Note that the system of manifolds $G_{r_1, \dots, r_n}$ is a complete partition of the manifold G realized by the elementary attributes $\varphi_1, \varphi_2, \dots, \varphi_n$. One has to keep in mind that some of the manifolds $G_{r_1, \dots, r_n}$ may be empty.

Let us set $S_{r_1, \dots, r_n}$ as the number of entrances into the training selection of the pairs $(x_i, f_R(x_i)), (1 \leq i \leq m)$ that satisfy condition $x_i \in G_{r_1, \dots, r_n}$.

Similarly, we shall set $S_{r_1, \dots, r_n}^j$, ($j = 0, 1, \dots, k-1$) as the number of entrances into the training selection of the pairs $(x_i, f_R(x_i)), (i = 1, \dots, m \leq m)$ that satisfy conditions $x_i \in G_{r_1, \dots, r_n}$ and $f_R(x_i) = j$.

Let us introduce the following quantities:

$$\delta_{r_1, \dots, r_n} = \frac{S_{r_1, \dots, r_n}}{m}; \psi_{r_1, \dots, r_n}^j = \frac{S_{r_1, \dots, r_n}^j}{S_{r_1, \dots, r_n}}; \rho_{r_1, \dots, r_n} = \max_j \psi_{r_1, \dots, r_n}^j. \quad (3)$$

Note that if $x_i \notin G_{r_1, \dots, r_n}$ for any $(i = 1, \dots, m)$, then $\delta_{r_1, \dots, r_n} = 0$ and $\psi_{r_1, \dots, r_n}^j = 0$, ($j = 0, \dots, k-1$).

The $\delta_{r_1, \dots, r_n}$ quantity characterizes the frequency of entrances of the members of sequence $x_1, x_2, \dots, x_m$ into the manifold $G_{r_1, \dots, r_n}$. The $\psi_{r_1, \dots, r_n}^j$ quantity characterizes the frequency of the object x belonging to the image (class) H_j , provided $x \in G_{r_1, \dots, r_n}$. It should be noted here that condition $x \in G_{r_1, \dots, r_n}$ is equivalent to the elementary attribute one: $\varphi_1(x) = r_1, \varphi_2(x) = r_2, \dots, \varphi_n(x) = r_n$.

The $\delta_{r_1, \dots, r_n}$ quantity characterizes the information efficiency of recognizing the belonging of x to one of the classes $H_0, H_1, \dots, H_{k-1}$, subject to the basic condition: $x \in G_{r_1, \dots, r_n}$, certainly.

Assume that at the next stage $x \in G_{r_1, \dots, r_n}$. Then the question arises, to what of the classes $H_0, H_1, \dots, H_{k-1}$ x could be related. Obviously, x must be related to the class H_l , for which the following relation holds true:

$$\rho_{r_1, \dots, r_n} = \psi_{r_1, \dots, r_n}^l, \{0 \leq l \leq k-1\}. \quad (4)$$

Note that this relation is a certain final classification rule. It is obvious that the larger is $\rho_{r_1, \dots, r_n}$, the higher is its general efficiency.

As mentioned above, the initial TS is meant a single information that represents the images $H_0, H_1, \dots, H_{k-1}$. Therefore, the class H_l means a manifold of all the TS pairs $(x_i, f_R(x_i))$ that satisfy the basic relation $f_R(x_i) = l$.

The mean efficiency of recognizing the images $H_0, H_1, \dots, H_{k-1}$ defined by the TS data with the help of the elementary attributes $\varphi_1, \varphi_2, \dots, \varphi_n$ shall be estimated by the following quantity:

$$o_S(\varphi_1, \varphi_2, \dots, \varphi_n) = \sum_{r_1, \dots, r_n} \delta_{r_1, \dots, r_n} * \rho_{r_1, \dots, r_n}. \quad (5)$$

The quantity $o_S(\varphi_1, \varphi_2, \dots, \varphi_n)$ could be considered the estimate of the TS approximation by a set of elementary attributes $\varphi_1, \varphi_2, \dots, \varphi_n$. From relation (5), the following expressions for $\delta_{r_1, \dots, r_n}$, $\psi_{r_1, \dots, r_n}^j$ and $\rho_{r_1, \dots, r_n}$ result:

$$\begin{aligned} \delta_{r_1, \dots, r_n} &\geq 0, \quad \sum_{0 \leq r_1, \dots, r_n \leq 1} \delta_{r_1, \dots, r_n} = 1; \\ \psi_{r_1, \dots, r_n}^j &\geq 0, \quad \sum_{j=0}^{k-1} \psi_{r_1, \dots, r_n}^j = 1; \\ \frac{1}{k} &\leq \rho_{r_1, \dots, r_n} \leq 1, (r_1, \dots, r_n \in \{0, 1\}). \end{aligned} \quad (6)$$

From relation (6), the following properties of the quantity $o_S(\varphi_1, \varphi_2, \dots, \varphi_n)$ result directly:

$$\begin{aligned} \text{i)} \quad &\frac{1}{k} \leq o_S(\varphi_1, \varphi_2, \dots, \varphi_n) \leq 1; \\ \text{ii)} \quad &o_S(\varphi_1, \varphi_2, \dots, \varphi_n) = 1 \Leftrightarrow \rho_{r_1, \dots, r_n} = 1; \end{aligned} \quad (7)$$

Note that the above properties are valid for any $r_1, \dots, r_n$ that satisfy relation $\exists i(1 \leq i \leq m, x_i \in G_{r_1, \dots, r_n})$. It follows from the second part of relation (7) that the set of elementary attributes $\varphi_1, \dots, \varphi_n$ shall then and only then realize a complete recognition of images $H_0, H_1, \dots, H_{k-1}$ (defined by the initial TS). This means that the set of elementary attributes will be a test given $o_S(\varphi_1, \varphi_2, \dots, \varphi_n) = 1$, while formula (5) enables one to find such algorithm sets.

Consider one more important peculiarity of the functional estimation of the set of elementary attributes with respect to the initial TS. Let us have a certain number

$b, (\frac{1}{k} \leq b < 1)$ and let M_b be a total number of entrances of the objects x_i into the sequence $x_1, \dots, x_m$, for which the following relations hold true:

$$\begin{cases} x_i \in G_{r_1, \dots, r_n} \\ \rho_{r_1, \dots, r_n} > b \end{cases}. \quad (8)$$

The above relations represent the recognition efficiency for the objects x_i by means of a certain set of elementary attributes $\varphi_1, \dots, \varphi_n$ larger than b .

The number $\gamma_b = \frac{m_b}{m}$ is a fraction of entrances of the objects x_i into the sequence $x_1, \dots, x_m$, for which the recognition efficiency by means of the elementary attribute set $\varphi_1, \dots, \varphi_n$ is larger than b .

The number γ_b is expressed as:

$$\gamma_b = \sum_{r_1, \dots, r_n > b} \delta_{r_1, \dots, r_n} . \quad (9)$$

Note that expression $\sum_{r_1, \dots, r_n > b}$ means summation over all $r_1, \dots, r_n$ that satisfy relation

$\rho_{r_1, \dots, r_n} > b$. Let us assume that $o_S(\varphi_1, \varphi_2, \dots, \varphi_n) = c$ and $\frac{1}{k} \leq b < c$. Then

from expression (5) and relation for $o_S(\varphi_1, \varphi_2, \dots, \varphi_n) = c$, we have:

$$c = \sum_{\rho_{r_1, \dots, r_n} > b} \delta_{r_1, \dots, r_n} * \rho_{r_1, \dots, r_n} + \sum_{\rho_{r_1, \dots, r_n} \leq b} \delta_{r_1, \dots, r_n} * \rho_{r_1, \dots, r_n} . \quad (10)$$

From relations (6) and (9), we obtain as follows:

$$\sum_{\rho_{r_1, \dots, r_n} \leq b} \delta_{r_1, \dots, r_n} = 1 - \gamma_b . \quad (11)$$

From relations (6), (9), (10) and (11), we have:

$$c \leq \gamma_b + (1 - \gamma_b) * b . \quad (12)$$

It results directly from relation (12) that:

$$\gamma_b \geq \frac{c - b}{1 - b} . \quad (13)$$

Let us substitute $c = 1 - \varepsilon^2$ and $b = 1 - \varepsilon$ into relation (13), where ε is an infinitely small number. Given such conditions, we have:

$$\gamma_b \geq 1 - \varepsilon . \quad (14)$$

It results from relation (14) that, if $c \rightarrow 1$, then there exists the quantity b (dependent on c) that:

$$\frac{1}{k} \leq b < c, \lim_{c \rightarrow 1} b = 1 .$$

Note that the quantity b in this case does not depend on the initial TS (and, thus, on its capacity and number m , as well).

5 A constrained method of classification tree construction on the basis of elementary attribute selection

Note that the main idea of the stage-to-stage elementary attribute selection is to maximize the value of the attribute $W_m(f)$ quality [2]. The latter means that such generalized attribute f must be found in the logic tree algorithms for the training selection (1), for which the value of $W_m(f)$ is as large as possible.

It should be noted that the elementary attribute importance (informativeness) means here the value that could be calculated, as a variant, by the following functionals [13].

It should be noted that each LCT vertex (Fig. 1) comprises either certain attribute (mark) φ_i^j or number m_i^j that belongs to the manifold $\{0, 1, \dots, k-1\}$. Since the vertex containing m_i^j is known as the final LCT vertex, it is also called usually the logic tree leaf.

Two guiding lines (arrows) denoted 0 and 1 go out from the vertex with the attribute φ_i^j . The guiding line 0 corresponds to the value $\varphi_i^j = 0$, whereas the guiding line 1 – to $\varphi_i^j = 1$. The logic tree is conditionally divided into the layers (levels), and the j -th LCT layer contains the corresponding attributes $\varphi_1^j, \varphi_2^j, \dots$.

Note that all the elementary attributes at all the LCT layers, beginning from the first one and ending by the n -th one, represent, in fact, the attributes obtained after the n steps (stages) of the LCT construction. The elementary attributes located at the n -th layer are those derived, respectively, at the n -th step (stage) of the logic tree construction process.

Note that such classification tree (Fig. 1) realizes (represents), in fact, some generalized attribute (x) defined at the manifold G that takes a value from the manifold $\{0, 1, \dots, k-1\}$. After constructing the attribute $f_i(x)$, a stage of a check-test begins. Thus, in the check-test regime, a total number S of all those pairs $(x_i, f_R(x_i))$ from selection (1), for which relation $f_R(x_i) = f_i(x)$ holds true, is calculated.

At the next stage, we have to check condition $\frac{S}{m} \geq \delta$. Here δ is a parameter that characterizes the estimate of training efficiency with respect to the current problem (TS). If this condition is satisfied, the classification tree construction process terminates, and the generalized attribute $f_i(x)$ represented by the structure tree (Fig. 1) ensures selection (1) approximation.

Let us emphasize that such scheme of constructing the logic classification tree has one principal disadvantage related to the fact that the number of elementary attributes φ_i^j in the tree increases considerably (here i is the elementary attribute number in

the set, j is that of the attribute location layer). Certainly, such complication of the resulting classification tree affects negatively the apparatus capabilities of the recognition system (i.e. memory, processor time).

To overcome the above negative issues, one may suggest the following modification of the TS data approximation method using the elementary attribute set.

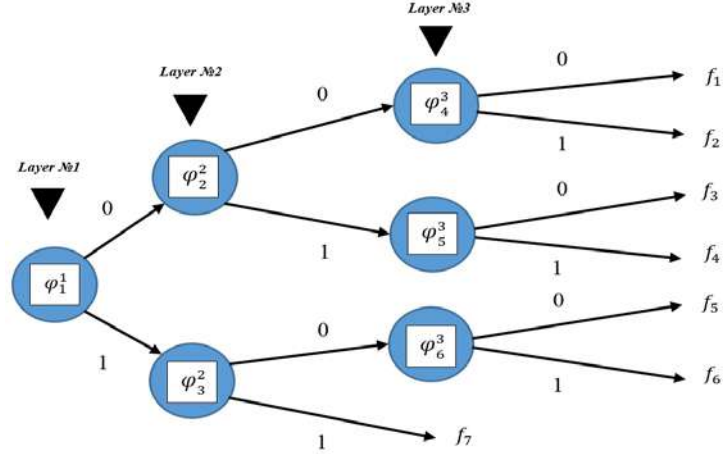


Fig. 1. Finished logic classification tree structure.

Let us fix some positive number Z . Consider a logic tree with structure shown in Fig. 1 that reflects a certain predicate (generalized attribute) $f_i(x)$.

So, at the test stage we calculated a certain number S that occurs in relation $\frac{S}{m} \geq \delta$. Now, besides the number S , for each unfinished path $r_1 r_2 r_3$ of the logic tree (Fig. 1), we calculate also the number $S_{r_1 r_2 r_3}$, where $S_{r_1 r_2 r_3}$ is a number of all pairs $(x_i, f_R(x_i))$ from TS that, in fact, belong to the path $r_1 r_2 r_3$ and satisfy relation $f_R(x_i) = l(r_1 r_2 r_3)$. Thus, $S_{r_1 r_2 r_3}$ is a number of all those errors made by a certain predicate $f_i(x)$ (generalized attribute) represented by the LCT (Fig. 1) on the path $r_1 r_2 r_3$.

At the next step, we choose the number S of such paths $(r_1 r_2 r_3)_1, \dots, (r_1 r_2 r_3)_Z$, for which the number $S_{r_1 r_2 r_3}$ will be maximal. For example, let $Z = 3$, and the following relation holds true: $S_{000} \geq S_{100} \geq S_{101} \geq S_{001} \geq S_{010} \geq S_{011}$.

Then the paths 000, 100, 101 will be chosen only. The further construction/selection of vertices (elementary attributes) $\varphi_{r_1 r_2 r_3}$ will be realized for the above paths only.

We shall call this classification tree scheme a constrained method of the LCT structure construction, because, according to this scheme, the paths with the maximal number of errors will continue only.

It should be noted here that, in case of using the above processes at the end of the paths $r_1 r_2 r_3$ that do not belong to the selected Z paths, the values $l(r_1 r_2 r_3)$ shall be preserved.

Note that the process of the constrained method of the LCT structure construction could be applied in case when the initial TS (1) is not fixed, i.e. when at each process step a separate selection is made.

6 Model selection

Note that the LCT construction scheme suggested above allows one to regulate the classification tree model complexity or to develop a model with prescribed accuracy. Here the task of choosing the classification tree model from a set of constructed LCTs for a certain problem is determined by an array of parameters having an essential importance with respect to a current applied problem (the TS data set).

Obviously, in order to compare the LCT models and select a particular one, we have to distinguish its principal parameters (i.e. the image space dimensionality, the number of vertices etc) and determine their errors with respect to the array of the input data.

It is principally important to consider, at this stage of study, the quality criteria of obtained information models dependent on the model errors, initial TS data array capacity (the number of training pairs and the problem attribute space dimensionality) and the number of the model parameters etc.

It is apparent that the model errors at the TS and ST data arrays for each of the classes defined by the initial condition of the current applied problem are critically important parameters of developed LCT model, and they must be necessarily minimized.

Note that reduction of the LCT structure complexity (i.e. the number of attributes in the LCT structure, the total number of vertices in the LCT model and that of transitions in the LCT structure), as well as the parameters of the information system total memory consumption and processor time remain here a principal moment. Thus, a total integral quality index in the following form:

$$Q_{Main} = \frac{Fr_{All}}{V_{All} \cdot \sum_i p_i} \cdot e^{-\frac{Er_{All}}{M_{All}}} . \quad (15)$$

is an important quality index of constructed model in a form of the classification tree with the parameters of the LCT model structure being taken into account.

Note that here the parameter Er_{All} is a total number of the LCT model errors at the data arrays of the initial test and training selections, respectively, while M_{All} is a total capacity of these two data arrays. The Fr_{All} parameter characterizes a number

of vertices in the obtained LCT model with the resulting values f_R (RF, i.e. the classification tree leaves), whereas the V_{All} parameter is a total number of all the types of vertices in the LCT model structure. A set of parameters p_i is the most important characteristic of the classification tree to be estimated (the number of elementary attributes used in the classification tree model, the number of transitions between the classification tree vertices, layers etc).

It should be noted that this integral LCT model quality index will take values from zero to unit. The less is this index, the worse is the quality of the constructed tree, and, in contrary, the larger is this index, the better the model obtained is.

7 Experiments and results

The suggested constrained method of constructing the LCT models (i.e. the modified method of elementary attribute selection) has been approved for the problem of classifying an array of geological data (an autonomous recognition system was constructed for it on the basis of a tree-like classification model), namely, a problem of oil-bearing formation separation. The initial parameters of the above applied problem are listed in the following table.

Table 1. Initial parameters of the problem.

Attribute space dimensionality – N	Initial TS data array capacity – M	ST data array capacity – S	Total class number over the TS data partition – l	Different TS class object ratio – (H_1 / H_2)
22	1,250	240	2	756/494

Information about partitioning into two classes is presented in the TS. At the stage of examination, the constructed classification system must provide effective recognition of objects with unknown classification with respect to these two classes. Note that at the initial stage the training and the test selections were automatically checked for the correctness (searching and deleting the similar objects of different belonging, i.e. the first and second-type errors).

It should be noted that the training pairs of the H_1 class (the oil-bearing formations) in the $\approx (1.5/1)$ proportion dominated in the training information array, while the TS itself consisted of 1,250 objects. The constructed recognition system efficiency was estimated at the test sampling of 240 objects, whereas the ST was a segregated TS part (consisting of the discrete objects of known classification). The training and the test selection array data were obtained on the basis of the geological survey on the territory of Transcarpathian province during the 2001–2011 period.

A fragment of the principal results of the above experiments (comparative tests of the LCT model construction methods at the applied problem data array) is presented

in Table 2. The constructed LCT models provided the desired level of accuracy preset by the problem condition, system response rate and operating memory consumption, but demonstrated different structural complexity of classification tree and generalized attribute set construction (in the case of the algorithmic classification tree model [7]).

Note that estimation of the classification tree model quality suggested here reflects the basic parameters (characteristics) of classification trees and could be used as the optimality criterion in the procedure of estimating an arbitrary tree-like recognition system, for instance, in case of the random LCT construction and selection methods taken from Ref. [17].

The constrained method of elementary attribute selection (the modified LCT construction method) suggested in this work was compared to the complete LCT method, the algorithm with a single (initial) estimation of the discrete attribute importance and the algorithmic classification tree method, and has shown, in general, a reasonable result.

Table2. Comparative table of classification tree models /methods.

No.	Logic tree structure synthesis method	Integral model quality index Q_{main}	Total number of errors of the model at the TS and ST Er_{All}
1	Complete LCT method on the basis of elementary attribute selection	0,004789	2
2	LCT model with a single attribute importance estimation	0,002263	3
3	Constrained LCT construction method	0,003181	2
4	Algorithmic tree method (type I)	0,005234	0
5	Algorithmic tree method (type II)	0,002941	0

And, finally, please note that the principal idea of classification tree method on the basis of the autonomous algorithms in the own structure is to provide the stage-to-stage approximation of the initial TS data array by the selected algorithm set (that is reflected in the tree construction itself). On the one hand, the ACT structure obtained is characterized by a high flexibility with respect to the applied problems and a relatively compact structure of the model itself, while, on the other hand, it requires the sufficient apparatus costs (i.e. memory and processor time) to save the generalized attributes (classification rule parameters) and the initial quality estimation of the fixed classification algorithms in accordance with the TS data. Therefore, as compared to the ACT concept, the LCT method (i.e. the constrained elementary attribute selection method) possesses high classification scheme response rate, relatively low apparatus costs for the tree structure (the data structure) saving and operating and high quality (regulated complexity) of the discrete object classification.

8 Conclusions

Thus, one may conclude that the problem of the LCT construction automation on the basis of a constrained method of the TS approximation by a set of elementary attributes has been solved and the scheme of the classification tree construction with the prescribed accuracy has been suggested. Present paper analyzes the constrained method of the LCT model construction that allows the prescribed-accuracy classification trees to be constructed by regulating the complexity of the scheme under generation. Such approach enables the optimal structure of the LCT model under construction to be achieved and provides the necessary and sufficient accuracy of the classification tree obtained.

The scientific novelty of the obtained results is that for the first time a simple constrained method/scheme of the LCT construction has been suggested for the first time on the basis of selecting the elementary attributes with the permanent estimation of their importance at each step of classification tree construction with the ability of a further priority-related construction of the fixed LCT construction blocks. Note that this method at each branching (LCT generating) step takes into account the influence of a certain object attribute value on the resulting RF value f_R in the classification tree structure.

Branching criteria in the LCT structure presented in this paper could be effectively used not only to estimate the informativeness efficiency of certain elementary attributes, but to calculate the importance of the attribute sets and combinations as well. This allows the question of specific features of the LCT model construction for less informative attributes to be raised [23]. We have also suggested a general integral index of the LCT model quality that allows the general classification tree characteristics to be represented effectively. Moreover, it can be used to select the most optimal LCT in case of the methods of random LCT construction [17].

The practical merit of the results obtained lies in the fact that suggested constrained LCT model construction method provides possibility to construct cost-saving and efficient classification models with prescribed accuracy. This method was realized in the ORION III system algorithm library to solve different applied classification problems. The above practical applications have confirmed the performance ability of the LCT models constructed and software developed. These studies may be addressed in future towards the further development of the LCT methods, optimization of the software realizations of the suggested constrained method of the LCT construction, as well as towards its practical approval for a number of real problems in the field of classification and recognition.

References

1. Quinlan, J.R.: Induction of Decision Trees. Machine Learning, no. 1, pp. 81-106 (1986).
2. Vasilenko, Y.A., Vasilenko, E.Y., Povkhan, I.F.: Conceptual basis of image recognition systems based on the branched feature selection method. European Journal of Enterprise Technologies, no. 7(1), pp. 13-15 (2004).
3. Povhan, I.: General scheme for constructing the most complex logical tree of classification

- in pattern recognition discrete objects. Collection of scientific papers. Electronics and information technologies, vol. 11, pp. 112-117. Lviv (2019).
4. Hastie, T., Tibshirani, R., Friedman, J.: The Elements of Statistical Learning. Springer, Berlin (2008).
 5. Srikant, R., Agrawal, R.: Mining generalized association rules. Future Generation Computer Systems, vol. 13, no. 2, pp. 161-180 (1997).
 6. Vasilenko, Y.A., Vasilenko, E.Y., Povkhan, I.F.: Branched feature selection method in mathematical modeling of multi-level image recognition systems. Artificial Intelligence, no. 7, pp. 246-249 (2003).
 7. Povkhan, I.F.: Features of synthesis of generalized features in the construction of recognition systems using the logical tree method. In: Materials of the international scientific and practical conference Information technologies and computer modeling ITKM-2019. Ivano-Frankivsk, pp. 169-174 (2019).
 8. Povhan, I.: Designing of recognition system of discrete objects, IEEE First International Conference on Data Stream Mining & Processing (DSMP), Lviv, 2016, Ukraine, pp. 226-231 (2016).
 9. Shai, B.: Decision Trees. Understanding Machine Learning. Cambridge University Press, (2014).
 10. Vasilenko, Y.A., Vashuk, F.G., Povkhan, I.F.: The problem of estimating the complexity of logical trees recognition and a general method for optimizing them. European Journal of Enterprise Technologies, 6/4(54), pp. 24-28 (2011).
 11. Vtogoff, P.E.: Incremental Induction of Decision Trees. Machine Learning, no. 4, pp. 161-186 (2009).
 12. Elmanova, N.A.: The construction of decision trees. Computer Press, no. 12, pp. 38-45 (2003).
 13. Povkhan, I.F.: The problem of functional evaluation of a training sample in discrete object recognition problems. Scientific notes of the Tauride national University. Series: technical Sciences, vol. 29(68), no. 6, pp. 217-222 (2018).
 14. Breiman, L.L., Friedman, J.H., Olshen, R.A., Stone, C.J.: Classification and regression trees. Boca Raton, Chapman and Hall/CRC (1984).
 15. Laver, V.O., Povkhan, I.F.: The algorithms for constructing a logical tree of classification in pattern recognition problems. Scientific notes of the Tauride national University, Series: technical Sciences, vol. 30(69), no. 4, pp. 100-106 (2019).
 16. Povkhan, I.F.: The concept of function and algebraic recognition scheme in classification problems of discrete objects. Scientific notes of the Tauride national University, Series: technical Sciences, vol. 30(69), no. 2, pp. 171-177 (2019).
 17. Povkhan, I.F.: Features random logic of the classification trees in the pattern recognition problems. Scientific notes of the Tauride national University, Series: technical Sciences, vol. 30(69), no. 5, pp. 152-161 (2019).
 18. Deng, H., Runger, G., Tuv, E.: Bias of importance measures for multi-valued attributes and solutions. In: Proceedings of the 21st International Conference on Artificial Neural Networks (ICANN), pp. 293-300 (2011).
 19. Kamiński, B., Jakubczyk, M., Szufel, P.: A framework for sensitivity analysis of decision trees. Central European Journal of Operations Research, 26 (1), pp. 135-159 (2017)
 20. Karimi, K., Hamilton, H.J.: Generation and Interpretation of Temporal Decision Rules. International Journal of Computer Information Systems and Industrial Management Applications, vol. 3, pp. 314-323 (2011).
 21. Kotsiantis, S.B.: Supervised Machine Learning: A Review of Classification Techniques. Informatica, no. 31, pp. 249-268 (2007).
 22. Bodyanskiy, Y., Vynokurova, O., Setlak, G. and Pliss, I.: Hybrid neuro-neo-fuzzy system and its adaptive learning algorithm. In: Xth Scientific and Technical Conference Computer Sciences and Information Technologies (CSIT), 2015, Lviv, pp. 111-114 (2015)

23. Subbotin, S.A.: Construction of decision trees for the case of low-information features. *Radio Electronics, Computer Science, Control*, no. 1, pp. 121-130 (2019).
24. Vasilenko, Y.A., Vasilenko, E.Y., Povkhan, I.F.: Defining the concept of a feature in pattern recognition theory. *Artificial Intelligence*, no. 4, pp. 512-517 (2002).
25. Subbotin, S., Oliinyk, A.: The dimensionality reduction methods based on computational intelligence in problems of object classification and diagnosis. *Recent Advances in Systems, Control and Information Technology* [eds.: R. Szewczyk, M. Kaliczyńska]. Springer, Cham, pp. 11-19 (Advances in Intelligent Systems and Computing, vol. 543). (2017).
26. Koskimaki, H., Juutilainen, I., Laurinen, P., Roning, J.: Two-level clustering approach to training data instance selection: a case study for the steel industry. In: *Neural Networks: International Joint Conference (IJCNN-2008)*, Hong Kong, 1-8 June 2008 - proceedings. Los Alamitos, IEEE, pp. 3044-3049 (2008). doi: 10.1109/ijcnn.2008.4634228
27. Subbotin, S.A.: Methods and characteristics of localitypreserving transformations in the problems of computational intelligence. *Radio Electronics, Computer Science, Control*, no. 1, pp. 120-128 (2014).
28. Subbotin, S.: The neuro-fuzzy network synthesis and simplification on precedents in problems of diagnosis and pattern recognition. *Optical Memory and Neural Networks (Information Optics)*, vol. 22, no. 2, pp. 97-103 (2013). doi: 10.3103/s1060992x13020082
29. Subbotin, S.A.: Methods of sampling based on exhaustive and evolutionary search. *Automatic Control and Computer Sciences*, vol. 47, no. 3, pp. 113-121 (2013). doi: 10.3103/s0146411613030073
30. Miyakawa, M.: Criteria for selecting a variable in the construction of efficient decision trees. *IEEE Transactions on Computers*, vol. 38, no. 1, pp. 130-141 (1989).
31. Painsky, A., Rosset, S.: Cross-validated variable selection in tree-based methods improves predictive performance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 11, pp. 2142-2153 (2017). doi:10.1109/tpami.2016.2636831
32. De Mántaras, R. L.: A distance-based attribute selection measure for decision tree induction. *Machine learning*, vol. 6 (1), pp. 81-92 (1991).
33. Deng, H., Runger, G., Tuv, E.: Bias of importance measures for multi-valued attributes and solutions. In: *21st International Conference on Artificial Neural Networks (ICANN)*, Espoo, 14-17 June 2011 - proceedings. Springer-Verlag, Berlin, vol. 2, pp. 293-300 (2011)
34. Lupei, M., Mitsa, A., Repariuk, V., Sharkan, V.: Identification of authorship of Ukrainian-language texts of journalistic style using neural networks. *Eastern-European Journal of Enterprise Technologies*, 1 (2 (103)), pp. 30-36 (2020). doi: <https://doi.org/10.15587/1729-4061.2020.195041>

Using the Elman neural network as an identity map in defect detection task

Volodymyr Khandetskyi¹[0000-0002-6386-4637], Nadiia Karpenko¹[0000-0003-4700-6357],

¹Oles Honchar Dnipro National University, Gagarin Avenue, 72, Dnipro, 49010, Ukraine
v.khandetsky@gmail.com, karpenko_n@ffeks.dnulive.dp.ua

Abstract. This article presents the results of a study of the Elman neural network to identify defects (delamination) in composite materials and estimate their area. The registration zone is a square matrix consisting of 100 elements. For training, we used images with different defect spot area, moving along the matrix and additively mixed with Gaussian noise of various intensity. The network structure optimal by the criterion of testing error/training time is determined. Testing showed that when the defect spot area changes by more than 8 times and with a noise level of 27%, the test error reaches 6%. This error is significantly decreased when narrowing the range of the defect area changing and reducing the noise intensity. A block diagram of the corresponding intelligent system is proposed.

Keywords: delamination, Elman neural network, images, distortions, noise, identity map, intelligent system

1 Introduction

At present in a number of industries composite materials are widely used due to their high mechanical and thermodynamic properties at low density [1]. Most of them are laminated materials, so the most common defect in them is delamination. To identify and evaluate parameters of such defects in composites, ultrasound diagnostics of products are most often used. With two-sided access it is reasonable to use a shadow testing method with the radiating and sensor elements are located on opposite sides, for example, the product wall. Delamination of the material is a barrier in the path of ultrasound.

A composite material is heterogeneous in structure. The spatial arrangement of reinforcing elements, such as fiber boundless, is often not ideal. Furthermore, in the bulk of the material and on its surface, especially at phase boundaries, there are accumulations of pores and various kinds of microdefects. Cracking of the composite matrix and fibers breaks are observed. The concentration of microdefects is especially high at the boundaries of delamination, i.e. in places where the material begins to disintegrate. All these structural imperfections distort the defect image fixed by the sensor element. Therefore, to identify the defect and estimate the true area of the delamination, it is necessary to process its noisy image.

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

The perspective of using neural networks for these purposes is based on the possibility of parallelizing the processing of information and their ability to learn, i.e. make generalizations. This makes it possible to identify images that have not been encountered in the learning process. In the problem of classification of defects images considered by us, the processing of information by a neural network in a number of key positions is close to the methods of non-parametric statistical training.

Recurrent neural network differs from the classical network of feed-forward by the presence of feedback loops from hidden or output neurons. Their presence has a direct impact on the ability of such networks to learn [2]. The Elman network contains recurrent links of hidden neurons with a layer of context units consisting of unit-delay elements. These elements save the outputs of hidden neurons for one-time step and then transmit them to the inputs of the neurons. This leads to the non-linear dynamic behavior of the network and allows implement a learning process that develops over time. Promising is the integration of a neural network with other information blocks into a single intellectual decision-making system [3].

2 Problem statement

The article aims to create an intelligent system with a core in the form of an Elman neural network to identify and estimate the area of distorted defect images moving within the control zone. In the process of solving this problem, we want to study the ability of recurrent neural networks to identify dynamic noisy images of various sizes defects in the process of scanning by the measuring transducer the product surface.

3 Literature review

Over the years, neural network image processing methods have been successfully used in the field of non-destructive testing, as well as in a number of other engineering fields.

In the article [4], the processes of teaching and testing of a back-propagation neural network to identify distorted defect signals have been analyzed. Using numerical simulation, including a network clustering mechanism, the authors found that clustering increases the probability of signal recognition.

The article [5] presents a deep-learning mechanism for classifying computer-generated images and photographic images. The proposed method accounts for a convolutional layer capable of automatically learning a correlation between neighboring pixels. The layer is designed to subdue the image's content and robustly learn the sensor pattern noise features as well as the statistical properties of images.

The authors of [6] analyze new trends in the digitization of complex engineering drawings. It includes symbol detection and symbol classification. The article [7] presents a symbol recognition method which applied an interactive learning strategy based on the recurrent training of a Hopfield neural network. This method was designed to find the most common symbols in the drawing, which were characterized by having a prototype pattern. The method recursively learns the features of the samples

to the detection and classification accuracy. However, the method can only identify symbols that are formed by a “prototype pattern”, which means that irregular shapes cannot be addressed through this framework.

Elman network [8, 9] has been successfully applied in many fields, regarding prediction, modeling, pattern recognition, and control. In [10] Elman network is trained to predict the future value of the residual time series. Finally, the network is used to capture the relationship among the predicted value of the original time series and residual time series.

The standard back-propagation algorithm used in Elman neural network is called Elman’s back-propagation algorithm. To increase the convergence, speed a new learning rate scheme is proposed [11].

Multilayer perceptron network (MLP) and Elman neural network were compared in four different time series prediction tasks [12]. Time series include load in an electric network series, which has a low-frequency trend, fluctuations in a far-infrared laser series, numerically generated series, and behavior of sunspots series. The time series is said to be stationary. MATLAB neural network toolbox training functions were used for training MLP and Elman network. Results show that the efficiency of the learning algorithm is a more important factor than the network model used. Elman network models load in an electric network series better than MLP. This network predicts the slope of the trend of the testing data more accurately than the MLP network. In other prediction tasks, it performs similarly to the MLP network.

One of the major problems facing researchers in the recurrent networks is the selection of the hidden neuron’s numbers in neural network layers [13].

Jinchuan and Xinzhe [14] investigated a formula: $N_h = (N_{in} + \sqrt{N_p}) \cdot L$, where L is the number of hidden layers, N_{in} is the number of input neurons and N_p is the number of input images.

Kallan R. [15] proposed to calculate the number of neurons in the hidden layer according to the formula: $N_h = 2iK + 1$, where K is the number of network inputs, $i = 1, 2, 3, \dots$.

Generalization performance varies over time as the network adapts during training. The necessary numbers of hidden neurons approximated in the hidden layer using a multilayer perceptron were found by Trenn [16]. The key points are simplicity, scalability, and adaptability. The number of hidden neurons is $N_h = n + \frac{(n_0 - 1)}{2}$, where n is the number of inputs and n_0 is the number of outputs.

Shibata and Ikeda [17] investigated the effect of learning stability and hidden neurons in neural networks. The simulation results show that the hidden output connection weight becomes small as a number of hidden neurons N_h becomes large. The formula of hidden nodes is $N_h = \sqrt{N_i \cdot N_0}$, where N_i is the number of input neurons and N_0 is the number of output neurons. A tradeoff is formed that if the number of hidden neurons becomes too large, the output of neurons becomes unstable, and if the

number of hidden neurons becomes too small the hidden neuron's output becomes unstable again.

Hunter et al. [18] developed a method in proper neural network architectures. The three networks: MLP, bridged MLP and fully connected cascades network are used. The implemented formula: as follows, $N_h = N + 1$ for the MLP network, $N_h = 2N + 1$ for bridged MLP network, where N is the number of input neurons.

The hidden layer nodes number was determined by using the empirical formula [19] $N_h \leq 2N_i + 1$, where N_h is the maximum number of nodes in the hidden layer and N_i is the number of inputs.

In accordance with [20], the optimal number of hidden nodes is found out by trial – and – error approach.

4 Neural network design method

For preliminary processing of information received from the sensor element of the ultrasonic transducer, we used the Elman network, which is an example of a feedback network. The feedback has a profound impact on the learning capacity and the challenge in the design of a neural network is the fixation of hidden neurons with minimal error. The accuracy of training is determined by the parameters: neural network architecture, number of hidden neurons in hidden layers, activation function, inputs number, and weights updating procedure.

In our case, the control zone of the ultrasonic transducer is displayed in the form of a square matrix of 100×100 elements. The Elman network, which performs image processing, has 100 inputs and 100 outputs. In the process of learning the network, we used 8967 images. The results of calculations of the number of neurons in the hidden layer are shown in Table 1.

Table 1. The number of neurons in the hidden layer of neural networks

Source number	Formula for calculating	Number of neurons in the hidden layer
[14]	$N_h = (N_{in} + \sqrt{N_p}) \cdot L$, for $L = 1$	195
[15]	$N_h = 2iK + 1$, for $i = 1$	201
[16]	$N_h = n + \frac{(n_0 - 1)}{2}$	150
[17]	$N_h = \sqrt{N_i \cdot N_0}$	100
[18]	$N_h = N + 1$	101
	$N_h = 2N + 1$	201
[19]	$N_h \leq 2N_i + 1$	201

The Elman architecture employs a context layer which makes a copy of the hidden layer outputs in the previous time step. The dynamics of the change of hidden state neuron activations in Elman style recurrent networks [13] are given by

$$y_i(t) = f \left[\sum_{k=1}^K v_{ik} y_k(t-1) + \sum_{j=1}^J \omega_{ij} x_j(t-1) \right], \quad (1)$$

where y_k and x_j represent the output of the context state neuron and input neuron, respectively; v_{ik} and ω_{ij} represent their corresponding weights; $f[\cdot]$ is the sigmoid transfer function (1).

Network training was conducted in sequential mode. This mode requires less internal memory for each synaptic link and is preferable for real-time processes. The initialization procedure was carried out without preliminary using the priory information. Herewith, the initial network parameters were set using a generator of uniformly distributed random numbers.

The number of examples for training the network should not be too large, as this can lead to overtraining of the network. In this case, the learning process can end only memorizing learning data. Overtraining network loses the ability to generalize. In accordance with Widrow's rule of thumb, the size of the training set necessary for a good generalization should be of the order of $N = O\left(\frac{W}{\varepsilon_0}\right)$, where W is the total number of free parameters, i.e. synaptic weights and thresholds, ε_0 is the permissible classification error. Let n_{in} be the number of input nodes of the network, and n_h be the number of neurons in the hidden layer. If the product $n_{in} \cdot n_h$ corresponds to the total number of free parameters W network [3], then N must have the order $\frac{n_{in} \cdot n_h}{\varepsilon_0}$.

If we take $n_{in} = 100$, $n_h = 200$ and $\varepsilon_0 = 0.02$, then $N = 10^6$. We used 8967 clean and noisy images for training, but each of them is repeated 100 times within the same epoch. Thus, the actually used volume of the training set is 896700, which is close enough to N .

Hidden neurons of the Elman network, learned by the method of back propagation, play the role of feature detectors. Therefore, it is advisable to use this network as a replicator or identity map. At the same time, the input and output layers of our network have the same size $n_{in} = n_{out} = 100$ neurons. The network is fully connected.

A sigmoidal logistic function is used as the transfer function of neurons of the hidden layer. One of the important advantages for us of this function is the high enough speed of calculating the derivative. The network training time largely depends on this. In the process of network learning, we used the conjugate – gradient method. This was due to the need to increase the sufficiently low rate of convergence of the quickest descent method and avoid computational difficulties caused by operations with the Hessian matrix in Newton's method. The computational complexity of the quasi-

Newton methods is estimated as $O(W^2)$. In contrast, the computational complexity of the conjugate – gradient method is estimated as $O(W)$. Thus, in our case when $W = 2 \cdot 10^4$, conjugate – gradient method is more preferable.

5 Flaw detection process simulation

The image of the defect (delamination) has the form of a continuous black spot, which is formed by the n matrix cells with a color level $L = 5$ adjacent to each other. This spot is surrounded by a layer of gray cells with $L = 4$, the next layer of cells corresponds to $L = 3$ and so on to $L = 1$ (white color). White color characterizes the areas of the solid defect-free material that are completely transparent for an ultrasonic signal.

The defect image (pattern) No.1 is a black square with the size 2×2 cells having $L = 5$. The spot area of the defect is $S_1 = 4$. This is the target image for the first class of images. Moving this image on a matrix of 10×10 cells is carried out line by line. Checkpoint of this pattern, like all subsequent ones, is its lower left cell. In the process of moving, the coordinates of this cell for the first row $j = 1$ change from $i = 1$ to $i = 9$, then the same is repeated for $j = 2$ and so on to $j = 9$. Herewith the black spot is not distorted at the edges of the matrix. The number of images used for network training in this case is $N_{1c} = 81$. Here, the index c means that the displayed images are clean, i.e. not containing noise.

In patterns No.2 and No.3, the left and right upper cells of the black square (No.1) are replaced by gray cells with $L = 4$. The number of training images does not change $N_{2c} = N_{3c} = 81$.

Image No.4 is a black square with the size 3×3 cells having $L = 5$. This is the target image for the second class of images. During the scanning process, the checkpoint of this square was moved along the coordinates $i = 1 \dots 8$, $j = 1 \dots 8$. The number of patterns used for network training in this case is $N_{4c} = 64$. Patterns No.5, 6 and 7 are formed similarly to No.2 and 3. For patterns No. 8 and 9 in the upper layer of the black square, two cells on the left and two on the right are replaced by grey ones.

Image No.10 is a black square with the size 4×4 cells having $L = 5$. This is the target image for the third class of images. The scanning process is similar to that described above. The number of training images is $N_{10c} = 49$. Patterns No.11, 12 and 13 differ from the target pattern No.10 by the alternate replacement of each of the black corner cells with the exception of the checkpoint on the gray. Patterns No.14 and 15 are formed similarly to patterns No.8 and 9 only for the target image No.10.

Image No.16 is a black square of 5×5 cells with $L = 5$. This is the target image for the fifth class of images. The scan coordinates for the checkpoint are $i = 1 \dots 6$, $j = 1 \dots 6$. The number of training images is $N_{16c} = 36$. Patterns No.17, 18 and 19 are formed similarly to patterns No.11, 12 and 13. For images No. 20, 21, 22, 23, 24 and

25 a pair of cells of the outer black layer, shifting along the periphery of the spot, is replaced by gray ones.

The distortion coefficient of the target image is determined by the ratio of the total area of the cells with the changed color to the area of the corresponding black spot (defect image).

The defect images described above were noisy. The absolute noise level P was set by us as follows:

$$P = \sum_{i=1}^M |b_{in} - b_{ic}|, \quad (2)$$

where b_{ic} is the elements of a clear image matrix and b_{in} is the elements of the noisy image matrix.

Noise is additively mixed with a clean image. We used $P = 10, 30, 50, 70, 90$ levels (2). Considering that the full image contains cells with a gray color gradation from $L = 1$ (white color) to $L = 5$ (black color), the question arises about the distribution of the distortion magnitude $\Delta L = 1, 2, 3, 4$ by the number of distorted image elements. We assume that this distribution is Gaussian with a zero mean value and a standard deviation which is equal to 1. Then the probability density is

$$f(\Delta L) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(\Delta L)^2}{2}}. \quad (3)$$

Take into account that $\Delta L = 0$ corresponds to the absence of distortion of the image element, we get the following:

$$P = f(|\Delta L| = 1) + f(|\Delta L| = 2) + f(|\Delta L| = 3) + f(|\Delta L| = 4). \quad (4)$$

The corresponding distribution of ΔL by the number of distorted image elements q is shown in Table 2 (3), (4).

Table 2. The number of distorted image elements

q						$\gamma, \%$
$\Delta L = -1$	$\Delta L = +1$	$\Delta L = -2$	$\Delta L = +2$	$\Delta L = -3$	$\Delta L = +3$	
4	4	1	1	-	-	10
12	12	3	3	-	-	17
20	20	5	5	-	-	22
28	28	7	7	-	-	27
36	36	8	8	1	1	31

The relative value of the noise intensity we calculated according to Table 2.

Herewith

$$\gamma = \frac{\sigma}{L_{\max}} \cdot k, \quad \sigma = \sqrt{\frac{1}{100} \sum_{i=1}^{100} (\Delta L_i)^2}, \quad L_{\max} = 5 \quad (5)$$

The value of the normalizing factor k was determined based on the following. If in all 100 cells of the matrix the black color with $L = 5$ (presence of a defect) is changed to white with $L = 1$ (defect-free material), then $\Delta L_i = 4$ and, in accordance with (5), $\sigma = 4$. Herewith $\frac{\sigma}{L_{\max}} = \frac{4}{5}$. Taking such a change in L in the entire matrix as a 100% error corresponding to $\gamma = 1$, we get $k = 5/4 = 1.25$. In the article, we used the following restrictions: if the total value of the image intensities and the noise in this cell of the matrix exceeds the maximum level $L_{\max} = 5$, then $L = 5$, if the summary value was below $L_{\min} = 1$, then $L = 1$ was kept.

6 Defect image capture

The outputs of the Elman neural network are connected to the inputs of a logical matrix connected in turn to a combinational circuit, which has four outputs by the number of classes of identifiable images. The functional scheme of the intelligent system for defect identification is shown in Figure 1.

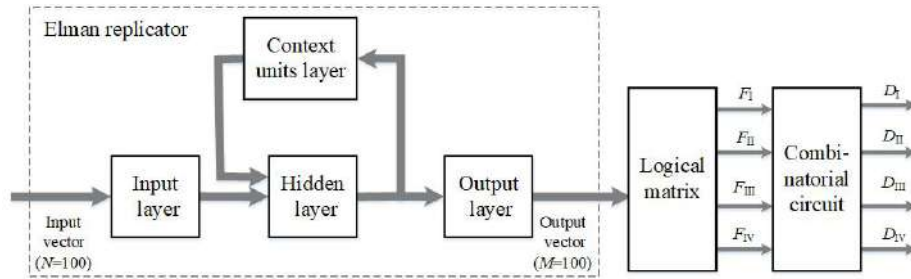


Fig. 1. Intelligent system for defect identification

To estimate the area of defect, we propose the following methodology: the matrix (10×10 elements) represents the control zone of an ultrasonic transducer operating in shadow mode and "sounding", for example, the product wall. The full image of this zone is captured by the sensor unit of the transducer.

During physical scanning of the product surface by the transducer, a noisy image of a defect in the form of a fuzzy dark spot may appear in the lateral part of the matrix. By a local displacement of the transducer, this spot must be moved to the central part of the matrix. The neural network allows you to significantly clear the image of the defect from the noise during its movement to the center of the matrix and make this image more distinct.

In the center of the matrix, there are positioned defect images capture zones, as shown in Figure 2. These zones correspond to images No.1 (2×2 elements), No.4 (3×3 elements), No.10 (4×4 elements), and No.16 (5×5 elements), which are targeted for the respected classes. The logic matrix has 100 inputs, corresponding to the number of outputs of the neural network that perform the function of a replicator. These inputs are numbered in accordance with the pair jk , where j denotes the row number and k denotes the column number of the matrix, as shown in Figure 2.

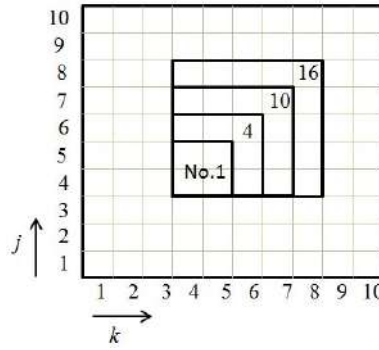


Fig. 2. Capture zones for defect images

The Boolean function F_I , implemented at the same output of the logic matrix, describes images of class I, to which images No. 1, 2, and 3 belong.

$$F_I = y_{44} \cdot y_{45} \cdot (y_{54} \vee y_{55}). \quad (6)$$

Herewith $F_I = 1$, if elements y_{44} and y_{45} have a level of $L = 5$ (black color), which corresponds to a logic-1 level, and at least one of the elements y_{54} or y_{55} has the same level (6).

The Boolean function F_{II} describes images of class II, to which images No. 4, 5, 6, 7, 8, and 9 belong.

$$F_{II} = y_{44} \cdot y_{45} \cdot y_{54} \cdot y_{55} \cdot y_{56} \cdot y_{65} (y_{64} \vee y_{66} \vee y_{46}). \quad (7)$$

Herewith $F_{II} = 1$, if all elements involved in conjunction have a level of $L = 5$, and at least one of the corner elements of the target image (No.4), i.e. one of the disjunctive members in brackets had the same level $L = 5$ (7).

The Boolean function F_{III} describes images of class III, to which images No.10, 11, 12, 13, 14, and 15 belong.

$$F_{III} = y_{44} \cdot y_{45} \cdot y_{46} \cdot y_{54} \cdot y_{55} \cdot y_{56} \cdot y_{57} \cdot y_{64} \cdot y_{65} \cdot y_{66} \cdot y_{67} \cdot y_{75} \cdot y_{76} \cdot (y_{47} \vee y_{74} \vee y_{77}). \quad (8)$$

Disjunctive terms of expression (8) correspond to the corner elements of the target image No. 10.

The Boolean function F_{IV} describes images of class IV, to which images No. 16 – No. 25 belong.

$$F_{IV} = y_{44} \cdot y_{45} \cdot y_{46} \cdot y_{54} \cdot y_{55} \cdot y_{56} \cdot y_{57} \cdot y_{64} \cdot y_{65} \cdot y_{66} \cdot y_{67} \cdot y_{68} \cdot y_{75} \cdot y_{76} \cdot y_{77} \times (y_{74} \cdot y_{84} \cdot y_{85} \vee y_{87} \cdot y_{88} \cdot y_{78} \vee y_{47} \cdot y_{48} \cdot y_{58}). \quad (9)$$

When forming the $F_I - F_{IV}$ functions, we made assumptions about the optional presence of all corner elements of the target images in the defect image capture zones. This is consistent with control practice.

The analysis of expressions (6) – (9) shows that the patterns described by them are sequentially nested one into another, namely F_I is included in F_{II} , F_{II} – in F_{III} , F_{III} – in F_{IV} . Thus, when fixing an image of a senior class, for example No. 16, logical unit will be present not only at the output F_{IV} of the logical matrix, but also at outputs F_{III} , F_{II} , and F_I . The combinational circuit shown in Figure 3 helps to eliminate this effect. The circuit uses inverters and four-input conjunctors and provides separate fixation of each class of image. Herewith, the appearance of a logical unit at one of the outputs $D_I - D_{IV}$ of the combinational circuit indicates the capture of the defect image of the corresponding class.

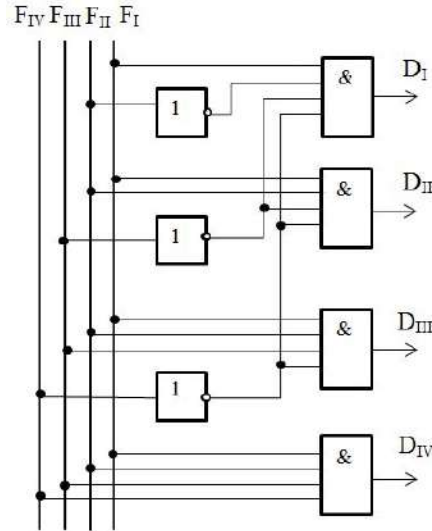


Fig. 3. The combinational circuit

The proposed system of the defect area estimation allows monitoring both in the manual scanning mode with indication of the defect magnitude by sound or light signals and automating the testing process. In the latter case, when scanning the product along parallel paths with a given step, it is easy to automatically register the area.

7 Experiments and results

To simulate the operation of a neural network, we developed the program functioning in the Matlab environment. Network training was carried out sequentially in three stages.

1. First, a clean image No. 1 located in the lower-left corner of the matrix is fed to the network input. This image then moves sequentially along the rows of the matrix within $i = 1 \dots 9$, $j = 1 \dots 9$. Here i and j are the horizontal and vertical coordinates of the matrix, respectively. The specified scan coordinates describe the movement of the lower-left cell of the image.
2. At the second stage, a clean image No.1 ($\gamma = 0$) was again fed to the network input and it was moved along the matrix as described in paragraph 1. Then, image No. 1, additively mixed with noise corresponding to $\gamma = 10\%$, was fed to the network input, and this noisy image was moved along the matrix. We repeated this process sequentially for all remaining noise levels, up to $\gamma = 31\%$. It should be noted that this scanning process was repeated 100 times for each value for the noise intensity.
3. After completing the second stage of training, in order to verify the absence of re-training effect, we repeated paragraph 1.

The network training process described above we fulfilled with each pattern from the training set. Table 3 shows the data on the image classes used for training.

Table 3. Image classes

Class number	Target image and its area S	Distorted images	The area of the undistorted part of the image	Distortion factor $\eta, \%$	Number of training images
I	Image No. 1, $S=4$	No. 2, 3	3	25.0	170100
II	Image No. 4, $S=9$	No. 5, 6, 7	8	11.1	268800
		No. 8, 9	7	22.2	
III	Image No. 10, $S=16$	No. 11, 12, 13	15	6.25	205800
		No. 14, 15	14	12.5	
IV	Image No. 16, $S=25$	No. 17, 18, 19	24	4.0	252000
		No. 20, 21, 22, 23, 24, 25	23	8.0	

Figure 4 shows a graph of the dependence of the training time, relative to one pattern α on the area of the black spot S , reflecting the area of the defect. But at the same time, α characterized training on clean images, without noise. Training on noisy images is carried out after the training on clean images. And here the training time relative to one pattern β is significantly less. So, for images of the first type,

$\theta_1 = \frac{\beta}{\alpha} = 0.045$. For images of the second, third and fourth types, we have: $\theta_2 = 0.027$, $\theta_3 = 0.032$, and $\theta_4 = 0.021$.

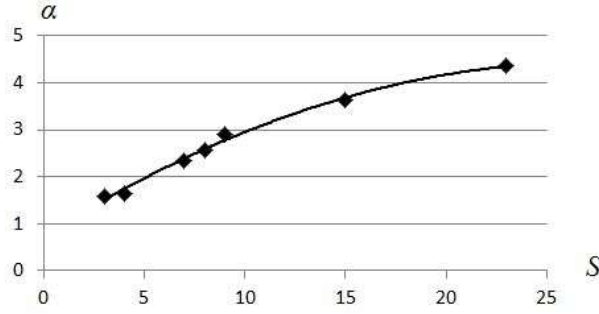


Fig. 4. Dependence of the training time for one pattern α on the area of the black spot S

Thus, we can state that the network training time at the second stage, assigned to one noisy image, is ten times less than the corresponding time spent at the first stage when training the network at the clean image.

The phenomenon of retraining the network was not observed.

Network testing was carried out by moving images from the set described above (No. 1 – No. 25) with noise levels from Table 2 along with the matrix. Testing results showed the following.

If testing was carried out after a full cycle of network training on patterns of one class using then a noisy image with $\gamma = 31\%$ from the same class, then the test error did not exceed 3%. This result is valid for all four classes.

When combining images of classes I and II, i.e. with the expansion of the defect spot area from $S = 3$ to $S = 9$, the testing error on noise patterns ($\gamma = 31\%$) increased to 8%. When combining all the images of I-IV classes in a single training set, the maximum testing error increased to 12%. And in the first and second cases, the parameter β increased significantly, its value became comparable with α . The reduction of the noise level during testing significantly decreases this error. So, with a reduction in γ from 31 to 27%, the test error decreases by 40-50%, from 31 to 22% – by 60-70% and so on.

The above results were obtained by using the Elman neural network as a replicator with $n_{in} = n_{out} = 100$, $n_h = 200$ and the number of layers $l = 1$. An increase in the number of neurons in the hidden layer to 400 on average by 25% reduces the testing error, however, the network training time increased by 30%. A network with two hidden layers, while reducing the test error by 20%, trained 4 times longer than a network with one hidden layer. Training a network with three hidden layers took 14 times more time than the single-layer network we used.

8 Conclusion

We investigated the Elman neural network as an identity map (replicator) for the task of identifying and estimating the area of a delamination in composite materials in the process of ultrasonic testing. Twenty-five images were used as training patterns. They were divided into four classes, according to the size of their defect area. Deterministic distortions caused by the nature of the testing object were introduced into the images of each class. Each image was moved step by step on a square matrix consisting of 100 cells, which represents the registration area of the ultrasound transducer. Images were additively mixed with white Gaussian noise, the intensity of which varied from 10 to 31%. At each scan point, the noise was superimposed 100 times.

The process of training the network began with presenting it with the clean (without any noise) patterns. In the process of training the network, it was found that the training time with clean patterns, assigned to one pattern fixed in a given position inside the matrix, increases with increasing the pattern area. Network training using noisy patterns was carried out after training with clean patterns. In this case, the training time, defined in the same way, turned out to be ten times shorter.

Network testing has provided the following. If testing was carried out after a full cycle of network training on patterns of one class using a test pattern with a noise level of 31%, then the test error did not exceed 3% for any class. When combining images of classes I and II, i.e. with the expansion of the defect spot area from $S = 3$ to $S = 9$, the testing error increased to 8%. When combining the images of all four classes in a single training set, i.e. with the expansion of the defect spot area from $S = 3$ to $S = 25$, the maximum testing error increased to 12%. The reduction of the noise level during testing significantly decreases this error. So, with a reduction in γ from 31 to 27%, the test error decreases by 40-50%, from 31 to 22% – by 60-70% and so on.

These results were obtained by modeling the Elman's neural network with the number of input and output neurons $n_{in} = n_{out} = 100$ and with the number of neurons in one hidden layer $n_h = 200$. This network structure turned out to be optimal by the criterion of error testing/training time.

To estimate the delamination area, four nested defect spot capture zones were implemented in the central part of the matrix. For this purpose, a cascade-connected logical matrix and the combinational circuit having four outputs by the number of classes of identifiable images are connected to the output of the neural network.

Estimation of the defect area serves as the basic for determining the residual strength of a given unit of a product and making a decision on the possibility of its further operation.

References

1. Composite Materials Handbook – 17: Polymer Matrix Composites, vol. 1-3. SAE International (2018)

2. Robinson, D.A.: Signal processing by neural networks in the control of eye movement. Computational Neuroscience Symposium, Indiana University – Purdue University at Indianapolis, pp. 73-78 (1992)
3. Naykin, S.: Neural Networks. A Comprehensive Foundation, second edition. Prentice Hall, New Jersey (2008)
4. Khandetskyi, V., Antonyuk, I.: Signal processing in defect detection using back-propagation neural networks. NDT&E International, 35, 483-488 (2002)
5. Chawla, C., Panwar, D., Anand, G.S., Bhatia, M.P.S.: Classification of computer generated images from photographic images using convolutional neural networks. International Journal of Computer and Information Engineering, vol. 12, No. 10, pp. 822-827 (2018)
6. Moreno-Garcia, C.F., Elyan, E., Jayne, C.: New trends on digitalization of complex engineering drawings. Neural Computing and Applications, June, pp. 1-18 (2018)
7. Ablameyko, A.V., Uchida, S.: Recognition of engineering drawing entities: review of approaches. International Journal Image Graph 07(04): 709-733 (2007)
8. Elman, J.L.: Finding structure in time. Cognitive Science, vol. 14, pp. 179-211 (1990)
9. Elman, J.L., Bates, E.A., Johnson, M.H., Karmiloff-Smith, A., Parisi, D., Plunkett, K.: Rethinking Innateness: A Connectionist Perspective on Development. Cambridge, MA: MIT Press (1996)
10. Ardalani-Farsa, M., Zolfaghari, S.: Chaotic time series prediction with residual analysis method using hybrid Elman-NARH neural networks. Neurocomputing, vol. 73, Iss. 13-15, Aug., pp. 2540-2553 (2010)
11. Ren, G., Cao, Y., Wen, S., Huang, T., Zeng, Z.: A modified Elman neural network with a new learning rate scheme. Neurocomputing, vol. 286, Iss. 19, April, pp. 11-18 (2018)
12. Koskela, T., Lehtokangas, M., Saarinen, J., Kaski, K.: Time series prediction with multilayer perceptron, FIR and Elman neural network. Tampere university of technology, Electronics laboratory, FIN-33101, Tampere, Finland (2017)
13. Sheela, K.G., Deepa, S.N.: Review of methods to fix number of hidden neurons in neural networks. Mathematical Problems in Engineering, vol. 12, article ID 425740, pp. 1-11 (2013)
14. Jinchuan, K., Xinzhe, L.: Empirical analysis of optional hidden neurons in neural network modeling for stock prediction. In Proceedings of the Pacific-Asia Workshop on Computational Intelligence and Industrial Application, vol.2, pp. 828-832, Dec. (2008)
15. Kallan, R.: Main Conceptions of Neural Networks. Williams Publ., New York (2001)
16. Trenn, S.: Multilayer perceptrons: approximation order and necessary number of hidden units. IEEE Transactions on Neural Networks, vol. 19, No 5, pp. 836-844 (2008)
17. Shibata, K., Ikeda, Y.: Effect of number of hidden neurons on learning in large scale layered neural networks. In Proceedings of the ICCAS-SICE International Joint Conference (ICCAS-SICE'09), pp. 5008-5013, Aug. (2009)
18. Hunter, D., Yu, H., Pukish, M.S., Kolbusz, J., Wilamowski, B.M.: Selection of proper neural network sizes and architectures: a comparative study. IEEE Transactions on Industrial Informatics, vol.8, No 2, pp. 228-240 (2012)
19. Gavin, I., Holder, R., Graeme, C.: Input detection for neural network models in water resources applications, part 2. Case study: forecasting salinity in a river. J. Hydrol 301 (1-4): 93-107 (2011)
20. Devi, R., Rani, B.S., Prakash, V.: Role of hidden neurons in an Elman recurrent neural network in classification of cavitation signals. International Journal of Computer Applications, vol.37, no 7, pp. 9-13, (2012)

Decision Support System for the Machine Learning Methods Selection in Big Data Mining

Nicolay Rudnichenko¹ [0000-0002-7343-8076], Vladimir Vychuzhanin¹ [0000-0002-5244-5808],
Igor Petrov² [0000-0002-8740-6198], Denis Shibaev³ [0000-0002-3260-5843]

¹Odessa National Polytechnic University, Odessa, Ukraine
nickolay.rud@gmail.com, vint532@yandex.ua

²National University "Odessa Maritime Academy", Odessa, Ukraine
firmness@list.ru

³Odessa National Maritime University, Odessa, Ukraine
denscreamer@gmail.com

Abstract. This article focuses on the aspects of the decision support system for the machine learning methods selection in big data mining development. The paper includes the results of the analysis of the problem of intellectual processing and analysis of big data. It describes proposal ways of using metadata as a basis for the formation of an analytical rating for evaluating machine learning methods. The paper presents the results of designing and using a decision support system for evaluating machine learning methods for solving data mining problems. The developed decision support system allows us to reduce the analysis time of suitable methods for solving machine learning problems by a data science analyst, taking into account the specifics of the input data arrays, their volumes, structure and other metadata.

Keywords: decision support system, big data, data mining, data science, machine learning, data analysis

1 Introduction

Currently, there is a steady trend of regular growth in the volume of data collected in the various business organizations of production, operational and research activities processes [1]. The sources of the such big data volumes (Big Data) appearance are often customers various behavioral factors, the frequency and size of payments for the services or goods, parameters and characteristics of installed technical equipment, medical indicators for diagnosing human health and others [2-4].

Due to the statistical visibility and representativeness, the value of such data lies in the possibility of using it to search for hidden and unobvious relationships between individual factors (attributes) and target actions of clients to adjust and formulate business development strategies [5].

In fact, the implementation of such tasks becomes possible based on the use of data mining methods in order to extract new knowledge by forming and proving hypothe-

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

ses about significant relationships between the individual attributes of data samples [6-8].

Thus, the constant companies need to ensure a sufficient level of quality in the provision of goods and services, due to the high level of competition in the organization's business targets requires usage of the data science, data mining methods and technologies in key business processes [9]. For this purpose, machine learning (ML) methods can be used to build various regression and predictive models for the business objectives.

2 Description of Problem

The concepts which are laid down in this approach allow data mining specialists to conduct a comprehensive phased Big Data analysis and processing, sequentially implementing the necessary processes, including distributed structuring of heterogeneous data, their consolidation, aggregation, cleaning and pre-processing, eliminating anomalies, omissions, errors side values [10-12].

However, all these processes are time-consuming for analysts on the experimental selection and mathematical model selection with the corresponding hyperparameters and quality assessment metrics. One of the key factors to provide a successful and prompt solution of a target problem by an analyst is his experience in building ML models, business problems vision depth and the software tools, technologies and libraries knowledge [13,14].

This can lead to subjectivity of the analysis results and affect the accuracy and generalizing ability of the formed ML models. At the same time, the computational costs of the computer equipment used in the data analysis process are also significant, which affects the total cost of developing ML models [15]

An additional complication is the correlation, evaluation, and required ML method or their combination selection processes for the effective solution of the data mining posed problem (with the achievement of the model created by a sufficient accuracy, adequacy level and generalizing ability) without lengthy computational experiments.

This problem becomes especially relevant in cases where the Big Data sample size exceeds the permissible amount of disk space in the used data warehouses [16-19]. If there are limitations in the throughput capacity Big Data transmission in serial or parallel mode in a local or global network, then effective analysis of such data becomes difficult [20-24]. A possible solution is to compress, transform or structure the data with the extraction of quantitative and qualitative meta-information from them [25-28].

In this regard, the urgent and relevant task is to automate the processes of selecting suitable ML methods based on the input Big Data volumes analysis, their various statistical and probabilistic characteristics, taking into account the subject area specifics and the ML problem type. This can be done by developing a decision support system (DSS) with a number of intellectual functions for the formation and accounting of meta-information about data attributes, their structure, level of generalization and significance.

Currently, there are various analytical systems on the market for comparing and comparing various ML methods for solving classification and regression problems on Big Data, however, their functionality is limited and does not allow to fully take into account the nature of the input data, their volume and subject area of analysis. DSS that can be adapted for such ML tasks are Wolfram Mathematica, EIDOS, Expert Choice and ViEA. But, there is no possibility of a flexible system configuration for the user's needs, filling with new functionality and updating dependencies is not performed regularly. This makes it impossible to use such solutions to obtain reliable and reliable data mining results [29]. The purpose of this article is to develop a DSS project for choosing ML methods to solve data mining tasks on user-specified data sets based on their structure and volume to reduce time spent on detailed experimental calculations.

3 Decision support system development

3.1 System concept

The functioning of the proposed system is carried out in several stages:

- data import;
- formation of meta-information from the downloaded data;
- Data Mining tasks type and methods selection;
- specification of criteria and metrics for ML model quality assessing;
- obtained ML models creation and evaluation;
- ML models obtained results visualization and the issuance of a methods ranked list which provides the highest quality solution to the problem.

The process of DSS in general is shown in fig. 1.

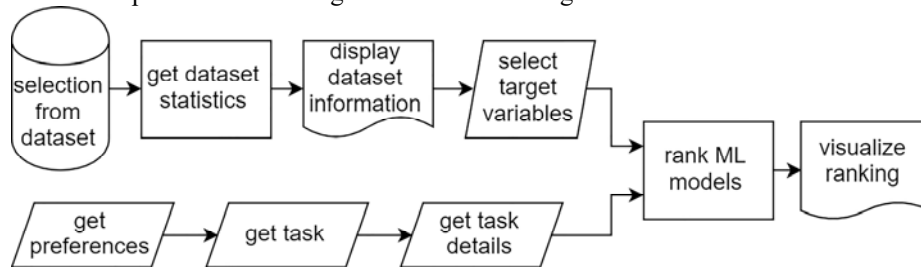


Fig. 1. General system operation

DSS supports the ability to import various training datasets and test ML models in *.csv, *.xls and *.json formats. If necessary, the processing of structured relational data supports the integration of SQL database management systems MySQL and Postgresql, as well as NoSQL MongoDB. Data loading is performed in multi-threaded

mode with blocks from 2 to 128 MB in each, which allows us to distribute computing processes and scale the system in the future.

After importing data, the system gives a message about the success of the operations performed and provides a brief meta-information, which includes: the total number of records in the loaded data set, signs number, signs and records number ratio, data amount. The user can set the block size for further analysis depending on the displayed data.

Since the size of the analyzed data volume in data mining tasks can reach large values, this can reduce the efficiency of computing operations in the analysis process. Therefore, the user is invited to choose one of the options: make a selection of a given size from the imported set from 0 to 100% or select a value randomly with the possibility of stratified sampling by a specific column to maintain proportions. This allows us to reduce the amount of RAM used to store data and speed up the computing analysis operations process. To ensure the analysis process flexibility the user has the opportunity to select the necessary features by disabling or removing unnecessary from the system, as well as specify one or more target output variables.

For each column from the dataset table, the selection indicates whether it is numeric or categorical. For columns with numerical values, the following statistical information is displayed: range of values, standard distribution, average value, median, asymmetry coefficient, kurtosis coefficient, chi-square test of the normal distribution test, Pearson correlation with the target variable (if one was specified) and its confidence. For columns with categorical values, the following is displayed: number of categories, relative mode frequency and Gini coefficient.

The system allows us to take into account the specifics of the data mining task being solved and set the priority of assessing the ML model quality based on the operation speed or the accuracy obtained. This is one of the target criteria for forming a ranked methods list. DSS implements support for 2 types of data mining tasks: classification and regression. Depending on the type of task, the user must specify the necessary metric for evaluating ML models.

For classification tasks, the following metrics are supported: accuracy (share of correct answers), recall (share of found objects of a positive class), precision (share of truly positive from classified positive objects), F1 (harmonious average), AUC. For regression tasks, metrics are supported: mean squared error (MSE), and mean absolute error (MAE). After the data import is completed, the analysis process is started in the background. The analysis consists in the phased creation of model instances with hyperparameter values in the ranges specified by the user (or in the default range specified for each ML method) and their assessment by the selected metrics. It is possible to view detailed logs of obtained metric estimates at individual iterations and epochs of training and testing ML models in *.txt format.

After the system performs data analysis procedures, the results are displayed in the form of a summary report on the most suitable ML methods for the criteria selected by users, which automatically saves to the *.xls file and is shown at the system's log screen. The structure of the general report has a tabular form and contains: the name of the method, the level of its adequacy for the selected data set and task (in relative units from 0 to 100), the approximate predicted RAM amount needed for training and

testing ML method on the sample, and the predicted time spent on carrying out computing processes.

User can select the criteria for ordering methods in a window for displaying analysis results. In particular, it is possible to sort the methods by individual metrics, speed, accuracy or ease of results interpretation (which depends on the data amount).

For a simpler and more understandable the analysis results interpretation DSS supports visualization using a bar chart, where each column reflects the quality of an individual model in the rating form for a given metric. It is possible to save the constructed graphic visualizations in *.png format.

Based on the developed concept of the system's functioning, its design and software implementation can be carried out next.

3.2 DSS project implementation

To display the relationship between users and the system, a diagram of use cases was compiled (fig. 2). The main unary scenarios of user interaction with the system are: selecting a data set, viewing data statistics, updating the task, viewing models rating. The server side provides support of optional data set selected part retrieval, imported data set statistics generation, computational processes for compiling the rating and its graphical visualization, saving results to a file.

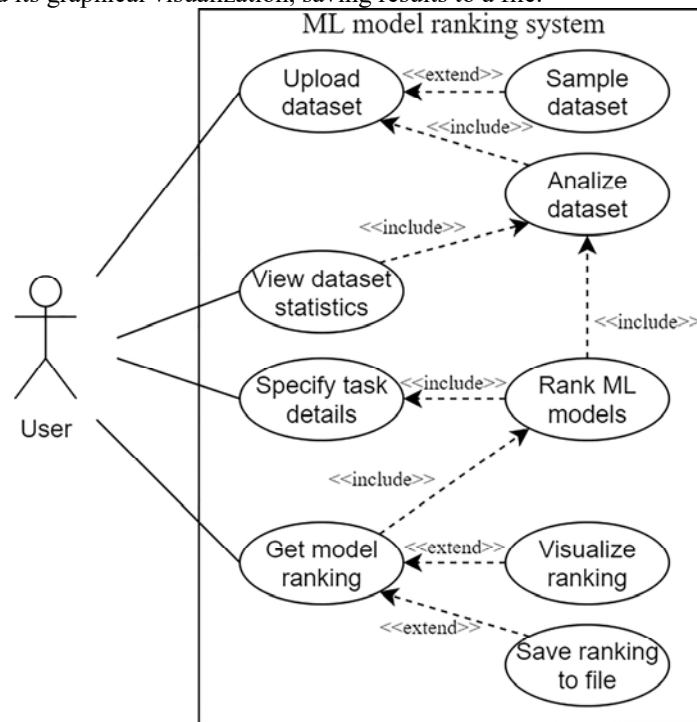


Fig. 2. DSS use cases diagram

The system is divided into some components that perform various parts of the task. These components, as well as the data exchange between them, are shown in fig.3. In this diagram, the `UIComponent` component is responsible for user interaction with the system, providing controls, as well as reflecting the selected table, its statistics and the ranking list generated. The `TableManager` component is responsible for loading and storing tabular data, it receives the path and data loading mode from the interface, after which it provides the other components with a dataset table.

The `TableAnalyzer` component is responsible for removing statistics from a table. After extracting statistical data and obtaining additional information from the user, the `ModelRanker` component determines which of the ML models are most likely to be suitable for this task. After that, the generated rating of models is displayed to the user, and is also saved in the file system by the `RankingSaver` component.

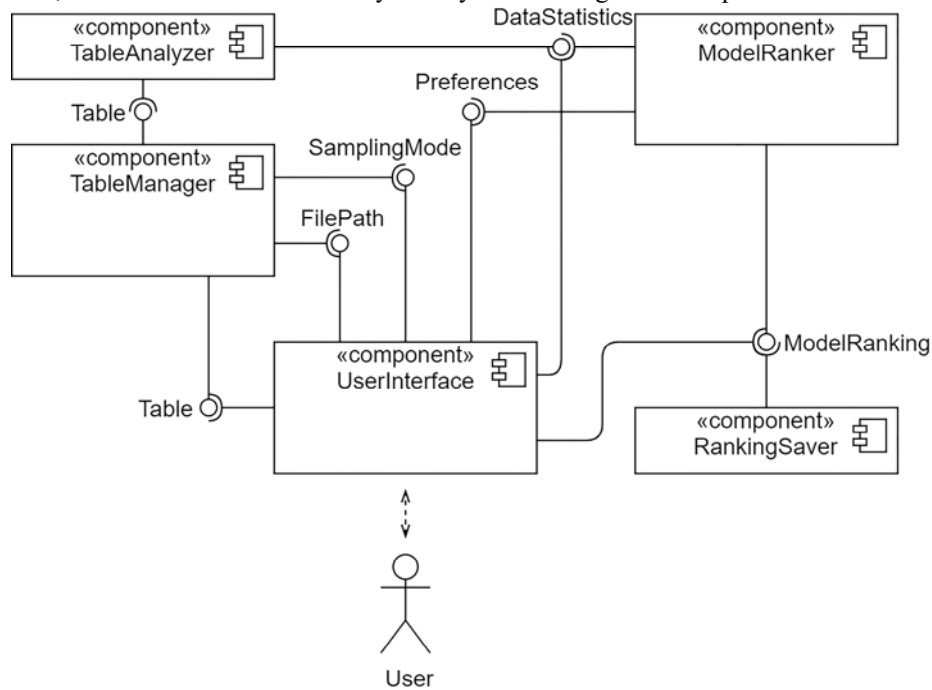


Fig. 3. DSS component diagram

To formalize the key functional processes that are carried out in the system and their relationship, a sequence diagram of actions has been drawn up, which reflects the basic call operations (fig. 4). The user through the interface sets the path to the data in the `TableManager`, which returns meta-information for the selected dataset table. After specifying the sampling mode and confirming the selected set, `TableAnalyzer` extracts statistical information, showing the result to the user. Next, user specifies the necessary ML methods, hyperparameters, model quality assessment metrics and selects the target and input columns (stored in the `Preferences` object), based on

which the required metadata is determined and computational operations are performed to evaluate the method.

ModelRanker object sequentially downloads the resulting ML models to generate a consolidated rating by storing a detailed calculation log in a file and displaying the output data in short form to the user.

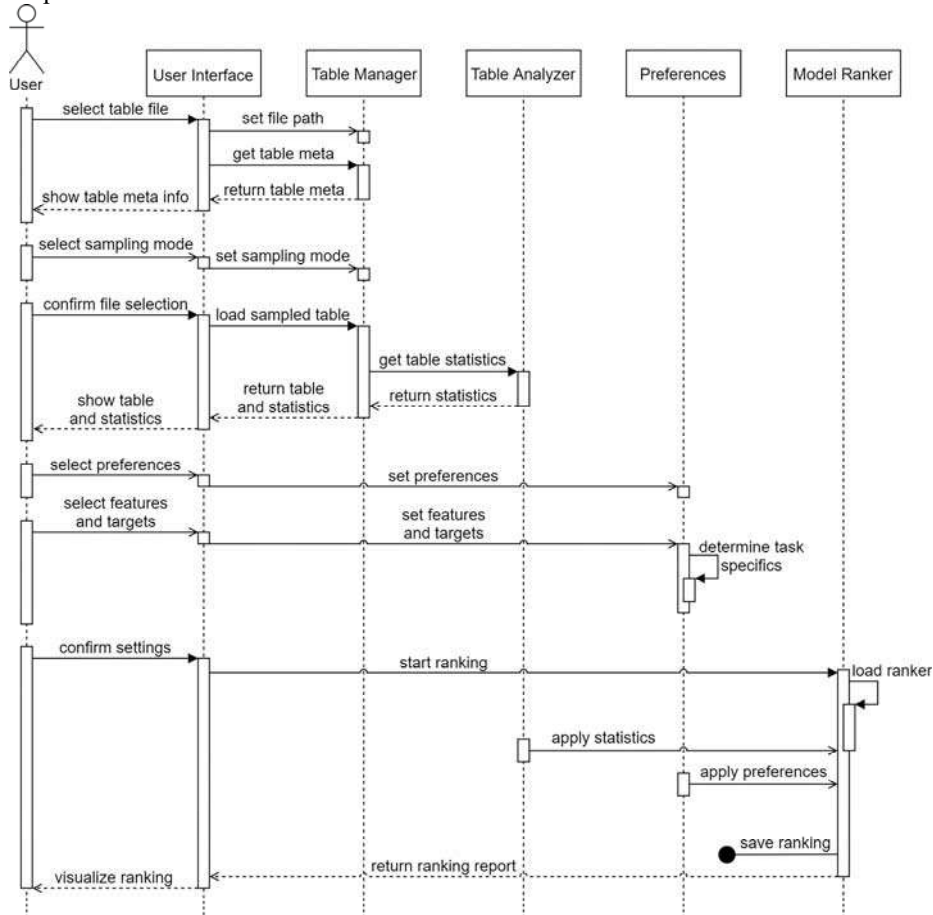


Fig. 4. DSS user sequence diagram

To visualize the system software implementation structure, a class diagram was created, a fragment of which is shown in fig. 5. This diagram shows the structural relationships between the various classes in the DSS. The user interface object has one-way associations with the table manager object (which in turn is associated with the table analysis object), the report saving object and the rating creating object. An analyzing table object creates an object containing metadata and statistical data for each column, which is a necessary process for rating models. After the user enters all the necessary options for data analysis, UserInterface creates an object of the preferences class.

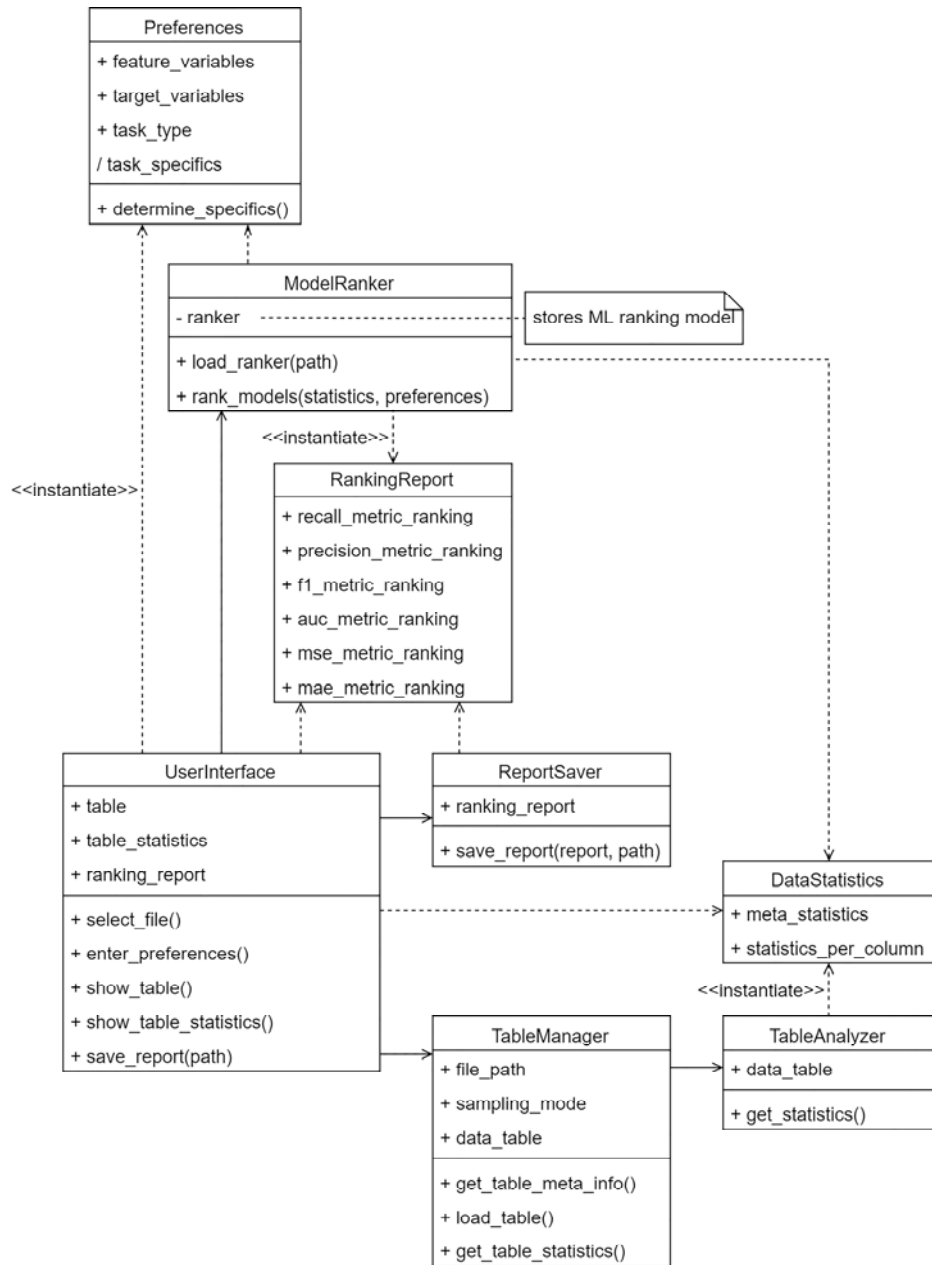


Fig. 5. DSS main classes diagram fragment

The functionality of this class is used as the basis for creating a rating. After creating a rating of models according to user-specified metrics, this object is used by the user interface to display and the ReportSaver class to save it to the file system. A

special method has been created in the ModelRanker class to load the ML model used to compile the rating along a given path. This allows us to simplify the process of choosing a suitable model for ranking in the ranked list form by quickly changing the model estimation algorithm used, which will allow analysts to train and use the created models in the future.

After user receiving the rating it's possible to estimate the approximate time of the ML model training process for each of the methods involved and the RAM amount. The approximate time of creating a model for non-iterative learning algorithms is estimated based on extrapolating the obtained model training duration on a given records number and attributes on an active workstation by the records number ratio and attributes with the data set used. The approximate model training time is calculated by multiplying the measured training time by the computational complexity with the substituted ratios. The training speed can largely depend on the user's workstation hardware therefore, it is not calculated accurately, the error can be up to 15-20%. The approximate spatial complexity is calculated in a similar way using the known memory usage and the table size and features ratio.

In the creating DSS process were used several technologies such as: Python 3.7 programming language, Pandas library for analyzing and manipulating data, library for working with ML sklearn models, XGBoost library, libraries for performing mathematical operations and scientific calculations NumPy and SciPy. PyCharm is used as an IDE. The matplotlib library was used to create the graph visualization interface.

The system user interface is implemented using block layout in the web application form. It includes 4 tabs with graphical components for managing data import processes (clicking on the corresponding button and selecting the desired data set in the dialog box), displaying data and statistical information in a tabular form, selecting and setting model parameters in text fields and drop-down lists and viewing the results of operations in the graphs and charts forms using Matplotlib and Chart.js libraries.

4 Experiments and results analysis

During the DSS creation several algorithm variants were considered for ranking the models, taking into account the possibilities of their training and the implementation complexity, due to the need in each sign of the dataset and metadata. Analysis of literary sources [15-19] let us to perform the following algorithm which based on the use of an artificial neural network (NN) multilayer perceptron:

1. For each attribute, the values of its statistics, metadata, and task type are used as input for NN. The NN model provides values that reflect the relative ranking of ML models by metrics for each attribute.
2. The generated ratings for each characteristic are taken into account by calculating the arithmetic average of all issued ratings.
3. The obtained models relative ratings values by metrics are reduced to a range from 0 to 1 so that 0 are not suitable models (with a low metrics rating), and 1 is the

most suitable. If a certain metric isn't compatible with the task type, it will not be displayed to the user.

This algorithm provides the possibility of combined accounting for the different task type, meta-information about data and statistical data for each attribute in the rating. If a single feature doesn't affect the quality of the compared models, NN is able to produce values close to 0 for the rating of such a feature that will not introduce significant errors and will not affect the final analysis result. Since all operations in this algorithm can be differentiated, it is possible to train the used ML model by the stochastic gradient descent method.

The model formation for ranking is based on the learning algorithm with a teacher. Metadata, task type, and attribute statistics taken from imported data sets are used as input for training.

As the target variables, model ratings obtained by experimental models comparison using various metrics on data sets are used. Ratings are converted to a range from 0 to 1 so that the value higher, the algorithm is better in evaluating relative to others. The loss function ensures the correspondence of the generated rating for each ML experimental rating model.

The developed DSS has been tested on several experimental model comparisons. The methods of linear regression, decision tree, random forest, support vector method (for classification and regression), xgboost, logistic regression, and naive Bayes classifier were evaluated.

The comparable metrics used to assess the models quality were AUC, F1, precision and recall for classification and MSE with MAE for regression. When creating models, the values of their hyperparameters are set by default in accordance with the predefined values by the library, except for random forest and xgboost algorithms, in which the number of trees was set to 64.

For experimental comparison, a cross-validation method was used on the basis of dividing the sample into 10 equal parts, with a stratified breakdown for the classification problem.

In the study of the system operation the Diabetes dataset was used for the regression task. The data set has 442 records and 10 attributes with real values. Signs indicate age, gender, weight to height, average blood pressure, and 6 blood counts. The target variable is the diabetes degree.

At the stage of data cleaning, objects in which at least one of the attributes departed by more than 3 standard deviations for all values of this attribute were deleted from the dataset.

After cleaning, 97.3% of the original objects remained. Due to the fact that during the experimental comparison some of the models used require input data normalization, all features were normalized to the range from -1 to 1, and the target variable to the range from 0 to 1.

After testing the models using the cross-validation method, regression model metrics (MSE, MAE) were obtained (Table 1). Results are sorted by MSE metric (best to worst).

Table 1. Experimental model metrics

<i>Model Name</i>	<i>MSE</i>	<i>MAE</i>
Linear Regression	0.030	0.141
Random Forest	0.033	0.149
Support Vector Regressor	0.034	0.144
XGBoost	0.034	0.149
Decision Tree	0.062	0.193

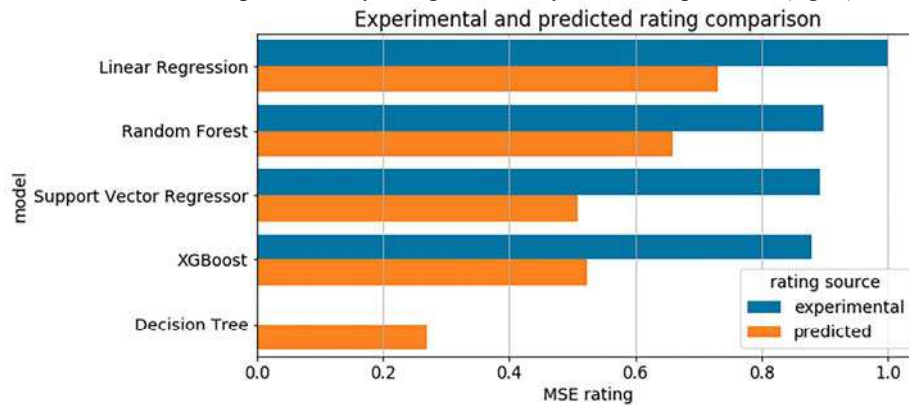
The experimentally obtained metrics were converted to the models ranking. A value 1 is the best result, 0 is the worst (Table 2). The results for experimental part were obtained using the developed software script in Anaconda and Jupiter notebook environment in usual "step by step" mode. The results for DSS part were obtained by importing dataset with default model's hyperparameter setting.

Further, using the collected statistics on the characteristics, metadata and a data set, the system generated a ML models rating for this task. As can be seen from the obtained results, the order of the best models in the rating issued by the system approximately coincides with the experimental rating.

Table 2. Experimental and analytical model ratings

<i>Model Name</i>	<i>Experimental</i>		<i>DSS</i>	
	MSE	MAE	MSE	MAE
Linear Regression	1.000	1.000	0.731	0.728
Random Forest	0.897	0.856	0.653	0.731
Support Vector Regressor	0.892	0.938	0.682	0.383
XGBoost	0.880	0.842	0.657	0.580
Decision Tree	0.000	0.000	0.368	0.368

For ease of comparison, a diagram was presented showing the comparative rating of models obtained experimentally and generated by the developed DSS (fig. 6).

**Fig. 6.** Comparison of the experimental and rated by the DSS

Based on this rating, the system was able to determine the best model (according to experimental comparison), but with less confidence. It is also noticeable that the evaluation of the support vector mechanism and XGBoost was put in the wrong order. Evaluation of a bad model, decision trees, was put below others, but far from experimental values. The reason for the decision tree low rating could be that in regression tasks the tree outputs only discrete values, which will increase the error of MSE and MAE.

5 Conclusion

Summarizing the comparisons of the estimated and experimental model ratings, we can say that DSS is capable of fairly accurate ML models estimation depending on the type of task, metadata, and data set statistical information, its level of accuracy reaches 70-75%. Errors sources in the ML model ratings are: a relatively small training set and the test data set similarity with the training one, which negatively affects the model's generalizing ability to analyze data dependencies.

Since the recommended ML methods ranked list issued by the DSS is not always accurate, in tasks with critical requirements for the model's reliability, completeness and accuracy it is advisable to conduct additional exploratory analysis on the selected fragment of the data set according to the first 3 methods issued by the system.

Due to the fact that the data set table's size may be larger than free memory, it was important to make possible records selection without loading the file completely into memory. The implementation of this functionality depends on which sampling mode is selected and in what format the table is saved. With stratified random sampling, memory usage is higher than in other modes, since it is necessary to calculate the different values number in the selected column. Therefore, there may be cases when some data are missing in the table; therefore, this problem must be taken into account when collecting statistics from the table.

When compiling statistics for some data sets, the missing fragments in the samples were ignored, which also introduced some errors in the data analysis process. The ability to fill in the missing data using various algorithms is one of the options for the future system development and upgrading.

Reference

1. Rudnichenko, N., Vychuzhanin, V., Shybaieva, N., Shybaiev, D., Otradskaia, T., Petrov, I.: The use of machine learning methods to automate the classification of text data arrays large amounts. Information management systems and technologies. Problems and solutions. Ecology, Odessa, pp.31-46 (2019).
2. Sandryhaila, A., Moura, J.M.: Big data analysis with signal processing on graphs: representation and processing of massive data sets with irregular structure. IEEE-Signal Process. vol. 31(5), pp. 80-90 (2014).
3. Dietrich, D., Heller, B., Yang, B.: Data Science & Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data. Wiley, Hoboken (2015).

4. Rudnichenko, N., Vychuzhanin, V., Shybaieva, N., Shybaiev, D.: Big data intellectual analysis in the diagnosis of the transportation systems technical condition. Systems and means of transport. Problems of operation and diagnostics. KSMA, Kherson, pp.57-69 (2019).
5. Jordan, M.I., Mitchell, T.M.: Machine learning: trends, perspectives, and prospects. Science, vol. 349(6245), pp. 255–260 (2015).
6. Chen, H., Chiang, R.H., Storey, V.C.: Business intelligence and analytics: From big data to big impact. MIS Q. vol 4, pp. 1165–1188 (2012).
7. Vychuzhanin, V.V., Shibaev, D.S., Boyko, V.D., Shibaeva, N.O., Rudnichenko N.D.: Big data mapping in the geopositioning systems for fishing industry. International Scientific and Technical Conference on Computer Sciences and Information Technologies. pp.28-31 (2017).
8. Phillips-Wren, G.: Ai Tools in Decision Making Support Systems: A Review. International Journal on Artificial Intelligence Tools. vol.21 (2012).
9. Rudnichenko N.D., Vychuzhanin, V.V., Shybaiev, D.S.: The use of cluster data analysis to highlight measures of factors affecting the performance similarity of complex technical systems. Informatics and mathematical methods in simulation. vol. 3, pp.214-219 (2017)
10. Zeng, X., Lu, J.: Decision Support Systems with Uncertainties in Big Data Environments. Knowledge-Based Systems. vol.143 (2018).
11. Rudnichenko, M.D., Gezha, N.I., Belyaev, K.O., Kuzmin, A.D.: Performance analysis of machine learning model ensembles. In III All-Ukrainian scientific-practical conference of young scientists, students and cadets “Information protection in information and communication systems”. Lviv. pp.259-260 (2019).
12. Gomez, D., Rojas, A.: An empirical overview of the no free lunch theorem and its effect on real-world machine learning classification (2015).
13. Han, J., Kamber, M., Pei, J.: Data mining: concepts and techniques, Morgan Kaufmann (2011).
14. Padhy, N., Mishra P., Panigrahi, R.: The survey of data mining applications (2012).
15. Sumiran, K.: An overview of data mining techniques and their application in industrial engineering (2018).
16. Ramageri, B. M.: Data mining techniques and applications (2010).
17. Engels, C., Bratsas, C., Koupidis, K., Musyaffa, F.: Requirements for statistical analytics and data mining (2016).
18. Shalev-Shwartz, S., Ben-David, S.: Understanding machine learning: from theory to algorithms. Cambridge University Press (2014).
19. Jordan, M. I., Mitchell, T. M.: Machine learning: trends, perspectives, and prospects (2015).
20. Ayon, D.: Machine learning algorithms: A Review (2016).
21. Chugreev, V.L.: Decision support systems using machine learning methods and predictive analytics. Problems of economic growth and sustainable development of the Vologda territories, pp.79-83 (2016).
22. Sinitsyn, E.V., Tolmachev, A.V.: Model of the system of decision support in the financial markets for enterprises on the basis of probabilistic analysis and machine learning. Herald UFU. Economics and Management Series. vol.18. pp.378-393 (2019).
23. Savenkov, P.A.: Using machine learning methods and algorithms in management decision support systems. Journal of Science and Education, vol. 55, pp.23-25 (2019).
24. Korneev, S.: Decision support systems in business. Networks and business. vol.25, pp.102-110 (2005).

25. Chichirin, E.N.: Intelligent methods in simulation of decision making processes. Computer tools, networks and systems, vol. 17, pp.86-94 (2018).
26. Kogalovsky, M.P.: Metadata, their properties, functions, classification and presentation tools. In the 14th All-Russian Scientific Conference "Electronic libraries: perspective methods and technologies, electronic collections". Yaroslavl (2012).
27. Kogalovsky M.P.: Metadata in computer systems. Programming, MAIK Science "Interperiodica", vol.39, pp.28-46 (2013).
28. Skvortsov, N. A., Bryukhov, D. O., Kalinichenko, L. A., Kovalev, D., Stupnikov. S. A.: Metadata on scientific methods to ensure their reuse and reproducibility of results.. RCDL. Yaroslavl (2014).
29. Kalinichenko, L. A., Stupnikov, S. A., Vovchenko, A. E., Kovalev, D. A.: Conceptual declarative problem specification and solving in data intensive domains. Informatics and Applications, vol. 7. (2013).

Neural Network Forecasting of Earth Globe Seismic Activity Level

Vadym Tiutiunyk¹[0000-0001-5394-6367], Leonid Chernogor²[0000-0001-5777-2392],
Olha Tiutiunyk³[0000-0002-3330-8920], Tural Agazade¹[0000-0002-0165-1118]

¹National University of Civil Defence of Ukraine,
Chernyshevska Str., 94, Kharkiv, 61023, Ukraine
tutunik_v@ukr.net, agazade.tural.2019@gmail.com

²V.N. Karazin Kharkiv National University,
Svobody Sq., 4, Kharkiv, 61022, Ukraine
leonid.f.chernogor@univer.kharkov.ua

³Simon Kuznets Kharkiv National University of Economics,
Science Ave., 9-A, Kharkiv, 61166, Ukraine
tutunik_o@ukr.net

Abstract. In order to develop the scientific and technical foundations of creating an artificial intelligence system for monitoring tectonic origin emergencies, the paper in the results of modeling and forecasting conducted on the basis of neural network technologies, total number of occurrences, with discretion in one month, earthquakes on Earth as one of the main characteristic indicators of the seismic activity of the Earth's globe are presented.

Keywords. emergency, earthquake, tectonic source emergencies monitoring, artificial intelligence system, artificial neural network, neural network simulation, seismic activity level prediction

1 Introduction

The tendency to an abrupt increase in the number and destructive power of natural disasters over the past few decades of the life of society leads to a deterioration of socio-economic and environmental consequences. It indicates the need to develop effective measures to prevent and eliminate emergencies (ES) of various nature on the Globe [1–3].

A promising direction for solving this problem is the development of an effective hazard detection system at the stage of their inception. Also, the causes will be establishing of the occurrence these factors manifestations and effects on them in order to prevent the occurrence of emergencies. This has been implemented on the basis of the classical control loop presented in Fig. 1 [4–7].

This article is part of a planned set of scientific studies aimed at developing a safety system. This eliminates or minimizes losses as much as possible under conditions of manifestation of an emergency. The work is focused on studying the proc-

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

esses of emergence and spreading of emergencies of lithospheric origin, which represent or may pose a serious danger to the life of society [8–12].

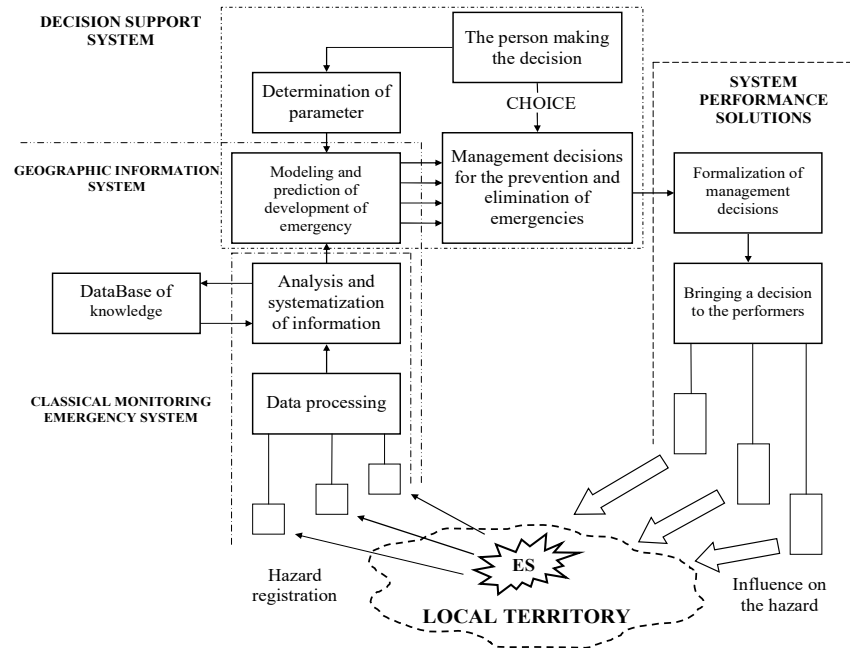


Fig. 1. Diagram of the emergencies monitoring structure as a means of control (Source: author's own elaboration)

When solving the problem of creating an artificial intelligence system for monitoring emergencies of the tectonic nature, there is a need to studying and simulation the processes of occurrence and propagation of the seismic activity level of the local territory under the conditions of seismic activity of the Earth's globe, as an element of the system of nonlinear energy interactions Sun – Earth – Moon.

2 Formal problem statement

Development of scientific and technical bases for creation of artificial intelligence system for monitoring of tectonic origin of the emergency, which is realized by constructing neural network models of forecasting the level of seismic danger of the Earth's territory by the amount and destructive energy of the emergency of tectonic origin.

The construction of neural network models for predicting the level of seismic danger in the Earth's globe was performed using the STATISTICA statistical package [13, 14].

3 Literature review

Today, the basis of predictive observations of Earth's seismic activity is knowledge of the physical laws of the earthquake mechanism and control of the physical fields in the seismic zone [15–19]. The basis of these studies is the idea of the absence of external factors affecting the study area that arise in the system of nonlinear energy interactions Sun – Earth – Moon.

The mechanistic approach, which has been developing in various countries for quite some time, has provided answers to many questions about the preparation of the earthquake. Currently, more than a dozen earthquake models have been developed, the most famous of which is following papers [20–22]:

- a model of avalanche-unstable crack formation (AUC), which consists in the rapid growth of the number of cracks, their interaction with each other and eventually the occurrence of a main or main rupture, the shift of which instantly drops the accumulated elastic with the formation elastic waves;
- dilatant diffusion model (DD), in which crack formation also plays a decisive role, but unlike the AUC model, fractures occur and the presence of water in the rocks of the epicentral region is essential;
- consolidation model of earthquake preparation (Dobrovolskiy's model) describes earthquake preparation as a process of emergence and evolution of rigid inhomogeneities in a continuous environment, as well as some other models that emerged as a generalization of the study of stress and fracture processes of solids samples in the laboratory and their subsequent transfer to natural seismic events.

At the same time, the following methods are now used to obtain a comparative assessment of the level of life-threat in the conditions of manifestation of a emergency [23–30]:

- statistical, based on the analysis of emergency statistics on local territories over several years to determine the risk indicators;
- probable, based on the application of mathematical models, which connects the prerequisites for the appearance of emergency with the possibility of their manifestation;
- expert based on expert judgment in combination with fuzzy set theory.

The advantage of the statistical method is objectivity. Probable and expert methods allow to take into account the sources of potential danger, which are rare in the form of emergencies, but the consequences of which are catastrophic. However, the probable method is extremely cumbersome and time-consuming, requiring a large number of outputs, which results in low accuracy of the results obtained. In the absence of tried and tested mathematical models and reliable enough initial data for them, it is advisable to carry out an expert evaluation of the impact on the possibility of implementing large-scale emergencies of a large number of difficultly formalized initial data.

The prospects of using a statistically probabilistic approach to predict earthquakes in a separate Earth globe, without taking into account the external factors (as an element of the Sun – Earth – Moon nonlinear energy interaction system), are given

in [31]. This paper presents the results of a study of earthquake prediction in the northwestern area of Vietnam using neural network technologies.

Thus, the use of neural network technologies is one of the promising directions for the development of approaches to predicting earthquakes across the globe into systems of nonlinear Sun – Earth – Moon interactions. This determines the direction of our scientific research in the field of monitoring of the tectonic origin emergencies [8–10, 32–34].

4 Particularities of the processes occurring in the Sun – Earth – Moon system and affecting the level of seismic hazard of the Earth’s globe functioning

The dynamics of the physical processes of the Sun – Earth – Moon system affecting the seismic hazard level of the local territory functioning can be schematically represented according to Fig. 2 – 4. This can be characterized by the following spatial constructions within the solar galaxy.

1. The axis of rotation of the Earth in the celestial sphere describes a complex wave-like trajectory. The points of the axis of rotation are at an angular distance of about from the pole ecliptic (Fig. 2). The vertex of the cone coincides with the Earth center. The points of equinoxes and solstices move along the ecliptic towards the sun. Moments of gravitational forces influence on the equatorial bulges and vary depending on the positions of the Moon and the Sun relative to the Earth. When the Moon and the Sun are in the plane of the Earth’s equator the moments of forces disappear. If tilts of Moon and Sun are the maximum, then the magnitude of the torque will be greatest. The nutations, owing to fluctuations in the moments of the forces of the axis of rotation of the Earth have been observed by consist of a series of small periodic oscillations. The main nutations have a period of 18.6 years - the time of the orbital nodes of the Moon. Movement with this period occurs on an ellipse. The major axis of the ellipse is perpendicular to the direction of the precessional motion and is equal to; small - parallel to it and equal. Next in magnitude of the amplitude are the components with a period of 0.5 year, 13.7 days., 9.3 years, 1 year, 27.6 days. etc., therefore the trajectory has the form of “thin laces” (shown on the enlarged fragment in the left part of Fig. 2) [35–38].

2. The pressure from the solid inner core and the surrounding melt (outer core) onto the mantle arises as a result of the eccentric revolution of the Earth’s shell around the displaced inner core, which squeezes the shell from the inside. The forces compressing the shell of the sphere (planet) and drawing it inward to the core arise in other parts of the planet. This process has two components: impact at the expense of the annual displacement of the center nucleus relative to the center of the globe (Fig. 2 – 4); impact at the expense of eccentric circulation of the core relative to the lower mantle, when due to the difference in angular rotation velocity of the core and lower mantle (ω_1 – angular velocity of rotation of the mantle; ω_2 – angular velocity of rotation of the outer core; ω_3 – angular velocity of rotation of the inner core;

$\Delta\omega = \omega_2 - \omega_1$ – angular velocity of rotation of the outer core relative to mantle (“western drift”)), therefore, there are zones of high pressure and vacuum ($P_1 \neq P_2$ where P_1 and P_2 are indicators of pressure of the inner core of the globe on its surface), affecting the level of seismic activity of the surface of the Earth (Fig. 3). As long as there is a difference in the angular velocity of rotation and displacement of the nucleus, the appearance of such zones will be maintained [39, 40].

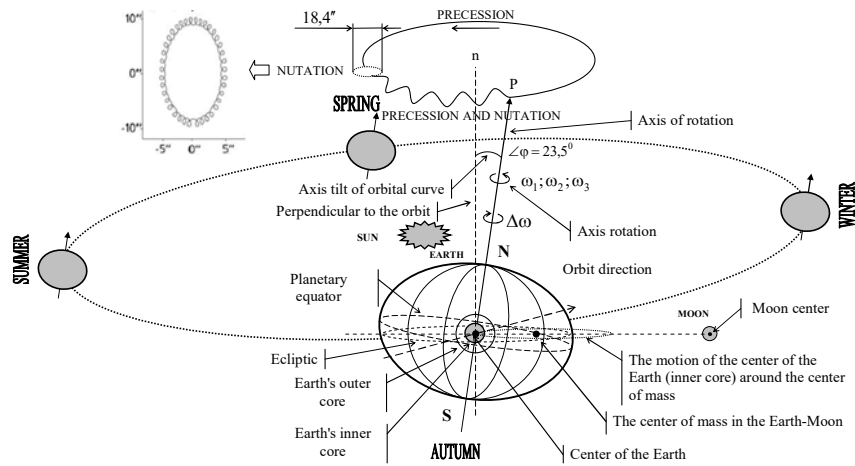


Fig. 2. Motion diagram of the inner core of the Earth in the Sun–Earth–Moon system (Source: author's own elaboration)

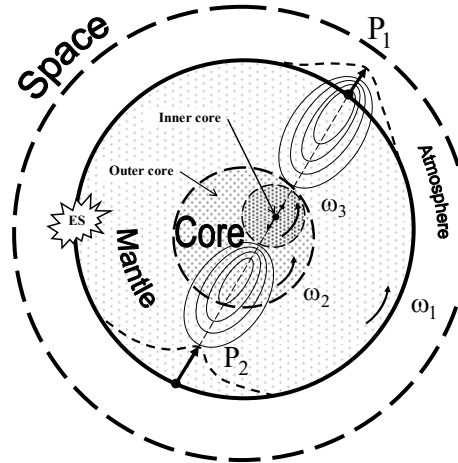


Fig. 3. Influence diagram of internal core oscillations on seismic activity (Source: author's own elaboration)

3. Internal elastic stresses arise in the process of moving lithospheric plates (Fig. 4), which are energy sources of earthquakes. The occurrence depth of elastic stresses depends on the nature of a movement plates. The relative motion of lithospheric plates leads to the emergence of shallow (not deeper than 20–25 km) earthquake sources and dipping of lithospheric plates into the mantle initiates the appearance of sources of deep (exceeding 70 km) earthquakes. The probability of elastic stresses - sources of earthquakes decreases with increasing distance from the interface of the lithospheric separation plates.

4. Surface and bulk seismic waves are the propagation factors of earthquake hazards Z_0 , that can cause secondary earthquakes [41, 42].

5. The probability of mutual amplification or weakening of bulk seismic waves increases in the process of spatial-vibrational movement of the Earth's internal core and its effect on the external core. Consequently, the possibility of secondary earthquakes Z' increases also [43].

7. The territorial-temporal changes in the intensity of the natural electromagnetic field pulses of the Earth initiating anomalous processes in the atmosphere occur due to the movement of the Earth's inner core has been established [44–46].

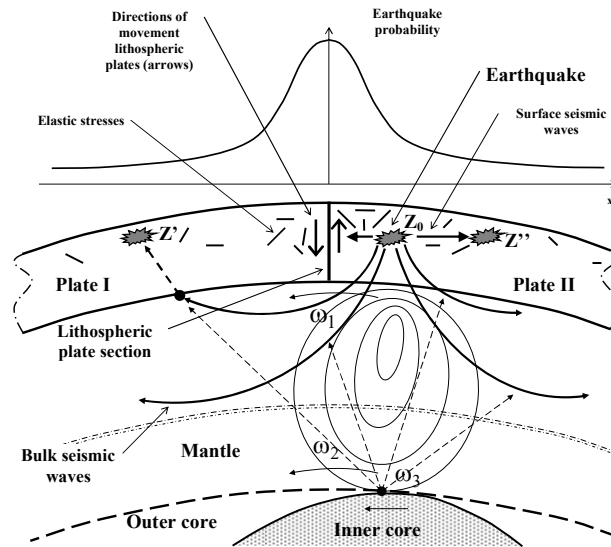


Fig. 4. Process diagram of earthquake and the spread of seismic activity (Source: author's own elaboration)

Thus, combining the analysis results of the impact dynamics and energy of the internal physicochemical processes of the Earth on the origin generating tectonic processes allowed to formulate an approach to studying the nature of seismic phenomenon. It is an important tool for analyzing the results of civil defense research on the development of models for the development of ES tectonic nature.

5 Experiments and results of neural network prediction of the level of seismic activity of the Earth

According to the purpose of the research, the solution of the scientific problem in the work is ensured by constructing an artificial neural network (ANN) model – a model for predicting the level of seismic activity of the planet from the conditions of the system functioning of nonlinear energy interactions Sun – Earth – Moon. This is a time series model, where the initial indicator, according to Fig. 5, is the total number of earthquakes in the world $N_{\Sigma}(t)$ (provided that $M \geq 5$ the magnitude of the earthquake), depending on the current time of analysis (t), as well as the distance of the Earth's inner core from the center of the planet $\Omega(t)$ and the change in the length of

day $LOD(t) = S(t) - 86400s$ (where $S(t) = \frac{r_0}{r(t)} 86400s$ – the length of day, $r_0 = 7,292115 \cdot 10^{-5}$ rad/s – constant (average) angular velocity of the Earth's own rotation).

For the development presented in Fig. 5 models have a multilayer perception (MLP) selected. The input parameters of this model are the results of the analysis of the monthly dynamics of indicators, $N_{\Sigma}(t)$, $\Omega(t)$ and $LOD(t)$ for 2009 – 2018, which are presented in [8–10, 32–34] and Fig. 6.

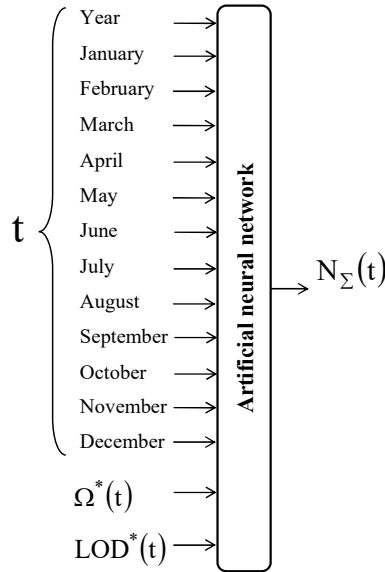


Fig. 5. Scheme of the ANN time series model for the monthly forecasting of the Earth's number of earthquakes, depending on the indices of the remoteness of the Earth's inner core from the center of the planet and changes in the length of day (Source: author's own elaboration)

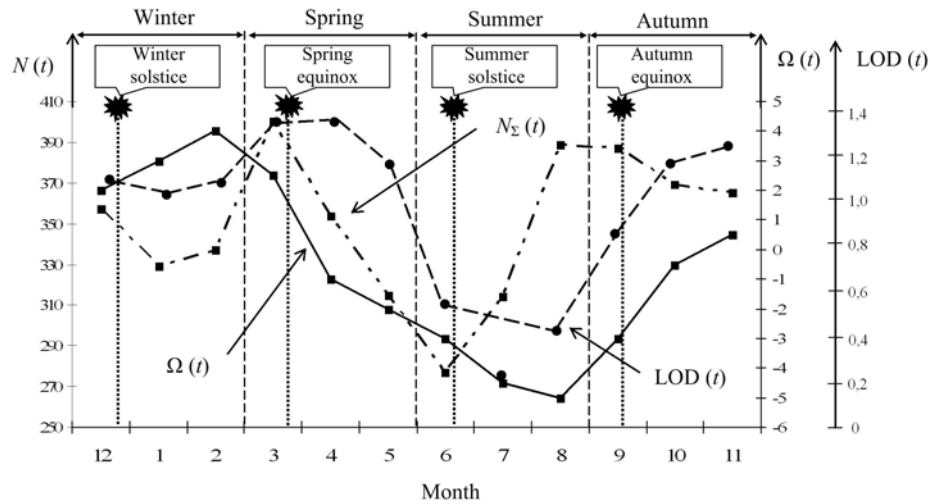


Fig. 6. Monthly dynamics of the variations in the rate of variation of the Earth's axial rotation ($LOD(t)$), the distance of the inner core from the center of the Earth ($\Omega(t)$) and the total seismic activity ($N_{\Sigma}(t)$) with magnitude during 2009 – 2018 are summarized on the Globe (Source: author's own elaboration)

When developing a mathematical model for predicting the level of seismic activity of the Earth ball, based on the decision parameter $N_{\Sigma}(t)$ on the complexity of the MLP architecture, it was based on five analyzes of the learning results of networks, which accidentally included five hundred neural networks. The criterion for choosing the optimal network was the relationship between the forecast error and the complexity of the architecture. The results of the analysis are presented in Table 1.

Table 1. Statistical characteristics of the three-layer perceptions suggested by the “decision-maker” as the best for neural network time series analysis of the forecast of the monthly dynamics of the Earth's seismic activity ($N_{\Sigma}(t)$)

№	Architecture	Study productivity	Control productivity	Test productivity	Activation hidden layer	Activation of yield
1.	MLP 15-199-1	0,944	0,859	0,976	Exp.	Logistic
2.	MLP 15-233-1	0,935	0,634	0,964	Identity	Exp.
3.	MLP 15-243-1	0,939	0,859	0,986	Exp.	Logistic
4.	MLP 15-187-1	0,938	0,884	0,985	Exp.	Logistic
5.	MLP 15-12-1	0,967	0,727	0,668	Tanh	Identity

For neural network training, all observations were divided into three samples. By default, random sampling was performed between samples to avoid retraining the network and to ensure quality generalization (forecasting). The first sample (Educational – 50% of observations) was used to train the network; the second (Control – 25% of observations) – for cross-validation of the learning algorithm during its operation; third (Test – 25% of observations) – for the final independent testing of a trained neural network. The training was done with speed $\eta = 0,01$.

Presented in Table. 1 networks are characterized by a more effective balance between modeling error and architecture complexity for time series analysis of the forecast of the monthly dynamics of Earth's seismic activity ($N_{\Sigma}(t)$). This was the basis for further construction of the three-layer MLP 15-12-1 network, which has fifteen inputs (Fig. 5), 12 elements in the hidden layer and one logical output function of thirteen inputs, presented in Fig. 7.

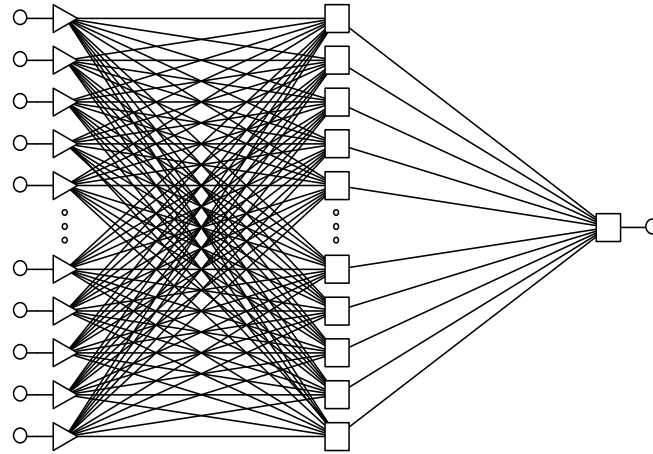


Fig. 7. MLP 15-12-1 Three-Layer Perceptron Architecture with Logical Signal Transmission for Time Series Prediction of Globe Seismic Monthly Dynamics ($N_{\Sigma}(t)$) (Source: author's own elaboration)

In this case, the use of logical activation functions, with scaling parameters, was based on a given fraction of span of a logical function equal 0,9 (respectively range $[0,05; 0,95]$) to the range of training of the neural network. The function of activating the hidden layer of the perceptron MLP 13-10-1 is hyperbolic tangent. This allows for a slight extrapolation of the data. In addition, the use of logical functions stabilizes learning.

The results of checking the adequate prognostic performance of MLP 15-12-1 network are presented in Fig. 8, where the dependence observed ($N_{\Sigma}^*(t)$) statistic the values of the Earth's seismic activity indicator via predicted ($N_{\Sigma}(t)$) values. The

coefficient of correlation between these indicators on the results of the training "Educational Choice" is equal $r_{N_{\Sigma}^*(t)N_{\Sigma}(t)}^2 \approx 0,967$.

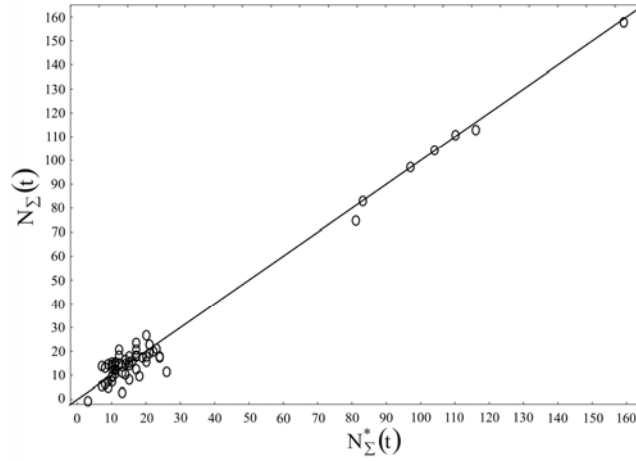


Fig. 8. Dependence of observed ($N_{\Sigma}^*(t)$) statistic values of the Earth's seismic activity indicator by via predicted values ($N_{\Sigma}(t)$) by MLP Network 15-12-1 (Source: author's own elaboration)

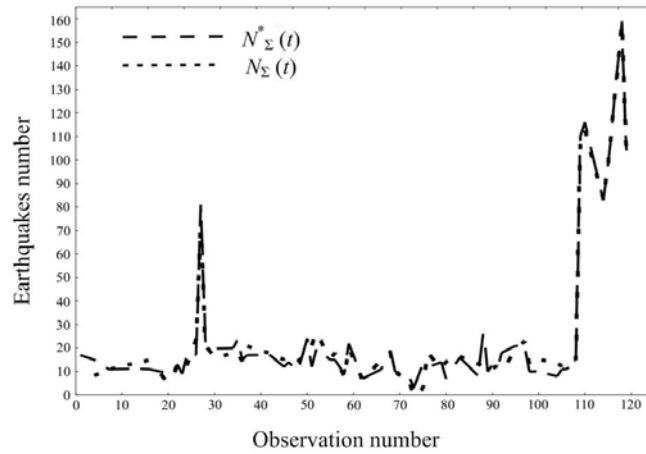


Fig. 9. Graph of the time dependence (of observation number) where the dependence observed ($N_{\Sigma}^*(t)$) statistic the values of the Earth's seismic activity indicator via predicted ($N_{\Sigma}(t)$) values by MLP network 15-12-1 (Source: author's own elaboration)

The graph of time dependence (from observation number), with the appropriate level of prognostic adequacy, where the dependence observed ($N_{\Sigma}^*(t)$) statistic the values of the Earth's seismic activity indicator via predicted ($N_{\Sigma}(t)$) values by the MLP 15-12-1 network is presented in Fig. 9.

According to the data analysis Table 2 it is necessary to state that obtained within the limits of the ideas about the dynamics of physical processes that occur in the system Sun – Earth – Moon and affect the level of seismic danger of functioning of the local territory of the planet (see Fig. 2 – 4), the MLP neural network 15-12-1 and the results of its forecast allow us to ascertain the adequacy, in accordance with the data of Figs. 8 and the $r_{N_{\Sigma}^*(t)N_{\Sigma}(t)}^2 \approx 0,967$ parameter figure presented in Fig. 5 models for predicting the level of seismic activity of the globe.

Table 2. The result of the prediction by the MLP network 15-12-1 level of seismic activity-those of the Globe

t	2019											
	01	02	03	04	05	06	07	08	09	10	11	12
$\Omega^*(t)$	3,00	4,00	2,50	-1,00	-2,00	-3,00	-4,50	-5,00	-3,00	-0,50	0,50	2,00
$LOD^*(t)$	0,81	1,00	0,96	0,98	0,70	0,19	0,01	0,05	0,17	0,27	0,37	0,24
$N_{\Sigma}^*(t)$	120	104	118	113	93	103	117	127	131	137	138	123
$N_{\Sigma}(t)$	116	113	132	120	83	97	125	141	144	124	147	138
$\delta_{N_{\Sigma}^*(t)}, \%$	3	9	11	6	10	6	7	11	9	9	7	12

6 Conclusion

The paper presents a neural network model of the time series for the monthly forecasting of earthquakes with magnitude $M \geq 5$. The obtained neural network allows us to predict the level of seismic activity of the globe with adequacy at the level $r_{N_{\Sigma}^*(t)N_{\Sigma}(t)}^2 \approx 0,967$.

The results obtained are the basis for the development of scientific and technical bases for the creation of a system of artificial intelligence for the performance of tasks of monitoring of tectonic origin.

References

1. First Report of the Chairman of the International Committee on Problems of Global Changes in the Geological Environment „GEOCHANGE”. <http://www.ru.geochngereport.org>
2. Barishpolets, V.A.: System analysis of disasters occurring in the world. Radio Electronics. Nanosystems. Information Technology, 2, 1 – 2, 162 – 176 (2010).

3. National report on the state of technogenic and natural security in Ukraine. <http://www.dsns.gov.ua/>
4. Code of Civil Protection of Ukraine. In: Voice of Ukraine, 220 (5470), 4 – 20 (2012).
5. Resolution of the Cabinet of Ministers of Ukraine of January 9, 2014, No. 11 “On Approval of the Regulation on the Unified State Civil Protection System”. <http://zakon5.rada.gov.ua/laws/show/11-2014-%D0%BF>
6. Kalugin, V.D., Tiutiunyk, V.V., Chernogor, L.F., Shevchenko, R.I.: Development of scientific and technical basis for establishment of monitoring, prevention and liquidation of emergency situations of natural and man-made nature, and also ensuring of environmental of ecological security. Information processing systems, 9(116), 204 – 216 (2013).
7. Andronov V.A., Divizynyuk, M.M., Kalugin, V.D., Tiutiunyk V.V.: Scientific and design bases of complex creation system of emergency monitoring situations in Ukraine. National university of civil protection of Ukraine (2016).
8. Tiutiunyk, V., Chernogor, L. V., Kalugin, Agazade, T.: Development of the bases of geoinformational the systems of emergency monitoring situations of tectonic origin. Applied Radio Electronics, 18, 1,2, 52 – 65 (2019).
9. Tiutiunyk, V., Kalugin, V., Pysklakova, O., Yaschenko, O., Agazade, T.: Hierarchical clustering of seismic activity local territories Globe. EUREKA: Physics and Engineering, 4: 41 – 53 (2019).
10. Tiutiunyk, V., Chernogor, L. V., Kalugin, Agazade, T.: Functional description of the zoning of local territories of the Globe by quantity and destructive energy of tectonic extreme origin situations. Municipal economy of cities, 1, 154: 272 – 287 (2020).
11. Chernogor, L.F.: Geomagnetic Disturbances Accompanying the Great Japanese Earthquake of March 11, 2011. Geomagnetism and Aeronomy, 59, 1, 62 – 75 (2019).
12. Chernogor, L.F.: Possible Generation of Quasi-Periodic Magnetic Precursors of Earthquakes. Geomagnetism and Aeronomy, 59, 3, 374 – 382 (2019).
13. Kim, J.O., Muller, C.U., Klekka, U.R.: Factor, discriminant and cluster analysis. Finance and Statistics (1989).
14. Khalafyan, A.A.: STATISTICA 6. Statistical Data Analysis. LLC Binom-Press (2007).
15. Sobolev, G.A., Ponomarev, A.V.: Earthquake Physics and Harbingers. Science (2003).
16. Guglielmi, A.V.: Foreshocks and aftershocks of strong earthquakes in the light of disaster theory, Advances in Physical Sciences, 185, 4, 415 – 429 (2015).
17. B.V. Levin, M.V. Rodkin, I.N. Tikhonov, Great Japanese earthquake, *Nature*, 10: 14 – 22, 2011.
18. Tertyshnikov, A.V., Platonov, V.V.: Prospects for monitoring seismic conditions from space. Electronic scientific journal "Investigated in Russia. <http://www.zhurnal.ape.relarn.ru/artecles/2007/031.pdf>
19. Smirnov, V.M., Smirnova, E.V.: Study of the possibility of using satellite navigation systems for monitoring seismic phenomena. Electromechanical issues, 105, 94 – 104 (2008).
20. Romashov, A.N., Gypsy, S.S.: In search of a generalized geotectonic concept. Geotectonics, 4, 3 – 12 (1996).
21. Gufeld, I.L.: Seismic process. Physicochemical aspects. TsNIIMash (2007).
22. Mishin, S.V., Panfilov, A.A., Hasanov, I.M.: Earth's gravity is the cause of earthquakes. Geophysical Journal, 41, 6, 213 – 222 (2019).
23. Kotovenko, O.A., Sobolevska, L.I., Miroshnichenko, O.Yu.: Stochastic modeling in the study of processes under the influence of environmental management in the region. East European Journal of Advanced Technology, 2/14, 37 – 41 (2012).
24. Akimov V.A., Radiev, N.N.: Determination of the relative hazard of the territories. Emergency Security Issues, 6, 129 – 140 (2000).

25. Lepikhin, A.M.: Integrated Territory Security Indicators. *Emergency Security Issues*, 5, 93 – 98 (2008).
26. Vladimirov, V.A., Kulba, V.V., Malinetsky, G.G., Makhutov, H.A.: Risk management. *Science* (2000).
27. Reinschke, K., Ushakov, I.A.: Graph reliability assessment using graphs. *Radio and communication* (1988).
28. Ryabinin, I.A.: Reliability and safety of structurally complex systems. *Polytechnic* (2000).
29. Ryabinin, I.A.: The concept of logical probability theory of security. *Devices and control systems*, 10, 6 – 9 (1993).
30. Y.D. Vishnyakov, N.N. Radaev. *General risk theory*, Publishing Center "Academy", 2008.
31. Pupkov, K.A., Chong, Khao Dinh: The use of neural network technologies in earthquake prediction problems (for example, the north-western region of Vietnam). *Bulletin of the N.E. Bauman Moscow State Technical University, Series "Instrument Making"*, 2, 70 – 78 (2012).
32. Tiutiunyk, V.V., Chernogor, L.F., Kalugin, V.D., Agazade, T.: Assessment of influence power effects in system the Sun-Earth-Moon on the level of seismic activities the territory of the Globe. *Management, navigation and communication systems*, 6(46), 238 – 246 (2017).
33. Tiutiunyk, V.V., Chernogor, L.F., Kalugin, V.D., Agazade, T.: Assessment dynamic and power effects on the Earth and their influences on ratios between seismic activity levels of Globe hemispheres. *Scientific Bulletin: Civil protection and fire safety*, 2(4), 101 – 117 (2017).
34. Tiutiunyk, V.V., Chernogor, L.F., Kalugin, V.D., Agazade, T.: Influence estimation of rotation speed variation of the Earth on level of seismic activity of local Globe territory. *GEOINFORMATIKA*, 3(67), 36 – 48 (2018).
35. Sidorin, A.Ya.: Influence of the Sun on seismicity and seismic noise. *Seismic instruments*, 40, 71 – 80 (2004).
36. Levin, B.V., Sasorova, E.V., Domansky, A.V.: Properties of “critical latitudes”, rotation variations and seismicity of the Earth. *Bulletin of the Far Eastern Branch of the Russian Academy of Sciences*, 3, 3 – 8 (2013).
37. Wiemer, S., Wyss, M.: Mapping spatial variability of the frequency-magnitude distribution of earthquakes: An overview. *Advances in Geophysics*, 45, 259 – 302 (2002).
38. Atef, A.H., Liu, K.H., Gao, S.S.: Apparent weekly and daily earthquake periodicities in the Western United States. *Bull. Seismol. Soc. Amer.*, 99, 4, 2273 – 2279 (2009).
39. Klimenko, A.V.: Global properties of Earth's seismic activity and their relationship with its rotation. The dissertation ... of the candidate Phys.-Math. of sciences (2005).
40. Malyshkov, Yu.P., Malyshkov, S.Yu.: Periodic variations of geophysical fields and seismicity, their possible connection with the motion of the Earth's core. *Geology and geophysics*, 2, 152 – 172 (2009).
41. Berezniakov, A.I., Niemets, K.A.: *Earth Physics*. V.N. Karazin Kharkiv National University (2010).
42. Chernogor, L.F.: *Physics and ecology of disasters*. V.N. Karazin Kharkiv National University (2012).
43. Malyshkov, Yu.P., Dzhumabaev, K.B., Malyshkov, S.Yu.: Earthquake prediction method. Patent of the Russian Federation, No. 2238575, MPK G01V3/00 (2004).
44. Tertyshnikov, A.V.: Harbingers of Severe Earthquake in the Ozonosphere, *Heliogeophysical research*, 2, 54 – 59 (2012).

45. Garcia, R., Crespon, F., Ducic, V., Lognonne, P.: Three-dimensional ionospheric tomography of post-seismic perturbations produced by the Denali earthquake from GPS data. *Geophys. J. Int.*, 163, 1049 – 1064 (2005).
46. Heki, K., Ping, J.: Directivity and apparent velocity of the coseismic traveling ionospheric disturbances observed with a dense GPS array. *Earth Planet. Sci. Lett.*, 236, 845 – 855 (2005).

Information Technology of Hot-metal Ladle Car Handwritten Numbers Recognition from Photo Image

Yevhen Chychkarov¹[0000-0002-4362-5129], Anastasiia Serhiienko²[0000-0003-1328-2572], Iryna
Syrnamiikh³[0000-0003-0524-6731]

1,3 Donetsk State University of Management, Mariupol, Ukraine
influence@meta.ua, siv.2017@ukr.net

2 Pryazovskyi State Technical University, Mariupol, Ukraine
invitery@i.ua

Abstract. The paper presents the results of the construction and testing of an information system for recognizing ladle car numbers drawn up of handwritten digits. The latest publications on solving problems of recognition and analysis of texts by their images were analyzed, a selected range of application libraries for creating a program for experiments are Keras, TensorFlow, Theano, Pandas, Matplotlib and others.

Two variants of recognition technology were analyzed and tested: using fuzzy logic and using deep learning neural networks. To construct and train neural networks, a well-known and rather complete base of handwritten digits MNIST was chosen. Two types of neural networks were considered: sequential and convolutional. The training of neural networks was carried out using a variable number of steps (epochs). Recognition images were scaled to a size of 28x28 (784 cells in a one-dimensional representation). Preliminary processing of images (filtering, scaling, etc.) was carried out using the OpenCV library. For recognition, each image of the digit was converted to a 28x28 size and fed to the input of a pre-trained neural network.

A technique to select the area of interest in photographs containing hand-written digits for further recognition has been worked out. For handwritten digit recognition, the best recognition accuracy is provided by a convolutional neural network, with which 97.6% of car ladle digits were correctly recognized.

Keywords: Keras; TensorFlow; MNIST; Python; Deep learning; Neural networks; Digit recognition.

1 Introduction

Currently, automatic recognition systems for numbers of moving objects (the license plate of the vehicles, wagon numbers, etc.) are demanded in a wide range of fields.

In this paper, a method of recognizing the digit of a moving hot-metal ladle car by

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

a photo using various mechanisms (fuzzy output and deep learning neural networks) is proposed. Hot-metal ladle cars are used to transport cast iron from blast furnaces to the mixer department of the oxygen-converter plant (OCP). The number of the hot-metal ladle car is its unique identifier. Therefore, the ladle car number is quite important information about it, and it is actively used in information systems operating at integrated iron-and-steel works. For inventory and control of their condition, an automated monitoring system is used to track the movement of platforms carrying hot-metal ladle cars.

The digits of car ladle number are applied manually (without stencil) with lime. During transportation and while conducting technological operations, the ladle cars are exposed to high temperatures, and the digits may be distorted due to dust, cast iron drips, etc.

The image of the hot-metal ladle car is obtained by shooting with the camera. The initial data for solving the problem of recognizing the ladle car number is a fragment of the photo image of the ladle car containing its number.

Recognizing handwritten digits or letters is obvious enough for a person, but for computers, the identification of such characters is a difficult task. Various methods are often used to fulfill this task: neural networks, classification algorithms, etc.

Recently, deep learning neural networks have been widely used to identify range of interest and recognize image elements.

2 Analysis of literature data and formulation of the problem

Currently, there are many areas of science and technology, which are largely focused on the development of systems that analyze the information presented as images. However, most of the tasks associated with image processing and recognition are classified as difficult to formalize.

The task of segmentation and recognition of image areas containing handwritten or printed characters is relevant due to the presence of a large number of technical applications.

A number of works on the recognition of license plates of cars [1-4] using neural network technologies have been published. A similar approach is used in the systems of registration and recognition of wagons and tanks numbers [5-6]. The functionality of all these systems is approximately the same – they automate the process of reading digits and save the received information.

Differences in the digit recognition systems associated with the format of car and wagon numbers, as well as with the observation conditions, are discussed in works [8–9]. In particular, according to [8], weather conditions and poor illumination are obstacles for the qualitative identification of digits. To eliminate these obstacles, additional lighting and image filtering should be used.

Various methods and algorithms to recognize digits are used: a modified algorithm for comparing an object with a standard in Hausdorff distance [10], a perceptron neural network with sigmoid activation functions [8], and a scale-invariant transformation method [5].

According to [11], the support vector method (SVM) is the most effective for recognizing license plate numbers. The advantage of the SVM method is that a

relatively small training sample is sufficient for constructing a classifier for character recognition. In addition, it has a low probability of error.

In work [12], the effect of preliminary processing and segmentation of license plate numbers images on the results of their recognition was noted. For any recognition methods, incorrect character segmentation does not allow to achieve accurate results. It was also noted there that license plate recognition is easily implemented in relation to standard plates. But in different countries the sizes and contents of the license plate differ, so there is a need to solve recognition problems without sticking to strict rules.

In recent years, much attention has been paid to the recognition of digits on various visual objects using deep learning neural networks [13-16]. For example, in work [16], a method for detecting and segmenting car license plates in real time on the basis of image analysis and processing methods is presented. According to the authors, the computation effort and accuracy degree, taking into account the proposed approach, are acceptable for real-time applications with a run time of less than 1 second.

A deep learning neural network is an artificial neural network with several hidden layers [17]. Additional layers make it possible to build abstractions of ever higher levels, making it possible to form models for recognizing complex real-life objects. Deep direct distribution networks are commonly used, however, recent studies have shown the successful use of deep architectures in recurrent networks [18]. In tasks related to image processing, convolutional neural networks (CNN) are predominantly used, taking into account their high efficiency for this range of tasks.

Deep learning technologies are used to solve a range of practically important tasks: tasks related to the natural language (grammar check, recognition of printed or handwritten texts, etc.), search for information on the Internet, image processing and recognition, etc. [20, 21]. Machine learning methods are used to build algorithms that provide self-learning on certain datasets, and then reliable forecasting and classification of test data [22-26].

To solve the problems of recognition and analysis of texts by their images, various types of neural networks are used: sequential neural networks [24], convolutional neural networks [21, 22, 24, 25], impulse neural networks [26].

To solve the problems of machine learning and building neural networks, the Python programming language and a number of applied libraries (Keras, TensorFlow, Theano, Pandas, Matplotlib) [27] are widely used.

TensorFlow is a specialized library for calculations using graphs of data streams (graph nodes – mathematical operations, edges – multidimensional data arrays). In TensorFlow, parallel computing on the GPU is available after installing necessary drivers and configuring special parameters.

Keras is a library for building Deep Learning neural networks, which provides a high-level API that uses TensorFlow (or Theano) as a backend.

Optimization of the structure of deep learning neural networks and optimization of their parameters are of considerable interest for the researchers.

Advantages related to recognition accuracy close to the human level when processing large data arrays have led to an increase in the use of CNN in recent years [22-23]. But the picture of the effect of CNN's hidden layers on overall network performance remains unclear.

According to [24], in practice, impulse neural networks with multilayer learning

turned out to be difficult to learn.

3 Goal and objectives of the research

Goal of the research: to develop an information technology for recognizing the handwritten digit of a hot-metal ladle car number by its photo image, capable of fast localizing the area of interest, identifying the digit, and testing this technology.

To solve this practical problem, it is necessary to solve several subproblems:

- development and testing of an algorithm of preliminary image processing and selection of areas containing a digit for recognition, as applied to the images of hot-metal ladle cars;
- setup and optimization of an artificial neural network to recognize handwritten or printed digits;
- development and adjustment of the car number recognition method by means of a fuzzy comparison with the sample.

The image of the ladle car with the number is obtained using the camera in grayscale.

The images under consideration (see the example in Fig. 1) have a number of features that significantly distinguish their processing from the well-known solutions on recognizing wagon or car numbers:

1. hot-metal ladle cars work in difficult conditions around the clock, therefore, processed images may be distorted by the presence of dust, drips and spray of cast iron, changing the lighting conditions;
2. the processed image contains non-informative areas (sky, wall, elements of nearby metal structures) with high or low brightness;
3. digits to be recognized are applied manually on an uneven porous surface, therefore, even on freshly applied numbers there are heterogeneities;
4. there may be drips when applying numbers, as well as areas with a light background at the edges of the image.

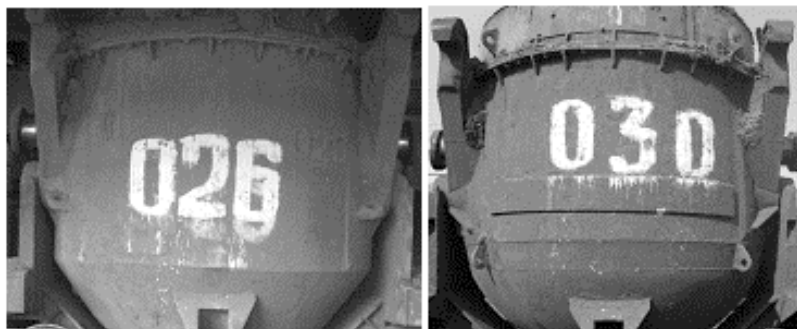


Fig. 1. Example of original photo images with distortions

4 Description of the proposed algorithm of selecting recognition area

All operations of selection a region of interest (see list below) were performed in a program in Python using functions of the OpenCV library. An illustration of the process of selecting recognition areas is presented in Fig. 2.

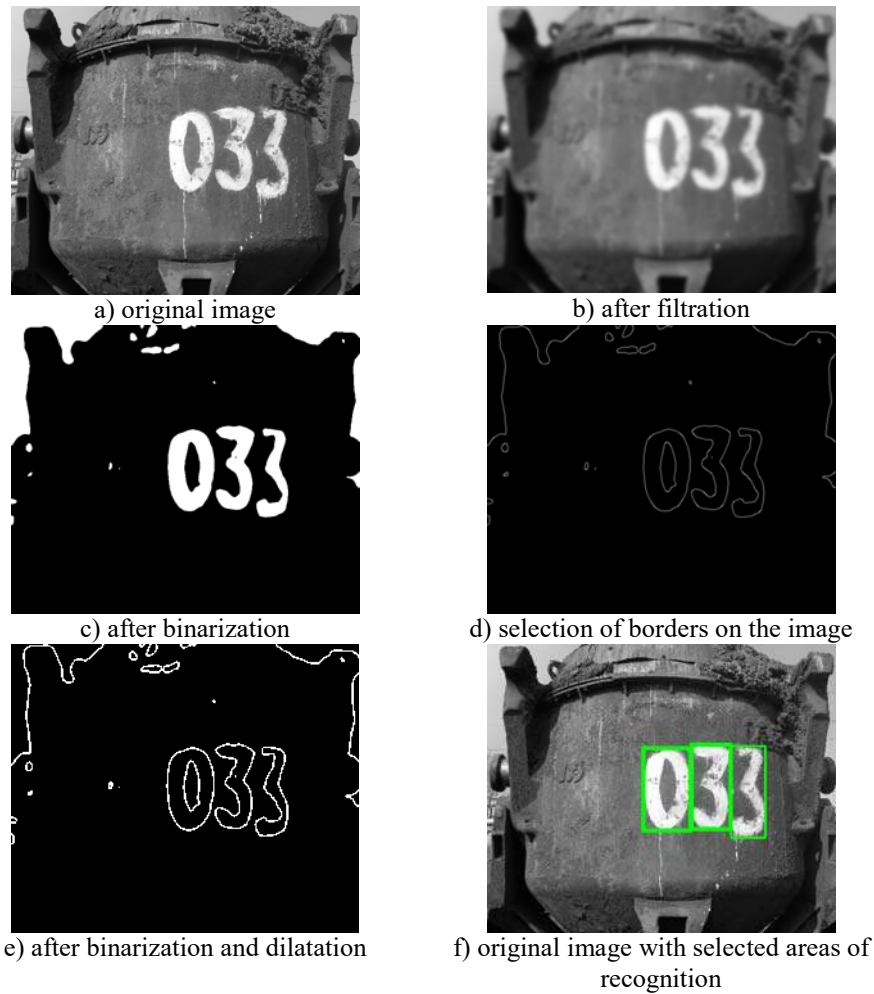


Fig. 2. Operating sequence of selection recognition areas algorithm

The basis of automatic recognition of car numbers is based on the following procedures [27-28]:

1. image filtering to reduce the noise level (a Gaussian filter was used - `cv2.GaussianBlur` function);
2. binarization of the image to cut off noise (the `cv2.threshold` function was used, its

- parameters were adapted to reliably select the outlines of the digits);
3. highlighting the contrasting borders on the image (the Canny edge detector was used - cv2.Canny function, for which the values of the maximum and minimum values of the gradient were preliminarily selected);
 4. filtering to reduce the effect of image heterogeneity (the median filter was used - cv2.medianBlur function);
 5. image binarization (cv2.threshold function was also used);
 6. morphological transformation (dilatation - function cv2.dilate);
 7. selection of contours and their sorting (selection of contours was carried out using the cv2.findContours function);
 8. image segmentation, i.e. selection of recognition areas in the form of rectangle set containing previously selected contours of digits (cv2.boundingRect functions were used).

Computer experiments showed that the proposed method of localization of informative is effective and has a reliability of localization of 97.6 %.

5 Description of digit recognition algorithms

The process of recognizing text areas in images can be divided into three stages: the stage of preliminary processing of the image, the stage of selecting an area containing characters, and the stage of their recognition.

5.1 Digit identification using fuzzy logic

Localization of the area of interest described in [29] is performed using clustering by nearest neighbors method. This method is effective and has a localization reliability of 96%. However, the algorithm for identifying the recognition area described above can also be applied in this information technology for recognizing a handwritten digit of a hot-metal car by its image, since it has a higher localization reliability of 97.6%.

The number of the hot-metal car is identified using the original method, based on a fuzzy representation of the images and comparing them with the standard segments [30]. The idea of the identification method is as follows. First it is necessary to find the common area of two comparable segments of different sizes, called the core. Associate a coordinate system with this core and recalculate the pixel representation of both images into one coordinate system.

Place the origin of the core coordinate system at the characteristic point of the digits image representation. A point has been adopted, which is calculated as the center of gravity of the image. The cores of the analyzed segment and prototype $\{A''\}, \{B''\}$ are divided into blocks $\Omega_{i,j}$ of size $n \times n$. The number of blocks for sets $\{A''\}, \{B''\}$ is calculated: $C = \text{round}(W/n)$; $D = \text{round}(H/n)$.

The set of units of image and prototype cores are denoted as:

$$\Omega^A = \{\Omega_{r,s}^A \mid r = \overline{1, C}; s = \overline{1, D}\}, \quad \Omega^B = \{\Omega_{r,s}^B \mid r = \overline{1, C}; s = \overline{1, D}\}.$$

Separate units $\Omega_{r,s}^A$ and $\Omega_{r,s}^B$ of sets Ω^A and Ω^B , respectively, are set by two characteristics: the set of pixels coordinates included in the unit:

$$\Omega_{r,s}^A = \Omega_{r,s}^B = \{(i, j) | i = \overline{(C-1) \cdot n + 1, C \cdot n}, j = \overline{(D-1) \cdot n + 1, D \cdot n}\},$$

and set of binarized brightness functions: $g_A : \Omega_{r,s}^A \rightarrow [0,1]$, $g_B : \Omega_{r,s}^B \rightarrow [0,1]$.

Two fuzzy sets $\tilde{\Omega}^A$, $\tilde{\Omega}^B$, which are defined on universal sets Ω^A , Ω^B are introduced. Membership functions are given in formula 1:

$$\mu_{\tilde{\Omega}^A} \left(\Omega_{r,s}^A \right) = \frac{\sum_{i,j \in \Omega_{r,s}^A} g_A(i, j)}{|\Omega_{r,s}^A|}, \quad \mu_{\tilde{\Omega}^B} \left(\Omega_{r,s}^B \right) = \frac{\sum_{i,j \in \Omega_{r,s}^B} g_B(i, j)}{|\Omega_{r,s}^B|}. \quad (1)$$

As a result fuzzy sets are received (formula 2):

$$\tilde{\Omega}^A = \left\{ \Omega_{r,s}^A \mid \mu_{\tilde{\Omega}^A} \left(\Omega_{r,s}^A \right) \right\}, \quad \tilde{\Omega}^B = \left\{ \Omega_{r,s}^B \mid \mu_{\tilde{\Omega}^B} \left(\Omega_{r,s}^B \right) \right\}. \quad (2)$$

Then Hamming distance is determined (formula 3):

$$\rho \left(\tilde{\Omega}^A, \tilde{\Omega}^B \right) = \frac{\sum_{r=1}^C \sum_{s=1}^D \left| \mu_{\tilde{\Omega}^A} \left(\Omega_{r,s}^A \right) - \mu_{\tilde{\Omega}^B} \left(\Omega_{r,s}^B \right) \right|}{|\Omega^A|} \quad (3)$$

For the analyzed segment and all prototypes available in the database, the values of the relative Hamming distance are calculated $\rho \left(\tilde{\Omega}^A, \tilde{\Omega}^B \right)$ and sorted in descending order: $\rho \left(\tilde{\Omega}^{A_1}, \tilde{\Omega}^{B_1} \right) \leq \rho \left(\tilde{\Omega}^{A_2}, \tilde{\Omega}^{B_2} \right) < \dots < \rho \left(\tilde{\Omega}^{A_N}, \tilde{\Omega}^{B_N} \right)$.

5.2 Digit identification a using a neural network deep learning

An important point for deep learning neural networks training is building a dataset to train the model.

To construct and train the model, a well-known and rather complete database of handwritten digits MNIST was chosen [18-24, 30]. This database (MNIST - Modified

NIST) is a subset of the larger NIST database [31], which contains handwritten images segmented by images of specially prepared templates filled by respondents from the Census Bureau and students of US educational institutions. MNIST consists of a training (60,000 images) and test (10,000) parts, in addition, images obtained from different authors were placed in different parts to increase uniqueness.

The Keras library contains numerous implementations of widely used building units of neural networks, such as layers, target and transfer functions, optimizers, and many tools to simplify the work with images and text.

The model for recognizing digits included the input and output layers, as well as one or more hidden layers. The model was trained using a variable number of steps (epochs). Recognition images to be scaled to a size of 28x28 (784 cells in a one-dimensional representation), so the number of neurons in the input and hidden layers was assumed to be 784.

The output level is a layer with 10 nodes `tf.nn.softmax`, which returns an array of ten probability estimates, the sum of which is 1. Each node contains an estimate that indicates the likelihood that the current image belongs to one of 10 classes.

For recognition, each image of the digit was converted to a size of 28x28, and then fed to the input of a pre-trained neural network.

Variants of structures of the deep learning neural network based on Keras framework, which were used to recognize car number elements, are shown in Fig. 3. The structure of simpler version of a fully connected neural network is shown in Fig.3a. The structure of a more complex convolutional neural network (CNN) is shown in Fig. 3b.

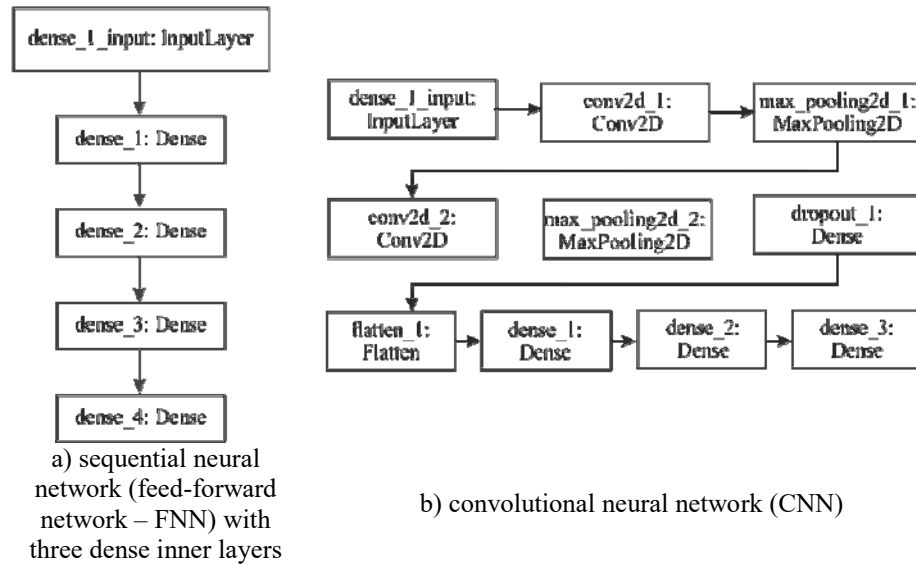


Fig. 3. Variants of structures of the deep learning neural network based on Keras framework, which were used to recognize car number elements

The stages of constructing and training a model and recognition of practical samples are easily separated due to the possibility of exporting models. To do this, the

model.save method (filepath) was used to save the Keras model in one HDF5 file, which contains:

- architecture of the model, allowing to restore the model;
- model weight;
- training configuration (loss calculation function, optimizer)
- the state of the optimizer, which allows to resume training exactly where it was stopped.

To restore the model, the function `keras.models.load_model (filepath)` was further used. This function also allows to build a model using the saved training configuration (if the model has never been compiled).

6 Discussion

The stepwise processing of the input image with an illustration of the achieved results is performed. The processing of ladle car images with numbers 026 and 030 is presented. During identification with different standards, the ladle car image segment with #026 was compared. Examples of original images with a highlighted recognition area are shown in Fig. 4.

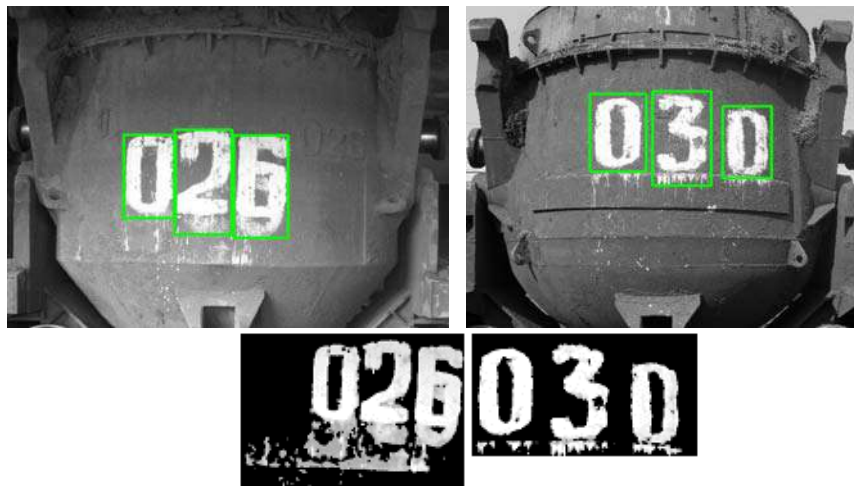


Fig. 4. Examples of original images with a selected recognition area

A comparison of the obtained car segments with all the prototype segments stored in the database was made. When highlighting the elements of the digit, it is clear that due to the distortion of the number by the traces of previous numbers in the image of ladle car #26, the recognition area is slightly increased. When using this recognition method, the resulting area (full image of the digit) was obtained by drawing a rectangle covering all areas with digits. The smallest value of the Hamming distance was used to determine the standard and the corresponding digit in the car image. From Fig. 5, it is seen that the analyzed segment corresponds to the reference #026, i.e.

identification is successful.

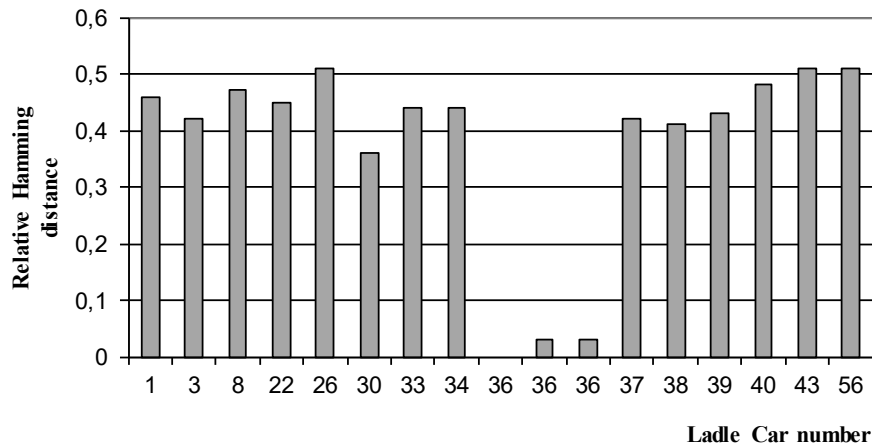


Fig. 5. Relative Hamming distance when comparing ladle car segment #026 and ladle car standards with numbers: № 01, 03, 08, 022, 026, 026, 030, 033, 034, 036, 037, 038, 039, 040, 043, 056

A simpler version of a fully connected neural network in the test set provided a recognition error level of 1.3-1.5%, but in real examples, the recognition accuracy did not exceed 70% (see Fig. 6). For elements of the ladle car number (see the example in Fig. 6), binarization and morphological “closing” operations (MORPH_CLOSE in terms of the OpenCV library) were performed before recognition.

Highlighted igits of number			
Recognized by FNN	0	3	5
Recognized by CNN	0	3	8

a) example of recognition of ladle car number 038

Highlighted digits of number			
Recognized by FNN	0	2	5
Recognized by CNN	0	2	8

b) example of recognition of ladle car number 026

Fig. 6. Examples of recognition of car number digits by various neural networks

The disadvantage of this type of networks is a significantly longer duration of training. On all versions of test photographs of cars with a qualitatively applied numbers, all digits were accurately recognized. The exception was ladle car number

026 (see Fig. 6b), where one of the digits was recognized incorrectly due to a defect in number application.

In the sample of 15 ladle car numbers, which included a total of 42 digits (12 three-digit and 3 two-digit numbers), 41 digits were correctly recognized using a convolutional neural network, so that the recognition accuracy was 97.6%. Recognition errors are connected with poorly written numbers.

When setting up the model according to the MNIST test data or a set of images for further recognition, it was found that the recognition accuracy increases slightly with the increase in the number of hidden layers. The model setup is much more sensitive to the choice of the type and parameters of the optimizer and the number of training epochs.

7 Conclusions

1. Two versions of information technology for recognizing the manually written numbers of hot-metal ladle cars from their photographs were built.
2. A technique to select the area of interest in photographs containing handwritten digits for further recognition, based on the use of the OpenCV library has been worked out.
3. A computer experiment to identify the numbers of car ladles using fuzzy comparisons according to the technology proposed in the article confirmed a rather high processing efficiency. Thus, the relative Hamming distance for comparing images of photo segments of the same car for different conditions was 0.01-0.16, for different cars – 0.26-0.49, which is sufficient to decide whether the proximity estimate is in the range of external variations of conditions for shooting the car number corresponding to the standard or does not belong to this standard. Errors of the first kind are in the range 0-0.07, and errors of the second kind are in the range 0-0.01. The average identification error is 4%.
4. The possibility of using deep learning neural networks constructed with the help of Keras library to solve the problem of recognizing the numbers of hot-metal cars, including handwritten digits, is shown. For handwritten digit recognition, the best recognition accuracy is provided by a convolutional neural network, with which 97.6% of car ladle digits were correctly recognized.

References

1. González, A., Bergasa, L.M.: A text reading algorithm for natural images. *Image and Vision Computing*. **31**(3), 255–274 (2013)
2. Ubozhenko, N.V.: Analysis of the effectiveness of character recognition methods in the framework of the license plate recognition task of vehicles. Prospects for the development of information technologies. **12**, 41–45 (2013)
3. Kachanovsky, Yu.P., Yavtukhovich, A.G.: Development of a license plate localization algorithm for use in a distributed hardware-software complex for license plate recognition. *Information Technologies for Modeling and Control*. **39**, 508–516 (2007)
4. Petrov, S.P.: Convolutional neural network for recognition of license plate symbols of a car. *System analysis in science and education*. **21**, 66–73 (2013)

5. Sanver, U.: Identification of train wagon numbers. In: Proceedings of the 2014 IEEE NW Russia Young Researchers in Electrical and Electronic Engineering Conference, pp. 63–68. St. Petersburg (2014)
6. Lisanti, G.A., Karaman, S., Pezzatini, D., Del Bimbo, A.: Multi-Camera Image Processing and Visualization System for Train Safety Assessment. *Multimedia Tools and Applications*. **77**(2), 1583–1604 (2018)
7. Shin, Shi: *Emgu CV Essentials*. Packt Publishing (2013)
8. Chernopyatov, A.V.: Neural network recognition system for railway car numbers. *Perm University Herald. Mathematics. Mechanics. Computer science*. **4** (23), 96–102 (2013)
9. Kazansky, N.L., Popov, S.B.: Distributed system of technical vision for registration of trains. *Computer Optics*. **36**(3), 419–428 (2012)
10. Volotovskiy, S.G., Kazansky, N.L., Popov, S.B., Khmelev, R.V.: Vision system for recognition of railway tank numbers using a modified correlator in the Hausdorff metric. // *Computer Optics*. **27**, 177–184 (2005)
11. Obukhov, A.V., Lyasheva, S.A., Shleimovich, M.P.: Methods of automatic recognition of license plates. *Bulletin of the Chuvash University*. **3**, 201–208 (2016)
12. Surekha, P., Pavan Gurudath, P., Prithvi, R., Ritesh Ananth, V.G.: Automatic license plate recognition using image processing and neural network. *ICTACT journal on image and video processing*. **8**(4), 1786–1792 (2018)
13. Hashmi, S., Kumar, K., Khandelwal, S., Lochan, D., Mittal, S.: Real Time License Plate Recognition from Video Streams using Deep Learning. *International Journal of Information Retrieval Research*. **9**(1), 65–87 (2019)
14. Björklund, T., Fiandrotti, A., Annarumma, M., Francini, G., Magli, E.: Automatic license plate recognition with convolutional neural networks trained on synthetic data. In: Proceedings of the 2017 IEEE 19th International Workshop on Multimedia Signal Processing (MMSP), Luton, UK, 16–18 October 2017, pp. 1–6 (2017)
15. Khan, J.A., Shah, M.A., Wahid, A., Khan, M.H., Shahid, M.B.: Enhanced car number plate recognition (ECNPR) system by improving efficiency in preprocessing steps. In: 2017 International Conference on Communication Technologies (ComTech), pp. 156–161. Rawalpindi (2017)
16. Lofrano Corneto, G.A., Assis da Silva, F., Pereira, D.R., et al.: New Method for Automatic Vehicle License Plate Detection. *IEEE Latin America Transactions*. **15**(1), 75–80 (2017)
17. Schmidhuber, J.: Deep Learning in Neural Networks: An Overview. *Neural Networks*. **61**, 85–117 (2015)
18. Mikolov, T., Karafiat, M., Burget, L., Cernocky, J., Khudanpur S.: Recurrent neural network based language model. In: 11th Annual Conference of the International Speech Communication Association. Japan, pp. 1045–1048 (2010)
19. Nikolenko, S., Kadurin, A., Arkhangelskaya, E.: *Deep Learning*, Peter, St. Petersburg (2018)
20. Vyugin, V.V.: *Mathematical foundations of the theory of machine learning and forecasting*. Moscow (2013)
21. Arif, R.B., Siddique, M.A., Khan, M.M., Oishe, M.R.: Study and Observation of the Variations of Accuracies for Handwritten Digits Recognition with Various Hidden Layers and Epochs using Convolutional Neural Network. In: 4th International Conference on Electrical Engineering and Information & Communication Technology (iCEEICT), pp. 112–117. Dhaka, Bangladesh (2018)
22. Siddique, F., Sakib, S., Abu Bakr Siddique, Md.: Recognition of Handwritten Digit using Convolutional Neural Network in Python with Tensorflow and Comparison of Performance for Various Hidden Layers. In: 2019 5th International Conference on

- Advances in Electrical Engineering (ICAEE) (2019). <https://arxiv.org/pdf/1909.08490.pdf>. Accessed 16 Dec 2019
23. Brownlee, J.: Handwritten Digit Recognition using Convolutional Neural Networks in Python with Keras. <https://machinelearningmastery.com/handwritten-digit-recognition-using-convolutional-neural-networks-python-keras/>. Accessed 16 Dec 2019
 24. Arif, R.B., Siddique, A.B., Khan, M.R., Oishe, M.R.: Study and Observation of the Variations of Accuracies for Handwritten Digits Recognition with Various Hidden Layers and Epochs using Convolutional Neural Network. In: 2018 4th International Conference on Electrical Engineering and Information & Communication Technology (iCEEICT), pp. 112–117. Dhaka, Bangladesh (2018)
 25. Isogawa, K., Ida, T., Shiodera, T., Takeguchi, T. : Deep Shrinkage Convolutional Neural Network for Adaptive Noise Reduction. *IEEE Signal Processing Letters*. **25(2)**, 224–228 (2018)
 26. Tavanaei, A., Maida, A.S.: Multi-layer unsupervised learning in a spiking convolutional neural network. In: 2017 International Joint Conference on Neural Networks (IJCNN), pp. 2023–2030. Anchorage, AK (2017)
 27. Sable, A.: Building Custom Deep Learning Based OCR models. <https://nanonets.com/blog/attention-ocr-for-text-recognition>. Accessed 16 Dec 2019
 28. Gonzalez, R., Woods, R.: Digital image processing. 3rd ed., Rev. and add. Technosphere. Moscow (2012)
 29. Serhiienko, A.V.: Method for the localization of informative areas using threshold image processing and clustering. *Railway information and control systems: Collection of scientific papers*, Issue. **4 (101)**, 8–14 (2013)
 30. Kargin, A.A., Serhiienko, A.V.: Identification of the number of iron bucket by image segment by comparing it with the reference segment. *Problems of Information Technologies, Collection of articles, scientific proceedings*. **7**, 114–121 (2010)
 31. LeCun Y. The MNIST database of handwritten digits. <http://yann.lecun.com/exdb/mnist>. Accessed 16 Dec 2019
 32. Grother P.J. Nist special database 19 – handprinted forms and characters database // National Institute of Standards and Technology (NIST), Tech. Rep. 1995. <https://www.nist.gov/srd/nist-special-database-19>. Accessed 16 Dec 2019

Information-Extreme Machine Learning On-Board Recognition System of Ground Objects with the Adaptation of the Input Mathematical Description

Anatoliy Dovbysh^{1[0000-0003-1829-3318]}, Igor Naumenko^{2[0000-0003-2845-9246]},
Mykyta Myronenko^{1[0000-0001-5005-1672]}, Taras Savchenko^{1[0000-0002-9557-073X]}

^{1,3,4}Sumy State University, Rymskogo-Korsakova st. 2, Sumy, 40007, Ukraine,
²Scientific-research center of missile troops and artillery, Gerasim Kondratyev st, 165, Sumy,
400021, Ukraine,
¹a.dovbysh@cs.sumdu.edu.ua, ²790905@ukr.net, ³nikitam1996@ukr.net,
⁴taras.savchenko01@gmail.com

Abstract. The method of machine learning of the on-board unmanned aerial vehicle system for autonomous recognition of natural and infrastructural terrestrial objects with the optimization of the frame size generated by the optical digital image of the region is considered. The problem of information synthesis of the onboard recognition system is being solved in the framework of information-extreme intellectual data analysis technology. As part of the functional approach to modeling the cognitive processes inherent in man in the formation and making of classification decisions, the method of information-extreme machine learning onboard recognition system, which allows to adapt the input mathematical description to the maximum full probability of correct recognition of terrestrial features. In accordance with the proposed categorical model, an algorithm of information-extreme machine learning was developed and programmatically implemented, during which the parameters of the system functioning according to the modified Kullback information criterion are optimized. Implementation of the proposed machine learning algorithm was carried out on the example of recognition of natural and infrastructural land objects, among which roads were considered as a zone of interest.

Keywords. extreme machine learning, onboard recognition system, digital image of the region, information criterion, optimization.

1 Introduction

The main way to increase the functional efficiency of onboard systems of unmanned aerial vehicles designed for autonomous recognition of ground objects is the application of ideas and methods of machine learning [1,2]. Thus there is a need to overcome a number of complications of a scientific and methodological nature, due to the following main reasons:

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

- arbitrary initial conditions for the optical formation of digital images of terrestrial objects;
- significant intersection of the recognition classes that characterize terrestrial objects;
- the multidimensionality of the feature dictionary and alphabet of recognition classes;
- strict requirements for the efficiency of machine learning and the operation of onboard recognition system (ORS) directly in working mode.

Overcoming the above complications depends on both the functionality of the simulator, close to real conditions, and the method of machine learning ORS. In the works [3 – 5], the detection of terrestrial objects is carried out according to its dimensions using descriptive methods or the determination of special (invariant) points by spline methods of different order. The main disadvantage of such methods is poor results with the same or approximate geometric parameters of the objects. It is clear that in order to increase the functional efficiency of machine learning, the ORS needs to increase the informativeness of the recognition feature dictionary by scanning the entire digital image of the object. But it is important to choose the method of machine learning ORS, capable of analyzing large amounts of data at high power alphabet a priori fuzzy recognition classes. The main disadvantage of applying neural structures is the dependence of their functional efficiency on the many dimensions of data and the flexibility to retrain [6, 7]. Therefore, a promising area of informational synthesis of ORS is the application of ma-

chine learning methods built within the functional approach to modeling the cognitive processes inherent in humans in the formation and decision making of classification decisions. This area includes methods developed in the framework of the so-called information-extreme intellectual technology (IEI technology) data analysis, which is based on maximizing the information capacity of the recognition system in the process of machine learning [8, 9]. In the methods of information-extreme machine learning, practical invariance of the many dimensions of data is ensured by the construction of decisive rules within the geometric approach. In addition, the decision rules constructed in the framework of the geometric approach are characterized by high efficiency of making classification decisions, which is especially relevant for the ORS. Extreme machine learning by analogy with human learning is considered as a process of optimization of parameters that affect its functional efficiency. In this case, the information criterion of optimization as a measure of diversity is considered as a generalized criterion for the validity of partitioning of the feature space into recognition classes, since it is functional both from a remote measure of proximity and from the exact characteristics of the classification decisions. But the functional efficiency of information-extreme machine learning, as well as other methods, is significantly influenced by the method of forming an input mathematical description of the recognition system. The paper [10] considered the problem of information-extreme machine learning of the onboard system for recognizing terrestrial objects at a given frame size of the digital image of the observed region. The article proposes a method of information-extreme machine learning of the ORS with the adaptation of the frame

sizes formed by the optical channel of digital image of the region to the maximum likelihood of making the right classification decisions.

2 Formulation of the problem

Let's consider the scheme of algorithm of information-extreme machine learning with optimization of the size of frames of digital image of the region according to the procedure (3): Consider the formalized formulation of the problem of information synthesis capable to learn ORS with optimization of the size of frames of digital image of the region. Let the alphabet of the $\{X_m^o \mid m = \overline{1, M}\}$ classes of recognition characterize the frames of the image of natural and infrastructural objects. For each recognition class, a three-dimensional training matrix $\|y_{m,i}^{(j)}\|$ of the pixel brightness of the receptor frame of frames is formed, in which a row $\{y_{m,i}^{(j)} \mid i = \overline{1, N}\}$, where N is the number of recognition features, is a structured vector of the recognition class features X_m^o , and the column of matrix $\{y_{m,i}^{(j)} \mid j = \overline{1, n}\}$ is a random learning sample of the i -th values n . In accordance with the concept of IEI technology, the input training matrix Y is transformed into a working binary matrix X , which is changed by optimal coding of recognition features by the level of control tolerances in the process of machine learning. Therefore, in the binary Hamming space is given a vector of parameters of operation that affect the functional efficiency of machine learning ORS to recognize the vectors of features of recognition class X_m^o :

$$g_m = \langle x_m, d_m, \delta, r \rangle, \quad (1)$$

where x_m is a vector averaged over the ensemble of realizations, the apex of which defines the center of the hyperspherical container of recognition class X_m^o and the size of which is determined by the number of recognition features; d_m is the radius of the hyperspherical container of recognition class X_m^o , $d_m < d\{x_m \oplus x_c\}$, where $d(x_m \oplus x_c)$ is the inter-center distance between vector x_m and vector x_c of the nearest neighboring class X_c^o ; δ is a parameter whose value is equal to half the symmetric field of control tolerances for recognition features, $\delta < \delta_H / 2$, where δ_H is the normalized tolerance field for recognition features; r is the side of a square frame of a digital image of a region.

It is necessary to optimize the parameters of the vector (1) in the process of machine learning of the ORS, which provide the maximum value of the information optimization criterion in the working (allowable) region of determining its function:

$$\overline{E}^* = \frac{1}{M} \sum_{m=1}^M \max_{G_E \cap \{k\}} E_m^{(k)}, \quad (2)$$

where $E_m^{(k)}$ is the value of the information criterion for optimization of vector parameters calculated at the k -th step of machine learning (1); G_E – work area of calculation of information criterion; $\{k\}$ – set of learning steps.

When operating the ORS in exam mode, it is necessary to check the functional effectiveness of machine learning.

Thus, the problem of information synthesis capable of learning ORS is to optimize the parameters of its machine learning by approaching the global maximum of information criterion (2) to its maximum limit value.

3 Mathematical model of machine learning

Within the functional approach to the modeling of cognitive processes, a categorical model of information-extreme machine learning is constructed in the form of an oriented graph. The input mathematical description of the categorical model is presented as a structure

$$I_B = \langle G, T, \Omega, Z, K, Y, X; f_1, f_2 \rangle,$$

where G is the set of factors that influence the ORS; T is the set of moments in which information is received; Ω – recognition features space; Z – the space of states of the system, (alphabet of recognition classes); K – multiple frames of digital image of the region; Y – the input training matrix of the pixel brightness of the image frame; X – working binary learning matrix; f_1 is the operator of forming the input training matrix Y ; f_2 is the operator of converting the matrix Y into a working binary matrix X .

A categorical model of information-extreme machine learning of the ORS with the optimization of the coordinates of the structured vector (1) is shown in Fig. 1.

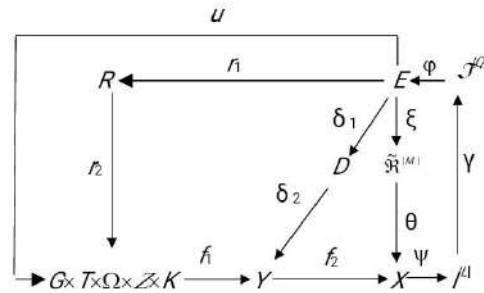


Fig. 1. Categorical model of machine learning of the ORS

In Fig. 1 Cartesian product $G \times T \times \Omega \times Z \times K$ sets the test universe, which is the source of information. The thermal set E of the information criterion values is common to all contours of machine learning parameter optimization. Operator

$\xi: E \rightarrow \tilde{\mathfrak{R}}^{[M]}$ builds at each step of learning partition $\tilde{\mathfrak{R}}^{[M]}$, which is displayed by operator Θ on the distribution of binary feature vectors. Next, operator $\psi: X \rightarrow I^{[S]}$, where $I^{[S]}$ is the set of S hypotheses, tests the main statistical hypothesis $\gamma_1: x_{m,i}^{(j)} \in X_m^o$. The operator γ determines the set $\mathfrak{Z}^{[Q]}$ of the precision characteristics of the classification solutions, where $Q = S^2$, and the operator $\wp$ calculates the set of values E of the information optimization criterion, which is a functional of the precision characteristics. The loop of optimization of control tolerances for recognition features is closed through a term set of D elements which is the value of the system of control tolerances for recognition features. An optimization loop that includes a term set of R values of the sides of a square image frame optimizes the area of the frame. The operator r_1 resizes the frame, and the operator r_2 resizes the space of the recognition features. The operator u governs the process of machine learning.

4 Information-extreme machine learning algorithm

According to the categorical model (Fig. 1), the information-extreme algorithm of machine learning of the ORS with optimization of the frame size of the digital image of the region will be presented in the form of an iterative procedure for finding the global maximum of alphabet-averaged classes of recognition of the information criterion in the working (allowable) area of determining its

$$r^* = \operatorname{argmax}_{G_r} \{ \max_{G_\delta} \{ \max_{|G_E \cap \{r\}|} \bar{E}^{(r)} \} \} \quad (3)$$

where r^* is the optimal side size of the square frame of the digital image; $\bar{E}^{(r)}$ – the information criterion for optimization of vector parameters calculated on the r -th step of machine learning averaged by the alphabet of recognition classes (1); G_δ is a valid value range for parameter δ of the tolerance field.

The input to the machine learning algorithm is a training matrix in the form of a three-dimensional array of feature vectors $\{y_{m,i}^{(j)}\}$ and a system of fields of normalized tolerances $\{\delta_{H,i}\}$ for recognition traits that specifies the range of values of the corresponding control tolerances.

The main stages of information-extreme machine learning of the ORS are:

1. definition for a given alphabet of recognition classes $\{X_m^o\}$ of base class X_1^o , relative to which a system of control tolerances for recognition features is specified;
2. optimization by the information criterion (2) of the parameters of machine learning of the ORS, which are included in the structure (1);
3. construction of decisive rules for optimal geometric parameters of containers of recognition classes, which were restored in the process of machine learning in the radial basis of the binary space of recognition features;
4. check the functional efficiency of machine learning ORS in exam mode.

The definition of base class X_1^o is made according to the scheme:

1. the counter of base recognition classes is reset: $b := 0$;
2. the counter of base recognition classes is initialized: $b := b + 1$;
3. the counter of the recognition classes is reset: $m := 0$;
4. $m := m + 1$;
5. the counter of steps of changing the radii of containers of recognition classes is reset: $k := 0$;
6. the averaged y_m vector is defined for array $\{y_{m,i}^{(j)}\}$;
7. if $b = b + 1$ and $m = m + 1$, then $y_m := y_b$, that is, vector y_m is taken as the base and the point 8 is fulfilled, otherwise – the point 9;
8. are calculated for each i -th sign of vector y_b lower $A_{HK,i}[b]$ and upper $A_{BK,i}[b]$ control tolerances by the formulas

$$A_{AK,i}[b] = y_{b,i} + \delta ,$$

9. where $y_{b,i}$ is the value of the i -th sign of the average vector of class y_b recognition X_m^o ;

10. formed a three-dimensional array of binary learning matrix $\{x_{m,i}^{(j)}\}$, elements of which are calculated by the rule

$$x_{m,i}^{(j)} = \begin{cases} 1, & \text{if } A_{HK,i}[b] < y_{m,i}^j < A_{BK,i}[b]; \\ 0, & \text{if else.} \end{cases} \quad (4)$$

11. for array $\{x_{m,i}^{(j)}\}$ is determined by the average binary vector x_m ;
12. if $m \leq M$, then item 4, is fulfilled, otherwise – item 12;
13. for the set of vectors $\{x_m\}$, a code distance matrix is constructed and pairs of nearest neighbors for which hyperspherical containers of recognition classes are restored in the process of machine learning;
14. $m := m + 1$;
15. the counter of steps of change of radius of containers of recognition classes is initialized: $k := k + 1$;
16. the information criterion $E_m^{(k)}$ of optimization of machine learning parameters is calculated according to the learning matrices of class X_m^o and its closest neighbor, for example, in the form of a modified Kullback information measure:

$$E_m^{(k)} = \frac{1}{n} \{n - [K_{1,m}^{(k)} + K_{2,m}^{(k)}]\} \log_2 \left\{ \frac{2n - [K_{1,m}^{(k)} + K_{2,m}^{(k)}] + 10^{-p}}{[K_{1,m}^{(k)} + K_{2,m}^{(k)}] + 10^{-p}} \right\}, \quad (5)$$

where $K_{1,m}^{(k)}$ is the number of events in which “their” recognition attribute vectors did not belong to class X_m^o ; $K_{2,m}^{(k)}$ is the number of events in which “alien” feature vectors belonged to class X_m^o ; n is the number of feature vectors in the learning matrix of each recognition class; 10^{-p} is a sufficiently small number that is entered to avoid division by zero (in practice, p is selected from the interval $1 < p \leq 3$).

17. if $k < d(x_m \oplus x_c)$, then item 14 is fulfilled, otherwise – item 17;

18. in the work area G_E is determined by the maximum value $E_m^{*(k)}$ information criterion (5);

19. if $b \leq M$, then item 2 is fulfilled, otherwise – item 19;

20. according to formula (2) the average maximum value $\bar{E}^*$ of criterion is calculated (5);

21. the recognition class for which the value of criterion $\bar{E}^*$ (5) is maximum is taken as the base;

22. STOP.

After determining the base class, the ORS machine learning procedure (3) is started with the optimization of the control tolerance system for the features of recognition and frame size. The main functions of the inner cycle of procedure (3) are:

- mation criterion (5) at each step of machine learning with values of control tolerance field δ and frame size of region image set in corresponding external cycles;
- search for the global maximum of the information criterion for optimization of machine learning parameters in the working (valid) area of determining its function;
- determination of the geometric parameters of recognition classes optimal in the information understanding.

The input for the implementation of procedure (3) is the alphabet of recognition classes $\{X_m^o\}$, in which class X_1^o is the base, the corresponding three-dimensional array of the training matrix $\{y_{m,i}^{(j)}\}$, parameter δ_H , which determines the system of normalized tolerances for recognition features and the maximum permissible size $r_{\max}$ side square frames of digital image of the region.

Consider the scheme of the algorithm of information-extreme machine learning with optimization of the frame size of the digital image of the region by the procedure (3):

1. resetting the step counter of resizing frames in the region image: $r := 0$;
2. $r := r + 1$;
3. reset the counter of recognition classes: $m := 0$;
4. $m := m + 1$;

5. resetting the training frame resizing step counter: $r := 0$;
6. $r : r + 1$;
7. resetting the step counter to change the tolerance field parameter: $\delta := 0$;
8. $\delta := \delta + 1$;
9. calculation of the lower and upper $\{A_{B,i}\}$ control tolerances for recognition signs according to the rules

$$A_{H,i} = y_{1,i} - \delta; \quad A_{B,i} = y_{1,i} + \delta, \quad (6)$$

10. where $y_{1,i}$ is the i -th sign of the averaged vector y_1 of the base recognition class X_1^o ;

11. resetting the counter of the steps of changing the radius of the hyperspherical container: $k := 0$;

12. $k := k + 1$;

13. formed a three-dimensional array of binary learning matrix, the elements of which are calculated by the rule (4);

14. the formation of an array of averaged vectors of signs $\{x_m\}$ whose elements are determined by the rule

$$x_{m,i} = \begin{cases} 1, & \text{if } \frac{1}{n} \sum_{j=1}^n x_{m,i}^{(j)} > \rho_m; \\ 0, & \text{if } \text{else}, \end{cases}$$

15. where ρ_m is the quantization level of the coordinates of binary vector x_m , which by default is 0,5.

16. splitting the set of vectors $\{x_m\}$ into pairs of nearest “neighbors”

$\mathfrak{R}_m^{[2]} = \langle x_m, x_c \rangle$, where x_c is the averaged vector of the nearest neighbor class X_c^o ;

17. the information criterion (2) is calculated;

18. if $k \leq N$, then item 11 is executed, otherwise – item 17;

19. if $\delta < \delta_H$, then item 8 is executed, otherwise – item 18;

20. the maximum value of the criterion in the workspace of determining its function is determined, where the first and second reliability are greater than 0.5;

21. if $m < M - 1$, then item 4, otherwise – item 20;

22. if $r \leq r_{\max}$, then item 6 is executed, otherwise – item 21;

23. determines the global maximum of average information criterion $\overline{E}^*$ in the work area of determining its function;

24. determine the optimal values of parameter δ^* , lower $A_{H,i}^*$ and upper $A_{B,i}^*$ control tolerances for all recognition features and size r^* side square frames of the digital image of the region;

25. STOP.

According to the optimal geometric parameters of the containers of recognition classes obtained in the process of machine learning, productive decision rules were constructed in the form

$$(\forall X_m^o \in \mathfrak{R}^{[M]})(\forall x^{(j)} \in \mathfrak{R}^{[M]})[\text{if } (\mu_m > 0) \& (\mu_m = \max_{\{m\}} \mu_m) \\ \text{then } x^{(j)} \in X_m^o \text{ else } x^{(j)} \notin X_m^o], \quad (7)$$

where $x^{(j)}$ is a recognizable vector; μ_m is a function of belonging to vector $x^{(j)}$ of a recognition class container X_m^o .

In expression (7), the membership function for a hyperspherical container of recognition class X_m^o is determined by the formula [3]

$$\mu_m = 1 - \frac{d(x_m^* \oplus x^{(j)})}{d_m^*},$$

where x_m^* , d_m^* – the optimal parameters of machine learning: averaged feature vector and radius of hyperspherical container recognition class X_m^o respectively.

Thus, in the exam, it is determined by decisive rules (7) that the recognizable feature vector belongs to one of the classes in the given alphabet. At the same time, the decisive rules, built within the geometric approach, are characterized by low computational complexity, which provides high efficiency of making classification decisions in the operation of ORS in the operating mode.

5 Simulation results

An algorithm of information-extreme machine learning of ORS is implemented with the purpose of optimizing the image of the region in understanding the information, the receptor field of which is shown in Fig. 2 [11].



Fig. 2. Image of the region obtained by aerial photography

The input training matrix was formed by processing the image frames shown in Figure 3 in the polar coordinate system according to [12]. As recognition classes, the

frames of the sections shown in Fig. 2 images of the region: class X_1^o – highway; class X_2^o – liquid forest; class X_3^o – plowing field; class X_4^o – sown field.

The selected frames are shown in Fig. 3.



Fig. 3. Picture frames: a – class X_1^o ; b – class X_2^o ; c – class X_3^o ; d – class X_4^o

In the previous stage of machine learning, the results of the implementation of the above algorithm were identified as the base class of recognition X_1^o – highway. Therefore, the system of control tolerances was determined relative to the averaged vector of features of recognition class X_1^o .

In the process of machine learning ORS according to procedure (3), the value of parameter r increased from one to 71 pixels of the receptor field of the video card.

Table 1 shows results of machine learning using side of the square frame optimization.

Table 1.– Results of machine learning

r	35	37	39	41	43	45	47	49	51	53	55
$\bar{E}$	0,59	0,62	0,63	0,64	0,61	0,67	0,70	0,67	0,66	0,64	0,63

The analysis of Table 1 shows that, in the process of machine learning, the optimal frame size of the image value is equal to $r^* = 47$ pixels with a maximum value of $\bar{E}^* = 0,70$ information criterion.

Fig. 4 shows a graph of the averaged normalized information criterion (5) from parameter of control tolerances system δ using optimal side of the square frame

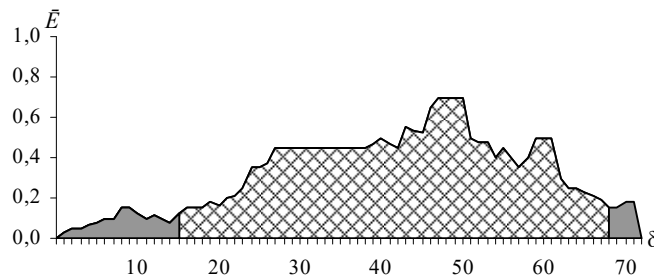


Fig. 4. Graph of information criterion dependence on parameter of control tolerances system

Fig. 4 shows a dark area on the graph of the working (admissible) area of determining the function of the information criterion of optimization, in which the first and second reliability exceed respectively the errors of the first and second kind.

The analysis of Fig. 4 shows that, in the process of machine learning, the optimal value of parameter of control tolerances system is equal to $\delta^* = 50$ (scale of pixel's brightness) with a maximum value of $\bar{E}^* = 0,70$ information criterion. Fig. 5 shows graphs of information criterion (5) dependence on the radii of containers of recognition classes the optimal values of which allow us to construct the decisive rules (7).

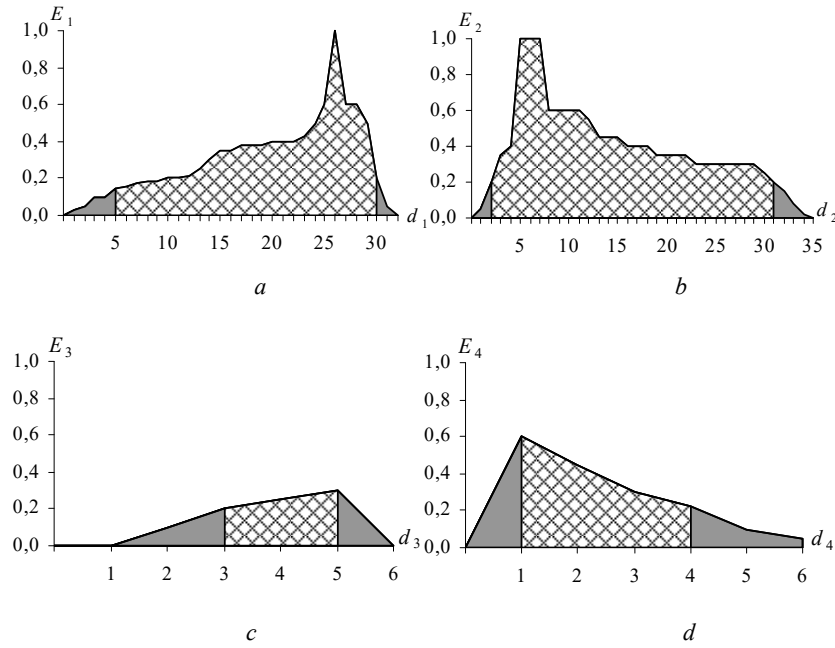


Fig. 5. Graphs of the criterion (5) on the radii of containers of recognition classes: a – class X_1^o ; b – class X_2^o ; c – class X_3^o ; d – class X_4^o

The analysis of Fig. 5 shows that the optimum values of the radii of the containers of the recognition classes are: for the class X_1^o – $d_1^* = 26$ (hereinafter in code units); for class X_2^o – $d_2^* = 6$; for class X_3^o – $d_3^* = 5$ and class X_4^o – $d_4^* = 1$.

Fig. 6 shows a digitized image (Fig. 3), obtained by the results of the identification of frames by the decisive rules (7) with the optimal frame size of the pixel image. The numbers in the frames correspond to the numbers of the recognition classes.



Fig. 6. Results of the identification of frames

Visual analysis of Fig. 6 shows that highways that are of interest for vehicle recognition are identified with a sufficiently high accuracy when the frame size is optimal.

6 Conclusion

1. Within the framework of functional approach to modeling of cognitive processes of formation and decision-making of classification decisions, a catheter model is proposed, which is considered as a generalized structural diagram of the algorithm of information-extreme machine learning of the image recognition system
2. On the basis of the proposed categorical model, an algorithm of information-extreme machine learning of the onboard recognition system was developed with optimization according to the information criterion of the frame size of the digital image of the region. The algorithm allows for a given flight altitude of an unmanned aerial vehicle complex to determine quickly and with high accuracy the area of interest in which the wanted land object may be located.
3. Since the decisive rules are not infallible by the learning matrix, to increase the functional efficiency of ORS, it is necessary to increase the depth of machine learning by optimizing other parameters of operation, including the parameters of the formation of the input mathematical description.

References

1. Muller, K.R., Mika S., Ratsch, G.: An introduction to kernelbased learning algorithms. IEEE Transactions on Neural networks, vol.12(2), pp. 181-202 (2001).
2. Duda, R.O., Hart, P.E., Stork, D.G.: Pattern Classification: second ed. John Wiley & Sons, New York (2001).
3. Jauregi Iztueta, E., Lazkano, E., Martinez-Otzeta, J.M., Sierra, B.: Visual Approaches for Handle Recognition. Springer Tracts in Advanced Robotics, vol. 44, pp.313-322 (2008).

4. Huang, H., Lu, L., Yan, B., Chen, J.: A new scale invariant feature detector and modified SURF descriptor. In: Proceedings Sixth International Conference on Natural Computation (ICNC 2010), Yantai, Shandong, China, pp. 3734-3738 (2010).
5. Moskalenko, V., Moskalenko, A., Korobov, A., Boiko, O., Martynenko, S. Borovenskyi, O.: Model and Training Methods of Autonomous Navigation System for Compact Drones. In: Proceedings Second International Conference on Data Stream Mining & Processing (DSMP), Lviv, IEEE, pp. 503-508 (2018). doi: 10.1109/DSMP.2018.8478521
6. Artale V., Collotta, M., Milazzo, C., et al.: Real-Time System based on a Neural Network and PID Flight Control. Appl. Math. Inf. Sci., vol. 10(2), pp. 395-402 (2016).
7. Jafari, M., Xu., H.: Intelligent Control for Unmanned Aerial Systems with System Uncertainties and Disturbances Using Artificial Neural Network. Drones, vol. 3, pp. 24-36 (2018).
8. Subbotin, S.: The neuro-fuzzy network synthesis and simplification on precedents in problems of diagnosis and pattern recognition. Optical Memory and Neural Networks (Information Optics), vol. 22(2), pp. 97-103 (2013). doi: 10.3103/s1060992x13020082
9. Dovbysh, A.S., Martynenko, S.S., Kovalenko, A.S., Budnyk, M.M.: Information-extreme algorithm for recognizing current distribution maps in magnetocardiography. Journal of Automation and Information Sciences, vol. 43(2), pp. 63-70 (2011). doi: 10.1615/JAutomatInfScien.v43.i2.60
10. Dovbysh, A.S., Moskalenko, V.V., Rizhova, A.S.: Information-Extreme Method for Classification of Observations with Categorical Attributes. Cibernetica and Systems Analysis, vol. 52(2), pp. 45-52 (2016). doi: 10.1007/s10559-016-9818-1
11. The world's most detailed globe, <https://www.google.com.ua/intl/en/earth/>
12. Naumenko, I., Myronenko, M., Piatachenko, V.: Information-extreme learning of on-board system for recognition of ground vehicle. CEUR Workshop Proceedings, p. 121–132 (2019).

Development of a Network Attack Detection System Based on Hybrid Neuro-Fuzzy Algorithms

Olexander Belej¹ [0000-0003-4150-7425], Liubov Halkiv² [0000-0001-5166-8674]

¹ Lviv Polytechnic National University, 5 Mytropolyt Andrei str., Building 4, Room 324, Lviv 79013, Ukraine

Oleksandr.I.Belei@lpnu.ua

² Lviv Polytechnic National University, 12, Stepan Bandera str, 79013, Lviv, Ukraine

Lubov.I.Halkiv@lpnu.ua

Abstract. The imperfection of existing intrusion detection methods, as well as the changing nature of malicious actions by the attacker, make computer systems unsafe, therefore it is important to identify new types of attacks and respond to them on time. The developed hybrid scheme for detecting and classifying network attacks based on a combination of adaptive classifiers. The study proposed a generalized scheme for combining classifiers to detect network attacks. Based on it, a software tool has been developed that allows you to analyze network traffic for the presence of abnormal network activity. To reduce the number of features used, it is proposed to use the method of principal components. The main feature of the proposed approach is a multilevel analysis of network traffic, as well as the use of various adaptive modules in the attack detection process. Computational experiments were carried out on two open data sets using various methods of combining classifiers. The developed modules can be used to process data received from sensors of the information and security events management system.

Keywords: network attack, intrusion detection system, algorithm, signal, traffic, protocol, transmission.

1 Introduction

The detection of network attacks is one of the urgent problems of information security due to the rapid development of computer technology and the imperfection of existing methods for detecting attacks. To correct the situation in this area, the current efforts of researchers are aimed at finding and applying new, including hybrid and adaptive, detection schemes.

An intrusion detection system (IDS) is an important component of computer network protection. Its main task is to monitor the network or system for malicious activity. Although the problem of detecting network attacks is quite old, it is still relevant.

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

A few years ago, network intrusions were rare because they required extensive knowledge of operating systems, network protocols. Today, any user can carry out malicious actions on the network by downloading one of the exploit programs available on the Internet.

One of the most popular areas of development of computer security systems today is the technology of artificial immune systems [1]. It is assumed that the use of methods and techniques for protecting the body from germs and viruses will overcome the problems of classical means of ensuring information security.

In this paper, an attempt is made to create an intrusion detection system based on the immune model. Based on the architecture of the IDS developed by Kim and Bentley, which is called the immune model of intrusion detection [2]. Among others, it most closely mimics the behavior of a biological immune system. However, according to the authors of this article, some properties of immunity are absent in the Kim and Bentley model, which negatively affects its effectiveness.

The detection of network attacks is one of the urgent problems of information security due to the rapid development of computer technology and the imperfection of existing methods for detecting attacks. To correct the situation in this area, the current efforts of researchers are aimed at finding and applying new, including hybrid and adaptive, detection schemes.

Formal problem statement

In this study, an attempt is made to identify network attacks based on the analysis of connections characterizing the activity of remote hosts. In this case, monitoring network activity usually consists of the following steps. At the first stage, the filtering of incoming and outgoing packets collected with the help of sensor agents installed by controlled network nodes is performed. At this stage, packet aggregation also takes place to form signs of an established session concerning a unique pair of sockets. At the next stage, the data collected from the first level is stored in a repository for their subsequent processing and interpretation. At the last stage, the administrator is notified of the detected changes in the security policy and offers possible solutions to prevent the further development of the invasion.

Network IDSs are based on the analysis of network traffic that passes through the sensors and, in the event of anomalies, is sent to the analyzers for further decision making. In large information systems (IS), the problem of a large amount of network traffic arises, since standard approaches to data processing are no longer effective. To increase the efficiency of identifying situations associated with a possible intrusion, recently, modern technologies of data mining have been widely used. One of the most effective approaches to classifying a large amount of data is the use of neural networks. This method allows you to detect not only already known network attacks, but also to identify new ones.

The main task of processing network data lies with the intrusion detection subsystem, which is responsible for analyzing network traffic. The effectiveness of IDS directly depends on the chosen method of constructing an algorithm for detecting attacks and processing a large amount of network data. The aim of this work is to develop an intrusion detection algorithm by processing and filtering a large amount of network traffic, based on an artificial neural network.

2 Literature review

Currently, many works investigate the applicability of the methods in question to detect attacks. We will present some of these works in more detail. As a training set for artificial neural networks, various data can serve. So, in [3, 4], to solve this problem, it is proposed to use a multilayer neural network trained on data from a variety of Snort signatures. In works [5, 6] many KDD Cup records are used as training data.

Modeled connections are another source of data. An example of a work in which the author used a three-layer direct distribution neural network as a binary classifier of network connections is an article [7]. These works are united by the general idea of using one classifier aimed at identifying illegitimate traffic.

Other authors [8, 9] propose a model of the immune system with the life cycle of T-lymphocytes, which took into account the process of their maturation in the thymus, as well as activation, the formation of immune memory cells through a stimulation signal and the death of immune cells. The lightweight immune system (LISYS) developed is designed to detect intrusions in a distributed environment. Although this system has several advantages, including relatively low computational overhead, reliability, and scalability, the authors note some of the disadvantages. The inability to detect network attacks using the UDP protocol and the ability to trick the system by conducting distributed attacks over time such as slow-scanning [10]. It provides an overview of a two-level intrusion detection system, which combines the advantages of Kohonen's immune systems and networks. At the first stage, the signs of network connections are filtered using immune detectors trained by the negative selection method, thereby eliminating those samples that correspond to normal connections. In the second stage, abnormal instances are processed by self-organizing cards and grouped into separate clusters with similar features. It also presented a two-level model in which immune systems were used to detect anomalies, and neural networks with the principal component method were used to detect abnormalities [11].

The application of neuro-fuzzy systems to intrusion detection problems was discussed in [12]. Their authors used the parameters of the KDD Cup 99 and DARPA 1998 datasets as input for fuzzy classifiers based on the fuzzy inference.

There are also various hybrid approaches, for example, a combination of the apparatus of immune systems and neural networks [13]. Multilayer neural networks that are generated using the clonal selection method are selected as immune detectors. The experiments were carried out on the KDD Cup 99 dataset; they proved the high ability of the detectors to adapt to new types of attacks.

Another hybrid solution to intrusion detection is to combine several neural networks into a single classifier. So, work [14] is devoted to the application of the IDS method and radial basis networks for the classification of records from the NSL-KDD data set. The final classifier is presented as a composition of classifiers sequentially built on different samples and simple voting procedures. As a result, the classification level was increased by about 1.6% compared with the classifiers taken separately. Other authors [15] propose submitting output values from several neural networks trained by various algorithms to the input of the weighted voting and majority voting. On a test sample of 6890 samples, a classification accuracy of over 99% was

achieved.

Article [16] describes a two-level scheme for detecting and classifying attacks. Several adaptive neuron-fuzzy modules are combined. Each of them is designed to detect only one class of compounds, processes the parameters of the KDD Cup 99 records. The final classification is performed by the fuzzy decision module, which implements the neuro-fuzzy inference system with two membership functions. This module is responsible for determining how abnormal the record being processed is. Its class corresponds to the class of the fuzzy module of the first level with the highest output value.

Previous work [17, 18, 19] has focused more on the transmission of chaotic signals using dynamic models. The offered algorithm of transfer images can be used for the decision of control tasks for integrity data and transfer of the information in modern telecommunication and information systems [17]. In the article [18] considered an algorithm of constructing an electronic-digital message signature based on encryption using elliptic curves. The article [19] is devoted to the calculation of the characteristics of dynamic chaos based on the traffic of the corporate computer network.

This article proposes an approach that develops the considered work. It is based on the implementation of multi-level and adaptive detection of network attacks through a combination of neural networks. Adaptive detectors perform parallel processing of input features of connections. The detectors are trained on various data samples and are designed to classify a single compound.

3 Hybrid signature model and adaptive approach

The proposed hybridization scheme for signature and adaptive approaches is presented in Fig. 1. To analyze packets and generate session parameters, the IDS system was used additional scripts were added for processing network events. The system has a built-in script interpreter that provides processing of various network events.

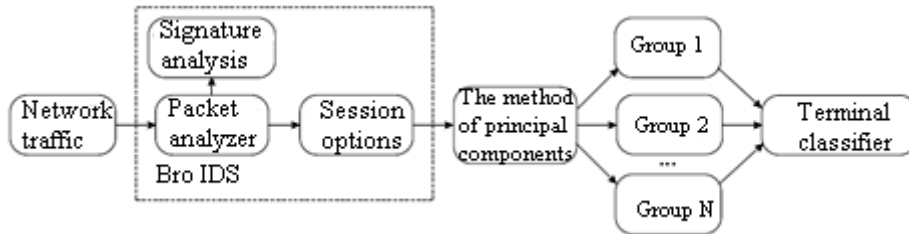


Fig. 1. Hybrid signature model and adaptive approach.

When creating a record about each formed connection, the parameters necessary for the classification of compounds are selected and calculated from the data field and header of each packet. Statistics are also collected on currently open connections. Optionally, to reduce the dimension, the input vector is compressed by the principal component method. Each group of detectors trained to recognize one particular type

of connection processes the last set of features. The final classification result is issued by the terminal classifier.

Within this scheme, several attack detection techniques can be used. The first of them implies that each group consists of several classifiers of the same type, which are trained on different samples from the training set. This allows you to create detectors that differ from each other in adjusted weights, and thereby increase the quantitative indicators of attack detection. The second technique involves the use of heterogeneous classifiers within each group. Here, instances of each of the three models are used to recognize a specific type of attack. In both cases, the output values from all classifiers can be considered as an input vector for the terminal classifier.

Let's take a closer look at the attack detection process. It can be divided into three stages.

At the first stage, the IP-defragmentation of raw packets and the assembly of TCP session segments are performed. Each connection (session) is a sequence of TCP packets for a certain time interval, during which data is transferred between a pair of remote hosts $\langle \text{addr\_src}, \text{port\_src} \rangle$ and $\langle \text{addr\_dst}, \text{port\_dst} \rangle$ using a specific protocol. In the process of analyzing network traffic, packets that initiate the start of a session (SYN) and serve as a sign of its termination (FIN, RST, timeout) are monitored. Each session can be characterized as a set of parameters, the elements of which are conditionally divided into two groups: attributes obtained from the packet headers, and statistical data. Also, this stage is responsible for the initial detection of network anomalies by verifying that the contents of individual packets match the specified regular expressions in the signature set.

The second step involves applying the principal component method and converting the output parameters of each session into a compressed set of attributes. Each element of the new vector is a linear combination of the elements of the old vector with coefficients, which are elements of the eigenvectors of the covariance matrix of the source data.

The third stage is the application of several groups of adaptive classifiers. Each group consists of detectors responsible for recognizing one particular type of connection. To increase the classification rate, each detector processes an incoming set of parameters in parallel with the others.

A terminal classifier can be represented in several ways. The simplest of them is the majority voting procedure. In this case, the connection class is the one for which a larger number of detectors within a certain group voted. Another method for implementing the terminal classifier is a weighted vote and its modification. In this case, each classifier will be assigned an appropriate coefficient assigned to it in the learning process. This will allow more reasonable use of one or another classifier, depending on its indicators of the correctness of recognition of attacks calculated on the samples of the training set. The next approach is to choose the best classifier for each particular record and ignore the rest. The main difficulty here is the construction for each classifier of a given area of competence in the space of attributes. It is also proposed to use output values from first-level classifiers as a training set for a terminal classifier. This technique, known as a multi-tiered generalization, can be built based on data mining methods.

To achieve the goal, it is proposed to use three artificial neural networks (ANNs) with various topologies: a multilayer perceptron (MLP), a network of radial basis functions (RBF) and a self-organizing Kohonen map (SOM). The same traffic is sent to the input of each ANN for the analysis of IP packet headers together with an inaccurate search for intrusion signatures directly in the applied contents of IP packets. To train an ANN, it is necessary to use both packets containing known signatures and normal IP datagrams [8]. In the process of detection, each ANN determines how much the network packet submitted to the input corresponds to a normal or abnormal situation.

At the output of each ANN, the results are presented as a multidimensional vector, each coordinate of which corresponds to one of the types of attacks. If the ANN detects an attack, then the corresponding coordinate takes a single value, otherwise, it takes zero.

The use of three ANNs with different topologies necessitates the conversion of network traffic to a single format so that each of them can equally analyze the data supplied to it at the input. The most suitable format is NLS-KDD, which is a standard dataset for IDS. The NLS-KDD dataset contains raw network packets that are independent of the operating systems used. In total, this data set includes 4898431 samples. NLS-KDD includes 37 different network attacks that can be divided into four large groups: denial of service, privilege escalation, remote access, and network scanning. All types of network attacks are presented in Table 1.

Table 1. Types of network attacks from the NLS-KDD dataset

Group of attacks	Attacks
Denial of service	Back, Land, Neptune, Pod, Smurf, Teardrop, Mailbomb, Processtable, Udpstorm, Apache2, Worm
Network scan	Satan, IPsweep, Nmap, Portsweep, Mscan
Getting Remote Access	Guess_password, Ftp_write, Imap, Phf, Multihop, WarezmaterXlock, Xsnoop, Snmpguess, Snmpgetattack, Httpunnel, Sendmail, Named
Privilege escalation	Buffer_overflow, Loadmodule Rootkit, Perl, Sqlattack, Xterm, Ps

The NLS-KDD format contains 41 fields. Bringing IP packets to NLS_KDD consists of extracting some features from network traffic. Such signs include connection duration, connection protocol, connection type, service and destination port, successful connection flag, number of failed authorization attempts, number of connection requests with administrator rights. The process of converting network packets to the NLS-KDD format is shown in Fig. 2.

Upon completion of converting network data to a single NLS-KDD format, they are fed to the input of the ANN. After analyzing the next portion of network traffic, the results of the ANN are recorded in multidimensional vectors. The resulting vector is found as the sum of the output vectors of the three ANNs. If the coordinate is zero or one in the resulting vector, then there is no attack. A unit means that a false positive occurred on one of the ANNs. If the value of the vector coordinate is 2 or 3, then an attack of the corresponding type has occurred. In Fig. 3, the process of obtaining the resulting vector is simplified.

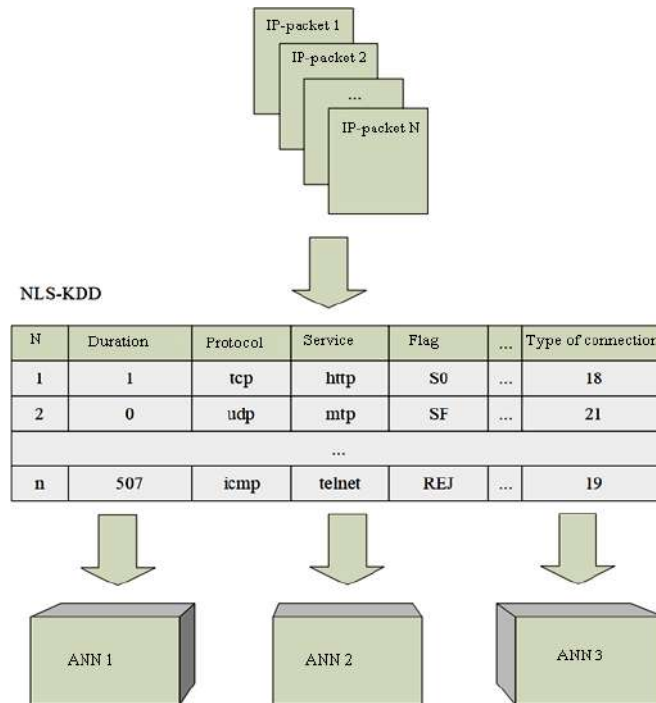


Fig. 2. The process of converting network packets to the NLS-KDD format for further submission to the input of the ANN of various topologies.

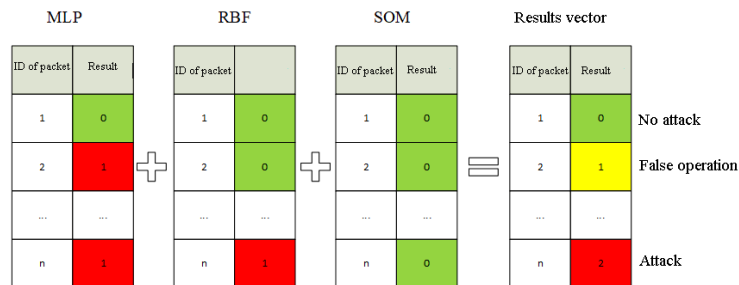


Fig. 3. Simplified representation of the addition of resulting vectors.

Thus, the algorithm of IDS (Fig. 4), which processes and filters a large amount of network traffic based on the ANN, will consist of the following: receiving suspicious network traffic from the sensor subsystem; converting network packets to the NLS-KDD format; analysis of a portion of n network packets using three neural networks with different topologies of MLP, SVM, and SOM; analysis results are entered into multi-dimensional vectors; the resulting vector is obtained by adding the three output vectors MLP, SVM, and SOV. After that, it is transferred to the data presentation subsystem.

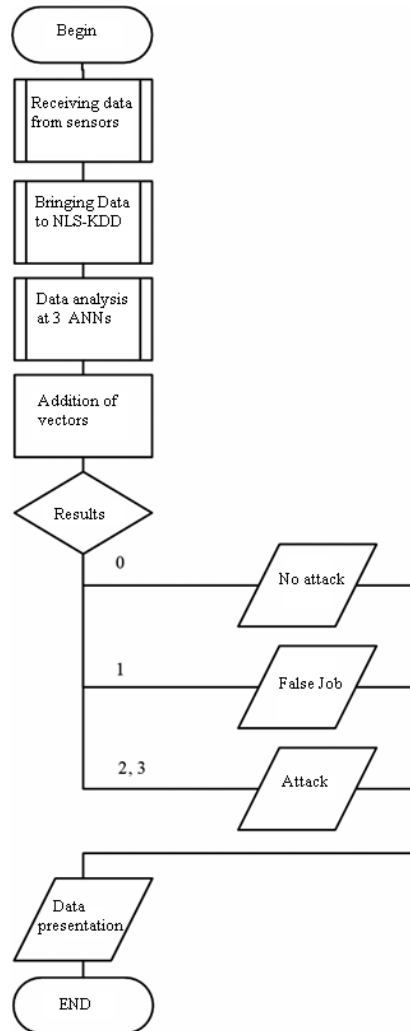


Fig. 4. Algorithm for detecting network attacks using three ANNs with different topologies.

The capabilities of the proposed algorithm can be expanded both by increasing the number of ANNs used and by using other types of ANNs depending on the task.

4 Experiments and results

To test the described models, two open data sets were selected: KDD Cup 99 and NSL-KDD. For these sets, each record represents an image of a real network session, described as a set of 41 parameters, and is marked as an attack or normal connection. The developed intrusion detection system monitors the level of data obtained from packet headers and statistical information generated by the sliding window method.

Therefore, 29 parameters were selected that satisfy this requirement. Also, for the classification of connections, 5 types of DDoS attacks and 4 types of attacks such as port scans were selected. The data received from the Bro system is processed by modules that implement a mechanism for classifying compounds using the proposed detectors. The following numerical indicators were selected to test the effectiveness of the functioning of the models:

1. Percentage of samples of normal compounds recognized as attacks (false positive):

$$FP = \frac{n_{FP}}{n_{FP} + n_{TN}} \cdot 100\% . \quad (1)$$

2. The percentage of correctly recognized samples of abnormal compounds (true positive):

$$TP = \frac{n_{TP}}{n_{TP} + n_{FN}} \cdot 100\% . \quad (2)$$

3. The percentage of network connection patterns whose class was correct classification:

$$CC = \frac{n_{CC}}{n} \cdot 100\% . \quad (3)$$

The threshold value for immune detectors was chosen experimentally in such a way that

$$TP - F + CC \rightarrow \max \quad (4)$$

was fulfilled on the training data.

The experimental results for each approach with the indicated values of these parameters, which were used as the terminal classifier by the majority voting procedure, are presented in Table 2.

Table 2. Indicators FP, TP, CC

An approach	KDD Cup 99			NSL-KDD		
	FP	TP	CC	FP	TP	CC
Neural networks	3,56	98,95	73,56	7,08	98,24	56,86
Immune detectors	2,26	91,16	94,82	12,51	93,06	65,65
Neuro-fuzzy classifiers	11,65	98,29	88,89	16,94	96,37	83,06
Combination of approaches	1,31	99,98	77,04	5,11	99,85	57,96

As can be seen from the results obtained, neural network and neuron-fuzzy classi-

fiers, in comparison with immune detectors, have a greater number of correctly recognized compounds when analyzing network parameters. So these classifiers are a more preferable tool for detecting network attacks. At the same time, immune detectors due to the dynamic configuration of the balance can modify their structure in response to detected attacks. Therefore, in the process of analyzing network traffic, the performance of recognition of network attacks by immune detectors increases over time. This evolutionary mechanism, as well as an updated set of training data, contributes to the recognition of previously unknown types of attacks.

Among the proposed three models, neural networks showed the best degree of recognition of network connection patterns. Neuron-fuzzy classifiers are characterized by a long learning process. This is explained by the complexity of the calculations associated with the multi-level structure of the network, and the setting of the parameters of membership functions in fuzzy rules. Nevertheless, the attack detection rate for these classifiers is high and comparable to the indicator of neural networks.

To evaluate the effectiveness of the proposed intrusion detection approaches, a series of experiments was conducted. The KDD Sup 99 database was used to train and test neural network models. This is one of those few intrusion detection bases that has attracted the attention of researchers due to its well-designed structure and accessibility. To study the characteristics of the proposed systems, we set three main indicators: the share of detected, the share of recognized attacks for each class and the number of false positives of the system. The share of detected attacks is defined as the number of attack images of a particular class detected by the system divided by the total number of records of attacks of this class in the database. Similarly, the share of recognized is determined. False positives indicate the total number of images of normal network operations that are erroneously classified as attacks.

The results of testing the recognition of an attack class for the operation of an intrusion detection system for sensor wireless networks are shown in Table 3.

Table 3. Model 1 test results for intrusion detection

Class	Total	Discovered	Recognized
DoS	391458	391441 (99.99%)	370741 (94.71%)
U2R	52	48 (92.31%)	42 (80.77%)
R2L	1126	1113 (98.85%)	658 (58.44%)
Probe	4107	4094 (99.68%)	4081 (99.37%)
Normal condition			
Normal	97277	---	50831 (52.25%)

Thus, the best result was achieved for DoS and Probe attacks. U2R and R2L are determined slightly worse, respectively, 80.77% and 58.44%. Also, there is a percentage of false positives for the system. A summary of each of the options for building an attack detection system is shown in Table 4.

Thus, model 3 is characterized by high accuracy (93.21%) and the least number of false positives. Using model 1, 86.3% of the input images were recognized, and model 2, 92.97%. Models 2 and 3 can be successfully used to work with large sets of complex data structures.

Table 4. Test results of proposed intrusion detection models in intrusion detection systems for sensor wireless networks

Model	Detected attacks	Recognized Attacks	False positives	Total share recognized, %
Model 1	396696 (99.98%)	375522 (94.65%)	46446 (47.75%)	86.30%
Model 2	395949 (99.80%)	375391 (94.61%)	13398 (13.77%)	92.97%
Model 3	396549 (99.95%)	375730 (94.70%)	12549 (12.90%)	93.21%

Also, for each type of attack, the detection rates of various detectors were calculated. So, Table 5 shows indicators of the effectiveness of recognition of attacks on a set of KDD Cup 99 combined classifiers. In the presented table, the left column is the type of connection, the top row is the type of classifiers, their intersection is the percentage of correctly recognized attacks.

Table 5. Attack recognition performance indicators with combined classifiers for sensor wireless networks, %

	back	neptune	pod	smurf	teardrop	ipsweep	nmap	ports-p	satan
back	100	0	0	0	0	0	0	0,46	3,37
neptune	0	99,98	0	0	0	0	0,56	99,99	99,93
pod	0	0	60,98	0	34,09	1,89	1,89	100	37,88
smurf	0	0	0	99,92	0	0	0,05	0	0,09
teardrop	0	0	56,63	0	100	0,2	0	100	100
ipsweep	0	0	0,4	0	0	100	91,95	1,69	0,4
nmap	0	0	0,43	0	0	44,16	100	44,59	44,59
ports-p	0,1	3,09	0,1	0	0	67,6	0,39	100	89,68
satan	0	88,64	0	0	0	9,28	0,25	88,26	99,87
normal	0,06	0,03	0,05	0,03	0	0,1	0,44	0,25	0,56

By combining individual classifiers, it was possible to increase the detection rates by 3.85 and 3.96% compared with the average value of this value for individual classifiers for KDD Cup 99 and NSL-KDD, respectively. At the same time, the indicator of correct classification remained above this indicator for neural networks.

The constructed intrusion detection system can be used in real computer networks after preliminary training on the traffic characteristic of this network. In particular, at present, the study is aimed at conducting experiments related to the newly generated data. For this purpose, a software stand was created to simulate the computer network of a large enterprise used for experiments with the developed SIEM system, taking into account various vulnerabilities in software and hardware.

5 Conclusion

An implementation of a hybrid intrusion detection system, designed to analyze traffic in computer networks using the TCP/IP protocol stack, is presented. One of their advantages is the flexible extensibility of functionality by writing additional scripts for processing network events. This allows you to embed the necessary scripts designed to generate session parameters. To analyze such parameters, three modules are proposed that are based on various adaptive models: neural networks, immune detectors, and neuron-fuzzy classifiers. The experiments were conducted on two data sets: KDD Cup NSL-KDD. The results showed that, in comparison with the existing detection approaches evaluated on these sets, the proposed integration scheme allows a compromise between the recognition of unknown types of threats and false positives.

As we noted above, there are a lot of ways to detect an intrusion into the system and no fewer ways to hide it. We examined only one of them and proposed their mechanism for detecting network attacks. Today, hackers are becoming smarter, so you need to do a dump of RAM and its analysis. After that, we do a selective analysis of the files on the hard drive. And only at the end - a full analysis of the hard disk, if necessary.

Further research will focus on finding and applying other hybrid approaches to detecting attacks, creating experimental data sets and conducting experiments.

References

1. Jing, Y., Feng, W.: Simulation Modeling of Network Intrusion Detection Based on Artificial Immune System. In: Life System Modeling and Intelligent Computing. Communications in Computer and Information Science, v. 97. Springer, Berlin, Heidelberg, pp. 1-4 (2010)
2. Kim, J., Bentley, P.: An Artificial Immune Model for Network Intrusion Detection. Computer Applications in Engineering Education, pp. 1-4 (2001).
3. Kumar, V.: Signature-Based Intrusion Detection System Using SNORT. International Journal of Computer Applications & Information Technology, pp. 1-7 (2012).
4. Guan, X., Li, Y.: Notice of Retraction: An New Intrusion Prevention Attack System Model Based on Immune Principle. 2nd International Conference on E-business and Information System Security, Wuhan, pp. 1-4 (2010).
5. Tavallaei, M., Bagheri, E., Lu W., Ghorbani, A.A.: A detailed analysis of the KDD CUP 99 data set. IEEE Symposium on Computational Intelligence for Security and Defense Applications, Ottawa, pp. 1-6 (2009).
6. Siddique, K., Akhtar, Z., Aslam Khan F., Kim, Y.: KDD Cup 99 Data Sets: A Perspective on the Role of Data Sets in Network Intrusion Detection Research. In Computer, vol. 52, no. 2, pp. 41-51 (2019).
7. Lei, J.Z., Ghorbani, A.A.: Improved Competitive Learning Neural Networks for Network Intrusion and Fraud Detection. Neurocomputing, vol. 75, Iss. 1, pp. 135-145 (2012).
8. Meng, Q.: An immune-neuroendocrine-inspired inspired artificial homeostatic security-coordination model for E-Government system. 2nd International Conference on Artificial Intelligence, Management Science and Electronic Commerce (AIMSEC), Dengcheng, pp. 6960-6963 (2011).

9. Giatzitzoglou, D.G., Sotiropoulos D.N., Tsihrintzis, G.A.: AIRS-x: An eXtension to the Original Artificial Immune Recognition Learning Algorithm. 2019 International Conference on Computer, Information and Telecommunication Systems (CITS), Beijing, China, pp. 1-5 (2019).
10. Yang, H., Yiwen, L., Tao, L., Ting, Z.: A model of Collaborative Artificial Immune System. 2nd International Asia Conference on Informatics in Control, Automation and Robotics (CAR 2010), Wuhan, pp. 101-104 (2010).
11. Hamid, M.B.A., Rahman, T.K.A.: Short Term Load Forecasting Using an Artificial Neural Network Trained by Artificial Immune System Learning Algorithm. 12th International Conference on Computer Modelling and Simulation, Cambridge, pp. 408-413 (2010).
12. Orang, Z. A.: Using Adaptive NeuroFuzzy Inference System in Alert Management of Intrusion Detection Systems. Intern. Journal of Computer Network and Information Security, vol. 4, N 11, p. 32–38 (2012).
13. Komar, M., Golovko, V., Sachenko, A., Bezobrazov, S.: Development of Neural Network Immune Detectors for Computer Attacks Recognition and Classification. IEEE 7th Intern. Conf. on Intelligent Data Acquisition and Advanced Computing Systems (IDAACS), vol. 2, pp. 665–668 (2013).
14. Govindarajan, M., Chandrasekaran, R.M.: Intrusion Detection Using an Ensemble of Classification Methods. Proc. of the World Congress on Engineering and Computer Science, vol. 1, pp. 459–464 (2012).
15. Almeida, L.A. Álvarez, Santos, J. Carlos Martinez: Evaluating Features Selection on NSL-KDD Data-Set to Train a Support Vector Machine-Based Intrusion Detection System. IEEE Colombian Conference on Applications in Computational Intelligence (ColCACI), Barranquilla, Colombia, pp. 1-5 (2019).
16. Syarif, I., Zaluska, E., Prugel-Bennett, A., Wills, G.: Application of Bagging, Boosting and Stacking to In-trusion Detection. Machine Learning and Data Mining in Pattern Recognition, vol. 7376, pp. 593-602 (2012).
17. Belej, O., Lohutova T., Banaś, M.: Algorithm for Image Transfer Using Dynamic Chaos. 2019 IEEE 15th International Conference on the Experience of Designing and Application of CAD Systems (CADSM), Ukraine, 2019, pp. 1-5 (2019).
18. Belej O.: The Cryptography of Elliptical Curves Application for Formation of the Electronic Digital Signature. Advances in Intelligent Systems and Computing, vol 938, Springer, Cham, pp. 43-57(2020).
19. Belej O.: The Controlling of Transmission of Chaotic Signals in Communication Systems Based on Dynamic Models. CEUR Workshop Proceedings, Vol. 2353, pp. 664-673 (2019).

Cattle Breed Identification and Live Weight Evaluation on the Basis of Machine Learning and Computer Vision

Oleg Rudenko¹[0000-0003-0859-2015], Yuriy Mege²[0000-0003-0072-0756],
Oleksandr Bezsonov¹[0000-0001-6104-4275], Antonina Rybalka¹[0000-0002-4148-2443]

¹ Kharkiv National University of Radio Electronics, Nauky Ave. 14, Kharkiv, 61166,
Ukraine

oleh.rudenko@nure.ua
oleksandr.bezsonov@nure.ua
antonina.rybalka@nure.ua

² Kharkov National Technical University of Agriculture named after P. Vasilenko,
Alchevskih str., 44, Kharkiv, 61002, Ukraine
megelye@gmail.com

Abstract. The problem of the cow's live weight estimation is considered. A convolutional neural network based method for animal recognition and its breed identification in combination with epipolar geometry approach for object's size measurement is proposed. Information regarding animal's size and its breed is further used for LW estimation by multilayer perceptron based predictive model. This approach can be used to replace traditional methods of direct observation and measurement. The proposed system can be widely used in the management of a modern farm. Accuracy and performance of the proposed method has been tested with the participation of the experts.

Keywords. Convolutional neural network, epipolar geometry, image processing, computer vision, cow, live weight, mask-rcnn

1 Introduction

Computer-based image analysis and various predictive applications are commonly used in different fields of human activity, one of which is agriculture. The number of farms in Ukraine has increased significantly and their productivity is growing, so the importance of computer technology in the automation of agricultural processes is

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

gradually increasing. When raising cows, the relationship between live weight (LW), milk yield and feed intake can be taken as criteria for organizing the care and nutrition of animals in modern keeping conditions. These parameters are quite important and must be strictly controlled. When they go beyond permissible limits, this significantly affects the immune system of cows, and, accordingly, the economic efficiency of the farm. Negative changes in live weight may indicate animal's health problems, inappropriate environmental conditions, and nutritional errors. Therefore, a parameter such as live weight (LW) is certainly important for dairy cows [1]. It should be noted that at present the process of measuring and servicing cattle is still carried out manually and is very expensive.

2 Formal problem statement

The aim of this study is the estimation of cows LW using several neural network models: the convolution artificial neural network for recognition of a cow at the pictures and its breed identification with subsequent determination of the size of its body by the stereopsis method with subsequent utilization of multilayer perceptron for LW estimation of a cow on the basis of information regarding its breed and size.

For more accurate estimation of the animal's physical parameters 3D camera (Intel RealSense D435i) was additionally used. It should be noted that the use of a 3D camera alone does not yield to good results due to its low resolution. Thus, cows images taken at different angles are used to determine parameters of bodies cows using photogrammetric method. Parameters such as the withers height (WH), hip height (HH), the body length (BL) and the hip width (HW) of cows were obtained via photogrammetry.

Model estimation based on the ANN has been developed using those parameters (input parameters WH, HH, BL, HW, and the output parameter - LW).

Cow's body dimensions are determined from the analysis of animal images taken synchronized cameras from different perspectives. Initially, a cow is identified at the image and its breed is determined by using Mask-rcnn convolution neural network. Then withers height, hip height, length and width of a cow determined via stereopsis method which allows to obtain geometric parameters of the objects at digital images and perform their measurements. Digital imaging and photogrammetric processing include several completely certain steps that can generate three-dimensional or two-dimensional digital model of the animal's body. Then obtained data about the species and its size are fed to the predictive model, which determines the estimated weight of the animal.

3 Literature review

In the literature different estimation of animals LW for various purposes is typically performed by using regression equations. However, in recent years the methods and means of computational intelligence based on ANN are increasingly used.

Prediction of LW of bulls' slaughter value from growth data by using ANN carried out in [2], an artificial intelligence technology in dairy industry [3], a comparison of an artificial neural network and a method of linear regression for prediction of the LW of hair goats [4], comparison of ANN and decision tree algorithms used for prediction of LW at post weaning period [5], weight prediction of broiler chickens with the help of 3D-computer vision is performed in [6], prediction of the goats masses [7], artificial neural network to predict rabbits body weight [8], prediction of carcass meat percentage in young pigs using linear regression models and artificial neural networks [9], weighing pigs using machine vision and artificial neural networks [10].

The purpose of this study is to estimate cows LW by using artificial neural networks. Body dimensions (BD) were determined by the photogrammetric method on images in which the cow is identified and classified by the convolution neural network.

4 The traditional methods of measuring the mass of cows

4.1 The method of Trukhanovskii

This method is used to determine LW of the adult cattle by the following formula

$$LW = \frac{A \cdot B}{100} \cdot K, \quad (1)$$

where A – chest girth behind the shoulders, cm; B – direct length of the trunk, measured with a stick, cm; K – a correction factor (2 - for dairy cattle and 2.5 - for milk-meat and beef breeds).

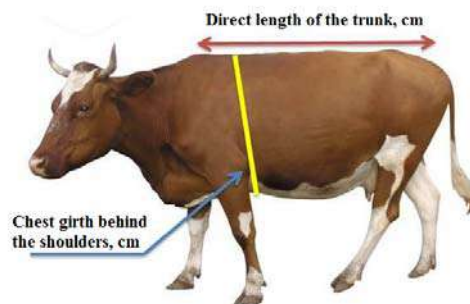


Fig. 1. The measurement method of Trukhanovskii

To determine the approximate body weight special tables are used. For these tables the initial data is measurements of animals taken at the correct animal's position (feet must stand upright, a head at the level of the back).

4.2 Other methods

There are some other methods of LW and cow's body size estimation that could be applied, in particular Pin Bone method [11]

$$BW = \frac{((HG)^2 \times BL)}{300}, \quad (2)$$

where BW – Body Weight estimation (pounds); HG - heart girth (inch); BL - Body Length (inch).

Thus, in estimating of cattle LW the important task of determining its size measure arises. This problem is complicated especially when measures are made in real conditions. The traditional animal measuring process is shown in Fig. 2.



Fig. 2. The traditional measuring process

5 Determination of the size of the cattle using the stereopsis method

The first step in the proposed algorithm is detection of a cow.

This step is based on the use of convolution neural network for the cows detection at the picture and stereopsis method, which allows the system to obtain measurements of the real world objects, located at different distances from the cameras.

5.1 The stereopsis principle

This principle can be explained as follows:

Suppose there are two cameras, defined by their matrices P and P' in some coordinate system. In this case we say that there is a pair of calibrated cameras. If the cameras centers do not match, the pair of cameras can be used to determine the three-dimensional coordinates of the observed points.

Often, the coordinate system is chosen in such a way that the cameras matrix are of the form $P = K[I | 0]$, $P' = K'[R' | t']$ (this is always possible, if we choose the origin coinciding with the first camera's center, and direct a Z-axis along the optical axis).

Consider a point P in the real three-dimensional space projected simultaneously in two image points p and p' through the two camera projection center (C and C'). The points P , p , p' , C and C' lie in a plane, called the epipolar plane. Epipolar plane inter-

sects each image forming the intersection lines. These lines (l and l') correspond to the projection ray through the p and P , and p' and P , and are called epipolar lines. This projection epipolar geometry is described in [12] (Fig. 3).

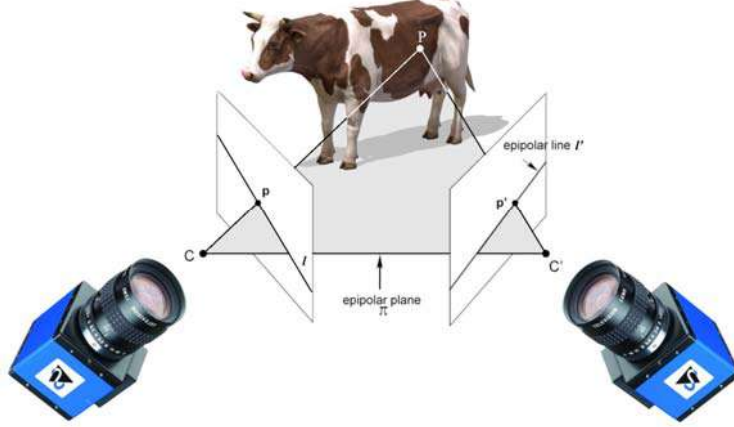


Fig. 3. Epipolar geometry

Epipolar geometry is used to search for stereo pairs and for verifying that a pair of points can be a stereo pair (i.e. the projection of a point in space).

Epipolar geometry has a very simple description in the coordinates. Suppose there is a pair of calibrated cameras, and let p is homogeneous coordinates of the point at the image from the first camera, and p' - from the second camera. A 3×3 matrix F exists, such that the pair of points p, p' is a stereo pair if and only if:

$$p'^T F p = 0. \quad (3)$$

The matrix F is called a fundamental matrix. Its rank is 2 and it is determined up to a nonzero factor that only depends on the original matrix cameras P and P' .

In the case where the matrix cameras are of the following form $P = K[I | 0]$, $P' = K'[R | t]$ the fundamental matrix may be calculated with the formula:

$$F = (K'^{-1})^T R K^T [K R^T t]_x, \quad (4)$$

where the e vector notation $[e]_x$ calculated as

$$[e]_x = \begin{bmatrix} 0 & -e_z & e_y \\ e_z & 0 & -e_x \\ -e_y & e_x & 0 \end{bmatrix}. \quad (5)$$

In [13] the various calculation algorithms of F with using a set of points are considered. In particular, gradient descent algorithm, Newton's method and the Levenberg-Marquardt algorithm are described.

Epipolar line equations are calculated with the help of the fundamental matrix. For the point x , the vector that defines the epipolar line is of the form $l' = Fp$, and the equation of the epipolar line itself is: $l'^T p' = 0$. Similarly, for a point p' , the vector defining the epipolar line is of the form $l = F^T p'$.

5.2 Construction of the depth map

Depth map - is an image in which each pixel, instead of color, keeps a distance from the camera. To this end, for each point in one image its pair is searched on the other image. A pair of corresponding points can be used for triangulation and determination of their prototype coordinates in three dimensions. Knowing the three-dimensional coordinates of the prototype image, the depth is calculated as the distance to the camera's plane [13].

Thus it is possible to determine the size of the object.

6 Convolutional Neural Network (CNN)

CNN was first proposed in [14] as the development of the neocognitron model [15], intended for effective image recognition. This network uses pattern recognition technology based not on hardcoded by developers algorithms, but on system training, which involves a consistent identification of a huge number of images.

Subsequently, the R-CNN (Regions With CNNs) networks were built based on CNN. R-CNN were used for detecting all objects of the given classes and determining the bounding box for each of them (object detection). R-CNN creates bounding boxes for each object in the image or suggestions regions using selective search process. Fast R-CNN increases productivity of R-CNN and classifies objects of each region together with tighter bounding boxes. Next network, Faster R-CNN, improved generation mechanism of used therein candidate regions by computing the regions not on the original image but on the features map derived from CNN. For this purpose the module called Region Proposal Network (RPN) was added.

Finally, the Mask R-CNN network [16] improves Faster R-CNN architecture by adding one more sub-module, which predicts the position of the detected object covering mask, and thus solves the task of instance segmentation. After image processing, the network outputs objects bounding boxes (bbox), their classes (class) and masks (mask). It worth to note, that Mask R-CNN is one of the fastest network at the moment. The structure of this network is shown in Fig. 4.

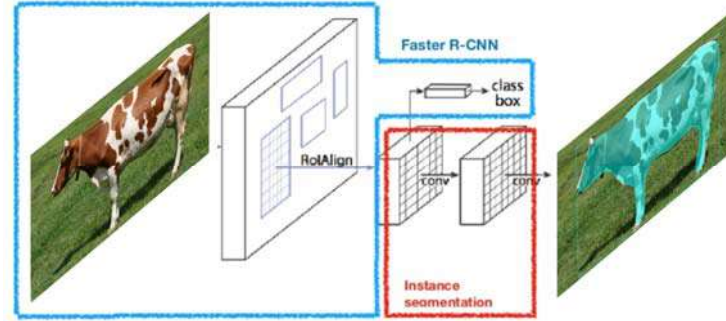


Fig. 4. Mask R-CNN structure

6.1 The architecture of the neural network convolution

Initially, the convolution neural network structure was created taking into account the structural features of some parts of the human brain responsible for vision. Three mechanisms are laid into the foundation for the development of such networks:

- local perception;
- building a set of layers in the form of characteristics maps (shared weights);
- sub-sampling.

In accordance with these arrangements, three main layers are used to build a convolutional neural network: convolution, pooling (otherwise subsampling or down-sampling), fully connected layer.

6.2 Convolution layer

Convolution equation for the l -th ($l=1, \dots, L$) network layer has the following form [17]

$$x_{ij}^l = \sum_{a=-\infty}^{+\infty} \sum_{b=-\infty}^{+\infty} w_{ab}^l \cdot y_{(is-a)(js-b)}^{l-1} + b^l. \quad (6)$$

It reflects movement of the w^l core along the image or map of input attributes for the layer y^{l-1} . Here $y^{l-1} = f(x_{ij}^{l-1})$ - image after $(l-1)$ -th layer; $f(\bullet)$ - used activation function; b^l - offset. Indices i, j, a, b - indices of the elements in the matrices, S - the value of the convolution step.

As seen from (3) convolution are performed for each element i, j of image matrix x^l .

Convolution preserves spatial relationships between the pixels.

Each convolutional layer is followed by subsampling or computational layer that is serving to reduce the dimension of the image by averaging the values of the local output neurons.

6.3 Subsampling Layer (pooling, MAX-pooling)

Subsampling layer zooms planes by local averaging of the output neurons values . Thus, the hierarchical organization is achieved. Subsequent layers are extracted more common characteristics that less depend on the image distortion.

The difference between a subsampling layer and a convolution layer is that in the convolution layer regions of the neighboring neurons overlap, which does not occur in the subsampling layer.

Pooling layer operates independently of the input data depth and scales the spatial volume by using a maximum function.

The architecture of the convolution network assumes that the presence of a sign is more important than information about its location. Therefore, the maximum one is selected from several neighboring neurons in the feature map and its value is considered as a single neuron in the feature map of lower dimension.

In addition to maximum subsampling, pooling layers can perform other functions, such as averaging subsampling or even L2-normalized subsampling.

6.4 Non-linear activation layer

On these layers within the network, a nonlinear activation function $f(\bullet)$ is applied to all input values and the result is sent to the output. Thus, the activation layer does not change the input resolution.

Usually due to significant positive properties for hidden layers, ReLu(ReLU (x) = max (0; x)) and its various modifications (Leaky ReLU, Parametric ReLU, Random-

ized ReLU) [18] are used. SoftMax function $f_j^L = e^{x_j^L} \left(\sum_{i=1}^{N^L} e^{x_i^L} \right)^{-1}$ (for solving clas-

sification problems) or linear function (for regression tasks) are used for fully connected layer.

6.5 Dropout layer

Different regularization techniques are used to avoid network retraining.

Dropout is a simple and effective regularization method and consists in the fact that in the process of training a network, a subnet is randomly allocated from its aggregate topology, i.e. part of the neurons is turned off from the process, and the next update of the scales occurs only within the allocated subnet. Thus, only weights of remaining neurons are changed. Each neuron is excluded from the total network with a certain probability, which is called the dropout rate.

This layer reduces the time of one training epoch due to the smaller number of optimized parameters, and also allows to better deal with retraining of the network compared to standard regularization methods.

6.6 Normalizing layer

The standard normalization of inputs occurs on this layer (the sample average of their values is subtracted, and the result is divided by the root of the sample variance). Sampled values are calculated taking into account the values at the inputs of this layer at previous training iterations. This approach allows to increase the speed of learning the network and improve the final result.

6.7 Fully connected layer

This layer is an ordinary multilayer perceptron, the purpose of which is classification. It models a complex nonlinear function, the optimization of which improves the quality of recognition.

The neurons of each map of the previous subsample layer are associated with one neuron of the hidden layer. Thus, the number of neurons in the hidden layer is equal to the number of maps in the subsample layer.

As in ordinary neural networks, in a fully connected layer, neurons are connected to all activations in the previous layer. Their activations can be calculated by multiplying matrices and applying bias.

The differences between fully connected and convolutional layers are that the neurons of the convolutional layer

- 1) connected only with the local area of the input;
- 2) can share parameters.

6.8 The choice of training criteria

ANN training is an iterative process. At each iteration, the network outputs for one (or more) samples in the training set are calculated, and the network weights are adjusted to reduce the error between the actual network output $(y_{i,p}^L)(i=1,...,N^L)$ and target output for a given sample $d_{i,p}$. Therefore, training is reduced to minimizing some error function.

In practice, a quadratic function, cross entropy, or some combined functional are used as a criterion for the error function.

7 Backpropagation Neural Network (BPNN)

Backpropagation neural network (BPNN) is a multilayer feedforward neural network that uses a supervised learning algorithm known as error back-propagation algorithm. Errors accumulated at the output layer are propagated back into the network for the adjustment of weights.

Multilayer perceptron (MLP) is a neural network with several layers, each of which consists of computing nodes (neurons). The topology of the MLP is shown in Fig. 5 [19]. Network inputs are connected to each neuron in the first layer. The outputs of

the neurons of the first layer then become inputs to the neurons of the second layer and so on. The last layer is the output layer, all other layers between the input and output layers are called hidden layers. The architecture of the multi-layer perceptron can be conveniently written as $n_0 - n_1 - \dots - n_l$ where n_0 is the network's input vector dimension, and n_i , $1 \leq i \leq l$ denotes the number of nodes in the respective layers.

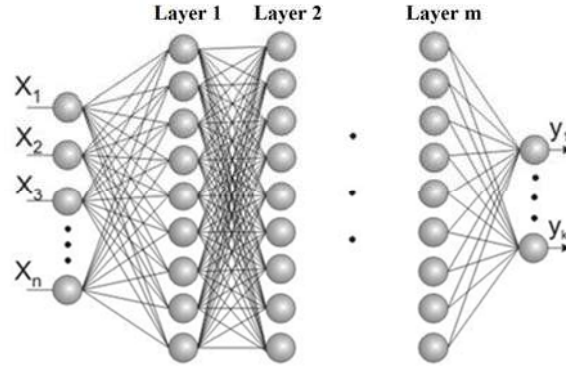


Fig. 5. Backpropagation neural network

Thus, MLP uses the following approximation of the nonlinear operator $f(\bullet)$:

$$\hat{y}(k) = \hat{f}(k) = f^q \left[\left(W^q \right)^T f^{q-1} \left[\left(W^{q-1} \right)^T f^{q-2} \left[\dots f^1 \left[\left(W^1 x(k) + b_1 \right)^T \right] \dots \right] \right] \right] + b_q, \quad (7)$$

where W^i - vector of weight parameters of neurons in the i -th layer of the network; $f^i[\bullet]$ - activation function (AF) of i -layer, b_i - bias of the i -th neuron.

Since in practical applications it is necessary to perform various operations not only with the activation function itself, but also with its first derivative, it is necessary to use a monotonous, differentiable and limited function as an activation function. A particularly important role is played by such functions in modeling nonlinear relationships between input and output variables. These are the so-called logistic or sigmoidal (S-shaped) functions.

In general, the representation of the input-output can be expressed as $\hat{f} : R^{n_l} \rightarrow R^m$.

Approximate capacity of MLP have been studied and described by many authors [19-23]. The basic idea is that every continuous function $f : D_f \subset R^{n_l} \rightarrow R^m$ can be uniformly approximated with arbitrary precision function $\hat{f}$ by D_f , where D_f is a

compact subset of R^n , provided that there is a sufficient number of hidden layers in the network. This is true even for networks with only one hidden layer. A typical assumption for the activation functions in the hidden layer is the following: $f(\bullet)$ is a continuous, non-constant and limited.

MLP is in general function approximator and it ensures that the network with one hidden layer will always be enough to represent any arbitrary continuous function. But this statement says nothing about the number of neurons in the hidden layer, which provide a given accuracy of approximation. It is also important that the proof of the possibility of approximation by means of MLP suggests that weights are set correctly. However, the question of a learning algorithm choice remains open.

The pseudocode algorithm for BPNN is given below [24].

- (I) Network initialization: randomly choose the initial weights
- (II) Select the first training pair
- (III) Forward computation that includes the following steps:
 - (A) Apply the inputs to the network
 - (B) Calculate the output for every neuron from the input layer, through the hidden layer (s), to the output layer
 - (C) Calculate the error at the outputs
- (IV) Backward computation
 - (A) Use the output error to compute error signals for preoutput layers
 - (B) Use the error signals to compute weight adjustments
 - (C) Apply the weight adjustments
- (V) Repeat Forward and Backward computations for other training pairs.
- (VI) Periodically evaluate the network performance. Repeat Forward and Backward computations until the network converges on the target output.

8 The structure of the developed system

Fig. 6 shows the structure of the developed intellectual breed recognition and evaluation of DA cattle.

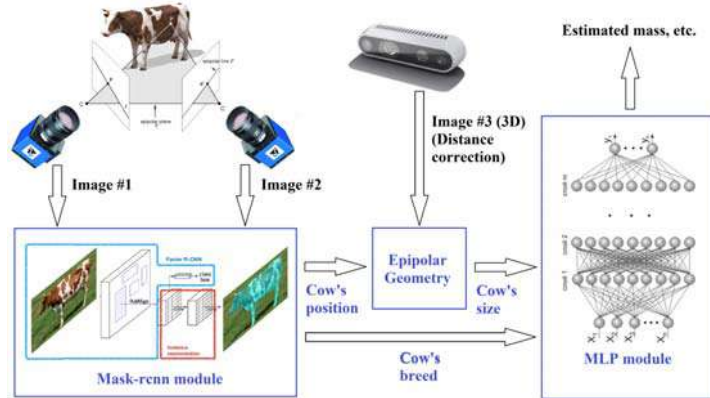


Fig. 6. The structure of the developed intellectual system for breed recognition and cattle LW evaluation

9 Experiments and results

The problem of recognizing various breeds of cows and assessing their linear sizes with the subsequent determination of mass using a neural network predictive model based on a multilayer perceptron was solved. To recognize the breed of a cow, the Mask-RCNN network trained on the COCO sample was used, the source code of which is freely available on the Internet [25]. The network weights were fine-tuned on 250 images of representatives of each of the considered breeds of cows (Ayrshire, Holstein, Jersey, Red Steppe). To increase the size of the training and test samples, the augmentation method was used [26], which made it possible to obtain additional images from the original ones. The network was retrained over 5000 epoches using the SGD algorithm. To train the network, we used the Nvidia GeForce 1060 graphics card, which allowed us to speed up the learning process many times as compared to learning on the CPU. The results of recognition and evaluation of the size and weight of cows are given in Table 1. Examples of correct recognition are presented in fig. 7. It should be noted that, despite the rather extensive training sample and the achieved high recognition accuracy (92%), as a result of the network, false recognition occurs. An example of such recognition is shown in Fig. 8.

The obtained masks of recognized objects were used to determine their linear dimensions using the triangulation method (epipolar geometry). As additional information that allows us to adjust the obtained sizes, we used images from a 3D camera, which allow us to estimate the distance to the recognized object. Examples of such images are presented in Fig. 9.

Table 1. The obtained results

Breed	Ayrshire	Holstein	Jersey	Red Steppe
Average calculated dimensions (length of the body / chest girth) (cm)	149/171	155/180	132/160	157/189
Qualification weight (kg)	610	745	422	490
The standard deviation of obtained sizes	1.3	1.6	2.4	1.8
The mean absolute error (cm)	0.9	1.2	2.5	1.5
The standard deviation of absolute error (cm)	0.5	1.7	4.5	3.8
The average relative error	0.1	0.12	0.17	0.15
The standard deviation of the relative error	0.11	0.15	0.09	0.09
Number of correct recognition	461	418	424	439
Number of false positives	39	82	76	61
Accuracy	0.92	0.84	0.85	0.88
The ratio of false positives to the correct ones	0.09	0.2	0.18	0.14

**Fig. 7.** Examples of correct recognitions



Fig. 8. Examples of false recognitions

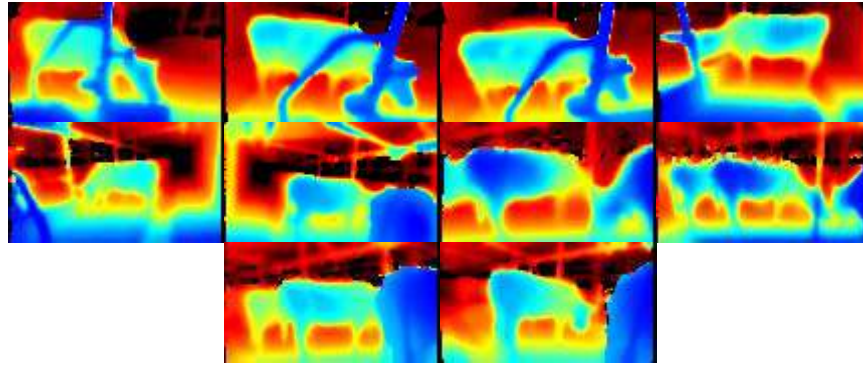


Fig. 9. Examples of 3D images from the camera

Acknowledgement

This project has been funded with support from the European Commission. This publication reflects the views only of the author, and the Commission cannot be held responsible for any use which may be made of the information contained therein.



10 Conclusions

The proposed method for measuring cattle using neural network image processing algorithms allows modern farmers to quickly and accurately measure the weight of the animal, as well as recognize its breed, saving time and reducing effort without compromising the habitat and disturbing livestock growth. This is possible because the method allows studying cows both in corrals and pastures, without interfering with their normal behavior. The proposed approach can be used to replace traditional methods that use direct observation and measurement, which adversely affect the behavior of animals. The system can be widely used in the management of modern farming. The accuracy and performance of the proposed methodology were tested

with the participation of the experts. The same measurements were carried out by the farm staff who confirmed the effectiveness of the proposed system.

References

1. Tasdemir, S.: Determination of Body Measurements On the Holstein Cows by Digital Image Analysis Method and Estimation of Their Live Weight. Ph. D. thesis, Selcuk University, Konya, Turkey (2010).
2. Adamczyk, K., Molenda, K., Szarek, J., & Skrzyński, G.: Prediction of Bulls'slaughter Value From Growth Data Using Artificial Neural Network. *Journal of Central European Agriculture*, 6 (2), pp. 133-142 (2005).
3. Akilli, A., Atil, H.: Artificial Intelligence Technologies in Dairy Science: Fuzzy Logic and Artificial Neural Network. *Hayvansal Uretim*, 55 (1), pp. 39-45 (2014).
4. Akkol, S., Akilli, A., Cema, I.: Comparison of Artificial Neural Network and Multiple Linear Regression for Prediction of Live Weight in Hair Goats.YYU Yuzuncu Yıl Universitesi *Journal of Agricultural Sciences*, 27 (1), pp. 21-29 (2017).
5. Ali, M., Eydurani, E., Tariq, M.M., Tirink, C., Abbas, F., Bajwa, M.A., ... & Shah, S.H.: Comparison of artificial neural network and decision tree algorithms used for predicting live weight at post weaning period from some biometrical characteristics in Harnai sheep. *Pakistan J. Zool.*, Vol. 47 (6), pp. 1579-1585 (2015).
6. Mortensen, A.K., Lisowski, P., Ahrendt, P.: Weight prediction of broiler chickens using 3D computer vision. *Computers and Electronics in Agriculture*, 123, pp. 319-326 (2016).
7. Raja, T.V., Ruhil, A.P., Gandhi, R.S.: Comparison of connectionist and multiple regression approaches for prediction of body weight of goats. *Neural Computing and Applications*, 21(1), pp. 119-124 (2012).
8. Salawu, E.O., Abdulraheem, M., Shoyombo, A., Adepeju, A., Davies, S., Akinsola, O., Nwagu, B.: Using artificial neural network to predict body weights of rabbits. *Open Journal of Animal Sciences*, 4 (04), 182 p. (2014).
9. Szyndler-Nęda, M., Eckert, R., Blicharski, T., Tyra, M., Prokowski, A.: Prediction of carcass meat percentage in young pigs using linear regression models and artificial neural networks. *Annals of Animal Science*, 16 (1), pp. 275-286 (2016).
10. Wang, Y., Yang, W., Winter, P., Walker, L.: Walk-through weighing of pigs using machine vision and an artificial neural network. *Biosystems Engineering*, 100 (1), pp. 117-125 (2008).
11. McNitt, J.I.: *Livestock Husbandry Techniques*, Low priced edition. Granada publishing company limited, 280 p. (1983).
12. Hartley, R.I., Zisserman, A.: *Multiple View Geometry*. Cambridge University Press (2004)
13. Bradski, G., Kaehler, A.: *Learning OpenCV: Computer Vision with the OpenCV Library*. O'Reilly Media, 580 p. (2008).
14. LeCun, Y., Boser, B., Denker, J.S., Henderson, D., Howard, R.E., Hubbard, W., Jackel, L.D.: Backpropagation Applied to Handwritten Zip Code Recognition. *Winter, Neural Computation*, 1 (4), pp. 541-551 (1989).
15. Fukushima, K.: Neocognitron: A Self-organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position. *Biological Cybernetics* 36(4) (1980).
16. He, K., Gkioxari, G., Dollar, P., Girshick, R.: Mask r-cnn.arXiv: 1703.06870 (2017).
17. LeCun, Y., Bengio Y.: Convolutional networks for images, speech, and timeseries. *The Handbook of Brain Theory and Neural Networks*, pp. 255-258 (1995).

18. Nair, V., Hinton, G.E.: Rectified linear units improve restricted Boltzmann machines. In ICML, pp. 807-814 (2010).
19. Bodyanskiy, Ye.V., Rudenko, O.G.: Iskusstvennyye neyronnyye seti: arkhitektura, obucheniye, primeneniye. Khar'kov: TELETEKH, 372 p. (2004).
20. Uossermen, F.: Neyrokomp'yuternaya tekhnika. M.: Mir, 184 p. (1992).
21. Ham, F.M., Kostanic, I.: Principles of Neurocomputing for Science and Engineering, NY: Mc Graw-Hill Inc., 468p. (2001).
22. Patterson, D. Artificial Neural Networks, Theory and Application. Singapur: Prenice Hall Inc., 497p. (1996).
23. Rudenko, O.G., Bessonov, A.A.: Mnogokriterial'naya optimizatsiya evolyutsioniruyushchikh setey pryamogo rasprostraneniya. Problemy upravleniya i informatiki, №6, pp.29-41 (2014).
24. Cilimkovic, M.: Neural Networks and Back Propagation Algorithm, Institute of Technology Blanchardstown, Dublin, Ireland (2015).
25. Mask R-CNN for object detection and instance segmentation on Keras and TensorFlow. https://github.com/matterport/Mask_RCNN
26. Image augmentation for machine learning experiments. <https://github.com/aleju/imgaug>

Physiological Signals Analysis, Recognition and Classification Using Machine Learning Algorithms

Iurii Krak^{1,2} [0000-0002-8043-0785], Anatolii Pashko¹ [0000-0001-6944-8477],
Oleg Khorozov² [0000-0001-8812-323X], Oleg Stelia¹ [0000-0002-1453-501X]

¹Taras Shevchenko National University of Kyiv, 64/13 Volodymyrska str., 01601, Ukraine
yuri.krak@gmail.com, aap2011@ukr.net, oleg.stelya@gmail.com

²Glushkov Cybernetics Institute, Kyiv, 40 Glushkov avenue, 03187, Ukraine
oleh753@hotmail.com

Abstract. The spatial and time-frequency domain features analysis are major approaches for signal classification. For non-stationary signals classification, the selection of features is paramount to the robustness of the recognition systems. Among machine learning algorithms, convolutional neural networks (CNNs) have achieved significant performance in many pattern recognition tasks, along with traditional approaches such as Random Forest or Gradient Boosting. Wavelet transformations allow utilizing spectral components that are useful in signal processing tasks. The performance of the wavelet applications for signal classification was evaluated. The experiments show that wavelet transformations can achieve better accuracy than existing algorithms while having significantly fewer parameters than conventional CNNs.

Keywords: machine learning, classification, signal analysis, expert systems, wavelet transformation

1 Introduction

The primary diagnosis of cardiovascular disease involves expert systems based on the rules of solving medical problems. Rule-based Expert systems are inefficient using big data because they are hard to scale in the presence of new data. Therefore, automating data analysis and recognition procedures is required.

Electrocardiography (ECG) as a non-invasive method for diagnosing the cardiovascular system, involves the identification of morphological features characterizing the types of a heart disability. The data getting from patients monitoring potentially contain a lot of information used inefficiently due to the heterogeneous data structure, which is difficult to interpret. One solution is to use a deep learning model to classify the type of arrhythmia along with the choice of cardiac cycle characteristics.

The deep learning architecture model with the wavelet convolutional layer for signals classification is presented in this work. The reliability of the results verified on the basis of the ECG database, which was used to evaluate the effectiveness of the pro-posed method. The article proposes to automate the process of features selected

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

from the time-frequency domain for ECG signals online classification using machine learning models. ECG signals from mobile devices can be used for person authentication in remote healthcare systems [1, 2]. The purpose is to recognize the cardiovascular disease and estimate the accuracy of predictions using deep learning methods for ECG signals classification. A software framework that facilitates integration of the mathematical models and computer simulations makes it possible to efficiently classify and predict the status of patients. The proposed solution is to use the wavelet convolutional layer for feature extractors and the neural network is used for making a final decision.

Most biological systems reflect their normal or abnormal processes by nonstationary signals and thus joint time-frequency analysis of the physiological signals has potential applications.

2 ECG signals analysis using wavelet transformations

The ECG signals contain many artifacts that need to be filtered. The main sources of ECG noise are baseline deviation (breathing bias), muscle activity (contraction), poor electrode contact, and high-frequency network interference from equipment. Baseline wander elimination is considered as a classical problem in ECG signal filtering during data collection. The ECG data cleaning step is necessary since a signal with noise reduces the accuracy of diagnosis. The wide variety of ECG signals waveforms. The purpose of signal processing is to eliminate low-frequency drift of the signal baseline and high-frequency noise. Many studies have developed linear or non-linear filters to remove additive noise and interference, such as moving averaging filters, low-pass filters, band-pass filters, non-linear filters and Savitsky-Golea smoothing filters. Wavelet-based noise removal methods are connected by wavelet coefficients for filtering high-frequency noise and baseline drift effects [3].

To analyse and verify the reliability of the results, the PhysioNet ECG dataset [4] was used, which contains ECG measurements for individuals labelled Normal Sinus Rhythm (NSR) and those with arrhythmia (ARR) or congestive heart failure (CHF). This data set contains 96 ARR, 36 NSR and 30 CHF records, which were used to evaluate the effectiveness of the proposed method. Based on ECG data, a classification over three groups of people with different pathologies: cardiac arrhythmia, congestive heart failure and healthy people was made.

2.1 ECG signals analysis using discrete wavelet transformations

The wavelet transform is used to analyse data, search for similarity in time series, classify and cluster time series, identify anomalies in the behaviour of time series.

Discrete wavelet transform operates with discrete parameter values. The basis of space are defined as

$$\psi_{m,k}(t) = a^{\frac{m}{2}} \psi(a^m t - k), \quad m, k \in \mathbb{Z},$$

where $\psi(t)$ - parent wavelet.

For practical use, there are convenient wavelets built on the basis of Gauss function and its derivatives. They have the best localization in both the time and frequency domains.

The value of a can be arbitrary, as a rule, use $a = 2$. In this case, the transform is called the dyadic wavelet transform. Forward wavelet coefficients:

$$W_{m,k} = \int_{-\infty}^{\infty} s(t) \psi_{m,k}(t) dt,$$

$s(t)$ - continuous signal.

The inverse discrete wavelet transform has the form $s(t) = \sum_{b=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} W_{m,k} \psi_{m,k}(t)$.

The number of m - parameter coefficients determines the level of signal decomposition. For the inverse wavelet transform, use the following representation:

$$s(t) = \sum_{k=-\infty}^{\infty} C_k \varphi_k(t) + \sum_{b=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} W_{m,k} \psi_{m,k}(t),$$

where C_k - scaling coefficients or approximation coefficients, $W_{m,k}$ - wavelet coefficients of signal detail.

The wavelet analysis algorithm:

Let a time series be given: $s_k = s(t_k)$, $k = 0, 1, \dots, N-1$.

1. Centre the row and exclude trends.
2. Assess the variance, evaluate the correlation function.
3. The values of the coefficients are determined by the formulas.

$$W(a_i, b_j) = \frac{1}{M(a_i, b_j)} \sum_{k=0}^{N-1} s_k \psi\left(\frac{t_k - b_j}{a_i}\right),$$

$$M(a_i, b_j) = \sum_{k=0}^{N-1} \exp\left(-\frac{1}{\sigma} \left(\frac{t_k - b_j}{a_i}\right)^2\right).$$

4. Discretization of arguments:

$$a_i = a_{\min} + i\Delta a, \quad i = 0, 1, \dots, N_a - 1$$

$$b_j = j\Delta b, \quad j = 0, 1, \dots, N - 1.$$

5. The value of the scalogram density in each node is calculated by the formula:

$$S(a_i, b_j) = \left| W(a_i, b_j) \right|^2.$$

Theoretical aspects of wavelet transform have been considered in works [5-7], the use of wavelet transform in vibration studies in works [8-9], identification of systems [10], fuzzy models [11-12].

In [13-19] various problems were analyzed when processing signals using wavelet transform and convolution neural networks.

Let us applying discrete wavelet transform (DWT) as a filter cascade for ECG signals.

A filter bank means dividing a signal into several frequency sub bands for use in applications. First, we apply a small scale corresponding to high frequencies. Then the scale increases by a factor of two (frequency decreases by a factor of two) and so on. We are using «pywt.dwt» library for decomposition signal into frequency bands. The DWT used to split a signal into different frequency sub-bands, as many as needed.

If the different types of signals exhibit different frequency characteristics, this difference should be exhibit in frequency sub-bands. If we generate features from each of the sub-band, set the collection of features as an input for Logistic Regression, Random Forest or Gradient Boosting classifiers and train it, the classifier should be able to distinguish between the different types of signals:

2.2 ECG signals analysis using continuous wavelet transformations

The continuous wavelet transform (CWT) of a signal can be using for analysing content of the signals because they are localized in both frequency and time. They allows to identify areas where the signal changing frequency content. It is can be useful for identifying signals with low-frequency components or frequency localization when changing frequency content. The Wavelets are using to localize transients when changing the content of signals components.

Wavelet transform provides a general method that can be used to process signals. Various properties can be calculated and processed using wavelet coefficients. We can apply the wavelet transform to the ECG signal and convert it to wavelet coefficients. The coefficients characterize the behaviour of the ECG signal, and the number of these coefficients is less than the amount of the original signal. This reduction in feature space is important for recognition and diagnosis.

A classification algorithm using a wavelet transformation for ECG signals is proposed. The wavelet transform is used to analyse spatial signal and identify anomalies in the behaviour of time series. The algorithm used morphological filtering and wavelet transform with a selected mother wavelet. The wavelet coefficients denotes the correlation between the signal and the wavelet in the given scale and position. The optimal wavelet should have a shape identical to the signal being analyzed. Continues wavelet transform is used to evaluate how much a signal resembles a scaled and translated mother wavelet.

The wavelet transform of a time-continuous signal $x(t)$ is defined as:

$$CWT(a, b) = \frac{1}{\sqrt{a}} \int x(t) \psi^* \left(\frac{t-b}{a} \right) dt ,$$

where $\psi^*(t)$ is the complex conjugate of wavelet function, a – scale, b - translation parameters respectively. The function $\psi(t)$ compresses or dilates depending on a , which enables the CWT to extract the low- and high- frequency components of $x(t)$. The wavelet scale a characteristic frequency's of the wavelet such as the spectral peak frequency, pass band centre, and central frequency. The original signal $x(t)$ may be reconstructed using an inverse transform.

The frequency associated with a wavelet of arbitrary a scale is given by $f = \frac{f_c}{a}$,

where the characteristic frequency of the mother wavelet (scale $a = 1$ and location $b = 0$), f_c becomes a scaling constant and f is the frequency for the wavelet at arbitrary a -scale. The contribution to the signal energy at the specific a scale and b location is given by the two-dimensional wavelet energy density function known as the scalogram.

$$S(a, b) = |W(a, b)|^2.$$

The scalogram can be integrated across a and b to recover the energy in the signal. The wavelets can be using for analyzing content of the signals because they are localized in both frequency and time. They allows to identify areas where the signal changing frequency content. The wavelets are using to localize transients when content of ECG signals components are changed.

3 Wavelet transform for filtering and analysis ECG signals with CNN architecture

We utilize Wavelet transform for filtering and analysis ECG signals with CNN architecture to develop a predictive model to indicate normal or abnormal behavior. A signal $x(t)$ transformed Wavelet into a 2D space, are contained continuous mother wavelet $\psi(t)$ with a - scale factor and b - translation factor.

The kernel at scales = [3, 4, 5, 6] with Morlet wavelet are plotted (Fig.1).

The frequency spectrum of ECG signal is in the range from 3 to 45 Hz. Compute CWT transform using the complex Morlet wavelet and 32 scales between 3Hz and 45Hz. We can get 32 or 64 channels with base line flattening.

A visualization of the CWT coefficients - 2D scalogram can be used to improve the distinction between varying types of a signal. To evaluate and analyze a signal morphology the Morlet wavelet transform was used, as show at the plotter below for fragment of the ECG signal (Fig.2).

Let presented the example of the spectrograms segment of the cardiac cycle with a stride in time every 10 samples with scale 32 wavelet Morlet and Baseline wander elimination (Fig. 3).

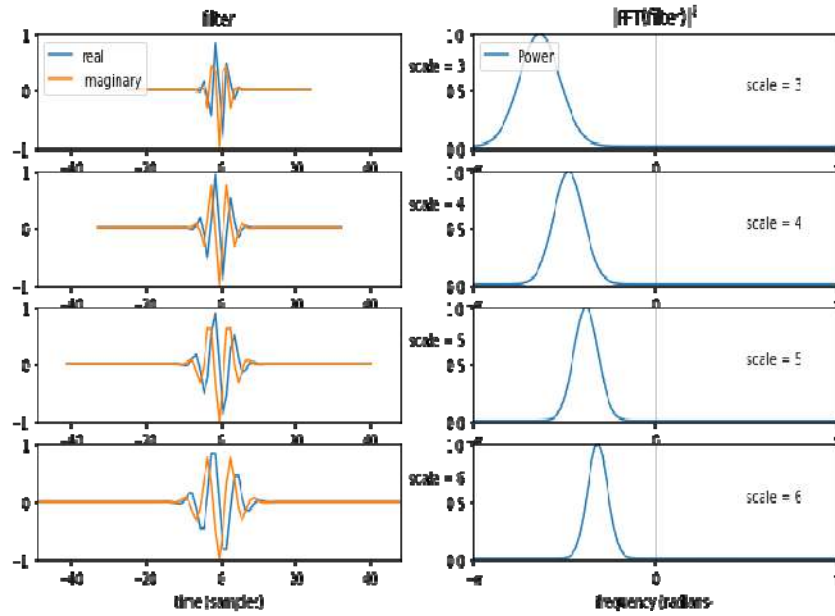


Fig.1. Morlet filters

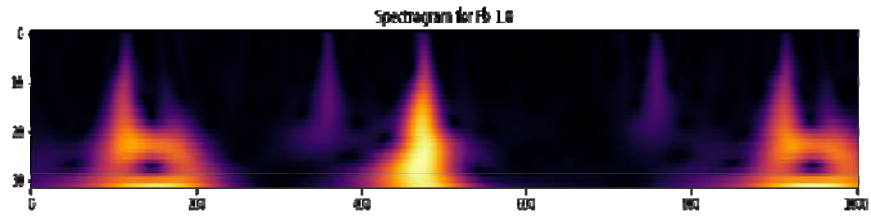


Fig.2. Segment of the cardio signals

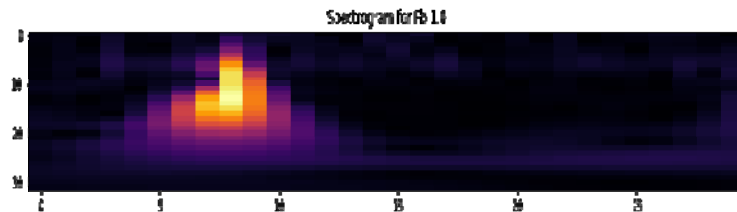


Fig. 3. ECG Cycle

The CWT consists of convolution of filter with the signal. The convolution kernels are defined by means of scales for mother wavelet. A multiscale approximation is the design method of most of the practically relevant wavelet transforms.

A time-frequency transform method and convolutional neural network was used as filtering. This formulation allows us to connect CNN with a multi-resolution analysis. We use a single-channel data set and wavelet transformation with multiple channels for using CNNs model.

4 Signal classification based on the continuous wavelet transform

A scalogram is the absolute value of the continuous wavelet transform coefficients of a signal. Since ECG signals are sensitive to noise, studies have been conducted by transforming signals into a frequency domain that is efficient for analysing noisy signals. The signal transforming from time domain to frequency domain, the 1-D signal becomes a 2-D matrix, and it could be analysed at multiresolution. This process makes signal analysis morphologically complex and existing simple classifiers could perform poorly. We investigate the possibility of using the scalogram as input to convolutional neural networks, which exhibit optimal performance for the classification of morphological imagery. When training data is small transfer learning can be used with pretrained deep models to decision making and prognoses with scale range 32 wavelet. Time-frequency representation of the cardiac cycle using Morlet Wavelets wavelet with 120 scales. A smaller range of scales (32) enables more focus of abrupt changes in signal and a wide range of scales (64 or 128) provides more information, which can provide a better classification accuracy.

Visualization of CWT coefficients allows to distinguish cardiac cycles and QRS complexes of the ECG signal analysing [11,15]. The scalogram with 128 range of scales of the cardiac cycle data. The plots of cardiac arrhythmia and normal sinus rhythm represent below.

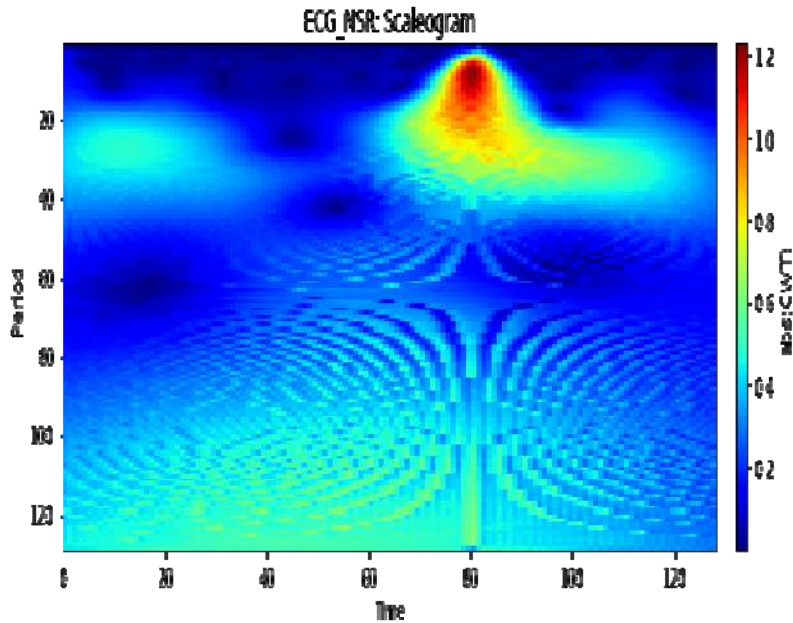


Fig. 4. Cmor0.5-1.0 ECG_NSR cycle

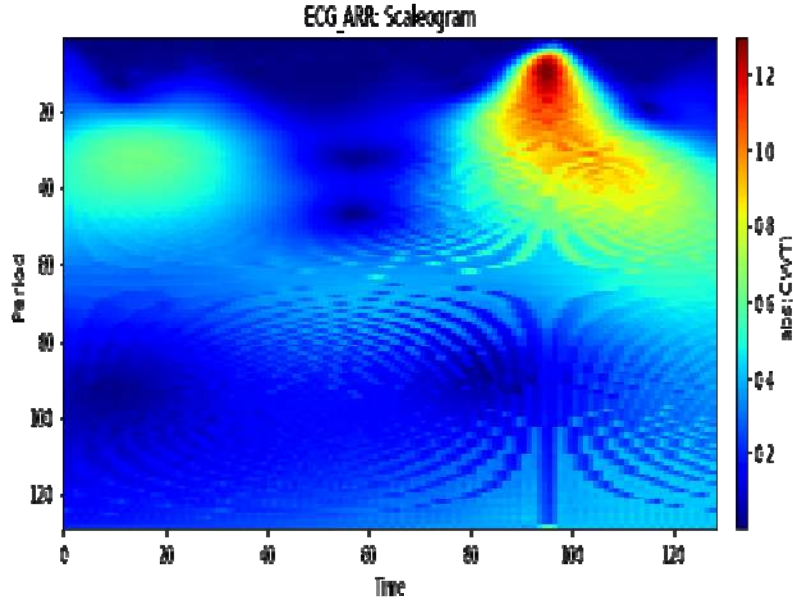


Fig.5 Cmor0.5-1.0 ECG_ARR cycle

A horizontal characteristic of the scalogram is signal frequency. The scalogram can be understood as an image and applied to a neural network model to train a classifier. The model parameters are shown in Table 1. Sequential convolutional neural network model with trainable parameters: 58,553,203

```
num_filter, num_classes = 3, 3
model = keras.models.Sequential([
    keras.layers.Flatten(input_shape=[fs-1, fs-1, num_filter]),
    keras.layers.Dense(300, activation="relu"),
    keras.layers.Dense(100, activation="relu"),
    keras.layers.Dense(num_classes, activation="softmax")
])
model.compile(loss="sparse_categorical_crossentropy",
              optimizer="sgd", metrics=["accuracy"])
```

Table 1 Model Parameters.

Layer (type)	Output Shape	Param
flatten (Flatten)	(None, 195075)	0
dense (Dense)	(None, 300)	58522800
dense_1 (Dense)	(None, 100)	30100
dense_2 (Dense)	(None, 3)	303

A 'Flatten' layer transforms a two-dimensional matrix of wavelet convolutional features into a vector that can be fed into a neural network classifier. Training the ECG data set segments are demonstrated on Fig. 6.

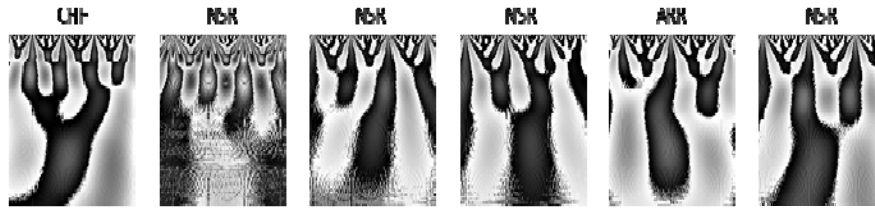


Fig.6. ECG signal segments

After running 10 epochs using a stochastic gradient descent as optimizer, and computing the loss the accuracy metrics shows a good performance. We got accuracy of 96% on the ECG dataset. Predicted label (loss: 0.1401 - accuracy: 0.9573). The results of the investigation are show on the Fig. 7.:

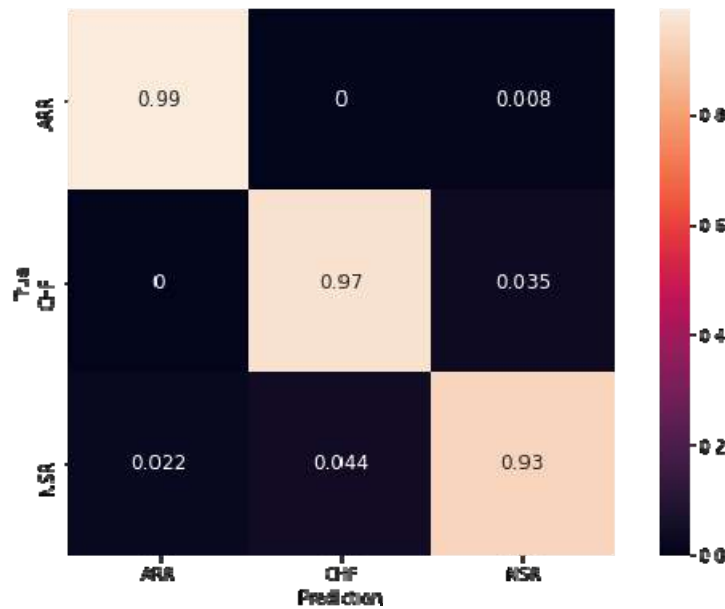


Fig. 7. Confusion matrix

With a wavelet convolution neural network we get a precise model. With a simple convolution neural network we were able to get a precise model which quickly allows us to detect a healthy person from others with heart disease. Based on ECG data a classification over three groups of patient with cardiac arrhythmia, congestive heart failure, healthy person was made. Automatic ECGs classification is useful for emergency medicine and remote monitoring as a method of disease identification.

The Wavelet transform technique allows us to explore the scalogram of non-stationary signals and then take advantage of image classification techniques.

5 Conclusion

There are many configurable model of the neural networks configurations that affect the result. Continuous and discrete wavelet transformations are superior to other non-stationary signal processing methods and classification accuracy.

The Wavelet transform technique is the most performant if the composite components of the signal vary in time. Wavelets transform signals from the time-domain to the frequency-domain and gives the spectrum. The Wavelet transform technique allows to explore the frequency domain of the non-stationary signals. The results of the CWT in combination with CNN prove that these approaches are a good choice for the classification of time series.

A neural network with wavelet cores is mainly designated for pattern recognition. Wavelet coefficients are used as input to the network.

A wavelet network usually consists of a neural network with direct connection, with one hidden layer and activation functions of which are taken from the wavelet family. Since wavelet coefficients quite effectively reflect the essence of the signal, dimensionality reduction can be achieved by wavelet processing.

Wavelet-neural networks combine the wavelet transform (decomposition-deconvolution) into one neural networks.

References

1. Tan, R., Perkowski, M.: Toward Improving Electrocardiogram (ECG) Biometric Verification using Mobile Sensors: A Two-Stage Classifier Approach. Published online 2017 Feb 20. doi: 10.3390/s17020410 (2017)
2. Hussein, A.F., AlZubaidi, A.K., Al-Bayaty, A., Habash, Q.A.: An IoT Real-Time Biometric Authentication System Based on ECG Fiducial Extracted Features Using Discrete Cosine Transform. Available from: <https://www.researchgate.net/publication/319327157>. [Accessed: 2020-03-03]
3. Borries, R.F., Pierluissi, J.H., Nazeran, H.: Wavelet Transform-Based ECG Baseline Drift Removal for Body Surface Potential Mapping. Proceedings of the 2005 IEEE. DOI: 10.1109/IEMBS.2005.161531. (2005)
4. Physionet ECG data. https://github.com/mathworks/physionet_ECG_data/ [Accessed: 2020-03-03]
5. Daubechies, I.: The wavelet transform, time-frequency localization and signal analysis. IEEE Transactions on Information Theory 36(5): 961-1005 (1990)
6. Baim, D.S., Colucci, W.S., Monrad, E.S., Smith, H.S., Wright, R.F., Lanoue, A., Gauthier, D.F., Ransil, B.J., Grossman, W., Braunwald, E.: Survival of patients with severe congestive heart failure treated with oral milrinone. J American College of Cardiology 7(3): 661-670 (1986)
7. Mallat, S.G.: A theory for multi resolution signal decomposition: the wavelet representation. IEEE T PAMI 11(7): 674-693 (1989)
8. Yan, Z., Miyamoto, A., Jiang, Z., Liu, X. An overall theoretical description of frequency slice wavelet transform. Mech. Syst. Signal Process. 24(2): 491-507 (2010)
9. Yan, Z., Miyamoto, A., Jiang, Z.: Frequency slice wavelet transform for transient vibration response analysis. Mech. Syst. Signal Process. 23(5): 1474-1489 (2009)
10. Billings, S. A., Wei, H.-L. A new class of wavelet networks for nonlinear system identification. IEEE Transactions on neural networks 16(4): 862-874 (2005)
11. Stelia, O., Krak, I., Potapenko, L.: Controlled Spline of Third Degree: Approximation Properties and Practical Application. In: Lytvynenko V., Babichev S., Wójcik W., Vynokurova O., Vyshemyrskaya S., Radetskaya S. (eds) Lecture Notes in

- Computational Intelligence and Decision Making. ISDMCI 2019. Advances in Intelligent Systems and Computing, vol 1020. Springer, Cham, pp.215-224 (2020)
12. Khorozov, O., Krak, I., Kulias, A., Kasianiuk, V., Wojcik, W., Tergeusizova, A.: Monitoring vital signs using fuzzy logic rules. In: Information technology in Medical Diagnostics II - - Proceedings of the International Scientific Internet Conference on Computer Graphics and Image Processing and 48th International Scientific and Practical Conference on Application of Lasers in Medicine and Biology, 2018 CRC Press/Balkema. pp. 237–244 (2019)
 13. Nagendra, H. et al.: Application of Wavelet Techniques in ECG Signal Processing: An Overview. *International Journal of Engineering Science and Technology* 3(10): 7432-7443 (2011)
 14. Umamaheswara, G.R., Murlidhar, M., Varadarajan, S.: ECG De-noising using improved thresholding based on Wavelet transforms, *International Journal of Advanced Computer Science and Network Security* 9(9): 221-225 (2009)
 15. Akshay, N., Vamse, N.A., Yeddanapudi, N.D.: ECG noise Removal and QRS complex detection using UWT, *International Conference on Electrical and Information Engineering (ICEIE)*, IEEE vol.2, pp.438-442 (2010)
 16. Zhongyao, Z., Chengyu, L., Yaowei, L., Yixuan, L., Jingyu, W., Lin, B.-S., Jianqing, L.: Noise Rejection for Wearable ECGs Using Modified Frequency Slice Wavelet Transform and Convolutional Neural Networks. *IEEE Access*. pp. 1-1. 10.1109/ACCESS.2019.2900719 (2019)
 17. Yu, S.N., Chen, Y.H.: Electrocardiogram beat classification based on wavelet transformation and probabilistic neural network, *Pattern Recogn. Lett.* 28: 1142–1150 (2007)
 18. Rai, H.M., Trivedi, Y.H., Shukla, S.: ECG signal processing for abnormalities detection using multi-resolution wavelet transform and Artificial Neural Network classifier. *Measurement* 9: 3238-3246 (2013)
 19. Martis, R.J., Acharya, U.R., Min, L.C. ECG beat classification using PCA, LDA, ICA and discrete wavelet transform. *Biomedical Signal Processing and Control* 5: 437-448 (2013)

Intelligent Forecasting in Multi-Criteria Decision-Making

Vladyslav Borysenko^[0000-0002-9381-3080], Galyna Kondratenko^[0000-0002-8446-5096],
Ievgen Sidenko^[0000-0001-6496-2469], Yuriy Kondratenko^[0000-0001-7736-883X]

Intelligent Information Systems Department, Petro Mohyla Black Sea National University,
68th Desantnykiv Str., 10, Mykolaiv, 54003, Ukraine,
kalambyr11@gmail.com, halyna.kondratenko@chmnu.edu.ua,
ievgen.sidenko@chmnu.edu.ua, yuriy.kondratenko@chmnu.edu.ua

Abstract. In this paper, the subject of the research is the intelligent forecasting in multi-criteria decision-making, in particular, methods of prediction of time series: traditional models of autoregression, smoothing techniques and machine learning methods with the use of artificial neural networks and deep learning. Time series of Amazon share prices from the official sources over the past years serve as a base for exploration. The purpose of the work is to find out the parameters that influence the efficiency and accuracy of models for analysis and forecast the share prices. Such assessment is complicated and vital because various types of criteria, in particular, sales and profitability, market indexes, exchange rates and general trends, etc., usually influence the decision-making process. The methods of research include the consideration and analysis of prediction methods using specific metrics. The relevance of the topic assumes that the accurate prediction at the financial market may contribute to the financial benefits for companies, government and other players on stock exchanges. A comparative analysis of the considered forecasting methods was conducted. It allows choosing the most appropriate intellectual method for increasing the efficiency of a specific share price prediction. The further development of the subject of research includes ensemble-learning methods for neural networks, feature engineering, collecting a more extensive data set for forecasting.

Keywords: time series, autoregression, smoothing, artificial neural networks, convolutional neural networks, recurrent neural networks, decision-making, multi-criteria approach, stock market, share price.

1 Introduction

The analysis of the behavior of stock prices is characterized by the ambiguous behavior of the process, which is usually affected by many factors (trend, seasonality, the geopolitical situation, etc.). Forecasting is a key point when making investment decisions. The ability to predict the behavior of stock for making final decisions allows you to make the best choice which otherwise might be unsuccessful [1].

However, not all investors successfully profit from their investment. This is because the stock price is constantly fluctuating, and at any moment the price may fall

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

below the price at which it was purchased. Therefore, to predict how the financial market will behave is one of the most difficult tasks in the economy. In this prediction, it is necessary to consider various factors, such as physical, psychological, rational and irrational behavior, and the like. All these aspects lead to the conclusion that stock prices are very volatile and it's difficult to predict them with a high degree of accuracy. However, this task is urgent for the world and the whole international economy, since the possibility of accurate prediction of the value of the stock is closely linked to the financial gain of the companies, the government or personal capital and the formation of the more rational financial behavior [2-3].

Accurate predicting the value of assets on the exchange will also help to reduce investment risk and protect the investment income from the market volatility.

2 Related Works and Problem Statement

A stock market (hereinafter the SM) is an organized market where the securities owners make the agreements of purchase and sale with the members of the SM as intermediaries. The prices of these securities are determined by supply and demand, and the process of sale is governed by the rules and regulations [4].

There are three different approaches to research and predict the asset prices on the stock market: technical, technological and fundamental analysis. In this paper, the combination of the first two approaches was used.

Technical analysis is the study of the dynamics of the main indicators of the market with the help of graphical methods to predict the future direction of their movement. A significant number of participants in stock and OTC markets use technical analysis [5]. A successful trader O. Elder has spoken figuratively on this subject: "Technical analysis is related to public opinion polls. It is a combination of science and art. The scientific part consists in using statistical methods and computers; the creative part is the interpretation of the data."

Technological analysis originated in the era of the application of computer technology in business and other fields. The success of its application in solving complex tasks is largely determined by the possibilities of modern information technologies. The result of technological research is usually a selection of well-defined alternatives. Therefore, the origins of this analysis, its methodological concepts lie in those disciplines that deal with decision-making: Theory of Operations and General Management Theory.

The technological analysis includes the ability to analyze, to predict, and to design decision-making in complex systems of various nature, which are based on the time series.

A time series can be represented by four components: *trend*, *seasonal variations*, *cyclical variations* and *irregular factors* [6]. The classical approach to the construction of the time series model is to schedule it into several components, each of which examines specific methods. A trend is a general systematic linear or non-linear component that can change over time. The seasonal and cyclical components are periodi-

cally repeating components. These components are not always present simultaneously in the time series. In our case, there are no seasonal and cyclical components.

3 Basic Concepts and Intellectual Forecasting Methods for Stock Prices Prediction

Since a prediction based on statistical data collected with the same interval is required, we are dealing with the time series. Therefore, the time series models will be considered, and those that suit the task of forecasting data will be selected. The most popular forecasting time series models are *autoregressive models*, *autoregressive models with a moving average* and *models derived from them*. Thus, we will focus on these models [3].

After selecting the specific methods of solving tasks, it is necessary to directly conduct a prediction of selected characteristics. Constructing the desired models will be implemented by their gradual complication, in other words, we will move from simple to complex. In this paper, models and methods are evaluated using MAE and MSE metrics [1, 7].

MAE measures the average absolute value of errors in a set of predictions for continuous variables. Assuming that $\hat{y}_t$ is the values of the time series forecast in period t , the metric is given by (1):

$$MAE = \sum_t |y_t - \hat{y}_t|. \quad (1)$$

MAE is a linear score, which means that all individual differences are resolved to the same average. MSE is the difference between the forecast and corresponding observed values squared at each iteration. Because errors rise to the square before they are averaged, MSE has a relatively high weight to large bias. This means that MSE is most useful when large errors are particularly undesirable, which is consistent with the objectives of this work:

$$MSE = \sum_t (y_t - \hat{y}_t)^2. \quad (2)$$

Both metrics, MAE and MSE, can take values from 0 to ∞ and do not take the direction of errors into account. The smaller is the value the index takes, the more accurate is the forecast.

Smoothing is an important and widespread method of financial market forecasting. Smoothing methods are used to reduce the influence of random components (random fluctuations) in time series. They provide an opportunity to obtain more “pure” values that consist only of deterministic components. Some of the methods are aimed at recovering some of the components such as trend.

The authors present 5 smoothing methods typically found when forecasting financial data: the simple moving average; the weighted moving average; the exponential

moving average; the double exponential moving average; the triple exponential moving average [8].

The basic assumption of these methods is that the fluctuations in past values represent random deviations from a smooth curve, which can be extrapolated to create a forecast.

The autoregressive model is another effective tool for understanding and predicting future values of the time series, which includes devolution of a variable on values of series in the past. The importance of ARMA models lies in their flexibility and in their ability to describe almost all features of the stationary time series. Autoregressive of these models describe how consecutive observations in time affect each other, while parts of the moving averages capture some possible unobserved upheavals, and that allows simulating various phenomena that can be observed in a variety of fields from biology to finance [1-3].

The main idea of autoregression methods is that future values of the time series cannot deviate to higher or lower than the previous values of the time series whatever the reasons that caused those deviations are. The paper has presented such models of autoregression as simple AR (Autoregression Model), ARMA (Autoregressive Moving Average Model) and ARIMA (Autoregressive Integrated Moving Average Model) [1, 8].

We've also proved that Artificial Neural Networks (ANN) have a significant advantage in time series forecasting because they are endowed with the capacity to solve complex problems of forecasting [2, 9, 10].

The output value of the neural network is defined mathematically as:

$$y_t = \alpha_0 + \sum_{j=1}^q \alpha_j g \left(\beta_{0i} + \sum_{i=1}^p \beta_{ij} y_i \right) + \varepsilon_t, \quad (3)$$

where p is the number of the input variables, q is the number of the hidden nodes, α_j and β_{ij} are the weights, ε_t is the random noise.

As a function of g the following functions can be used [2, 11, 14].

Sigmoidal function:

$$f(x) = \frac{1}{1 + e^{-x}}. \quad (4)$$

In some literature, this is called a logistic function. This nonlinear function is one of the most common activation functions for deep learning [12].

Hyperbolic tangent:

$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}. \quad (5)$$

This function gives the best result for multilayer neural networks compared with the sigmoidal function. However, the function does not solve the problem of the vanish-

ing gradient. The main benefit offered by this function is that it is centered relative to zero, which helps in the process of error backpropagation [11].

Softmax function:

$$f(x) = \frac{e^{x_i}}{\sum_j e^{x_j}}. \quad (6)$$

The Softmax function is another type of activation function used in neural computing. It is used to calculate the probability distribution of a vector of real numbers and gets the value range from 0 to 1.

ReLU function:

$$f(x) = \max(0, x). \quad (7)$$

This function is considered to be the most successful and widely used transfer function. ReLU shows the best productivity in deep learning compared to sigmoid and hyperbolic tangent. ReLU is a nearly linear function and therefore preserves the properties of linear models, which make them easily optimized by methods of gradient descent. The main advantage of using ReLU in calculus is that this function does not require computing the exponent or dividing, therefore it guarantees a more rapid execution [11, 13].

ANN has been and continues to be actively used in the financial markets; one of the main advantages of ANN that make them so popular as harbingers of the market is the natural nonlinearity that allows them to learn nonlinear mapping and correlation of data. ANN also works on data, can be trained in real-time; they are highly adaptive, easy to retrain in the event of market fluctuations and, finally, do well with data that contain a certain number of errors [14-17]. ANN mechanism implies minimum participation of the analyst in the model shaping, as far as the learning ability is characteristic of all neural network models and learning algorithms adapt (adjust) the weights according to the structure of the data presented for training.

Consider the most common optimization techniques.

Gradient descent [18]:

$$x_{t+1} = x_t - \alpha \cdot f'(x_t). \quad (8)$$

According to this method, the steps proportional to the opposite gradient value are taken for minimizing functions. The parameter to this method is a descent speed α . When reaching a large value α will possibly make large steps for finding the minimum, but there is a risk of skipping the lowest point. At a very low speed of learning, the algorithm will move confidently in the direction of the negative gradient, but the implementation, in this case, will take a long time.

The method of stochastic gradient descent [19]. This is a type of gradient descent, which handles 1 element for learning at each iteration. So, parameters are updated even after one iteration, which processed only one value of the variable. This allows optimizing the target function much faster than ordinary gradient descent. But if the

number of training data is very large, the number of iterations is therefore sufficiently large.

Mini-batch gradient descent [20]. This is the type of gradient descent, which is faster than batch and stochastic gradient descent. Let there be the m number of input data, then in one iteration of this method $b < m$ elements will be processed. Consequently, even if the number of training data is large, it is processed in smaller training groups at once. Thus, the method works for a large amount of training data by optimizing an objective function with less number of iterations.

Gradient descent with momentum is the method that helps to accelerate a descent in the corresponding direction and dampens the oscillations, approaching the local minimum. This is implemented by adding a part of the update vector from the previous step to the vector of the current update.

Adam (Adaptive Moment Estimation) [21]. Adam optimizer is one of the most popular optimization algorithms of gradient descent since it is computationally efficient and has very little memory requirements. This method calculates an individual adaptive learning rate for each parameter from the evaluation of the first and second moments of the gradients.

In the simulation of isolated time series using ANN models a transformation of the original data for increasing the number of input neurons and, consequently, increasing predictive ability is allowed. Modeling of the time series using the ANN mechanism consists in the formation of ANN as a certain structure, which describes the behavior of the system under study in points in time, and forecasting is the prediction of the future behavior of the system in the background.

To sum it up, this work explains the mathematical foundations of neural networks, namely multilayer perceptron of Rumelhart, convolutional neural networks and recurrent neural networks. The correctly chosen input is significant. In this paper, only the value of the stock price in the past was taken as the initial data.

4 Practical Implementation of the Described Methods and Techniques

This section describes the architecture of the developed program and the final comparative results of all models. Compared to the classical mathematical methods, this work has allowed confirming the viability and the feasibility of using artificial neural networks for further study of their application in the financial markets [1].

For this study, AMAZON stock statistics were taken from the official NASDAQ (National Association of Securities Dealers Automated Quotation) website for the past 5 years. Dataset (Fig. 1) is daily information on the price of shares at the beginning and at the end of the day, the minimum and maximum value of a share during the day, and the number of shares sold. The above parameters can be considered as criteria since they have an optimization direction (maximum or minimum). Moreover, the considered methods and approaches are also used to solve multi-criteria decision-making problems with a more complex structure and the number of criteria. There-

fore, this task of forecasting the stock price with the considered methods and approaches can be considered in the context of multi-criteria decision-making.

The moving average method was implemented for various values of the N parameter of the previous time points number that had to be taken into account when building a forecast (Fig. 2). It was checked that decreasing the N window length model shows the more accurate result on the test sample, indicating that the latest data is the most influential to the prediction, i.e. prediction considering the past 5 days is more accurate than the forecast, which takes the last 20 days into account.

	date	close	volume	open	high	low
0	5/10/19	1164.27	1314546	1162.38	1172.600	1142.50
1	5/9/19	1162.38	1185973	1166.27	1169.660	1150.85
2	5/8/19	1166.27	1309514	1174.10	1180.424	1165.74
3	5/7/19	1174.10	1551368	1189.39	1190.440	1161.04
4	5/6/19	1189.39	1563943	1185.40	1190.850	1166.26

Fig. 1. Dataset as input data

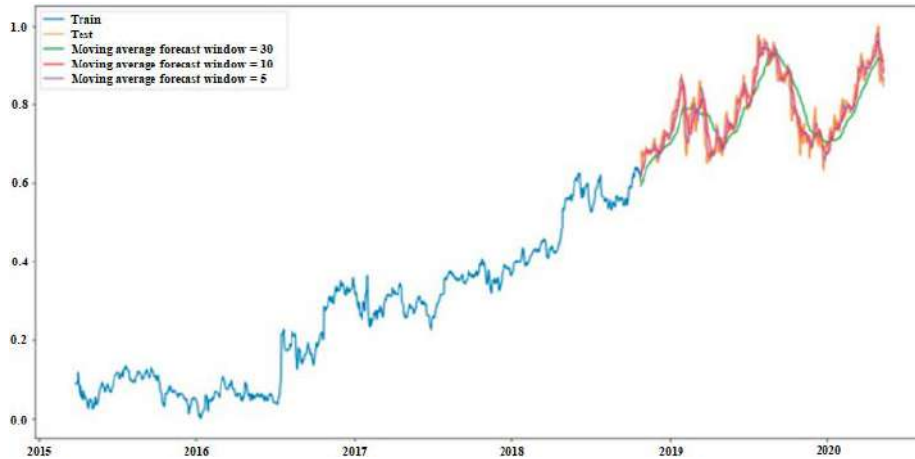


Fig. 2. The schedule of the simple moving average forecast

For the exponential moving average (Fig. 3) parameter is the level of the α smoothing coefficient, which represents the degree of reduction of weighting from 0 to 1. The lower the level of smoothing is, the more accurate is the forecast, as every previous reading weighs more [2-3].

This part of the paper also contains a comparison of the results of forecasting (by terms of MAE and MSE) stock prices by various types of neural networks (MLP, CNN, LSTM) with an additional comparison of their models, which have different structures and parameters. Therefore, a detailed review of the architecture of neural

networks, justification for choosing the number of hidden layers, the number of neurons in them, is not necessary.

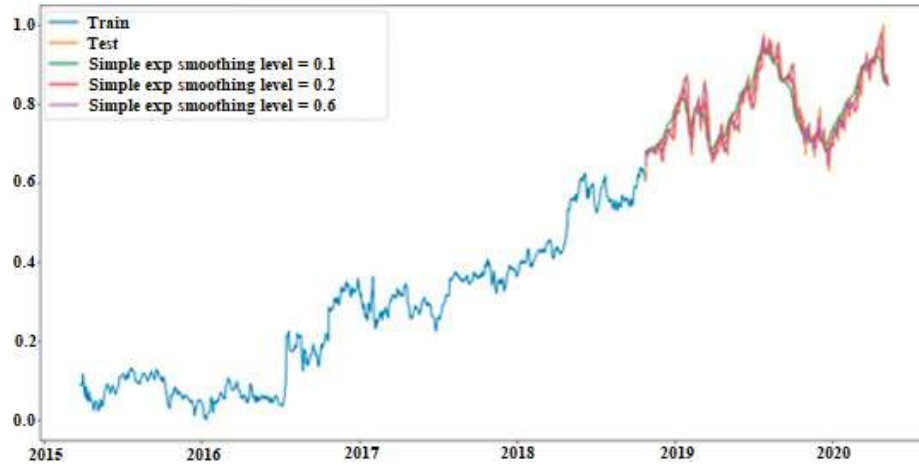


Fig. 3. The schedule of exponential moving average method forecast

For the data with a distinct trend, which corresponds to the input operation, the method of double exponential smoothing that is the recursive application of exponential filter twice, gives the best result. This method has provided an additional parameter β , which is responsible for the smoothing of the trend (Fig. 4). The combination of the α and β pair has adjusted the accuracy and the quality of the forecast [7].



Fig. 4. The schedule of the double exponential moving average method forecast

The results of implementation of all smoothing methods are summarized in Table 1.

The autoregressive model [22] is an effective tool for understanding and predicting future values of the time series, which includes the devolution of a variable on values

of series in the past. Autoregressive of these models describe how consecutive observations in time affect each other, while the parts of the moving averages capture some possible unobserved upheavals. For the considered models, the autoregressive Akaike coefficient allowed the software to choose the model that gives the best forecast. Therefore, in the case of the dataset, it was the model ARIMA (2,1,1).

Table 1. The smoothing methods implementation results

Method	Specific parameters	Test MSE	Test MAE
Simple moving average	Window $N = 30$	0.0029	0.0455
	Window $N = 10$	0.0012	0.0281
	Window $N = 5$	0.0005	0.0180
Exponential moving average	The level of smoothing $\alpha = 0.1$	0.0018	0.0340
	The level of smoothing $\alpha = 0.2$	0.0011	0.0259
	The level of smoothing $\alpha = 0.6$	0.00049	0.0165
Double exponential moving average	$\alpha = 0.1, \beta = 0.3$	0.00244	0.0388
	$\alpha = 0.2, \beta = 0.8$	0.00163	0.0322
	$\alpha = 0.6, \beta = 0.6$	0.00054	0.0177

The execution results of all three models were summarized in a single comparative Table 2.

Table 2. The autoregressive models implementation results

Model	Specific parameters	Test MSE	Test MAE
Simple autoregressive model	$p = 2$	0.000347	0.013322
Autoregressive moving average model	$p = 2, q = 1$	0.000348	0.013346
Autoregressive integrated moving medium model	$p = 2, q = 1, d = 1$	0.000345	0.013335

The simplest type of the neural network is a single-layer perceptron network (usually a multilayer perceptron of Rumelhart), which consists of a single layer of output nodes, and the inputs provided directly to the outputs via a series of weights. Universal approximation theorem for neural networks states that each continuous function, which maps intervals of real numbers to a certain initial interval of real numbers, can be arbitrarily closely approximated by a multilayer perceptron with only one hidden layer [2, 12]. The main characteristic of the multilayer perceptron is its architecture, namely the number of hidden layers, nodes and activation functions. Besides, since the study outcome partially depends on the initialization of variables, each model was trained separately 5 times, and all the characteristics for a comparative table are average values. For a more in-depth study of the model 5 architectures were implemented in the paper, the accuracy of each of them is represented in Table 3.

According to the comparative results (Table 3), the dependence of the quality of the forecast on the number of nodes in the hidden layer, the optimization method and

the activation function is completely traced. The best result was the MLP 20-100-1 model with the Adam optimization algorithm and hyperbolic tangent as an activation function. Here is a graph of the price of stocks sold predicted by this model (Fig. 5, a), as well as a graph of the forecast error (Fig. 5, b).

Table 3. MLP learning results

Model	Parameters number	Training MAE	Training MSE	Test MAE	Test MSE	Algorithm	Activation function
MLP 20-10-1	221	0.0645	0.0074	0.090	0.014	adam	relu
MLP 20-20-1	441	0.0462	0.0038	0.052	0.005	adam	relu
MLP 20-100-1	2201	0.0438	0.0038	0.037	0.002	adam	relu
MLP 20-100-1	2201	0.0035	0.0428	0.032	0.002	adam	tanh
MLP 20-100-1	2201	0.0668	0.0078	0.093	0.014	sgd	relu

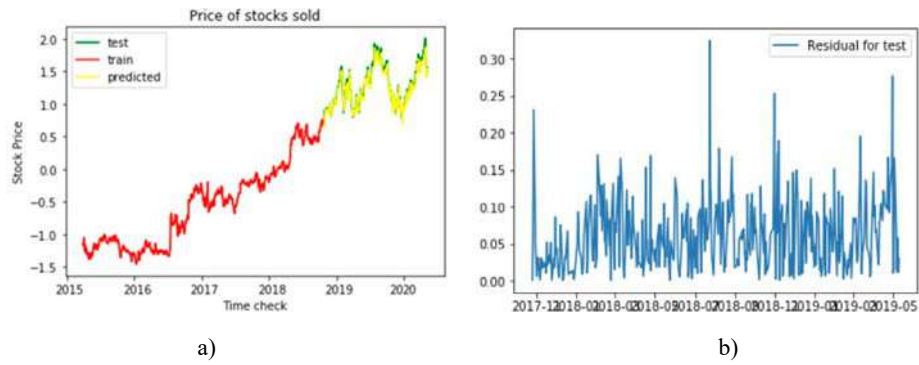


Fig. 5. MLP 20-100-1: a) price of stocks sold; b) forecast error

Convolutional neural network [23] is a type of artificial neural network, in which the pattern of connectivity between neurons is inspired by the organization of the visual cortex of animals, individual neurons that are arranged in such a way that they respond to overlapping regions of the field. The main advantage of convolutional neural networks is that we use the convolutional layers to identify the signs of a network that allows training of a neural network without complex pre-processing since useful features are learned during training [13]. Unlike multilayer perceptrons, convolutional neural networks have a complex structure, since they need to gauge the number and type of layers, the number of nodes in each layer, the optimization method, the activation function and the number of filters. So CNN-20-200-3 will designate 20 knots at the entrance as the first layer, 200 filters sized 3x3, MaxPooling layer and 2 fully-connected layers. In this work, 5 different CNN architectures were tested, the results are shown in Table 4.

The CNN-10-256-3 model with the Adam optimization algorithm and ReLU activation function showed the best result.

A recurrent neural network (RNN) is an artificial neural network, the neurons of which transmit feedback signals to each other [24]. The idea of RNN is to use consistent information. In traditional neural networks, we assume that all the inputs (and the outputs) are independent of each other. But for many tasks, this is not the best idea. This type of neural network is well suited to predict the value of the stock, as further steps may depend on the past.

Table 4. CNN learning results

Model	Parameters number	Training MAE	Training MSE	Test MAE	Test MSE	Algorithm	Activation function
CNN 10-256-3	198.657	0.0238	0.0011	0.096	0.013	adam	relu
CNN 20-256-5	201.367	0.0182	0.0074	0.174	0.038	adam	relu
CNN 20-256-3	199.937	0.0513	0.0050	0.290	0.110	sgd	relu
CNN 20-500-3	756.501	0.0515	0.0050	0.259	0.089	sgd	relu
CNN 20-300-3	273.901	0.0414	0.0030	0.255	0.079	sgd	tanh

For the implementation of the forecasting recurrent neural networks, the LSTM model (Long Short-Term Memory Units) was chosen [25]. LSTMs help to save the bug, which can be spread through time and layers. Maintaining a more constant error, they allow the networks to continue exploring for many time steps [16]. To analyze the effect of parameters on the prediction value of the shares 5 variants of the LSTM architecture summarized in the following Table 5 were implemented. LSTM-20-30-30-30-1 was used to identify neural LSTM networks with initial data of the previous 20 days, the number of nodes 30 on the first, the second and the third layers of LSTM and one fully connected layer with the resulting output of one element.

Table 5. LSTM learning results

Model	Parameters number	Training MAE	Training MSE	Test MAE	Test MSE	Algorithm	Activation function
LSTM 20-30-30-30-1	18,631	0.0074	1.0564e-04	0.0198	0.0005	adam	relu
LSTM 20-30-30-40-1	22,681	0.0079	1.8716e-04	0.0188	0.0005	adam	relu
LSTM 30-30-30-30-1	18,631	0.0106	1.9467e-04	0.0159	0.0004	adam	relu
LSTM 10-30-30-30-1	18,631	0.0099	1.7086e-04	0.0150	0.0003	adam	relu
LSTM 10-30-30-30-1	18,631	0.0082	1.2895e-04	0.0152	0.0003	sgd	relu

The best result was shown by the LSTM model 10-30-30-30-1 with the Adam optimization algorithm and ReLU activation function, and with the least lag window.

The performance of classical algorithms and machine-learning algorithms in section 4 can be summarized in a comparative Table 6.

Table 6. The comparative results of the implemented methods

Model/Method	Advantages	Disadvantages
Smoothing	Ability to process trends of varying levels and seasonality components	Vulnerability to extreme values
Autoregression	Easy to automation	Limitations in the assumptions
ANN	Ability to handle complex nonlinear patterns. High forecast accuracy	Require a large amount of data

5 Conclusions

For investors, capital management is becoming increasingly important today. Professional investment managers and individual investors tend to have effective tools for understanding the trends in the stock market, to minimize investment risk and increase their profits. Some people believe that it is very difficult to predict stock prices. However, in the real business world, successful traders carry out thousands of transactions every year.

Since the very beginning of the financial operations on the stock exchange, people developed methods to predict the value of assets in the future. With advances in computing technology and artificial intelligence, the accuracy of the methods is growing every day.

In this paper, the authors compared the indicators of forecasting between the neural network and the classic time series forecast method, namely, the value of the company's shares on the stock exchange. The analysis showed that the neural network models described in this study showed a very much better able to accurately predict, and therefore confirmed the viability and the feasibility of using artificial neural networks for further study of their application in the financial markets. The predictive power of the model was influenced by different factors, depending on the architecture and input data.

References

1. Kingdon, J.: Intelligent Systems and Financial Forecasting. Springer, London (1997).
2. Palit, A., Popovic, D.: Computational Intelligence in Time Series Forecasting. Theory and Engineering Applications. Springer, London (2005). doi: 10.1007/1-84628-184-9
3. Guerard, J.: Introduction to Financial Forecasting in Investment Analysis. Springer, New York (2013). doi: 10.1007/978-1-4614-5239-3
4. Pavlov, V., Pilipenko, I., Krivov'yazyuk, I.: Securities in Ukraine: Students Book. Condor, Kyiv (2004). (in Ukrainian)
5. Sokhatska, O.: Stock Exchange. Carte Blanche, Ternopil (2003). (in Ukrainian)

6. Box, G., Jenkins, G., Reinsel, G.: Time Series Analysis: Forecasting & Control. John Wiley & Sons, New Jersey (2008)
7. Wei, W.: Time Series Analysis: Univariate and Multivariate Methods. Pearson, New York (2005)
8. Sitte, R., Sitte, J.: Neural Network Systems Technology in the Analysis of Financial Time Series. In: Leondes, C.T. (eds.) Intelligent Knowledge-Based Systems, pp. 1564–1615. Springer, Boston, MA (2005). doi: 10.1007/978-1-4020-7829-3_43
9. Kondratenko, Y., Gordienko, E.: Neural Networks for Adaptive Control System of Caterpillar Turn. In: Annals of DAAAM for 2011 & Proceeding of the 22th Int. DAAAM Symp. "Intelligent Manufacturing and Automation", pp. 0305–0306 (2011)
10. Gomolka, Z., Dudek-Dyduch, E., Kondratenko, Y.P.: From homogeneous network to neural nets with fractional derivative mechanism. In: Rutkowski, L. et al. (eds.) International Conference on Artificial Intelligence and Soft Computing (ICAISC-2017), Part I, pp. 52–63, Zakopane, Poland (2017)
11. Nwankpa, C., Ijomah, W., Gachagan, A., Marshall, S.: Activation Functions: Comparison of Trends in Practice and Research for Deep Learning. arXiv:1811.03378 [cs.LG] (2018)
12. Robert, J.: The Application of Neural Networks in the Forecasting of Share Prices. Finance and Technology Publishing, New York (1996)
13. Di Persio, L.: Artificial Neural Networks architectures for stock price prediction: comparisons and applications. International journal of circuits, systems and signal processing **31**(10), 404–405 (2016)
14. Kushneryk, P., Kondratenko, Y., Sidenko, I.: Intelligent dialogue system based on deep learning technology. In: 15th International Conference on ICT in Education, Research, and Industrial Applications: PhD Symposium (ICTERI 2019: PhD Symposium), vol. 2403, pp. 53–62, Kherson, Ukraine (2019)
15. Sidenko, I., Filina, K., Kondratenko, G., Chabanovskyi, D., Kondratenko, Y.: Eye-tracking technology for the analysis of dynamic data. In: IEEE 9th International Conference on Dependable Systems, Services and Technologies (DESSERT), pp. 479–484, Kiev, Ukraine (2018). doi: 10.1109/DESSERT.2018.8409181
16. Cryer, J., Chan K.-S.: Time Series Analysis. With Applications in R. Springer, New York (2008). doi: 10.1007/978-0-387-75959-3
17. Beran, J.: Mathematical Foundations of Time Series Analysis. A Concise Introduction. Springer, Cham (2017). doi: 10.1007/978-3-319-74380-6
18. Paper, D.: Gradient Descent. In: Data Science Fundamentals for Python and MongoDB, pp. 97–128. Apress, Berkeley, CA (2018). doi: 10.1007/978-1-4842-3597-3_4
19. Ketkar, N.: Stochastic Gradient Descent. In: Deep Learning with Python, pp. 113–132. Apress, Berkeley, CA (2017). doi: 10.1007/978-1-4842-2766-4_8
20. Danner, G., Jelasity, M.: Fully Distributed Privacy Preserving Mini-batch Gradient Descent Learning. In: Bessani, A., Bouchenak, S. (eds.) Distributed Applications and Interoperable Systems. DAIS 2015. Lecture Notes in Computer Science, vol. 9038, pp. 30–44. Springer, Cham (2015). doi: 10.1007/978-3-319-19129-4_3
21. Jiang, Y., Han, F.: A Hybrid Algorithm of Adaptive Particle Swarm Optimization Based on Adaptive Moment Estimation Method. In: Huang, D., Bevilacqua, V., Premaratne, P., Gupta, P. (eds.) Intelligent Computing Theories and Application. ICIC 2017. Lecture Notes in Computer Science, vol. 10361, pp. 658–667. Springer, Cham (2017)
22. Olson, D., Wu, D.: Autoregressive Models. In: Predictive Data Mining Models. Computational Risk Management, pp. 55–69. Springer, Singapore (2017)
23. Ayyadevara, V.: Convolutional Neural Network. In: Pro Machine Learning Algorithms, pp. 179–215. Apress, Berkeley, CA (2018). doi: 10.1007/978-1-4842-3564-5_9

24. Hammer, B.: Learning with recurrent neural networks. Springer, London (2000)
25. Manaswi, N.: RNN and LSTM. In: Deep Learning with Applications Using Python, pp. 115–126. Apress, Berkeley, CA (2018). doi: 10.1007/978-1-4842-3516-4_9

Machine Learning Classification of Multifractional Brownian Motion Realizations

Lyudmyla Kirichenko^{1[0000-0002-2780-7993]}, Bulakh Vitalii^{1[0000-0002-9177-8787]}, Tamara Radivilova^{1[0000-0001-5975-0269]}

Kharkiv National University of Radio Electronics, Kharkiv, 61166, Ukraine
lyudmyla.kirichenko@nure.ua
bulakhvitalii@gmail.com
tamara.radivilova@gmail.com

Abstract. A comparative analysis of machine learning classification of stochastic time series based on their multifractal properties is proposed. Multifractal time series were obtained by generating realizations of fractional Brownian motion in multifractal time. The features for classification were statistical, fractal and recurrent characteristics calculated for each time series. The various machine learning classifiers were chosen for classification: bagging with classification and regression decision trees, random forest with classification and regression decision trees, fully connected perceptron and recurrent neural network. Both cumulative time series of multifractal Brownian motion and time series increments were carried out. It was shown that in general, classification accuracy is higher when using series of increments. When classifying realizations of multifractional Brownian motion, bagging and recurrent neural network showed the best accuracy.

Keywords: multifractal, multifractional Brownian motion, classification of time series, features, random forest, bagging, recurrent neural network

1 Introduction

Over the past decades, it has become apparent that many complex objects and systems have fractal (self-similar) properties. This applies to time series that reflect the dynamics of complex nonlinear systems. Numerous studies show that changes in the structure or state of a system lead to changes in fractal properties of the corresponding time series. The results of time series fractal analysis are widely used in practice, in particular, for analysis of information systems with self-similar data flows to prevent system overload and to analyze and predict financial markets [1-4].

In these cases, models of fractal processes that are used for forecasting, modeling, etc. play an important role. One of the interesting and applied in practice models is the fractional Brownian motion in the multifractal time proposed by Mandelbrot [5,6]. Currently, multifractional Brownian motion is used to model various phenomena [7–

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

9], among which the prevailing place is occupied by financial series [10-14] and self-similar infocommunication traffic [15].

In recent years, to solve practical problems associated with the analysis and recognition of dynamic phenomena, time series classification by machine learning has been used [16–20]. Typically, time series are collected in classes based on whether they have a common attribute or property. Often a change in the state of a system entails a change in its fractal structure. For example, telecommunication traffic under DDoS attacks change their fractal properties [21-23]. Thus, the task of classifying time series based on their fractal properties is relevant. However, the classification of time series by machine learning methods is a fairly new area and most of the studies do not take into account their fractal properties.

The objective of the work is a comparative analysis of the fractal time series classification carried out by machine learning methods. Time series are realizations of the fractional Brownian motion in the multifractal time and time series of their increments, which are divided into classes according to their fractal properties. The ensembles of decision trees and neural networks are considered as classification methods.

2 Fractal Random Processes and Models

A random process $X(t)$ is self-similar if the process $a^{-H}X(at)$ has the same finite-dimensional distribution laws with $X(t)$. The parameter H , $0 < H < 1$, is called the Hurst exponent. It is the self-similarity degree and the measure of the long-term dependence of the process. The moments of the self-similar process satisfy the scaling relation $E\left[|X(t)|^q\right] \propto t^{qH}$.

Multifractal random processes are inhomogeneous fractal processes and have more flexible scaling relation: $E\left[|X(t)|^q\right] \propto t^{\tau(q)+1}$, where $\tau(q)$ is a nonlinear function of scaling exponent [24].

One of the most used characteristics of multifractal processes and time series is the generalized Hurst exponent $h(q)$, which is associated with the function $\tau(q)$ by the ratio [25]:

$$h(q) = \frac{\tau(q) + 1}{q}.$$

The value $h(q)$ at $q = 2$ corresponds to the value of Hurst exponent H . The self-similar process are monofractal, their scaling exponent $\tau(q)$ is linear.

The popular models of the multifractal processes are the stochastic conservative binomial multiplicative cascades [24]. Such multifractal models are constructed using an iterative algorithm, where the values of the cascade realization are the values of some specially selected random variable. The conservatism of the cascade consists in

the fact that for any number of iterations, the sum of the cascade values remains the same.

B. Mandelbrot proposed a multifractal model of financial time series which is based on fractional Brownian motion in multifractal time by operation of subordination [5,6]. The subordination is a random substitution of time and it can be represented in the form $Z(t) = Y(T(t))$, where $T(t)$ is a nonnegative nondecreasing random process called subordinator, $Y(t)$ is a random process, independent of $T(t)$.

In [6] it is proved that if $X(t)$ is the process of subordination

$$X(t) = B_H(\theta(t)) \quad (1)$$

where $B_H(t)$ is fractional Brownian motion with Hurst exponent H and $\theta(t)$ is conservative binomial multiplicative cascades, then $X(t)$ is the multifractal process. The scaling function $X(t)$ is defined by

$$\tau_X(q) = \tau_\theta(Hq), \quad (2)$$

where $\tau_\theta(Hq)$ is the scaling function of the multiplicative cascade $\theta(t)$.

Quite often, for practical purposes, it is not interested in the time series itself, but in its increments. The series of increments $Xdif$ for time series $X(t)$ is determined by the formula

$$Xdif(t) = X(t) - X(t-1). \quad (3)$$

3 Features for classification by machine learning

One of the most important issues of the time series classification task is the selection of features by which the partition into classes is carried out. Changes the time series fractal properties entails the changes the statistical and correlation properties. Therefore, statistical, fractal and recurrent characteristics calculated from the time series were chosen as features.

The studies presented in [26, 27] showed that the statistical characteristics that reflect the change in the fractal properties of the time series are variance, coefficient of variation, median, asymmetry coefficient, etc. As features representing fractal properties, it is convenient to use the values of the Hurst exponent H and the generalized Hurst exponent $h(q)$ such as the mean and standard deviation of the generalized Hurst exponent, the specific values $h(q)$ and the range Δh .

A fairly new approach to the selection and use of time series features in machine learning is the calculation of recurrence characteristics. The recurrence plot of a time series $X(t)$ is a matrix, where an element with coordinates (i, j) characterizes the proximity of points $X(t_i)$ and $X(t_j)$ in phase space [28]. A numerical analysis of

recurrence plots allows calculating the quantitative characteristics of recurrence such as a measure of recurrence, a measure of determinism, a measure of entropy, etc. These characteristics are advisable to use as features in machine learning classification [23,29,30].

4 Classification Methods

The ensemble methods of decision trees bagging and random forest, as well as neural networks, were chosen as classifiers.

The ensemble of models is a complex model consisting of separate basic models. Component models can be of the same type, or different. One of the first and most famous ensembles is bagging, based on the statistical bootstrap aggregating: multiple sample generation based on a single sample [31]. In this classification method, all elementary classifiers are trained and work independently of each other. Several samples of the same size are extracted from a single training sample, each of which is used to train one of the ensemble models. The decision is made either by voting: the class that was chosen by a simple majority of models is selected by averaging, which is defined as the average of all outputs.

The decision tree method is one of the simple and effective solutions to classification tasks in many different areas. Decision trees change quite a lot with a small change in the data sample, however, when several trees are combined into an ensemble, the spread in the values of the target variable becomes much smaller. In this paper, as one of the methods for classifying time series, was used the bagging with classification and regression decision trees. When using regression trees, the result of the classification model is the probability of matching the time series with the class.

Random Forest is a bagging method with regression or classification decision trees. However, unlike its main version, it has several features, in particular, in addition to randomly selecting learning objects, features are also randomly selected [32]. In this work, for comparison, along with the bagging, a random forest was also used with classification and regression decision trees.

Neural networks are widely used as classifiers. There are many neural network architectures designed to classify various objects. During the experiment, different neural networks were investigated and two neural network architectures were selected.

The first neural network was a fully connected perceptron of seven large layers with activation function of the ReLU type. Using the ReLU activation function is less expensive and significantly accelerates the convergence of gradient descent. After each full connected layer, the level of regularization was included in the network. As a regularization method, batch normalization was chosen [33], which is used to stabilize the neural network and prevent the effect of overtraining. The second network contained recurrent levels to take into account the relationship between elements. For both neural networks, the Adam stochastic optimization method (Adaptive Moment Estimate) [34] was used.

5 Experiment Description

To simulate multifractional Brownian motion $X(t)$ realizations the generation of multifractal cascades $\theta(t)$ with $Beta(\alpha, \alpha)$ -distribution [27] were used. They were subordinators for the fractional Brownian motion $B_H(t)$ by (1). In the case when the cascade weight coefficients are the values of $Beta(\alpha, \alpha)$ -distribution, the scaling exponent $\tau(q)$ is uniquely determined by the value of the Hurst exponent H , $0.5 < H < 1$ [35].

When specific $Beta(\alpha, \alpha)$ - distribution for the multifractal cascades was set and the specific Hurst exponent of fractional Brownian motion was selected, subordinated processes $X(t)$ with needed Hurst exponent H which is determined by (2) was obtained.

In this paper, each class was a set of model time series of multifractional Brownian motion $X(t)$ with the Hurst exponent H belonging to the same range of values. For each time series, the Hurst exponent was chosen randomly within the appropriate range. The values of the Hurst exponent are changing in the range from 0.5 to 1 with a step 0.05. The minimum and maximum values of the Hurst exponent were selected 0.51 and 0.99, respectively. Thus, the training of models was carried out in 10 classes, where $H \in \{[0.51, 0.55), [0.55, 0.6), [0.6, 0.65), \dots, [0.9, 0.95), [0.95, 0.99]\}$.

Figure 1 shows the realizations of cascade processes $\theta(t)$ (top) and corresponding realizations of multifractional Brownian motion $X(t)$ of different classes (bottom).

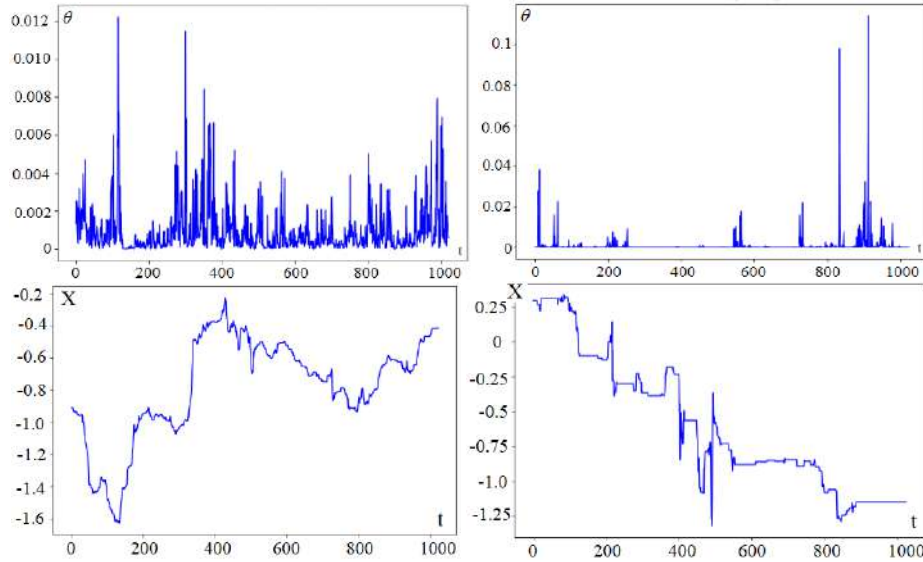


Fig. 1. Realizations of cascade processes (top) and corresponding realizations of multifractional Brownian motion (bottom)

For each multifractional Brownian motion realization, the realization of its increments was obtained by (3) and classification of increment time series was also carried out by a separate experiment. Figure 2 shows corresponding realizations of increments of multifractional Brownian motion, which shown in Fig. 1.

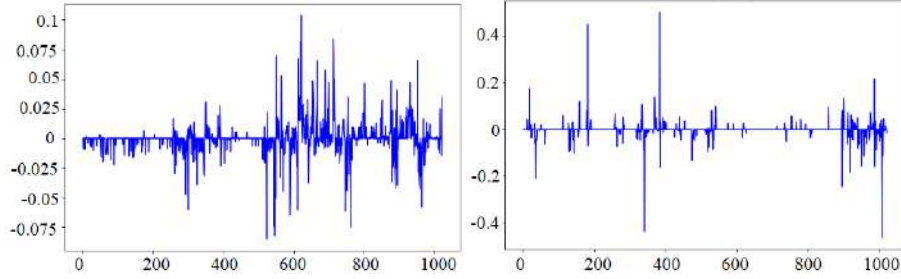


Fig. 2. Realizations of increments of multifractional Brownian motion

Thus, each class of time series is a set of realizations of multifractional Brownian motion or their increments with the same multifractal properties. To classify the statistical, fractal and recurrent characteristics of time series were used as features. The obtained features were the inputs of each of 6 classifiers: bagging with classification and regression decision trees, random forest with classification and regression decision trees, fully connected perceptron and recurrent neural network.

The research was conducted for a time series of different lengths: 512, 1024, 2048 and 4096 values. Such length is associated with the method of generating realizations of the binomial stochastic cascade. Model training for each class was carried out on 300 examples of training time series and was tested on 150 test ones.

6 Results and discussion

To implement bagging and random forest methods and neural networks, Python with libraries of machine learning methods was used [36]. The results of the classification of multifractional Brownian motion realizations indicate different classification accuracy for different classifiers. The best practice was the bagging with regression trees and a recurrent neural network. The worst results were shown by the Random forest with classification trees and a fully connected perceptron. Fig. 3 shows the histograms of the probability distribution of matching to class number for each value of the Hurst exponent for classifying time series with a length of 1024 values by the recurrent neural network. Such distributions are typical for all variants of classification.

Table 1 presents the average probabilities of class determining depending on the length of time series and the method of classification. The dependence of the classification accuracy on the length of the time series is obvious since the longer the series, the more accurately its fractal and recurrent characteristics are considered. Starting from the length of 2048 values, the probability of a correct class definition for bagging, random forest with regression trees and the recurrent neural network becomes greater than 0.9.

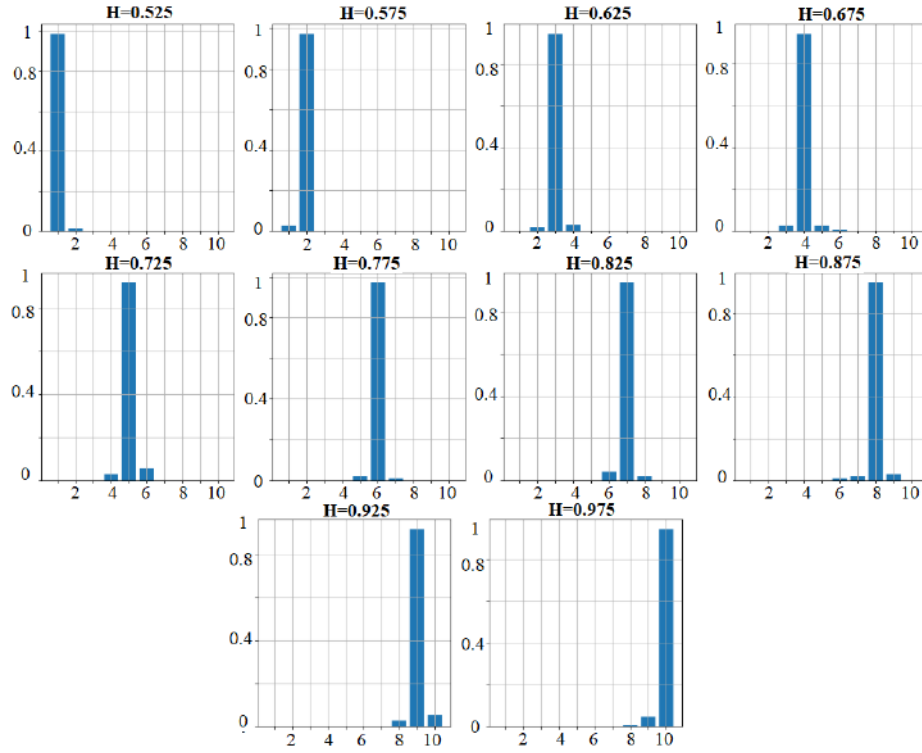


Fig. 3. Distribution of probabilities of the class number determining depending on the value of the Hurst exponent

The classification performed on the increments realizations of the multifractional Brownian motion showed better results. In this case, it was sufficient to use only the statistical and fractal characteristics of time series without building recurrence plots and calculating recurrence characteristics. This significantly reduces the training time and the structure of classifiers.

Table 1. The average probability of class determination for the cumulative realizations

Length of time series	Bagging		Random forest		Neural network	
	classification trees	regression trees	classification trees	regression trees	perceptron	recurrent
512	0.547	0.564	0.575	0.592	0.587	0.594
1024	0.781	0.796	0.631	0.685	0.627	0.785
2048	0.918	0.936	0.769	0.933	0.698	0.940
4096	0.956	0.961	0.957	0.962	0.764	0.969

Table 2 presents the average probabilities of class determination depending on the time series length and the method of classification for the increments realizations. The high probability of the correct class determination quickly is achieved even for relatively small lengths of time series.

Table 2. The average probability of class determination for the increments realizations

Length of time series	Bagging		Random forest		Neural network	
	classification trees	regression trees	classification trees	regression trees	perceptron	recurrent
512	0.899	0.900	0.903	0.899	0.791	0.744
1024	0.985	0.986	0.982	0.986	0.949	0.957
2048	0.995	0.995	0.996	0.995	0.970	0.973
4096	0.998	0.998	0.998	0.998	0.979	0.981

Conclusion

The multifractal time series were classified by machine learning methods. Time series were obtained by generating realizations of fractional Brownian motion in multifractal time. Both cumulative time series of multifractal Brownian motion and series of increments were carried out.

Time series were divided into classes according to their multifractal properties. The classification was carried out on the basis of quantitative features calculated for each time series. The features for classification were statistical, fractal and recurrent characteristics. Such classifiers were chosen for classification: bagging with classification and regression decision trees, random forest with classification and regression decision trees, fully connected perceptron and recurrent neural network.

The results of the research have shown that the accuracy of the classification is higher when the increments time series were classified. When classifying cumulative realizations of multifractional Brownian motion, method of bagging with regression trees and recurrent neural network showed the best accuracy.

In future research it worth to concentrate on the classification of real multifractal time series using different classification algorithms.

References

1. Brambila, F.: Fractal Analysis – Applications in Physics, Engineering and Technology. <https://www.intechopen.com/books/fractal-analysis-applications-in-physics-engineering-and-technology>

2. Shelukhin, O.I.; Smolskiy, S.M.; Osin, A.V.: Self-Similar Processes in Telecommunications. John Wiley & Sons, New York, USA, 320 p. (2007).
3. Peters, E. E.: Fractal Market Analysis: applying chaos theory to investment and economics. Wiley (2003).
4. Daradkeh, Y. I., Kirichenko, L., Radivilova, T.: Development of QoS methods in the information networks with fractal traffic. *International Journal of Electronics and Telecommunications* **64**(1), 227-232 (2018). doi: 10.24425/118142
5. Mandelbrot, B. B., Fisher, A., Calvet, L.: A multifractal model of asset returns Cowles Foundation Discussion. Sauder School of Business Working Paper, Paper **1164**, 1–33 (1997).
6. Calvet, L. E., Fisher, A. J., Mandelbrot, B. B.: Large deviations and the distribution of price changes. Cowles Foundation Discussion Paper, Yale University **1165**, 1–28 (1997).
7. Ayache, A., Cohen, S., V  hel, J. L.: The covariance structure of multifractional Brownian motion, with application to long range dependence. In: 2000 IEEE International Conference on Acoustics, Speech, and Signal Processing Proceedings (Cat. No. 00CH37100), vol. 6. IEEE (2000). doi: 10.1109/ICASSP.2000.860233
8. Pierre, R. B., Dury, M. E., Haouas, N.: Selection of sparse multifractional model. <https://hal.archives-ouvertes.fr/hal-01194347>.
9. Ahn, K. I., Lee, K.: Identification of nonstandard multifractional brownian motions under white noise by multiscale local variations of its sample paths. *Mathematical Problems in Engineering* **2013** (2013). doi: <https://doi.org/10.1155/2013/794130>
10. Corlay, S., Lebovits, J., V  hel, J. L.: Multifractional stochastic volatility models. *Mathematical Finance* **24**(2), 364-402 (2014). doi: <https://doi.org/10.1111/mafi.12024>
11. Bianchi, S.: Pathwise identification of the memory function of multifractional Brownian motion with application to finance. *International Journal of Theoretical and Applied Finance* **8**(02), 255-281 (2005). doi: <https://doi.org/10.1142/S0219024905002937>
12. Muniandy, S. V., Lim, S. C.: Modeling of locally self-similar processes using multifractional Brownian motion of Riemann-Liouville type. *Physical Review E* **63**(4), 046104 (2001). doi: 10.1103/PhysRevE.63.046104
13. Fauth, A., Tudor, C. A.: Multifractal random walks with fractional Brownian motion via Malliavin calculus. *IEEE Transactions on Information Theory* **60**(3), 1963-1975 (2014).
14. G  nay, S.: Performance of the Multifractal Model of Asset Returns (MMAR): evidence from emerging stock markets. *International Journal of Financial Studies* **4**(2), 11 (2016). doi:10.3390/ijfs4020011
15. Li, M., Zhao, W., Chen, S.: MBm-based scalings of traffic propagated in internet. *Mathematical Problems in Engineering* **2011** (2011). doi:10.1155/2011/389803
16. Esling, P.; Agon, C.: Time series data mining. *ACM Computing Surveys* **45**, 12:1–12:34 (2012).
17. Kirichenko, L., Radivilova, T., Zinkevich, I.: Comparative Analysis of Conversion Series Forecasting in E-commerce Tasks. In: Shakhovska N., Stepashko V. (eds) *Advances in Intelligent Systems and Computing II*. CSIT 2017. *Advances in Intelligent Systems and Computing*, vol 689. Springer, Cham (2018). doi: https://doi.org/10.1007/978-3-319-70581-1_16
18. Fawaz, H. I., Forestier, G., Weber, J., Idoumghar, L., Muller, P. A.: Deep learning for time series classification: a review. *Data Mining and Knowledge Discovery* **33**(4), 917-963 (2019).
19. Kirichenko, L., Radivilova, T., Tkachenko, A.: Comparative Analysis of Noisy Time Series Clustering. In: 3rd International Conference on Computational Linguistics and Intelli-

- gent Systems (COLINS-2019) Proceedings 2362, 2019 April 18-19, pp.184-196. Kharkiv, Ukraine (2019)
20. Buza, K.: Time series classification and its applications. In: 8th International Conference on Web Intelligence, Mining and Semantics Proceedings, pp.1-4 (2018). doi: <https://doi.org/10.1145/3227609.3227690>
 21. Kaur, G., Saxena, V., Gupta, J.: Detection of TCP targeted high bandwidth attacks using self-similarity. *Journal King Saud University Computer and Information Sciences* (2017).
 22. Deka, R., Bhattacharyya, D.: Self-similarity based DDoS attack detection using Hurst parameter. *Security Communication Networks* **9**, 4468–4481 (2016).
 23. Radivilova, T., Kirichenko, L., Ageiev, D., Bulakh, V.: Classification Methods of Machine Learning to Detect DDoS Attacks. In: 2019 10th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS) Proceeding Vol.1, pp. 207-210. IEEE, Metz (2019). doi: 10.1109/IDAACS.2019.8924406
 24. Riedi, R. H.: Multifractal Processes. https://www.researchgate.net/publication/2839202_Multifractal_Processes
 25. Kantelhardt, J.W.: Fractal and Multifractal Time Series. <http://arxiv.org/abs/0804.0747>
 26. Bulakh, V., Kirichenko, L., Radivilova, T.: Time Series Classification Based on Fractal Properties. In: 2018 IEEE Second International Conference on Data Stream Mining & Processing (DSMP) Proceedings, 21–25 August 2018; pp. 198–20. IEEE, Lviv (2018). doi: 10.1109/DSMP.2018.8478532
 27. Kirichenko, L., Radivilova, T., Bulakh, V.: Machine Learning in Classification Time Series with Fractal Properties. *Data* **4(1)** **5**, 1-13 (2019). doi:10.3390/data4010005
 28. Marwan, N., Romano, M., Thiel, M., Kurths, J.: Recurrence plots for the analysis of complex system. *Physics Reports* **438(5–6)**, 237-329 (2007). doi: <https://doi.org/10.1016/j.physrep.2006.11.001>
 29. Kirichenko, L., Radivilova, T., Bulakh, V.: Binary Classification of Fractal Time Series by Machine Learning Methods. In: Lytvynenko V., Babichev S., Wójcik W., Vynokurova O., Vyshemyrskaya S., Radetskaya S. (eds) *Lecture Notes in Computational Intelligence and Decision Making. ISDMCI 2019. Advances in Intelligent Systems and Computing* 1020, 701-711 (2020). doi: https://doi.org/10.1007/978-3-030-26474-1_49
 30. Kirichenko, L., Radivilova, T., Bulakh, V.: Classification of Fractal Time Series Using Recurrence Plots. In: 2018 International Scientific-Practical Conference Problems of Informatics. Science and Technology (PIC S&T), pp. 719-724. IEEE, Kharkiv (2018). doi: 10.1109/INFOCOMMST.2018.8632010
 31. Breiman, L.: Bagging predictors. *Machine Learning* **24 (2)**, 123–140 (1996).
 32. Breiman, L.: Random Forests. *Machine Learning* **45 (1)**, 5–32 (2001).
 33. Ioffe, S. Szegedy, C.: Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. In: 32nd International Conference on Machine Learning Proceeding, PMLR 37, pp.448-456. Lille, France (2015). <https://arxiv.org/abs/1502.03167>
 34. Kingma, D. P., Ba, J.: Adam: A Method for Stochastic Optimization. In: 3rd International Conference on Learning Representations (ICLR) Proceeding, San Diego, USA (2015). <https://arxiv.org/abs/1412.6980>
 35. Bulakh, V., Kirichenko, L., Radivilova, T.: Classification of Multifractal Time Series by Decision Tree Methods. In: 14th International Conference ICTERI 2018 ICT in Education, Research, and Industrial Applications Proceeding, 14–17 May 2018, pp. 1–4. Kyiv, Ukraine (2018).
 36. Cielien, D., Meysman, A., Ali, M.: Introducing data science: big data, machine learning, and more, using Python tools. Manning Publications Co Manning Publications (2016).

An Informetric View on Relations between Global Brands and Research Activity

Serhiy Shtovba^{1, 2} [0000-0003-1302-4899], Olena Shtovba¹ [0000-0003-1418-4907]

¹ Vinnytsia National Technical University, Khmelnytske Shose, 95, Vinnytsia, 21021, Ukraine
shtovba@vntu.edu.ua

olena.shtovba@yahoo.com

² Vasyl' Stus Donetsk National University, 600-richchia str., 21, Vinnytsia, 21021, Ukraine
s.shtovba@donnu.edu.ua

Abstract. The article identifies the relations between global brands and research activity during 2009 – 2018. Two types of relations are studied: support of the research by global brands and usage of the global brands in the research. The research support by 27 global brands and the global brands usage in the research are analysed over 2009 – 2018. The support is assessed by the number of papers in which the global brand is mentioned in the funding section. The usage is assessed by a number of papers in which the global brand is mentioned in the presentation part. The most generous and simultaneously popular brands are Google, Microsoft, IBM, Samsung and Facebook. In the last three years their support dynamics and usage dynamics are positive or strong positive.

Keywords: informetrics, global brands, research support, brands in research, Scopus, classification.

1 Introduction

Global brands rapidly penetrate into new spheres of influence, including research. The level of the infiltration can be assessed by the public “reports” of the scholars – by the scientific papers. Until now, such activity of the global brands has not been studied in a systematic way, only some particular studies were carried. For example, a bibliometric analysis of the scientific literature related to the use of Facebook in educational research is carried in [1]. The analysis is performed on Web of Science data. Scopus-based bibliometric analysis of the papers related to the use of Facebook and of YouTube in any research field is carried in [2] and in [3]. An evaluation of correspondence between Coca-Cola’s “Transparency List” of funded researchers and a list of papers with scientific research acknowledging funding from Coca-Cola is performed in [4]. A list of the supported papers is taken from Web of Science Core Collection database. Paper [5] describes the scope of partnerships between Coca-Cola and 74 health organizations in Spain, examining marketing strategies contained in scientific papers funded by Coca-Cola between 2010 and 2016. A list of the supported health organizations was formed through PubMed. Papers [6], [7] and [8] study vari-

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

ous conflicts of interest between donors (Coca-Cola Company, Mars Inc.) and results of the supported research.

The impact of the global brands on the research activity is proposed to assess by two indices. The first index is the number of papers, where the global brand is mentioned in the presentation part. This index roughly assesses how the global brand is connected with the research activity. This connection can be in the form of some brand tools, technologies, recipes, datasets etc. It is possible that the global brand itself acts as an object of research. The second index is the number of papers, where the global brand is mentioned in the funding research section. This index enables to estimate roughly the financial support of the research by the global brands. Such informetric approach to identifying the various brand relations is becoming more and more popular. For example, paper [9] analyses B2B-branding papers during 1972 – 2015, paper [10] analyses analysed publications related to researches in brand personality during the 1995 – 2017, paper [11] identifies the leaders and the trends in branding researches during 2000 – 2019.

Our paper studies not only current state of “brands – research” relations, not only statistics on certain date. Informative figure is also the dynamics – change of the indices during some time interval. For data acquisition the Scopus base with the corresponding search services is applied.

2 Studied global brands

The studied global brands are selected from Top-100 list of the most expensive global brands of 2018 according to “Best global brands rankings” by Interbrand. All the homonymous brands that do not have the unique names are rejected. For instance, the word “apple” in the scientific papers is used not only as the name of the most expensive brand but also as an ordinary apple. Caterpillar is not only the producer of powerful road construction machinery but also ordinary caterpillar. Nike is not only sportswear but also one of the species of fish and the name of laser. Amazon and Honda have the geographical homonyms. Some short names of the brands coincide with the specific abbreviations, widely spread in certain research fields. For instance, NIKE is used as the abbreviation for “Non-Interactive Key Exchange”, BMW – as “biomedical waste”. In case of homonymy, automatic processing the query outputs from Scopus is complicated. For our study we have chosen the following 27 brands with unique names: Google, Microsoft, Coca-Cola, Samsung, Toyota, Facebook, IBM, Cisco, Louis Vuitton, Budweiser, Accenture, Hyundai, Ebay, Volkswagen, Goldman Sachs, L'Oréal, Adidas, Hewlett Packard, Morgan Stanley, Harley-Davidson, Netflix, Huawei, PayPal, John Deere, Spotify, Johnnie Walker, and Nintendo.

3 Brands classification based on statistics over 10 years

We have collected statistics about support of research by global brands as well as using the global brands in research over 10 past years. For 2D-distribution of the brands in the axes “support – usage” the data for the period of 10 years must be ag-

gregated in some way. The simplest way is summation of values for the 10 years. But naturally, the fresh data have the greatest impact on the current perception of the brand. The old papers produce the least impact. That is why, during the averaging we will normalize the indices, taking into account the various importance of the data. For the normalization of the support and usage, the following formulas are proposed:

$$U_N = \frac{1}{10} \sum_{i=2009, 2018} w_i \cdot U_i, \quad (1)$$

$$S_N = \frac{1}{10} \sum_{i=2009, 2018} w_i \cdot S_i, \quad (2)$$

where S_i denotes the number of the papers of the i^{th} year in which the corresponding global brand is mentioned in the section Funding; U_i denotes the number of the papers of the i^{th} year, in which the corresponding global brand is mentioned in the presentation part; w_i denotes weight coefficient of the i^{th} year.

We use arithmetic sequence of weight coefficients with decay 0.1, namely: $w_{2018} = 1$; $w_{2017} = 0.9$; $w_{2016} = 0.8$, ..., $w_{2009} = 0.1$. Hence, the impact of the events of the year 2009 is 10 times less than the events of the year 2018. 2D-distribution of the brands according to the normalized indices is shown in Figure 1.

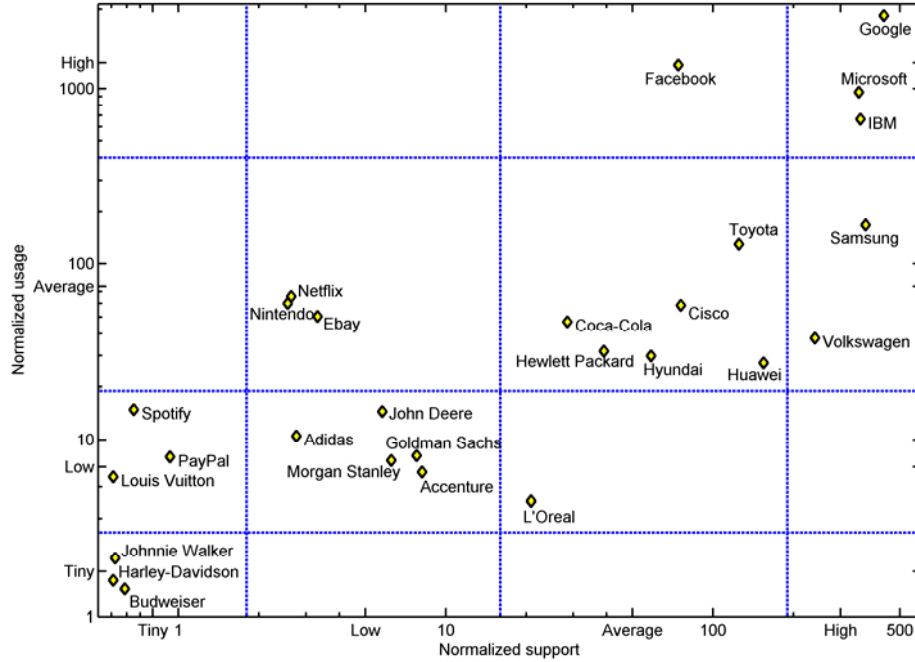


Fig. 1. Classification of the global brands in case of arithmetic sequence of weight coefficients with decay 0.1 (log-log scale)

By the ordinal scale {Tiny, Low, Average, High} 9 clusters are allocated on Figure 1. For example, the leader cluster High-High, the cluster with High Normalized Support and High Normalized Usage comprises the brands Google, Microsoft and IBM.

In Figure 1 all the brands are located within the vicinity of the diagonal – each brand has similar or neighboring linguistic values of the normalized support and usage. There is no brand with High Support level and Low Usage level or vice versa, Low Support level and High Usage level. Greater part of the brands – 17 out of 27 are in the clusters with the equivalent levels of the Normalized Support and Normalized Usage. Twelve brands locate in the clusters High-High, High-Average, Average-High or Average-Average. Eleven out of twelve of these brands are referred to technologies, business services, automotive, and only one brand – Coca-Cola belongs to the food industry. The most generous and popular brands are located in the right upper corner of Figure 1. These brands are Google, Microsoft, IBM, Samsung and Facebook.

To checking the reliability of brands classification let us plot 2D-distributions in case of other weight coefficients in formulas (1) and (2). Figure 2 shows 2D-distribution for a geometric sequence of weight coefficients with ratio 0.9, namely: $w_{2018} = 1$; $w_{2017} = 0.9$; $w_{2016} = 0.9^2 = 0.81$, ..., $w_{2016} = 0.9^9 = 0.39$.

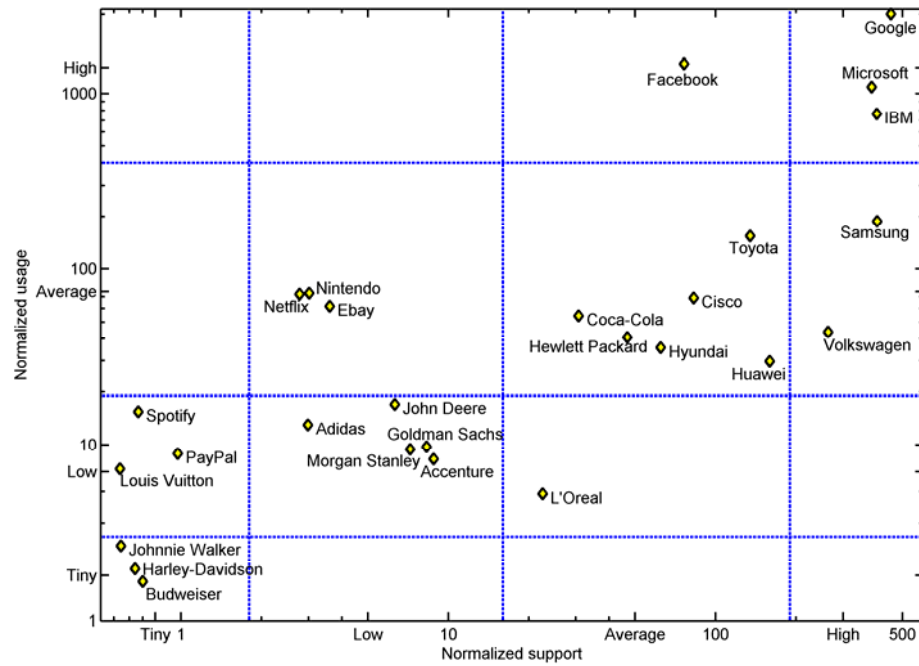


Fig. 2. Classification of the global brands in case of geometric sequence of weight coefficients with ratio 0.9 (log-log scale)

Figure 3 shows 2D-distribution for a geometric sequence of weight coefficients with ratio 0.8, namely: $w_{2018} = 1$; $w_{2017} = 0.8$; $w_{2016} = 0.8^2 = 0.64$, ..., $w_{2016} = 0.8^9 = 0.13$. Figures 1–3 show slight fluctuations of the brands without any change of the classifying results. Hence, the proposed classification is robust – it is not sensitive to reasonable change of weight coefficients in (1) and (2).

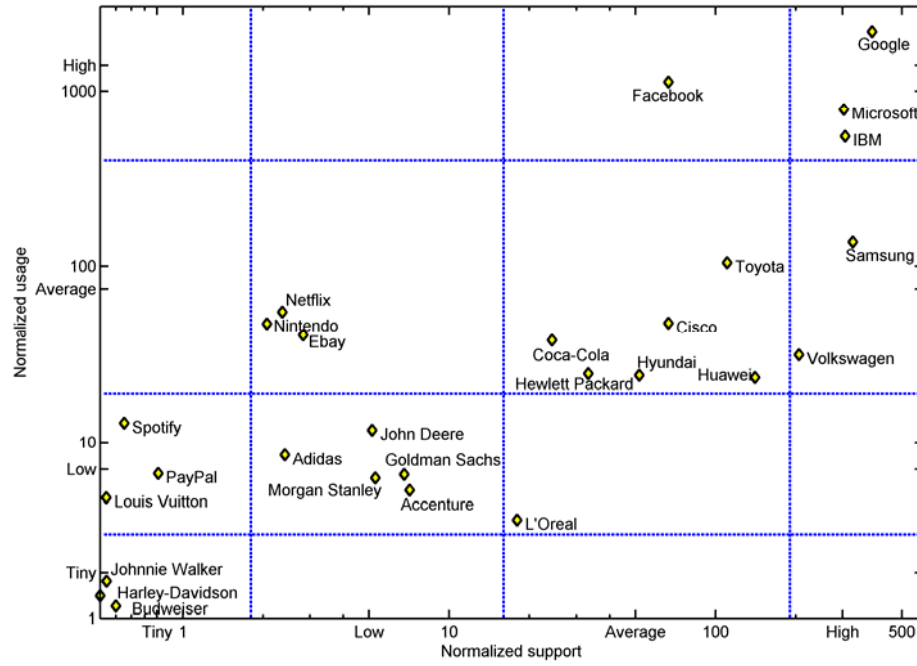


Fig. 3. Classification of the global brands in case of geometric sequence of weight coefficients with ratio 0.8 (log-log scale)

4 Support Dynamics

According to Figures 1–3 all the brands are divided into 4 groups by normalized support index. Let's analyse support dynamics of the brands from the groups with high, average and low support level. The dynamics of tiny support group is out of interest due to statistically insignificant number of cases.

Five brands Google, Microsoft, Samsung, IBM and Volkswagen are in the group with high support level. Time series of research support by those global brands are shown in Figure 6. There is no dominant leader among these brands. Just one brand – Google increased the number of supported papers every year. Google's support dynamics is very close to a cube function. Figure 4 shows that all the brands considerably increased their support during last 3 years. In 2018 Volkswagen supported 2.7

times more papers than in 2015, Samsung – 4.4 times, other brands – approximately three times. Average dynamics shows that 2015 looks like as a breaking point. An annual pace of support is equal to 22 during 2009 –2015 and has increased drastically up to 297 during 2015–2018.

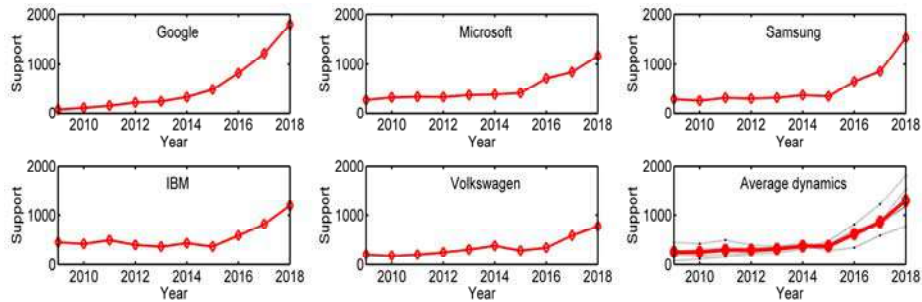


Fig. 4. Support dynamics for High support group

Eight brands are in the group with the average level of support. Annually they support 2–4 times less papers than the brands from the High support group. All the brands from Average group increased also their support in 2016–2018 but did it very unevenly (Figure 5). The slowest three-year pace demonstrates Coca-Cola – 2.3 times. The most rapid are Facebook – 4.7 times, and, especially, Huawei – 7.8 times. Huawei supported 727 papers in 2018, and only 93 papers in 2015. Huawei with such pace has a good change to move into the High support group during next year. Toyota is another brand in the candidate-list for the High support group.

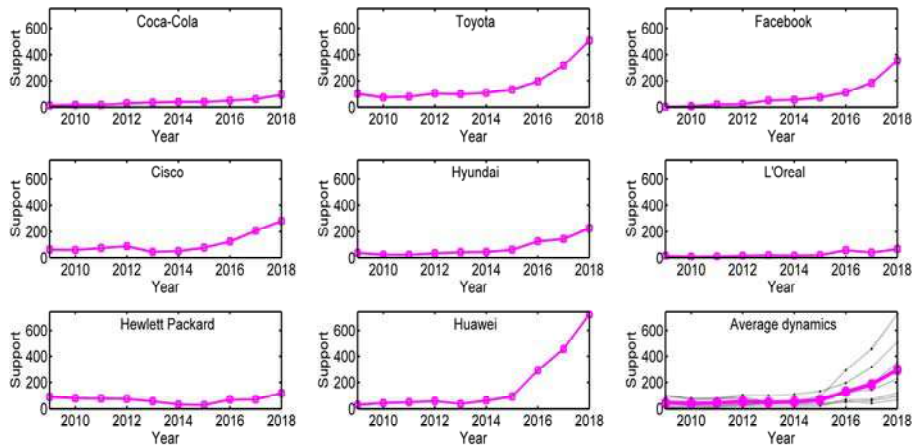


Fig. 5. Support dynamics for Average support group

On average, the brands of Average support group increased considerably the support during the last three years. We draw the attention to Average support group that

includes 2 brands from food industry and FMCG-industry. However, the level of their support is far less than IT, telecommunication and automotive brands.

Eight brands are in the Low support group (Figure 6). The greater part of the brands has zero-dynamic, i.e. the number of supported papers remains approximately the same. Exception is 2018, when all the brands besides Nintendo increased the number of supported papers significantly.

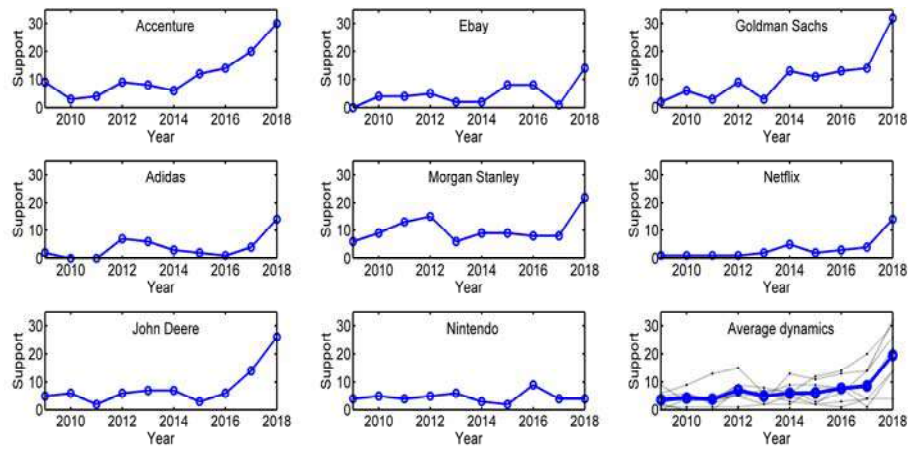


Fig. 6. Support dynamics for Low support group

5 Usage Dynamics

Among the analysed brands the researchers most often use Google. Since 2016 the number of papers, in the presentation part of which Google is mentioned, exceeded 5000 annually. Facebook, Microsoft and IBM are also among the leaders (Figure 7). All four brands stable growth of their usage in research during the last several years. However, Google has highest pace with almost quadratic dynamics during all the years.

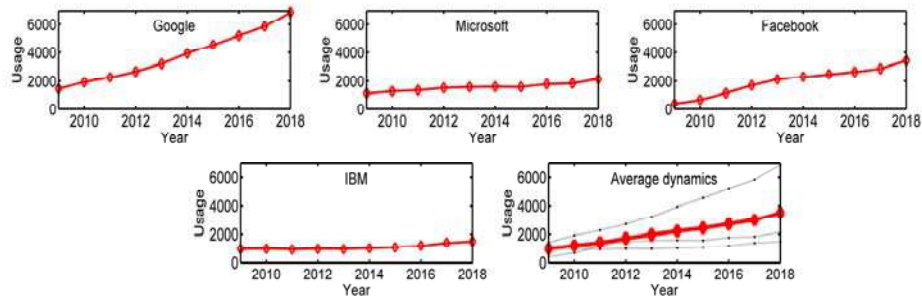


Fig. 7. Usage dynamics for High usage group

Average usage group comprises nine brands (Figure 8). Samsung and Huawei have good positive dynamics during almost all the years. The usage of remaining brands of this group is either stable or slightly decreases. As a result, the average usage is stable for this group.

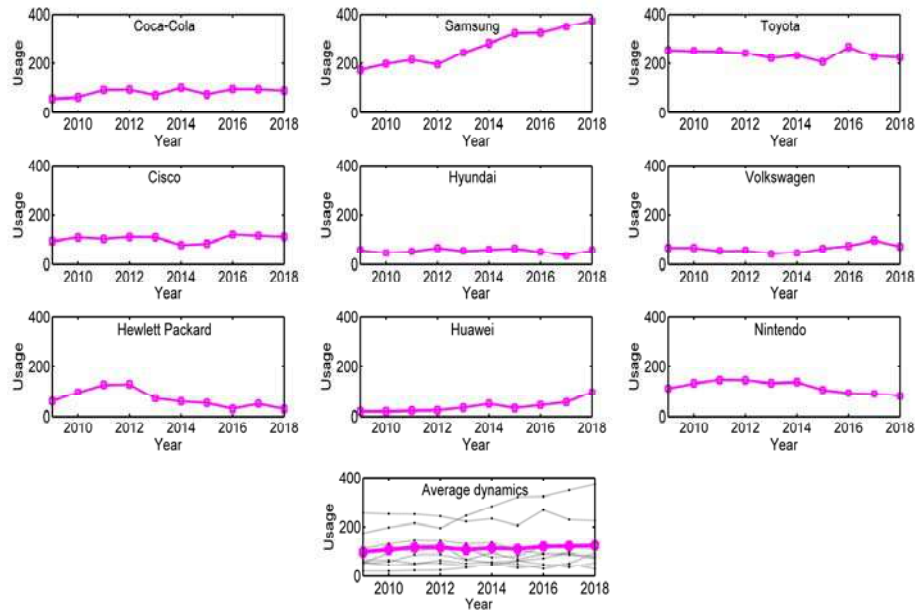


Fig. 8. Usage dynamics for Average usage group

Nine brands has formed Low usage group (Figure 9). Ebay demonstrates strong negative dynamics. Probably, Ebay-phenomenon as a research object loses the attractiveness. Spotify has reliably positive dynamics. Usage of the rest of the brands of this group during 10 years is more or less stable.

Resuming the data from all the groups, we conclude that strong positive dynamics of the usage is demonstrated only brands related with information services, information resources, media, and electronics.

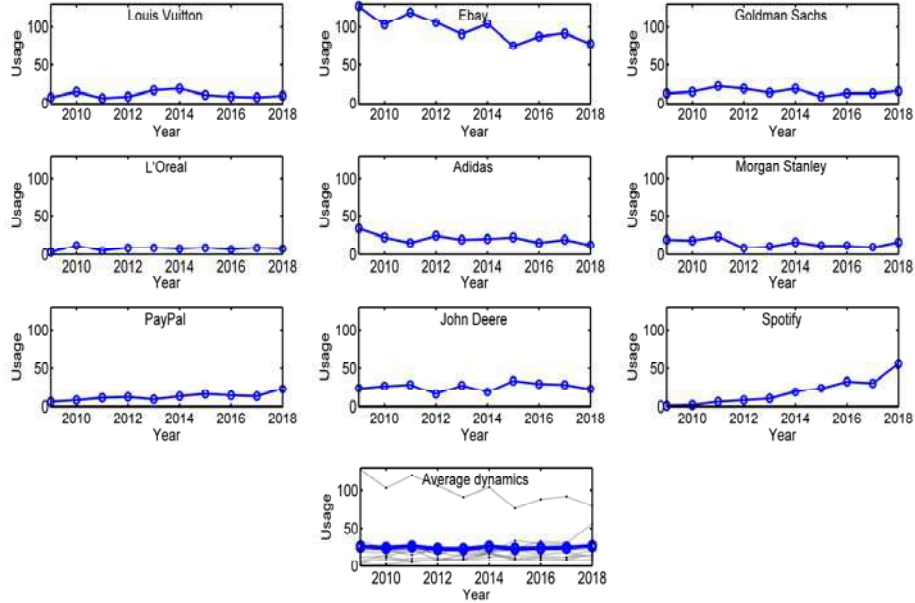


Fig. 9. Usage dynamics for Low usage group

6 Conclusion

The research support by 27 global brands as well as their global brands usage in the research during the last 10 years has been studied. The analysis is based on the informetric approach with Scopus data. The support was assessed by the number of papers where the global brand is mentioned in the funding section. The usage was assessed by the number of papers, where the global brand is mentioned in the presentation part.

The statistics about support of research by global brands as well as using the global brands in research over 10 past years were aggregated. The aggregation in form of weighted sum with the highest impact for data of 2018 and the lowest impact for data of 2009 is proposed. Nine clusters of brands are induced by 2D-distribution according to the normalized indices of support and of usage in log-log scale. The proposed classification is robust – it is insensitive to reasonable change of weight coefficients.

According to the normalized level of support, all the brands are divided into four groups with high, average, low and tiny levels. All the brands with high and average levels are characterized by the considerable growth of the number of the supported papers during 3 last years. This can be explained by the increase of funding and (or) growth of the requirements to the grant recipients, concerning the obligation to mention the source of funding in the corresponding section of the paper. Regarding to group with low level, the greater part of the brands has zero-dynamics. Exception is 2018, when all the brands besides Nintendo increased the number of supported papers significantly.

According to the normalized level of usage all the brands are divided into 4 groups with high, average low and tiny levels. For the group with high level stable considerable growth of the brand usage in the papers during all 10 year is characteristic. High usage group consists of Google, Microsoft, Facebook and IBM. Also Samsung and Huawei from Average usage group and Spotify from Low usage group have good positive dynamics during almost all the years. Hence, the strong positive dynamics of the usage is demonstrated only information-centered brands.

It is revealed that the most generous and simultaneously popular brands are Google, Microsoft, IBM, Samsung and Facebook. All these five top brands are connected with the information systems and services. In the last three years their support dynamics and usage dynamics are positive or strong positive. This enables to put forward the hypothesis that the strong connection of the brand with research stimulates further strengthening of these relations. That is, the effect that “strong becomes stronger” is observed also in the relations between global brands and research activity. For the verification of this hypothesis the additional statistics for several future years is required.

References

1. Lopes, R. M., Fidalgo-Neto, A.A., Mota, F.B.: Facebook in educational research: A bibliometric analysis. *Scientometrics*. **111**, 1591–1621 (2017). doi: 10.1007/s11192-017-2294-1.
2. Baran, K.S., Ghaffari, H.: The manifold research fields of Facebook: A bibliometric analysis. *Journal of Information Science Theory and Practice*. **5**, 33–47 (2017). doi: 10.1633/JISTAP.2017.5.2.3.
3. Noruzi, A.: YouTube in scientific research: A bibliometric analysis. *Webology*. **14**, 1–7 (2017).
4. Serôdio, P. M., McKee, M., Stuckler, D.: Coca-Cola – a model of transparency in research partnerships? A network analysis of Coca-Cola’s research funding (2008–2016). *Public health nutrition*. **21**, 1594–1607 (2018). doi: 10.1017/S136898001700307X.
5. Rey-López, J. P., Gonzalez, C.A.: Research partnerships between Coca-Cola and health organizations in Spain. *European journal of public health*. **29**, 810–815 (2018). doi: 10.1093/eurpub/cky175.
6. Steele, S., Ruskin, G., McKee, M., Stuckler, D.: Always read the small print: A case study of commercial research funding, disclosure and agreements with Coca-Cola. *Journal of public health policy*. **40**, 273–285 (2019). doi: 10.1057/s41271-019-00170-9.
7. Greenhalgh, S.: Science and serendipity finding Coca-Cola in China. *Perspectives in Biology and Medicine*. **62**, 131–152 (2019). doi: 10.1353/pbm.2019.0007.
8. Nestle, M.: Corporate funding of food and nutrition research: science or marketing?. *JAMA Intern Med*. **176**, 13–14 (2015). doi: 10.1001/jamainternmed.2015.6667.

9. Seyedghorban, Z., Matanda, M.J., LaPlaca, P.: Advancing theory and knowledge in the business-to-business branding literature. *Journal of Business Research*. **69**, 2664-2677 (2016). doi: 10.1016/j.jbusres.2015.11.002
10. Llanos-Herrera, G.R., Merigo, J.M.: Overview of brand personality research with bibliometric indicators. *Kybernetes*. **48**, 546-569 (2019). doi: 10.1108/K-02-2018-0051.
11. Shtovba, S., Shtovba, O., Filatova, L.: The current state of brand management research: An overview of leaders and trends in branding research over the past 20 years. *The Bottom Line*. **33**, 1-11 (2020). doi: 10.1108/BL-08-2019-0106

Decision Tree Based Targeting Model of Customer Interaction with Business Page

Hrystyna Lipyanina ^[0000-0002-2441-6292], Anatoliy Sachenko ^[0000-0002-0907-3682], Taras Lendyuk ^[0000-0001-9484-8333], Serhiy Nadvynychny ^[0000-0002-5567-1114], Sergii Grodskyi ^[0000-0002-9366-2243]

Ternopil National Economic University, Lvivska Str., 11, Ternopil, 46000, Ukraine
xrustya.com@gmail.com, as@tneu.edu.ua, tl@tneu.edu.ua, nadvynychny@tneu.edu.ua, sergij80@yahoo.com, www.tneu.edu.ua

Abstract: Company branding through social networks is most effective if it reaches the right customers. This study explores how to improve business page targeting on Facebook by customer behavior targeting, age and gender to form a more comprehensive contextual advertising strategy. Paper focuses on empirical modeling of targeting based on decision trees. The general practice of such models developing does not take into account business goals sufficiently. To correct this, we propose to create an algorithm for customer interaction simulating with a business page on Facebook based on decision trees within the control and optimization of a company's marketing strategy. The resulting algorithm combines statistical training principles and business goals in the form of campaign income maximizing. The basic approach to the marketing strategy formation is considered, the parameters of the algorithm and the algorithm of forming the client interaction targeting with the business page on the basis of decision tree are established. Based on the above algorithm, we build a model of customer interaction targeting with a business page based on the decision tree using the data of contextual advertising campaign on Facebook. Based on the simulation results, a re-formation of the advertising campaign and analysis with the input data were performed. The results of the study confirm the value of the proposed method, since the targeting model of customer interaction with a business page based on decision trees recommends significantly more profitable target groups than a few benchmarks.

Keywords: model, targeting, algorithm, method of decision tree, R language.

1 Introduction

One of the main tasks of modern business is to attract new customers and maintain relationships with existing customers. The importance of this problem is determined by the fact that social networks have become an integral part of modern life of most people whom the business views as potential customers. Social networks have become the landscape for business and attracting new customers. This space has its own features that set it apart from the traditional market.

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

All social networks are distinguished type. There are networks for people finding: classmates, colleagues and other people. There are business networks for searching of job, partners, professional communication and other business matters. Some networks are video-based, some are oriented on audio and specifically music and some are photo-based. There are also specialized networks that target a specific, well-defined group of people, such as the Habrahabr network that integrates everything related to IT. Advantages of social networks are the ability to use them to establish relationships with customers and monitor them [29].

All digital marketers must have the skills to build distributed computing systems, data warehouses, and electronic document management systems based on the client's network, as well as be able to register domain names, hosting and colocation traditionally targeted at large non-governmental enterprises and companies, government agencies, small businesses.

The global cost of online advertising has caught up newspaper advertising in 2013. The figures are huge: \$ 82.09 billion spent on digital advertising in the US only in 2017, and it is expected to cost more than \$ 113 billion in 2020 [1]. In March 2019, the most popular social network in Ukraine was Facebook, which was used by 50% of respondents. This is confirmed by the survey data from Research & Branding Group [2].

Facebook allows companies placing audience-targeted ads in the news feed, the right column of their account, show them on the mobile app, promote videos, etc. Properly setting and selecting the target group will save your campaign budget and time for setting your campaign up, so this is an issue for nowadays business community.

This work proposes Model of Customer Interaction Targeting with Business Page based on Decision Tree. The rest of the paper has the following structure: Section 2 examines existing methods and approaches to online advertising, including the use of targeting techniques to improve it. Section 3 describes the proposed algorithm for customer targeting interaction with a business page based on a decision tree. Section 4 discusses the findings of the study and Section 5 summarizes the findings.

2 Related Work

The study [3] investigates how to enhance geographical targeting by a suite of other targeting strategies, including behavioral targeting, temporal targeting, and use of discount in an online-to-offline commerce context, to form a more comprehensive contextual targeting strategy. However, there is no specification of the obtained results in this study, in particular the targeting group is not clearly defined.

Authors [4] stay that the level of focus on consumer advertising significantly affects the idea of ordering informativity, whereas specialized online remarketing advertising have a direct adverse effect on the customer, in addition, it increases the irritation caused by targeting behavior.

In general, work [5] can be considered as a first step in studying of large-scale, fine-grained digital traces of online human behavior and how they can be used to predict people's future marketing behavior.

Attention in articles [6, 7] focus on empirical targeting models. The papers argue that the general practice of such models developing does not sufficiently take into account business goals. The results of a comprehensive empirical study confirm that it recommends significantly more profitable target groups.

The authors of [8] use content analysis to study topics and formats of 5932 Facebook posts from leading US colleges and universities. The results show that there are content topics, such as athletics, that significantly increase engagement, and others tend to be less active. In addition, format, as well as user-generated content, is another contributing factor to engagement.

The study [9] outlines the risks of using Facebook for both users and marketers. The suggested scenarios will help marketers understand how Facebook marketing uses knowledge management tools like plan-do-check-act (PDCA) and root cause analysis (RCA).

A study [10] examines the effectiveness of advertising on Instagram and Facebook in terms of advertising attitude, advertising persistence and loyalty intentions. The results show that Instagram Stories not only improves consumer attitudes towards advertising, but also increases its responsiveness compared to Facebook Wall.

The results of a study [17] revealed that attitudes towards advertising on the social network, that is, any efforts to transmit product messages between network members, who are also consumers of different products, are formed and persuaded by hedonic motivation (HM), a source of derogation (SD), self-concept (SC), messages informality (MI), and experience messages (EM).

A study [18] examines Facebook advertising, namely social influence theory and regulatory focus theory. Structural equation modeling results show that in both developed countries (Australia) and developing countries (South Africa), there are significant relationships between the considered parameters in the model (privacy, trust, advertising, advertising, attitude to advertising, promotional values, Facebook advertising attitudes, and promotional behavior).

A study [19] proved that different advertising content in Facebook leads to different levels of recruitment and involvement of users in the advertisement. Different advertisements also change the choice of advertisement in terms of demographic and mental health characteristics.

Research findings [20] have shown that Facebook advertising has had a significant impact on brand image and price, both of which contribute to increased brand sales. The authors of a study [21] examined whether customer involvement in social media brands has a bearing on brand trust, brand loyalty, and brand creation, and the results indicated that motivation to participate in SNSs significantly influences customer engagement, which in turn, has a positive effect on brand credibility, brand loyalty.

The authors of [22] confirmed that Facebook, among users of social networks, is perceived as an effective means of advertising, and it is strongly associated with the benefits of "customer relationship management" and "promotion of new products".

Article [23] investigates the impact of two advertisement placements on Facebook, a sidebar ad, and a message bar to avoid advertising. The results indicate the crucial role of product engagement in Facebook ads targeting to the right audience and choosing the right ad placement.

Article [24] defines the influence of online advertising on students' decision-making and the choice of their universities. The results revealed that social media and websites have a positive effect on student decision making, which then has a significant impact on students' choices at a particular university.

The of the study result [25] revealed the two-sided nature of community size: universities with a strong reputation tend to have more Facebook fans, but the presence of numerous Facebook fans has a detrimental effect on attracting individual fans.

Article [26] analyzes the trends of video marketing. The algorithm of video content creation is defined and the ways of video marketing using in the activity of universities are suggested, namely: creation of high quality video, video format should demonstrate social proof of quality of educational services; universities' education video services to increase conversions.

We consider the work [11] as the most relevant one, where authors developed a two-step method based on the Gaussian filter and decision tree (M-GFDT). The Gaussian filter corrects the distribution of business data in the first phase, and classifies the decision tree to remove inefficient online advertising while achieving high accuracy in predicting effective advertising. The second step provides testing of the method experimentally with data from a cross-border e-commerce company. In our opinion such approach is too complicated.

Thus, the above mentioned works, on the one hand, mostly analyze user actions in response to online advertising, such as clicks and visits to brand sites, etc. On the other hand, a number of analogs require relatively sophisticated tools for their implementation, that is, the question remains the simplification of data flows intellectual processing, their interpretation, classification in the process of formation and, accordingly, the adoption of targeted management decisions in the formation of advertising strategy of customer interaction with the business side.

3 Algorithm for Customer Engagement Targeting with a Facebook Business Page

Facebook is suitable for finding customers among the most solvent customers in age 25–50. This segment of the audience responds well to ads that they find useful to both themselves and their friends. This is manifested in the “natural” user's activity: in likes, reposts, comments. Due to this activity, the ad can gain additional reach and, conditionally, “free” customers who have seen their friend's repost or comment and also are interested.

Currently, targeted Facebook ads have great business opportunities. Artificial Intelligence analyzes user reactions to promotional content and concentrates impressions only on those users who are more likely to become your customers. Advertising

works great even with small budgets. The algorithm is fast-paced and capable of delivering clients with limited budgets.

With a quality strategy, it is possible to achieve the desired result in social networks, because, quite often, marketers work with an audience that does not know the advertised brand. Based on the recommendations [12, 13, 27, 28] it is possible to form a strategy for advertising on social networks (Fig. 1).

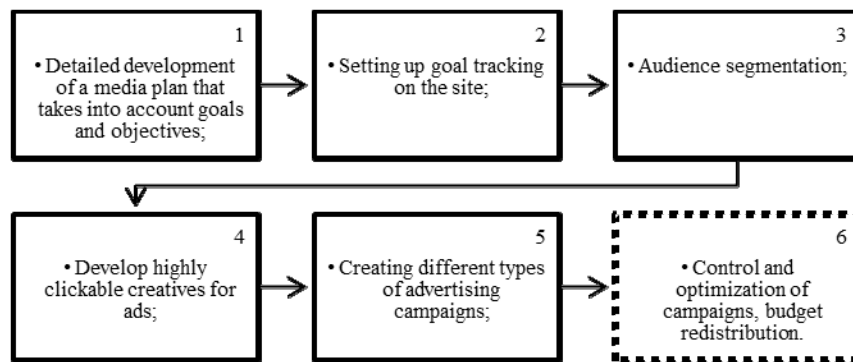


Figure 1. Step-by-step formation of advertising strategy on social networks

The last step (see Fig. 1) is worth considering in more details, as it takes a long time to process and is important for budget redistribution.

The objectives of controlling advertising include the following: analysis of the goals compliance and objectives of the advertising campaign with the goals and objectives of the company marketing strategy; Determining the difference between the planned and actual costs of the advertising campaign; determining specific results of advertising by certain time; development of measures for advertising activity improvement in future.

The control of advertising campaign results has seven stages [12, 14, 15]:

Stage 1. Conducting an audit, i.e. situational analysis.

Stage 2. Establishment of planned values and standards (goals and norms).

Stage 3. Measuring of actually achieved results for a certain period (day, week, month, quarter, year).

Stage 4. Comparison of actual values with planned and standard values.

Step 5. Analysis of the comparison results, which allows making changes in the planned values and standards or in the course of the advertising campaign.

Stage 6. Reflect on the effectiveness.

Stage 7. Ongoing project management.

The last two steps require the most time, so they will target the customer interaction with the business page based on the decision tree, and accordingly, it allows making changes in the advertising campaign.

The decision tree is a fairly common approach now to identify and visualize logical patterns in data. Dichotomous trees are used in this paper, here only two branches emerge from the top. Each node is mapped to a certain attribute, and branches to either specific values for qualitative features or a range of values for quantitative fea-

tures. The decision tree allows constructing a model of the dependence of many cases on many characteristic features.

The decision tree algorithm, first proposed by Quinlan, operates on the principle of recursively partitioning a dataset and incremental tree construction [16].

Table 1 lists the notations and descriptions used in the model tree targeting model. The table shows all the performance parameters that are displayed when targeting a Facebook ad campaign.

Table 1. Targeting options forming for customer interaction with a Facebook business page

Parameter	Value	Description
<i>Age</i>	Age	Age groups: 13-18, 18-25, 25-30, 30-35, 35-40, 40-45, 45-50, 50-55
<i>Sex</i>	Sex	female, male
<i>r</i>	Results	The number of times that an ad has reached a goal-specific result and setting.
<i>W</i>	Interaction with page	The total number of actions taken by people on a Facebook business page and posts on it as a result of viewing ads.

Let's build an algorithm for customer interaction targeting with a business page based on the decision tree (Fig. 2). The algorithm of customer interaction targeting with the business page on the basis of the decision tree will allow to reduce time-consuming and on the basis of it is possible to make changes in the advertising campaign.

For algorithm implementing the R programming language is used, which is free and has significant capabilities for statistical analysis, time series analysis, cluster analysis etc.

Initially, the data should be prepared for analysis (Block 1), with the R language requiring a database file (file with the extension "*.xlsx") (Block 2), where the model parameter values are located in the columns. Parameters are converted into data with factor values. In our case, there were parameters *Age*, *Sex* and *r*.

Next it is needed installation and running of the libraries (Unit 3) to build the model based on the decision tree.

The next step is to create control samples (Block 4). Splitting data from vector *r* into two sets in a predetermined ration, keeping the relative ratios of different labels in *r*. Also returned (Block 5) are subsets of vectors, matrices, or data frames that meet certain conditions.

In block 6, the model is constructed based on the recursive separation and regression tree. Due to the fact that the dependent parameter *r* is a factor, the construction of the model is determined by the class method. Next (Block 7) the targeting model of customer interaction with the business page is based on the decision tree, built on the breakdown of independent variables (*minbucket*, *minsplits*, *cp* (Complexity parameter)).

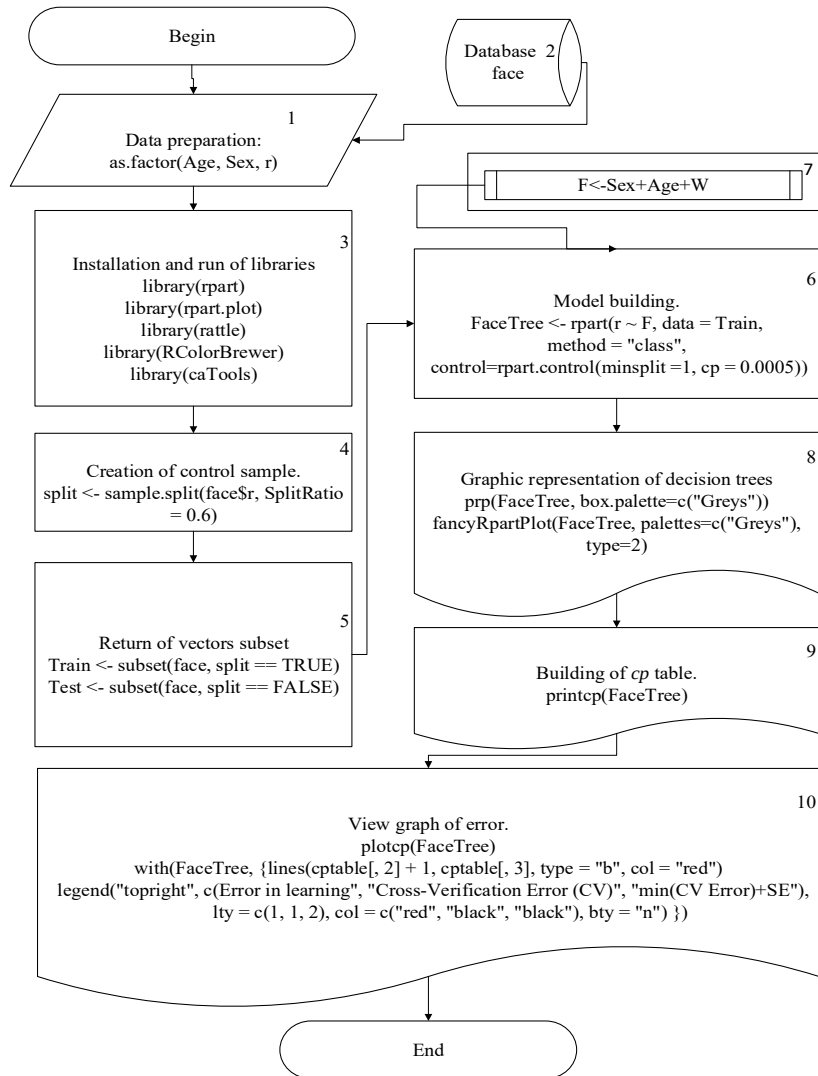


Figure 2. Algorithm for customer interaction targeting with business page based on decision tree

The graphs of the decision tree are then constructed, with the choice of the best representation (Block 8), the table of *cp* optimal segments based on the parameter of complexity (Block 9) and the graph of model error are presented.

4 Results and Discussion

Based on the above algorithm, we will build a model of targeting customer interaction with the business page based on the decision tree using data (more than 1 thousand indicators) of contextual advertising campaign on Facebook during the admission campaign, this is considered on the example of specialty “Computer Science” of Ternopil National Economic University (see Fig. 3).

Built model will reduce time spent, redistribute budget and make changes in the course of advertising campaign.

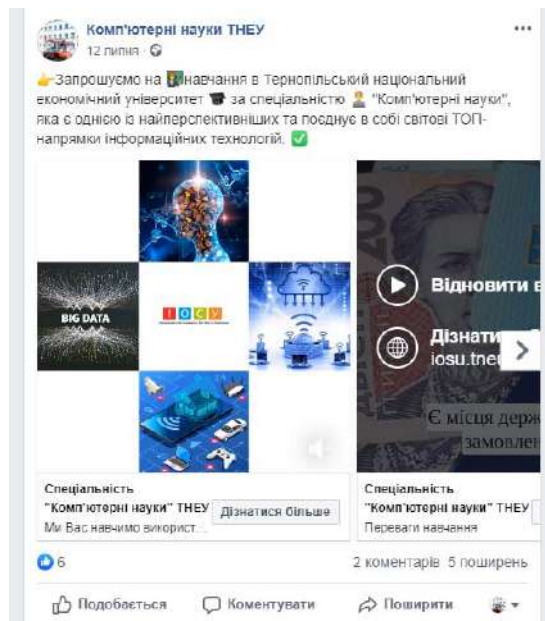


Figure 3. Advertising “Computer Science” of Ternopil National Economic University on Facebook

The targeting model of customer interaction with a business page based on a decision tree (see. Fig. 4) is usually the highest result (r) of a customer interaction with a Facebook business page “Computer Science TNEU” are for male and female in age group 40-55.

Fig. 5 shows the typical error behavior during ensemble training, the result is stabilized on the tree branch no. 5, the minimum relative error during cross-checking is $cp = 0.0021$.

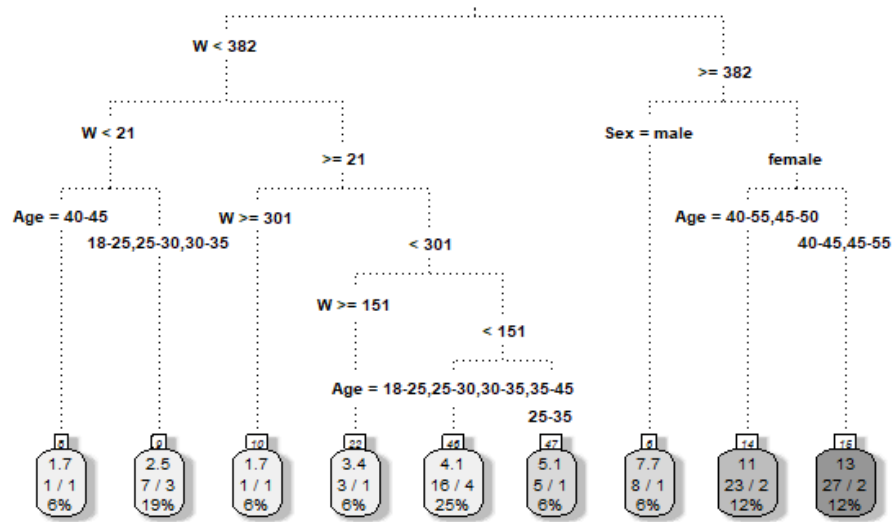


Figure 4. Targeting model of customer interaction with business page based on decision tree

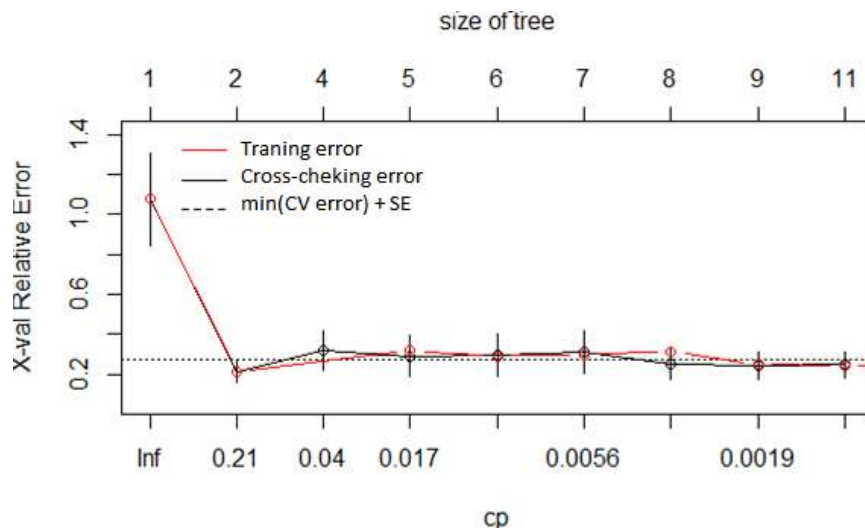


Figure 5. Errors in the targeting model for customer interaction with the business page based on the decision tree

Re-targeting results made it possible to improve advertising campaign results by 30% over first ad campaign option. The average r for all ad groups increased from 2.78 to 3.7. These results indicate the model adequacy.

The novelty of the work is following. In contrast to analogues [3, 11], a model of customer interaction targeting with a Facebook business page based on decision trees enables for the edges ("branches") of the tree to make changes to the advertising campaign strategy by the attributes on which the target function depends. Moreover, the

use of the rpart library in the R programming language makes it possible to clean and filter data quickly, which makes target groups forming easily comparing with [11].

The developed model of customer interaction targeting with a business page based on the decision tree has a practical importance for forming an advertising strategy of admission campaign in higher educational institutions.

5 Conclusions

The algorithm for forming the client interaction targeting with business page, based on the decision tree, is proposed. It enables to simplify the data flows intelligent processing, their interpretation, classification according to making the targeted management decisions.

Based on the algorithm, a model of customer interaction targeting with a business page based on the decision tree has built. It made it possible to evaluate the targeting of advertising campaign and form an advertising strategy of admission campaign.

According to the built model, the strategy of the advertising campaign for admission was formed as well as the basic rule was highlighted: the male and female clients in the age category 40-55 had the greatest interaction with the business page of Computer Science at Ternopil National Economic University. It enables to consider this rule to be taken into account during the advertising campaign for admission in higher educational institutions.

References

1. Zenith – <https://www.zenithmedia.com/>
2. Research & Branding Group – http://rb.com.ua/blog/praktika-polzovanija-socsetjami-v-ukraine/?fbclid=IwAR3LAJVpOwYPgDWdQ7uqvKWxrUnX8S_rkXry9nhmtuxsNUsinpNnn96zIvs
3. Lian, S., Cha, T., & Xu, Y.: Enhancing geotargeting with temporal targeting, behavioral targeting and promotion for comprehensive contextual targeting. *Decision Support Systems*. 117:28-37 (2019).
4. Ozcelik, A. B., & Varnali, K.: Effectiveness of online behavioral targeting: A psychological perspective. *Electronic Commerce Research and Applications*. 33:100819 (2019).
5. Ghose, A., Li, B., & Liu, S.: Mobile targeting using customer trajectory patterns. *Management Science*. 65(11): 5027-5049 (2019).
6. Lessmann, S., Haupt, J., Coussement, K., & De Bock, K. W.: Targeting customers for profit: An ensemble learning framework to support marketing decision-making. *Information Sciences*. in press. (2019).
7. Lo, Y. C., & Fang, C. Y. Facebook marketing campaign benchmarking for a franchised hotel. *International Journal of Contemporary Hospitality Management*. 30(3): 1705-1723 (2018).
8. Peruta, A., & Shields, A. B.: Marketing your university on social media: a content analysis of Facebook post types and formats. *Journal of Marketing for Higher Education*. 28(2) 175-191 (2018).

9. Jaman, S. F. I., & Anshari, M.: Facebook as marketing tools for organizations: Knowledge management analysis. In *Dynamic perspectives on globalization and sustainable business in Asia*. IGI Global: 92-105 (2019).
10. Belanche, D., Cenfor, I., & Pérez-Rueda, A.: Instagram Stories versus Facebook Wall: an advertising effectiveness analysis. *Spanish Journal of Marketing-Esic*. 23(1): 69-94 (2019).
11. Wang, H., & Hong, M.: Online ad effectiveness evaluation with a two-stage method using a Gaussian filter and decision tree approach. *Electronic Commerce Research and Applications*. 35: 100852 (2019). DOI: 10.1016/j.elerap.2019.100852.
12. Heinze, A., Fletcher, G., Rashid, T., & Cruz, A.: *Digital and social media marketing: a results-driven approach*. Routledge. (2016).
13. Paleha O.I., Horban Yu.I.: *Informational Business*. Kyiv National University of Culture and Arts. Kyiv: Lira-K. (2015). (in Ukrainian)
14. Mygal O. F., Gupalo O.V.: Features of Goods Promotion on Social Networks. *Proceedings of S.I. Yuriy third Scientific Readings, Ternopil, TNEU, 28 November 2017*, pp. 82-86 (2017).
15. Perna, S. N., Clifton, M. A., Jongjin, K. I. M., Varma, B. S., Piro, S. J., Mapen, B. E., & Davis, T. J.: U.S. Patent No. 10,025,982. Washington, DC: U.S. Patent and Trademark Office. (2018).
16. Shareef, M. A., Mukerji, B., Alryalat, M. A. A., Wright, A., & Dwivedi, Y. K.: Advertisements on Facebook: Identifying the persuasive elements in the development of positive attitudes in consumers. *Journal of Retailing and Consumer Services*. 43: 258-268 (2018).
17. Wiese, M., Martínez-Climent, C., & Botella-Carrubi, D.: A framework for Facebook advertising effectiveness: A behavioral perspective. *Journal of Business Research*. 109: 76-87 (2020).
18. Choi, I., Milne, D. N., Glozier, N., Peters, D., Harvey, S. B., & Calvo, R. A.: Using different Facebook advertisements to recruit men for an online mental health study: engagement and selection bias. *Internet Interventions*. 8: 27-34 (2017).
19. Dehghani, M., & Tumer, M.: A research on effectiveness of Facebook advertising on enhancing purchase intention of consumers. *Computers in Human Behavior*. 49: 597-600 (2015).
20. Kamboj, S., Sarmah, B., Gupta, S., & Dwivedi, Y.: Examining branding co-creation in brand communities on social media: Applying the paradigm of Stimulus-Organism-Response. *International Journal of Information Management*. 39: 169-185 (2018).
21. Ertugan, A.: Using statistical reasoning techniques to describe the relationship between Facebook advertising effectiveness and benefits gained. *Procedia computer science*. 120: 132-139 (2017).
22. Evert, V., Poels, K., & Walrave M.: An experimental study on the effect of ad placement, product involvement and motives on Facebook ad avoidance. *Telematics and Informatics*. 35: 470-479 (2018).
23. Jan M. T. & Ammari D.: Advertising online by educational institutions and students' reaction: a study of Malaysian Universities. *Journal of Marketing for Higher Education*. 26(2): 168-180 (2016). DOI: 10.1080/08841241.2016.1245232
24. Brech F. M., Messer U., Vander Schee B. A., Rauschnabel P. A. & Ivens B. S.: Engaging fans and the community in social media: interaction with institutions of higher education on Facebook. *Journal of Marketing for Higher Education*. 27(1), 112-130 (2017). DOI: 10.1080/08841241.2016.1219803

25. Kovpak, V., & Trotsenko, N.: The Pragmatic Potential of Native Advertising: Forms, Trends (on the Example of Native Content on the Social Network Facebook by the Brand of the Journalism Department of ZNU). *State and Regions. Series: Social Communications*. 1(41): 113-121 (2020).
26. Semeniuk, S. B.: Video Marketing in the Activities of Universities. *Marketing and Digital Tehnologies*. 3(1): 68-77 (2019). (in Ukrainian)
27. Drewniany, B. L., & Jewler, A. J.: *Creative strategy in advertising*. Cengage Learning. (2013).
28. Altstiel, T., Grow, J., & Jennings, M.: *Advertising creative: Strategy, copy, and design*. Sage Publications. (2018).
29. Bomba, A., Kunanets, N., Pasichnyk, V., & Turbal, Y.: Process modeling of message distribution in social networks based on socio-communicative solitons. *International Journal of Computing*. 17(4): 250-259 (2018).

Image Classification Based on the Kohonen Network and the Data Space Modification

Volodymyr Gorokhovatskyi^[0000-0002-7839-6223], Iryna Tvoroshenko^[0000-0002-7184-8143]

Kharkiv National University of Radio Electronics, Nauka Ave., 14, Kharkiv, 61166, Ukraine

gorohovatsky.vl@gmail.com,
iryna.tvoroshenko@nure.ua

Abstract. In this paper, we propose the solution of visual objects recognition in computer vision problems using the classification of descriptors of image key-points based on the training of Kohonen neural network on the description data of etalon images. According to the results of training within the set of etalons, the image classification method has been improved by defining a specific data space in the form of a statistical center for each etalon. We propose mathematical models for the bitwise analysis of multiple descriptors searching for the centers and the method for convolution of descriptions from multiple descriptors with the determining a posteriori probabilities for the system of bit centers. Methods of data space transformation of description bits are proposed for various options for Kohonen network training, processing and estimation of class centers. The software implementation of the changed classifier was performed as well as the processing time with different options for determining the space of training data was estimated. Experimental researches confirmed the high efficiency of classification preserving sufficient performance and the ability to use proposed methods in real-time applications.

Keywords: image classification, keypoint, ORB detector, descriptor, Kohonen network, network training, statistical center of the class, bit processing, key-points binary descriptors dataset space

1 Introduction

Accuracy and performance are the main indicators in computer systems when recognizing visual objects. Structural methods of image classification based on the use of neural networks as a way to identify patterns on the set of features of structural descriptions of the etalon base have become widespread in applied problems [1-6]. The result of training the network using the integrated feature space in the form of “data centers” is used for classification, which speeds up the calculation process [7].

Modern structural methods are based on the detection of keypoints of images and classification models in the space of binary vectors [3-6]. The keypoint is a coordinate-fixed numerical vector (descriptor), which reflects the properties of some of its surroundings. Many descriptors of keypoints provide the opportunity to recognize

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

images effectively and quickly. Modern keypoint generation methods (detectors), for example, ORB (Oriented FAST and Rotated BRIEF) [5] or BRISK (Binary Robust Invariant Scalable Keypoints) [6], also calculate keypoint descriptors in the form of binary vectors [1-6].

The main advantages of modern ORB and BRISK detectors are that they provide high performance with comparative or better accuracy than SIFT (Scale Invariant Feature Transform) and SURF (Speeded up Robust Features), because of the significant simplification of the processing using binary descriptors.

The investigation of the adaptive properties and parametric characteristics of neural networks for training in image classification problems with a variety of keypoint descriptors [8, 9] and almost unlimited variety of data being analyzed is important for data science.

It is also important to research the effectiveness of network training schemes that take into account the proximity of elements of various classes in the constructed feature space.

The implementation of Kohonen networks into the classification process makes it possible to tune methods to process arbitrary sets universally and successfully [10]. Classification quality directly depends on the results of training the network on a set of descriptors of the training sample [11].

Leading for the classification is the study of the possibilities and properties of the learning process of a neural network in terms of the most efficient use or transformation of the data space, and the research of in-depth training schemes.

The aim of the work is to improve the method for classifying images within the database of database of etalons on the detection of keypoint descriptors by defining a specific data space for network training and applying the classification to training results.

2 Formal problem statement

Let $W = \{x | x \in R^n\}$, $W \subseteq R^n$ be the space of structural features (descriptors of keypoints) of images. The base of descriptions of samples (etalons) $Z \subset W$ is given in the form of a set $Z = \{Z^j\}_{j=1}^J$, $s_j = \text{card } Z^j$, $s = \text{card } Z$, $s = \sum_j s_j$. Moreover,

each attribute $x_i^j \in Z^j$ is associated with a certain etalon Z^j of the base in terms of membership in it.

The main task of training the Kohonen network is to establish a classifier in terms of the formation of a system of centers for etalon data, the effective use of the etalon information placed in the descriptions to ensure high classification efficiency.

The purposes of the work are the use of the Kohonen neural network to train the classification system in a specific data space, analyze the options for training the network, study adaptive capabilities and assess the effectiveness of the functioning of the network and the effectiveness of classification by software modeling.

3 Literature review

We know the use of various types of neural networks for pattern recognition of visual objects for tasks such as getting key parameters or features of given patterns, classifying patterns or characteristics already got from them [12, 13], and solving optimization problems. An example of the use of a single-layer neural network on multivalued neurons is classification using the frequency characteristics of the image [14]. As a rule, applied experiments are based on information obtained from training a network in a fixed dataset.

It is known to use a multilayer neural network to classify images of faces (features – nose, mouth, eyes) [15]. The use of a neural network for classifying images was demonstrated when the results of decomposition by the principal component method are input to the network [16]. As a rule, a neural network to reduce the dimensionality of data is used precisely to calculate the key characteristics of the image which are used for further classification [17, 18].

A neural network detector was used to detect a human image [19]. The training was based on images that contain and do not contain images of people. To increase the reliability of detection, a complex of neural networks trained in different initial weights was used. Each of the networks gives its error, so the final decision was made by voting.

A unique effect on the classification problem is provided by networks that can self-organize – Kohonen neural networks, they provide a topological ordering of the input image space. Unlike other methods, the topological ordering of classes preserves the similarity of input images [20], which is especially useful in classifying many classes. The change of the Kohonen network weights is carried out by the competitive training method with a sufficiently high processing speed.

The application of the gradient descent algorithm is presented when setting weights in the Kohonen network [21]; the search for the minimum point is carried out in the opposite direction to the gradient of the optimized function. Such an algorithm is characterized by a “dip into the pit of a local minimum”, when the modification of the parameters practically ceases, despite the presence of another deeper extremum. This problem can be partially solved if we take into account the factors of change of each weight [22] or take into account the value of second-order derivatives [20]. An analysis of the sources [13-22] showed that the most common use of Kohonen neural networks is precisely when solving the classification problem without a teacher, for clustering.

The Kohonen network can recognize clusters in data and establish the proximity of classes [23]. Thus, it is possible to improve understanding of the data structure to further refine the neural network model by adjusting the existing rules for classifying objects. We can use the Kohonen network in classification problems where classes are already given [23]. It is important that, if, after recognition, the Kohonen network encounters a data set is unlike any of the samples known to it, then it classifies such a set as a new class [20]. The Kohonen network is studied by the method of successive approximations [23]. The training, in this case, does not consist in minimizing the

error, but in fine-tuning the internal network parameters for a maximum match with the input data.

The author's monograph [23] contains a variety of approaches to building and training networks, which allows you to choose the network structure by the data being analyzed. In general, the application of the Kohonen network in computer vision systems is aimed at identifying the most significant features of the image which are subsequently used for recognition [20-22].

The statement of the problem of cluster formation on a set of descriptors is studied [3, 4], and the classification efficiency based on cluster centers for the Leeds Butterfly application dataset is evaluated [20, 22].

Researches related to the formation of initial centers on the basis of statistical processing of multiple descriptors contribute to more efficient data self-organization [2, 7]. Learning options for learning a network on a set of descriptors related to the number of adapted neurons in the learning algorithm are studied [11]. With an increase in the number of such neurons, the quality of training improves by increasing the processing time.

In this paper, the keypoints binary descriptors dataset was generated using an ORB detector, justified the choice of which in [11].

4 Analysis and selection of learning options for the Kohonen network

The purpose of training the network in this formulation is to form a system of centers on a multitude of descriptors of the etalon base. Let us introduce the Kohonen network learning procedure as a sequence of stages [11, 20].

Stage 1. Choose the elements of the training set in the form of a set of $Z = \{Z^j\}_{j=1}^J$ descriptions of all elements of the database of etalons.

Stage 2. Initiate the matrix of weights $M = \{m_j\}_{j=1}^J$, where the rows are the formed vectors of neurons of m_j centers of classes, so $m_j = x_i^j$, $x_i^j \in Z^j$, where i is the number of an arbitrary vector from the class of samples Z^j .

Stage 3. Choose the current element $z \in Z$, for each $j \in [1, 2, \dots, J]$ we calculate the distance $q_j = \rho(z, m_j)$ and determine the class d of the inner neuron: $d = \arg \min_j q_j$.

Stage 4. Calculate the changes in the weights for the neurons of the output network layer

$$\Delta m_j = h(j, d, t) \cdot \eta \cdot (z - m_j) \quad (1)$$

where η is the learning rate, $h(j, d, t)$ – the value of the neighborhood function for neuron number j at the time of training t .

Usually, $h(j, d, t) = \exp[-\rho(mj, md) / \sigma(t)]$ is defined as a Gaussian function, and radius $\sigma(t) = 1 / \exp(t^{-2})$ of the outskirts decreases with increasing parameter t , $t = 1, \dots, s$.

Stage 5. Adjust the matrix of weights $M = M + \Delta M$ in step t .

Stage 6. Continue learning the Kohonen network (stages 3-5) until list Z is completed.

Stage 7. Verify the fulfillment of the termination condition of the Kohonen network. Traditionally, the criterion is the error value or the total distance between the centroid systems at steps t and $(t-1)$. If the stopping condition is not fulfilled, we continue training by stage 3, the selection of data from set Z is carried out in a fixed or random order.

In practice, we apply another common criterion for stopping network learning in the form of a fixed number of iterations [16]. This choice is made to evaluate and compare training time in existing data spaces. As a result of the training, we get a system of centers that is adapted to recognize arbitrary structural descriptions based on the training sample of the database of etalons.

The choice of metric in stage 3 is determined by the space of descriptors and the method of forming the centers. For binary ORB descriptors, we will use the Hamming metric. To carry out the learning of the Kohonen network, the square of the Euclidean metric was used in the experiments, even though for binary features the value of the Hamming metric and Euclidean distance is identical.

The quality of classification directly depends on the learning outcomes of the system and the available set of structural descriptions of the training sample.

Let us focus on possible applied versions of the modified definition of the space of educational data in the form of:

- a) multiple descriptors for each sample;
- b) sets of sample descriptors for which the statistical procedures for the formation of class centers have been previously applied [2, 11];
- c) systems of centers for each sample class. This option differs from b) in that a class of centers is formed based on each of the samples;
- d) implementation of the convolution procedure for the center system of each sample.

Options c) and d) carry out a more in-depth analysis, taking into account the characteristics of each of the samples as representatives of the class.

Training of the Kohonen network during the experiments, normalized data were used to maintain a sufficient distance between neurons to ensure effective separation of classes.

The effectiveness of the classification of visual objects according to a set of key-points using the Kohonen network directly depends on such fundamental interrelated factors: a dataset as a set of descriptors, a data space for training, pre-processing methods are applied (the method of forming descriptors or the initial choice of centers), a metric for comparing descriptors, the size of the tuple adapted during the training of neurons.

As a criterion for assessing the quality of classification (error value), we choose a value that calculates the proportion of the elements of the training sample Z , which, according to the results of the classification, fell into the wrong class [11, 24-26]. We set the criterion as

$$\beta = \sum_{j=1}^J (s_j - a_j) / s, \quad (2)$$

where a_j is the number of features from their total number s_j in the description of the etalon Z^j , classified in the classification process as class j . The value β reflects the level of erroneous decisions in the classification. The closer β is to zero, the higher is the quality of classification in the training set.

We analyze three basic options for constructing a classifier which differs in the number of neurons that are modified in the learning process:

- only the winning neuron is configured;
- three neurons that are closest in distance to the winner neuron are modified;
- the complete network of neurons is modified by the distance to the winner's neuron.

Let us evaluate the recognition performance using these training options for the Kohonen network depending on the data management parameters that are different in depth of data analysis.

The indicated training options for the classifier fundamentally differ only in stage 4, where function $h(j, d, t)$ is calculated differently. To achieve convergence, condition $h(j, d, t) \in [0, 1]$ must be fulfilled. However, the option when only the winning neuron is tuned implements rough recognition. The variant, where three neurons are modified that are closest in distance to the winner neuron, classifies each element under research to one of the three nearest centers. The variant where the complete network of neurons is modified in accordance with the distance to the winner's neuron performs the most complete processing, relating the considered descriptor to all class centers at the same time, but with different weights proportional to the distance [11].

5 Ways to transform a data space for learning

Given the binary representation of ORB descriptors, we apply bitwise processing and, at the pre-processing stage, for each etalon description Z^j , we determine the class center vector based on a logical rule that compares the total number of units for each of 256 bits of the entire set of description descriptors and half of their number [2, 7]:

$$m_j(b) = \begin{cases} 1, & \sum_{d=1}^{s(j)} x_d(b) \geq s(j)/2, \\ 0, & \sum_{d=1}^{s(j)} x_d(b) < s(j)/2, \end{cases} \quad x_d \in Z^j, \quad b = 1, \dots, 256, \quad (3)$$

where $x_d(b)$ is a bit with number b of the descriptor and number d in the etalons description.

According to (3), the values of each of the bits for center m_j are determined by the bits of the total set of descriptors belonging to etalon number j . Center (3) reflects the statistical properties of the sample or class. During the research, centers (3) can be used to define the learning data space in the variants of classification b) and c).

An important characteristic for options a) and b) of learning the Kohonen network is the training time to ensure the classification is effective. The potential need is to increase the number of keypoints because ORB and other detectors sometimes select keypoints that do not contribute to the efficient classification of the image which introduces the requirement to consider alternatives by method (3), which reduces to comparing the obtained centers. This enables the neural network to significantly reduce the time of training and classification. Another option is to increase the number of generated keypoints and use pre-processing methods to obtain image descriptions.

The main function of neural networks is a generalization of the system of attributes for an image [20]. Most of all, this concerns convolutional neural networks, whereby targeted processing, which is typical for computer image analysis methods, a significant reduction in the total number of parameters of the network model is achieved to minimize network retraining [23, 27].

We interpret the image described by bitmaps of the descriptors of etalon Z^j in the form of a probability map

$$p^j = \left[\sum_{d=1}^{s(j)} x_d(b) \right] / s(j), \quad (4)$$

that in the values of vector p^j integrates the spatial information of description Z^j by adding the values of the bits of the descriptors to obtain a shortened model of the input data [3, 28]. Vector p^j summarizes the characteristics of the image by supplying the frequency of occurrence of a unit in each of the 256 bits of the ORB descriptor.

We feed the normalized vectors (4) to the input of the Kohonen network and train the network only on them. The classification according to option d) is also based on the values of (4). Figure 1 shows an example of the values of the vector (4) for the test image.

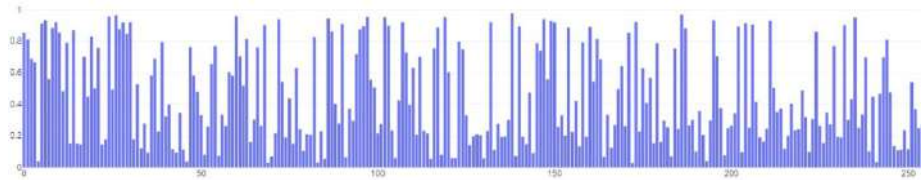


Fig. 1. An example of a vector p^j

Vectors p^j can be used for any type of network, which significantly reduces the learning time of the Kohonen network by processing a single vector where hundreds of vectors have traditionally been processed.

6 Experiments and results

To implement the proposed classifier modifications, the Visual Studio 2019 environment, and the OpenCV library tools were selected [29]. The OpenCV library has over 2500 software modules, which include a set of traditional and modern computer vision algorithms, as well as libraries of machine learning programs. For the experiment, four categories of images were selected, which are shown on the example of the Leeds Butterfly dataset [20, 30] (Fig. 2).



Fig. 2. Test images

Figure 3 contains an example of the coordinates of the prevailing keypoints shown as circles. Images of butterflies is a unique object for scientific research because visually different butterflies have both significantly different and a number of common properties.



Fig. 3. An example image with the coordinates of keypoints

We will evaluate the effectiveness of the classification with network training options and different data analysis depths. The performance criterion is selected (2). During the experiment, 2 variants of the number of descriptors on each etalon 400 and 100 were tested to study their effect on performance. In combination with the parameter of the number of descriptors per etalon, 2 options were tested for the number of epochs of the Kohonen network 100 and 200 for searching and achieving a balance of performance and accuracy. The increase in the number of iterations to 300 or 400 was tested, the result showed a lack of growth in the classification efficiency, while the learning time of the Kohonen network significantly increases.

It has been established that options with tuning exclusively the winner neuron or three neurons closest in distance to the winner neuron make it possible to perform recognition faster or more accurately [3, 28].

The proposed options a)-d) were researched for learning the Kohonen network and classifying it into 4 classes using 4 categories of etalons (Fig. 2).

It was revealed that the training time of the researched network depends on the following factors:

- the number of epochs;
- the number of descriptors per sample;
- the number of neurons that change during network learning.

A series of experiments with a fixed number of iterations of learning the Kohonen network was carried out. The number of descriptors for each etalon ranged from 375...400.

The results of the research of option a) for the set of ORB descriptors for the test case (Fig. 2) showed that this approach does not make it possible to classify etalons efficiently since the number of keypoints defined by the classifier from the class description is not the maximum. The value of criterion $\beta = 0.65$, while the estimated operating time of the improved method, was 14.94 s. This argues for the use of transforming the data space of the description in the form of pre-processing approaches or building centers to increase efficiency.

The training procedure for option b) was carried out on two different images of each of the 4 etalons (Fig. 2). The number of keypoints ranged from 378...400, and the number of iterations was 200. The total classification time was 4.98 s, the error value was $\beta = 0.31$. Table 1 contains the number of votes of descriptors for input images of given classes from the test sample which is assigned according to the results of classification using training method b) to certain etalons. Having analyzed the data in Table 1, a high degree of difference was revealed: the maximum values of the votes are on the diagonals and significantly exceed other elements, that is, all etalons are classified correctly. As we see, the use of the Kohonen network with the center parameter (3) significantly reduces the classification error: from 0.65 for method a) to 0.31 for method b).

Table 1. The result of the classification method b)

Etalons	Classes			
	1	2	3	4
Z^1	279	33	27	61
Z^2	58	288	25	29
Z^3	36	91	228	23
Z^4	88	9	3	300

We will analyze the effectiveness of the classification using network training options that differ in the depth of analysis depending on the number of neurons to be configured, test a different number of descriptors for each etalon, and a different number of iterations of the network learning.

For variants with 400 descriptors per etalon, the classification error ranged from 0.33-0.28. For variants of 100 descriptors, the error naturally increased and amounted

to 0.46-0.29. So, the main dependence is the effectiveness of the number of iterations of the network. The best results were obtained for 400 descriptors, where even with a few iterations of the Kohonen network (100); an error index of 0.28 was achieved.

A change in the number (1 or 3) of adapted neurons almost does not affect the learning indicators of the Kohonen network, but it is effective about the classification error and the number of erroneously classified data. Modification of one neuron is practical only for 200 iterations and a significant number of descriptors. Even under these conditions, the classification error is 10% larger than in the corresponding variants with the modification of 3 neurons, the number of errors increased from 0 to 2. The best result was obtained at 200 iterations of 3 neurons, while all classes of etalons with an error of 0.23 were correctly classified.

The estimated runtime for options with 400 descriptors was 13.7-16.21 s for options with 100 descriptors 6.2-7.9 s. It was found that method b) showed a high level of difference compared to others; however, it requires significant amounts of training time. In addition, experiments showed that the processing time in method c) was significantly reduced in comparison with other methods; it amounted to 2.27-2.77 s, the error ranged from 0.20-0.34.

It was revealed that the number of descriptors in the etalon and the number of iterations have the main influence on the effectiveness and time of classification. The best results were obtained at 200 iterations, even when using 100 descriptors, the error was 0.2.

It should be noted that a change in the number of adapted neurons practically does not affect either the operating time of the Kohonen network or the classification efficiency.

In variant d) we apply definition (4) and calculate for the descriptors of each etalon the probability vector of the appearance of a unit in each bit. The error for the best network operation options (400 descriptors and 200 iterations) significantly decreased to 0.06, only one etalon was incorrectly classified. It was found that with a decrease in the number of descriptors the error increases to 0.34, and a decrease in the number of iterations to 100 leads to a classification quality of 0.20.

Thus, the results showed a clear relationship between the classification indicators and the number of descriptors in the description of the etalon. The number of iterations is also an important parameter, enhances the classification capabilities of the network, and the number of adapted neurons under the given experimental conditions and the content of the analyzed images have almost no effect on the result.

Method d) based on (4) significantly reduces the operating time of the neural network compared to other methods without compromising quality. The classification quality for options c) and d) is almost the same. The processing time was reduced by almost 8 times (from 16.21 s to 2.82 s for options with similar quality), which allows us to use the proposed approach d) for large image datasets, in real-time applications, when processing video signals. The specified time potentially allows you to process every tenth frame of video with a refresh rate of 60 frames per second.

The obtained data on the network training time and the value of the classification error are summarized in Table 2. We see that the use of Kohonen network learning tools in the calculation of centers significantly reduces the classification error.

Table 2. Estimated Kohonen network learning time and mean error β

	Method			
	a)	b)	c)	d)
Learning time, s	14.94	4.98	2.52	2.82
Error mean β	0.65	0.31	0.27	0.06

The experiments confirmed the ability to successfully adapt the Kohonen network to arbitrary sets of visual data, especially in the case of the successful formation of the initial centers of the class or the process of descriptor pre-processing.

7 Conclusions

The paper describes the results of a comparative analysis of the developed training methods and classification of images based on it for experimental images of the Leeds Butterfly dataset for various parameters of the Kohonen classification network. The research confirmed the ability to adapt network parameters universally for arbitrary visual data; transformations using statistical class centers and data convolution at the preliminary processing stage are especially effective.

The methods with the formation of a system of centers and the use of data convolution showed the best results for each class sample, and the use of convolution significantly increases the speed of data processing compared to other options.

The contribution of the research is the improvement of the structural classification methods using the Kohonen network by introducing a new description data space based on a system of etalon centers and data convolution, which helps to ensure high classification efficiency with sufficient performance and makes it possible to use modified methods in real-time applications.

The practical importance of the work is made by the obtained software models for assessing the effectiveness of classifiers in computer vision systems; the effectiveness of the development in the examples of image datasets is confirmed.

Further performance improvements can be achieved through supervised training approaches since the dataset descriptor classes are known. However, it is clear that such accounting for additional information potentially complicates the processing and affects performance.

References

1. Gorokhovatskyi, V.A.: Image classification methods in the space of descriptions in the form of a set of the key point descriptors. *Telecommunications and Radio Engineering*. 77(9), pp. 787–797 (2018). doi:10.1615/TelecomRadEng.v77.i9.40
2. Gadetska, S.V., Gorokhovatsky, V.O.: Statistical measures for computation of the image relevance of visual objects in the structural image classification methods. *Telecommunications and Radio Engineering*. 77(12), pp. 1041–1053 (2018). doi:10.1615/TelecomRadEng.v77.i12.30

3. Gorokhovatskyi O., Gorokhovatskyi V., Peredrii O.: Analysis of Application of Cluster Descriptions in Space of Characteristic Image Features. *Data*. 3(4), 52 (2018). doi:10.3390/data3040052
4. Gorokhovatskyi, V., Putyatin, Y., Gorokhovatskyi, O., Peredrii, O.: Quantization of the Space of Structural Image Features as a Way to Increase Recognition Performance. In: *Proceedings of The Second IEEE International Conference on DataStream Mining & Processing, DSMP-2018, Lviv, Ukraine, 21–25 August 2018*, pp. 464–467 (2018). doi:10.1109/DSMP.2018.8478434
5. Rublee, E., Rabaud, V., Konolige, K., Bradski, G.: ORB: an efficient alternative to SIFT or SURF. In: *Proceedings of the IEEE International Conference on Computer Vision*, 6 November 2011, pp. 2564–2571 (2011). doi:10.1109/ICCV.2011.6126544
6. Leutenegger, S., Chli, M., Siegwart, R.Y.: BRISK: Binary Robust invariant scalable keypoints. In: *Proceedings of the IEEE International Conference on Computer Vision*, 6 November 2011, pp. 2548–2555 (2011). doi:10.1109/ICCV.2011.6126542
7. Gorokhovatskyi, V., Gadetska, S., Ponomarenko, R.: Recognition of Visual Objects Based on Statistical Distributions for Blocks of Structural Description of Image. *Lecture Notes in Computational Intelligence and Decision Making*. In: *Proceedings of the XV International Scientific Conference “Intellectual Systems of Decision Making and Problems of Computational Intelligence”, ISDMCI’2019, 21 May 2019*, pp. 501–512 (2019). Available online: https://rd.springer.com/chapter/10.1007/978-3-030-26474-1_35
8. Liu, Yu, et al.: Exploring disentangled feature representation beyond face identification. In: *Proceedings IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2080–2089 (2018). doi:10.1109/CVPR.2018.00222
9. Khan, A., et al.: Structural vibration-based classification and prediction of delamination in smart composite laminates using deep learning neural network. *Composites Part B: Engineering*. 161, pp. 586–594 (2019). doi:10.1016/j.compositesb.2018.12.118
10. Samanta, S., et al.: Fast Character Recognition Using Kohonen Neural Network. In: *2018 International Conference on Advanced Computation and Telecommunication (ICACAT)*, 28 December 2018, pp. 1–4 (2018). doi:10.1109/ICACAT.2018.8933627
11. Gorokhovatskyi, V.A., Pupchenko, D.V.: Classification of images of visual objects according to the set of descriptors of singular points based on the Kohonen neural network. *Control, navigation and communication systems*. 2(48), pp. 68–72 (2018). (in Russian) Available online: <http://dspace.ubs.edu.ua/jspui/handle/123456789/1356>
12. Kumar, V., Namboodiri, A., Jawahar, C.V.: Semi-supervised annotation of faces in image collection. *Signal, Image and Video Processing*. 12(1), pp. 141–149 (2018). doi:10.1007/s11760-017-1140-5
13. Hietanen, A., et al.: A comparison of feature detectors and descriptors for object class matching. *Neurocomputing*. 184, pp. 3–12 (2016). doi:10.1016/j.neucom.2015.08.106
14. Zhang, S., et al.: Constructing deep sparse coding network for image classification. *Pattern Recognition*. 64, pp. 130–140 (2017). doi:10.1016/j.patcog.2016.10.032
15. Liu, Y., Yan, J., Ouyang, W.: Quality aware network for set to set recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5790–5799 (2017). doi:10.1109/CVPR.2017.499
16. Miikkulainen, R., et al.: *Evolving deep neural networks. Artificial Intelligence in the Age of Neural Networks and Brain Computing*. Academic Press, pp. 293–312 (2019). doi:10.1016/B978-0-12-815480-9.00015-3
17. He, T., et al.: Bag of tricks for image classification with convolutional neural networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 558–567 (2019). doi:10.1109/CVPR.2019.00065

18. Stanley, K.O., et al.: Designing neural networks through neuroevolution. *Nature Machine Intelligence*. 1(1), pp. 24–35 (2019). doi:10.1038/s42256-018-0006-z
19. Islam, M., Chen, G., Jin, S.: An Overview of Neural Network. *American Journal of Neural Networks and Applications*. 5(1), pp. 7–11 (2019). doi:10.11648/j.ajinna.20190501.12
20. da Silva, I.N., et al.: Self-organizing Kohonen networks. *Artificial Neural Networks*. pp. 157–172 (2017). doi:10.1007/978-3-319-43162-8_8
21. Galutira, E.F., Fajardo, A.C., Medina, R.P.: A Novel Learning Rate Decay Function of Kohonen Self-Organizing Maps Using the Exponential Decay Average Rate of Change for Image Clustering. In: *Proceedings of the 2nd International Conference on Natural Language Processing and Information Retrieval*, 7 September 2018, pp. 55–59 (2018). doi:10.1145/3278293.3278299
22. Hu, Z., Bodyanskiy, Y., Tyshchenko, O.: Kohonen Maps and Their Ensembles for Fuzzy Clustering Tasks'. *Self-Learning and Adaptive Algorithms for Business Applications*. pp. 51–77 (2019). doi:10.1108/978-1-83867-171-620191004
23. Kohonen, T.: *Self-Organizing Maps*. Springer-Verlag, Berlin, Heidelberg (2001). doi:book/10.5555/558021
24. Ahmad M. Ayaz, Tvoroshenko Irina, Baker Jalal Hasan, Lyashenko Vyacheslav: Modeling the Structure of Intellectual Means of Decision-Making Using a System-Oriented NFO Approach. *International Journal of Emerging Trends in Engineering Research*. 7(11), pp. 460–465 (2019). doi:10.30534/ijeter/2019/107112019
25. Tvoroshenko, I.S., Kramarenko, O.O.: Software determination of the optimal route by geoinformation technologies. *Radio Electronics Computer Science Control*. 3, pp. 131–142 (2019). doi:10.15588/1607-3274-2019-3-15
26. Tvoroshenko, I.S., Gorokhovatsky, V.O.: Intelligent classification of biophysical system states using fuzzy interval logic. *Telecommunications and Radio Engineering*. 78(14), pp. 1303–1315 (2019). doi:10.1615/TelecomRadEng.v78.i14.80
27. Nasirahmadi, A., Ashtiani, S.H.M.: Bag-of-Feature model for sweet and bitter almond classification. *Biosystems engineering*. 156, pp. 51–60 (2017). doi:10.1016/j.biosystemseng.2017.01.008
28. Pathak, D., Krähenbühl, P., Donahue, J., Darrell, T., Efros, A.A.: Context encoders: Feature learning by inpainting. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2536–2544 (2016). doi:10.1109/CVPR.2016.278
29. OpenCV Open Source Computer Vision. <https://docs.opencv.org/master/index.html>
30. Yan, X., Yang, J., Sohn, K., Lee, H.: Attribute2image: Conditional image generation from visual attributes. In: *European Conference on Computer Vision*, 8 October 2016, pp. 776–791 (2016). doi:10.1007/978-3-319-46493-0_47

Image structural classification technologies based on statistical analysis of descriptions in the form of bit descriptor set

Volodymyr Gorokhovatskyi¹[0000-0002-7839-6223], Svitlana Gadetska²[0000-0002-9125-2363],
Natalia Stiahlyk³[0000-0002-0573-3508]

¹Kharkiv National University of Radio Electronics, Nauka Avenue 14, Kharkiv, 61166,
Ukraine
gorohovatsky.vl@gmail.com

²Kharkiv National Automobile and Highway University, Yaroslava Mudrogo str., 25, Kharkiv,
61002, Ukraine
svgadetska@ukr.net

³Kharkiv Educational and Research Institute of the University of Banking, Peremogy Avenue
55, Kharkiv, 61174, Ukraine
natastyaglick@gmail.com

Abstract. The problem of image recognition in computer vision systems is considered. We offer technologies for classifying visual objects using a statistical center based on a structural description of the image as a set of key point descriptors. The use of statistics for the bits of the description data helps to increase performance while providing sufficient classification performance. Results of experimental modeling and peculiarities of implementation of the developed approaches are discussed.

Keywords: computer vision, image classification, key point, descriptor, statistical center, nonparametric statistics, performance, classification results

1 Introduction

In today's computer vision systems, the use of data processing technology has become widespread, providing information about recognized visual objects as set of key points (KPs) of 100-1,500 pixels [1]. At the same time, each point reflects the properties of the image in the neighborhood of its coordinates and is described by a vector-descriptor with dimensions in 64... 512 components [2]. Due to the large volumes of visual data flows, there is an urgent need to use the mining apparatus to identify the most essential regularities in the content of the description of the visual objects in order to reduce computational costs while ensuring effective classification.

Considering the structure of the analyzed data, which contains numerous set of identical elements, it is promising to use the apparatus of mathematical statistics. Its tools are able to reveal the deep informational properties of descriptions, which are necessary for effective recognition of images [2-5].

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

The aim of the paper is to study the technologies of statistical analysis and nonparametric methods for the formation of innovative feature space, which provides effective classification of images by their structural description.

The tasks of the paper are the construction of mathematical models for the transformation of descriptions and the calculation of relevance in the synthesized space of features, the implementation of software modeling to evaluate the effectiveness of the developed modifications.

2 Formal problem statement

Traditionally, the system of structural pattern recognition $R = \langle V, T, U \rangle$ involves three main components [1]: V – used method (comparison with the etalon in the feature space), T – set of features used in training and recognition (set of descriptors of KPs), U – set of recognized situations (set of classes given by samples). One of the particular cases R is a "one-on-one" classification, when only one class is identified compared to others, that is $U = \{0,1\}$.

In our work, within the framework of recognition technology, effective modifications for features T and related method V have been proposed. If it is reliable assumption that the elements of description create the concentrated cluster structures, then it is possible to build a classification based on the cluster characteristics for the description data [6]. From the point of view of processing speed the most effective situation is when only cluster is created on a dataset and its properties are given by the statistical center [8]. The analysis and purposeful processing of the center proposed in the paper significantly increase the rate of performance.

The second area of research is the direct application of nonparametric statistics methods to the set of the bit description, which together with the acceleration of processing considers the peculiarities of the entire statistical picture of the data description.

3 Literature review

Detecting of deep regularities in the set of numerical data is a powerful tool for the robust distinguishing of dataset by their quantitative characteristics in the classification process.

Recent studies in the field of image analysis and processing have developed in the aspect of quantitative analysis regarding the structure of the analyzed data [7-11].

According to specific applications, authors of the contemporary research continue to offer effective measures to determine the similarity of object descriptions that are calculated and analyzed with the help of the key point descriptor sets obtained by detectors [8-14].

Particular attention is focused on the parameters of the proposed method performance [16-20]. Based on this, it is essential to use spatial analysis and

processing of descriptor values in the form of a block system model, which makes it possible to statistically summarize the information and accelerate the implementation of classification procedures [4, 8].

Powerful statistical analysis tools in the practice of computer vision are nonparametric statistics methods that do not rely on information about type of distribution [4-6]. It seems natural to use them to analyze the properties of a random vector system, which in this case acts as a model for the analyzed object description.

The essence of the BRISK detector for determining the KPs of the image and the procedures of comparison of descriptors are analyzed in [19-22]. The possible ways of forming the description centers are discussed in [2, 3], moreover, modifications of the space of KP descriptors are considered in [8, 25]. Studies related to the use of structural descriptions when processing video streams are contained in the monograph section [26].

Reference to the software used in modeling the developed methods is given in [27].

4 Analysis and construction of feature space

Let the vector space of binary vectors of dimension n defines the universe B^n , $\text{card } B^n = 2^n$. We will identify space B^n with the set of descriptors which are the vectors constructed by the KPs detector for the image [2-3].

Consider the description of the analyzed image as a fixed finite subset $Z \subset B^n$ which consists of the vectors from B^n . In general, we can consider Z as a multiset, since the description elements can be repeated or similar to each other. This approach promotes the representation of the visual object as a collection of its structural elements, and the value of each descriptor is the result of extracting meaningful features of the spatial fragments of the image. The set $Z = \{z_v\}_{v=1}^s$, $z_v \in B^n$, is composed of s descriptors of KPs with binary values, generated by detectors such as ORB, BRISK, AKAZE [19-21].

Consider the set Z in a structured matrix form $D = \left\{ \left\{ d_{i,j} \right\}_{i=1}^s \right\}_{j=1}^n$ for the fixed value n and specific sequence of s descriptors of KPs. For simplicity, we consider s to be the same for all etalons, whose finite set represents the recognized classes of images [2-3].

Suppose that the structural description Z is characterized by the compact placement of elements around a center in the space B^n . This makes it possible for the available data to estimate the allowable limit for the distance within the neighborhood of the determined center and classify the description by calculating the number of its equivalent elements and the refined center used the established threshold.

It is clear that in the aspect of multi-class classification, the effectiveness of this method naturally depends on how well the computed centers will differ from each other in the applied feature space.

Let's interpret the sequence of bits of each descriptor z_v (it is the row of the matrix D) as separate case of n – dimensional discrete signal. We also construct for Z the aggregated image $h(Z)$ of the “data center” in the form

$$h(Z) = (h_1, \dots, h_j, \dots, h_n) \quad h_j = \sum_{i=1}^s d_{i,j}, \quad (1)$$

where h_j are the sums of the elements of corresponding columns.

Let's call the vector $h(Z)$ the statistical center (SC) of the description and put its values at the basis of classification.

It is important for the practical implementation of recognition procedures that representation $h(Z)$ doesn't depend on the order of descriptors z_v in the set Z , that is, the model $h(Z)$ is invariant to mixing descriptors that corresponds to the arbitrary arrangement of rows in the matrix. This property makes it possible to receive descriptors in any order, taking into account the allowable geometric transformations of objects. In fact, the vector $h(Z)$ is a generalized image of an object in the n – dimensional space C^n of the vectors with integer nonnegative components since $h(Z) \in C^n$, $h_j \geq 0$. The range of values h_j in (1) is determined directly by the calculation procedure and the number of descriptors s ; due to the bit type of data, it can be considered as given: $h_j \in \{0, 1, \dots, s-1, s\}$.

In addition to the description, we will also refer to the normalized representation $h^s(Z) = (h_1^s, h_2^s, \dots, h_s^s)$, where $h_i^s = h_i / s$ are obtained by normalizing values h_i to the descriptor number s of the description, $h_i^s \in [0; 1]$. Vectors h and h^s can be considered as posteriori characteristics of the description, as they are calculated on the basis of input information.

5 Classification based on data center

Let's carry out a statistical analysis of the data of the binary matrix D and on the basis of the values of the vectors h or h^s set a logical procedure for determining the fact of belonging an arbitrary vector $b \in B^n$ to the given description $Z \in B^n$.

At the pre-processing stage, we calculate a vector whose components have the form $q_i = \rho(h^s, z_i)$, $i = 1, \dots, s$, where $\rho(h^s, z_i)$ is a distance between vector h^s and descriptor $z_i \in Z$. The range of values for $\rho(h^s, z_i)$ is known, since the space of vectors is determined by the metric and value of the data. For example, the value of the Manhattan distance $\rho(h^s, z_i) = \sum_{v=1}^n |h^s(v) - z_i(v)|$ belongs to the segment $[0; n]$.

Let's rank the values q_i by type $q_1 \leq q_2 \leq \dots \leq q_s$, then choose the median value $q_{s/2}$ from the sorted sequence. The distance value $q_{s/2}$ halves the whole set of

descriptors and half of the descriptors will be in the neighborhood of the center h^s with distance less than $q_{s/2}$ [2].

We choose the value $\alpha = q_{s/2}$ as the threshold for the distance $\rho(h^s, z_i)$ to classify the query for the assignment of an arbitrary vector b to the set Z . This choice of the value α under the condition $\rho(h^s, z_i) \leq \alpha$ provides equivalence with accuracy α for most vectors of description Z and its center h^s .

Determining the threshold α by moving the sorted sample to the left or right of its center allows further to adapt the data content to the center value and can contribute to a more reliable classification.

After that we define a refined SC h^{s*} for the subset of the descriptors selected by the threshold α with the help of the averaging the values of only the descriptors for which the condition $\rho(h^s, z_i) \leq \alpha$ is satisfied:

$$h^{s*} = \frac{1}{s^*} \sum_{i=1}^s z_i \chi(z_i), \quad \chi(z_i) = \begin{cases} 1, & \rho(h^s, z_i) \leq \alpha, \\ 0, & \rho(h^s, z_i) > \alpha, \end{cases} \quad (2)$$

where s^* – the number of such descriptors, $\chi(z_i)$ – characteristic function as a result of logical analysis of distance.

We now apply the refined SC h^{s*} to classify descriptors of the arbitrary description. The selected concentrated portion of the description data in the form (2) contributes to a more thorough identification of characteristics of data concentration in terms of differences with other descriptions.

In general, the classification using the SC for the image database will be represented by technology in the form of data processing steps:

- 1) analysis of the set of etalon descriptions of the image base to determine the SC and the threshold α for each etalon;
- 2) calculating the refined SC (2) for the set of etalons;
- 3) analysis of descriptor set that is the description of a recognized object for satisfaction of inequality $\rho(h^{s*}, z_i) \leq \alpha$ with respect to each etalon;
- 4) counting the number of voices of descriptors that satisfy the condition $\rho(h^{s*}, z_i) \leq \alpha$ for each etalon;
- 5) classification of the object on the basis of determining the etalon with the highest number of received votes.

The proposed technology gives the possibility of modification, for example, in stage 3 it can be competitively identified only one of the etalons based on the shortest distance to the refined center. Another development of the approach may be the use of weighting classification coefficients for single bits of the SC, which enhances the individual characteristics of the descriptions and aims to improve the quality of classification.

6 Application of nonparametric statistics apparatus

The vector $h(Z)$ is a characteristic of a structural description of a visual object. Another factor for distinguishing the descriptions presented by the matrix D can be directly statistical distributions of data [2-3] or the parameters of these distributions for individual bits or blocks D [3]. Such “indirect” approaches to performance analysis, instead of directly data values, often make it possible to successfully solve applied problems. The advantage is a significant reduction in the amount of computation. The disadvantages are related to some blurring in the image space when determining relevance.

Let's imagine the value $d_j = \{d_{i,j}\}_{i=1,\dots,s}$ of the j -column bits of the matrix D as the realization of some random variable b_j , $j = 1, \dots, n$, which takes the binary values $b_j \in \{0,1\}$. We consider the column sequence as a system of independent random variables. On the basis of sampling d_j , we determine experimentally the mathematical expectation and variance of each of the quantities $b_j \in \{0,1\}$.

We now apply a nonparametric statistical criterion based on the experimental values of mathematical expectation and variance to evaluate the relevance of the distributions of the corresponding bits for the compared descriptions. A distinctive advantage of nonparametric criteria using numerical characteristics of distributions is their independence on the type of statistical data distribution [4, 5].

If the statistical characteristics a_1, σ_1^2 and a_2, σ_2^2 for two random variables are known (a – mathematical expectation, σ^2 – variance), then the degree of closeness of these variables can be estimated on the basis of statistics [1, 3]:

$$\beta = \frac{\left[(\sigma_1^2 + \sigma_2^2)/2 + ((a_1 - a_2)/2)^2 \right]^2}{\sigma_1^2 \sigma_2^2}. \quad (3)$$

At full coincidence of parameters a_1, a_2 and σ_1^2, σ_2^2 we have $\beta = 1$, and as their differences increase, the value β increases indefinitely. Based on the introduction and application of the threshold δ_β for the criterion value β , we will construct a procedure for determining the relevance of the two descriptions and the classification.

Unlike parametric statistical criteria, the implementation of criterion (3) does not rely on statistical tables. In this case, there is a problem of determining the lower limit δ_β for the value β_j , which exceed indicates a significant difference between the studied object descriptions.

The value of expression (3) is calculated separately for each element d_j of the bit array. At the same time it can be considered in sense of multidimensional distribution, where the parameters a, σ^2 are vectors containing the characteristics of the

independent variable system. For a multidimensional situation, the criterion β is given by a vector of values.

The application of criterion (3) in the relevance analysis of the descriptions makes it possible to significantly simplify the classification of the object in comparison with the traditional apparatus for accepting statistical hypotheses [4, 5] and does not depend directly on the type of probability distribution.

For example, consider the value of a single bit from a set of descriptors as a discrete random variable, described by distributions p_0, p_1 , where p_0 – the probability of 0, p_1 – the probability of 1. Mathematical expectation of θ is defined as $M[\theta] = \sum_k \theta_k p_k = p_1$. The variance of θ is calculated as $D[\theta] = \sum_k (\theta_k - M[\theta])^2 p_k = p_0 p_1$ [4].

Based on the analysis of the values $b_1, b_2, \dots, b_n$ from the system of independent bits, we obtain for each etalon description its image in the form of a representation by the vector values of the mathematical expectation $M = (M_1, \dots, M_n)$ and variance $D = (D_1, \dots, D_n)$. Further, on the basis of values M, D by expression (3) and by applying the logical processing using a threshold δ_β for the value β , we can decide or reject the equivalence of the compared descriptions.

The most general model for a system of independent bits involves different probability values of occurrence of 1 in each single bit for a given etalon class. In this case, we have a distribution where in each bit values 1 or 0 appear independently. If p_j is the probability of occurrence of 1 in the j -th bit, then we have a system of random variables given by the set of proper probabilities: $p_1, p_2, \dots, p_n$. A specific set of values for these probabilities will determine the etalon class. That is, the etalon can be represented in a probabilistic way in the form $p_1, p_2, \dots, p_n$.

The probability of a specific sequence of 0s and 1s is described by the product of n probabilities p_j or $(1 - p_j)$ depending on the value at the j -th position: 1 or 0. The probabilities p_j are directly calculated based on the normalized distribution as $p_j = h_j / s$. For example, the probability of occurrence of a sequence (0,1,0,1,1,1) is a product $P(0,1,0,1,1,1) = (1 - p_1) \cdot p_2 \cdot (1 - p_3) \cdot p_4 \cdot p_5 \cdot p_6$.

A separate theoretical case of the analyzed distribution is binomial one when $p_1 = p_2 = \dots = p_n = p$, that is, the case of coincidence of values p_j for single bits from the list of descriptors. It means that the characteristic of the description is the value p . It is clear that, in such an idealized model, the different descriptions must have the different values p to distinguish them.

Then the descriptor values provide a random vector of dimension n with the parameter p of occurrence of 1 in every single bit. The probability of occurrence of 1s in some specific sequence of 0s and 1s can be written as $C_n^k p^k (1 - p)^{n-k}$, where C_n^k is a number of combinations. As we know, the binomial distribution is described

by the parameters of mathematical expectation $M = np$ and variance $D = np(1 - p)$. Values M, D are the integral characteristics of the whole image and they can be directly applied in expression (3).

The threshold value δ_β for logic processing can be found by the fraction of the number of bits in the chain representation of the description, whose distributions have significantly different mathematical expectations according to the Z-criterion [3, 4]. For example, 75% of the bits of two test objects have significantly different mathematical expectations (i.e. 25% of them have no significant difference). We calculate the values of statistics $\beta_j, j = 1, \dots, m$, rank the obtained values in ascending order, and reject the first 25%. The first remaining value is taken as the limit. Objects will be considered the different ones if at least 75% of the obtained statistic values β_j exceed the threshold.

7 Experiments and results

We have performed a software modeling of the proposed classification technologies based on structural data description in C # Visual Studio 2017 using Open CV library tools [2-4, 27]. To conduct the investigation we took the images (Fig. 1) for which the structural descriptions were calculated using the BRISK detector (parameters $s = 176, n = 512$). Neighborhoods of KPs are shown in Fig. 1 as small rings. The first two images are visually similar to each other ("angel and demon"), which always causes difficulties in recognition even by human vision.



Fig. 1. Images with neighborhood of KPs

The calculated Manhattan distances $\rho(i, k)$ between statistical centers (i, k - numbers of centers) for the images of Fig. 1 have the following values: $\rho(1, 2) = 23$, $\rho(1, 3) = 44$, $\rho(2, 3) = 37$. Generally, it indicates a significant similarity between the values of the descriptions in the investigated feature space of SCs, since all values of distance in this example belong to the interval $[0; 512]$. The corresponding

normalized values are $\rho^*(1,2)=0.044$, $\rho^*(1,3)=0.085$, $\rho^*(2,3)=0.072$. As we can see, even visually similar images in Fig. 1 can be distinguished by the value of the distance between the SCs.

Calculating the refined centers (2) on the basis of the threshold for half of the descriptors, we obtained: $\rho^*(1,2)=0.052$, $\rho^*(1,3)=0.087$, $\rho^*(2,3)=0.076$, therefore, the degree of differentiation between the SCs has obviously increased. This indicates a certain improvement in the effectiveness of classification.

The obtained number of votes for the elements of the etalons assigned to the corresponding class (Fig. 1) by comparing the distance between them and refined center with the boundary value is shown in Table 1.

Table 1. The number of classified elements in the description of etalons

Class number	1	2	3
1	44	29	11
2	36	44	20
3	32	27	44

As we can see, the number of votes on the diagonal of Table 1 is sufficiently greater than the values in the row and column, which confirms the effectiveness of the method when classifying the descriptors. The two visually similar images in Fig. 1 are also successfully differentiated. Thus, the formation of a concentrated subset (cluster) of the description using the refined SC with logical processing (2) provides effective classification.

The effectiveness of the application of nonparametric criterion (3) in the classification of visual objects was tested experimentally in the problem of distinguishing halftone silhouette images shown in Fig. 1. For each of the images, 500 ORB descriptors of dimension 256 have been generated. We can see that the images used in the experiment are visually similar to each other, which makes it possible to evaluate the sensitivity and performance of structural methods to distinguish such signals. It is clear that for significantly dissimilar images, recognition performance will be better.

Calculation of the criterion β by the expression (3) for one-bit distributions of the first two similar images in Fig. 1 showed results which were quite close to the minimum value $\beta=1$. All values for each of the 256 bits of the dataset were in the range 1.0–1.12. Choosing the threshold value $\delta_\beta=1.006$ the necessary decision value $n_0=128$ (half of the total number of bits) is achieved. It means the ability to distinguish these essentially similar images in the constructed space of statistical features, which is based solely on the values of mathematical expectation and variance of data. Under the condition $\delta_\beta \geq 1.006$, these images are considered different by the proposed classification method. In Fig. 2 the curve 1 shows the dependence of the value n_0 on the threshold value δ_β .

The effect of impulse noise on the input image was also simulated: with a given probability p , the pixels of the image were changed by a randomly generated noise value in the range $0, \dots, 255$ [1]. Experiments have shown that under the influence of impulse noise for images in Fig. 1 the curve of dependence of the number of equivalent bits n_0 required for statistical decision making on the threshold δ_β shifts to the right (Fig. 2, curve 2). Even for a significant noise value $p = 0.7$, each of the distorted images are recognized successfully by this measure when setting the threshold $\delta_\beta = 1.05$.

Software modeling has also shown that when calculating the criterion β as a vector for substantially dissimilar images, a slightly higher threshold value should be set, for example, $\delta_\beta = 1.1$. It makes it possible to reliably identify each of the images in the experiment. The dependence of the value on the threshold value has the form shown in Fig. 2.

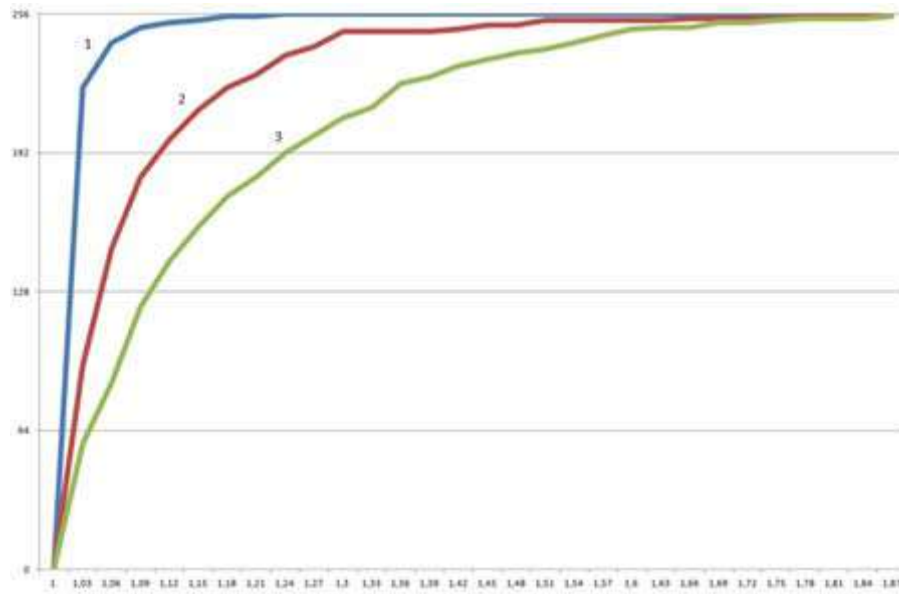


Fig. 2. Dependence of n_0 on threshold δ_β : 1 - for similar images, 2 - for noise-affected images, 3 - for visually different images

As we can see from the obtained experimental results, effective classification actions of a logical plan with a threshold value δ_β for the analyzed images can only be performed in a quite small size range $\delta_\beta \in [1; 1.1]$ (when choosing $n_0 = 128$). Obviously that closeness of δ_β to 1 is a criterion for identifying an etalon image among others, or when assigning a classified image to one of the etalons. In fact, the

choice of the threshold value $\delta_\beta = 1.1$, as can be seen from Fig. 2, makes it possible to reliably identify each of the images analyzed in the experiment, including visually similar ones. The analysis of the results of the modeling showed that for an arbitrary set of images it is possible to set a common or individual threshold δ_β for each etalon, which gives an opportunity to classify the images to one of the etalons.

8 Conclusion

The statistic center and criterion (3) constructed in the paper are based on the evaluation of the properties of the bit structure formed on the data description, as they obtained by independent bitwise processing of a set of descriptors.

The formation of a concentrated subset in the form of clusters for description elements using a refined statistical center with logical processing provides a reduction in computational cost and effective classification by revealing significant patterns of description of the visual objects.

The use of nonparametric methods for the statistical analysis of descriptor bits using the mathematical expectation and variance makes it possible to classify even visually close images.

The scientific novelty of investigation is improvement of the methods of structural classification of images by introducing a statistical methods apparatus for a system of bit description, which facilitates the processing and enhancement of classification performance without reducing the efficiency.

The practical value of the paper is the developed mathematical and software models for the implementation of image classifiers in the computer vision.

The further development of the study is the creation of applications of multidimensional or block data structure based on set of bits.

References

1. Gorokhovatskiy, V.A.: Compression of Descriptions in the Structural Image Recognition. *Telecommunications and Radio Engineering*. 70 (15), pp. 1363–1371 (2011) doi: 10.1615/TelecomRadEng.v70.i15.60
2. Gorokhovatskiy, V.O., Gadetska, S.V., Ponomarenko, R.P.: Logical analysis and processing of data for the classification of images on the basis of formation of statistical center description. *Control, Navigation and Communication Systems*, №4 (56), pp.43–48 (2019) (in Ukrainian)
3. Gorokhovatskiy, V.O., Gadetska S.V., Stiahlyk, N.I.: Study of statistical properties of the block supply model for a number of decorators of key points of images. *Radio Electronics, Computer Science, Control*, №2, pp. 100–107 (2019) (in Ukrainian)
4. Duda, R.O., Hart, P.E., Stork, D.G.: *Pattern Classification*. Wiley-Interscience Publication (2001)
5. Aivazyan, S.A., Yenyukov, I.S., Meshalkin, L.D.: *Applied Statistics: Bases of Modeling and Initial Data Processing*. Moscow, Finansy i Statistika (1983) (in Russian)

6. Gorokhovatskyi, O., Gorokhovatskyi, V., Peredrii, O.: Analysis of Application of Cluster Descriptions in Space of Characteristic Image Features. *Data*, 3(4), 52 (2018) doi: 10.3390/data3040052
7. Gadetska, S.V., Gorokhovatsky, V.O.: Statistical Measures for Computation of the Image Relevance of Visual Objects in the Structural Image Classification Methods. *Telecommunications and Radio Engineering*, 77 (12), pp. 1041–1053 (2018) doi: 10.1615/TelecomRadEng.v77.i12.30
8. Liu, H., Tan, T.H., Kuo, T.Y.: A Novel Shot Detection Approach Based on ORB Fused with Structural Similarity. *IEEE Access*, pp. 2472–2481 (2019)
9. Wang, J., Cheng, K., Ke, C., Liu, M., Cai, L., Shi, H.: Image Registration of Infrared and Visible Image Based on Geometric Constraint. *Journal of Physics, Conference Series* (2019)
10. Gayathiri, P., Punithavalli, M.: Partial Fingerprint Recognition of Feature Extraction and Improving Accelerated KAZE Feature Matching Algorithm. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)* Volume-8 Issue-10, pp. 3685-3690 (2019) doi: 10.35940/ijitee.J9653.0881019
11. Merino, I., Azpiazu, J., Remazeilles, A., Sierra, B.: 2D image features detector and descriptor selection expert system. Natarajan Meghanathan et al. (Eds) : *NLP, ARIA, JSE, DMS, ITCS*, pp. 51-61 (2019) doi: 10.5121/csit.2019.91206
12. Kamel, M.M., Taha, H.S., Salama, G.I., Elhalwagy, Y.Z.: Ground target localization and recognition via descriptors fusion. *18th International Conference on Aerospace Sciences & Aviation Technology*, pp. 1-11 (2019) doi:10.1088/1757-899X/610/1/012015
13. Xu, Y., Gao, F., Wu, T., Bennett, K.M., Charlton, J.R., Sarkar, S.: U-Net with optimal thresholding for small blob detection in medical images. *15th International Conference on Automation Science and Engineering (CASE) IEEE*, pp. 1761-1767 (2019)
14. Fawad et al.: Image Local Features Description Through Polynomial Approximation. *IEEE Access*, vol. 7, pp. 183692-183705, (2019) doi: 10.1109/ACCESS.2019.2959326
15. Liang, L., Pan, J.S., Zhuang, Y.: A Fast Specific Object Recognition Algorithm in a Cluttered Scene. *Journal of Internet Technology* Volume 20, pp. 2023-2031 (2019)
16. Bay, H., Ess, A., Tuytelaars, T., et al: Speeded-Up Robust Features (SURF). *Computer Vision and Image Understanding*, 110(3):346 (2008)
17. Patil, V.P., Singla, C.R.: Performance Evaluation of Various Image Features Detectors and Descriptors Used in the Development of Panorama from Real Time Video Files, *International Journal of Recent Technology and Engineering (IJRTE)*, Volume-8 Issue-3, pp. 7274-7279 (2019) doi:10.35940/ijrte.C6401.098319
18. Rosten, E., Porter, R., Drummond, T.: Faster and Better: A Machine Learning Approach to Corner Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.32, no. 1, pp. 105-119 (2008)
19. Leutenegger, S., Chli, M., Siegwart, R.Y.: BRISK: Binary Robust invariant scalable keypoints. *International Conference on Computer Vision, Barcelona*, pp. 2548-2555 (2011)
20. Wulff, D., Kuhlemann, I., Ernst, F., Schweikardm A., Ipsen, S.: Robust motion tracking of deformable targets in the liver using binary feature libraries in 4D ultrasound. *Current Directions in Biomedical Engineering*, 5(1), pp. 601–604 (2019)
21. Muja, M., Lowe, D.G.: Fast Matching of Binary Features. *Conference on Computer and Robot Vision*, pp. 404-410 (2012) doi: 10.1109/CRV.2012.60
22. Gorokhovatskyi, V.A.: Image Classification Methods in the Space of Descriptions in the Form of a Set of the Key Point Descriptors. *Telecommunications and Radio Engineering*, 77 (9), pp. 787-797 (2018) doi: 10.1615/TelecomRadEng.v77.i9.40

23. Orobinskyi, P., Petrenko, D., Lyashenko, V.: Novel Approach to Computer-Aided Detection of Lung Nodules of Difficult Location with Use of Multifactorial Models and Deep Neural Networks. In: 2019 IEEE 15th International Conference on the Experience of Designing and Application of CAD Systems (CADSM), pp. 1-5 (2019) doi: 10.1109/CADSM.2019.8779340
24. Mashtalir, S., Mashtalir, V.: Spatio-Temporal Video Segmentation. In: Mashtalir V., Ruban I., Levashenko V. (eds) Advances in Spatio-Temporal Segmentation of Visual Data. Studies in Computational Intelligence, vol 876. Springer, Cham. pp. 161-210 (2020) doi: 10.1007/978-3-030-35480-0_4
25. OpenCV Open Source Computer Vision. Available online: <https://docs.opencv.org/master/index.html>

Analysis of Voice Signal Phase Data Informativity of Authentication System User

Mykola Pastushenko¹[0000-0003-2664-1167], Yana Krasnozheniuk¹[0000-0001-9884-0275],
Oleksandr Lemeshko¹[0000-0002-0609-6520]

¹ Kharkiv National University of Radio Electronics, Kharkiv, 14 Nauky Ave., UKRAINE

mykola.pastushenko@nure.ua
yana.krasnozheniuk@nure.ua
oleksandr.lemeshko@nure.ua

Abstract. Directions of improving the quality characteristics specific to voice authentication systems in various access systems are analyzed and explored in the article. One of the main directions for improving the quality characteristics of the user authentication systems is the use of phase information of a voice signal. The urgent scientific task of studying new procedures is being solved to refine the estimates of the pitch frequency obtained on the basis of the amplitude-frequency spectrum analysis. The estimates were refined using phase data of the voice signal, as well as estimates of the pitch frequency in the process of obtaining cepstral coefficients. The results are obtained in the course of statistical analyzing the simulation results using experimental voice data of the authentication system user. Phase data of a voice signal allows obtaining adequate and reliable estimates in the process of spectral analysis. However, if there are errors associated with gross errors, for example, when taking the first or second formants for estimating the pitch frequency, preference should be given to the estimate obtained in the process of calculating cepstral coefficients. The presented research results should be used in voice authentication systems, improving speech recognition systems, as well as in solving speaker identification problems.

Keywords: amplitude, authentication, voice signal, information, spectrum, phase, pitch frequency.

1 Introduction

In recent decades, the achievements of science and the latest infocommunication technologies more than ever determine the dynamics of economic growth, the level of population's well-being, the competitiveness of the state in the world community, the degree of ensuring its national security and equitable integration into the global economy. The rapid development and widespread use of modern information and tele-

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

communication systems marked the transition of mankind from the industrial society to the information society that is based on the latest communication systems, the reliability of which does not always comply with increasing requirements. The quantity, technical level and accessibility of information systems, their reliability and performance stability already determine the degree of country's development and its status in the world community, and in the near future they will undoubtedly become a decisive mark of this status.

At the same time, the process of informatization of the world community generates a complex of negative phenomena, first of all, theft of financial, informational and computing resources. Indeed, the high complexity and vulnerability of all the systems on which regional, national and global information spaces are based, as well as the fundamental dependence of state infrastructures on their stability, lead to the emergence of principally new threats that in some cases can be solved by improving access systems.

Due to the wide spreading of distributed systems in all spheres of human activity, the task of ensuring information security in such systems is acute. One of the main measures to protect financial resources, information data and computing resources is to ensure reliable user authentication.

Currently, there are many approaches to authentication and even more implementations of these approaches. However, not all classical solutions to the authentication problem are suitable for implementation in distributed systems. In addition, various types of systems present their unique requirements for authentication subsystems. Moreover, the active development of computer technology makes it easy to crack authentication algorithms that were considered reliable 10-15 years ago. For example, in 2019, the total estimated income of fraudsters obtained using bank cards in Ukraine increased from UAH 245.8 million to UAH 361.99 million (an increase of 47.3%), as reported by the Deputy Director of the Ukrainian Interbank Payment Systems Member Association "EMA", Olesya Dalnichenko. This is largely caused by the insecurity of the password protection of bank cards.

In this regard, continuous work is ongoing in the field of research and development of authentication methods. New algorithms are constantly appearing and existing ones are being improved to ensure secure user authentication. The problem of authentication of users with access to public and personal information resources is becoming increasingly relevant. This problem is especially important for open, mass telecommunication and information systems. One of the most promising areas for protecting such systems from unauthorized influence is biometric methods for identifying users. However, despite all the attractiveness, this approach is fraught with a number of serious problems.

Initially, the development and implementation of biometric systems was associated with static biometric attributes of the user (face image, papillary finger pattern and iris), which have proven themselves in forensics. However, to date, these hopes have been destroyed, primarily because of the simplicity of the fake.

Therefore, in recent years, a lot of research has been carried out in the field of application of dynamic (behavioral) biometric authentication systems. Among these biometric systems, voice authentication takes a special place, which is simple and convenient. However, like all biometric systems, voice authentication has low quality

characteristics. In this regard, intensive research is being carried out in the field of voice authentication, as evidenced by the works [1-4].

In modern voice authentication systems (VASs), the amplitude information of a polyharmonic non-stationary voice signal of a user is recorded. User authentication is carried out mainly in the process of analyzing the amplitude-frequency spectrum of registration materials [2]. The main efforts of researchers in this case are focused on search for new or improvement of existing procedures for the formation (estimation) of templates (a set of attributes – pitch frequency, formant data, cepstral coefficients, mel-frequency cepstral coefficients, linear prediction coefficients and their dynamic characteristics, etc.) of the user, as well as the development of decision rules. The following decision-making procedures are the most popular among the latter – the methods of Gaussian Mixture Model (GMM) and Support Vector Machine (SVM). For these purposes, artificial neural networks and Hidden Markov Models (HMM) are also used.

The aim of this work is to study the influence of modern achievements in digital information processing on the accuracy of evaluating individual characteristics of the analyzed voice signal in the process of forming a user template. The object of study is the process of digital processing of voice signals.

2 General problem statement

In our opinion, an increase in the quality indicators of VASs is connected, first of all, with a change in the paradigm of digital processing of registration materials, which is associated with the addition of the amplitude-frequency spectrum analysis with modern advances in digital information processing, including algorithms for recording phase data of voice signals.

Currently, there is another way to improve the quality of the VASs, which is based primarily on the use of phase information of the user voice signal. It has long been known [5] that the phase is a more informative parameter of the signal, however, it is traditionally ignored in the VASs [2].

This is caused by the fact that to obtain phase information, additional computational and algorithmic resources are needed, which are not always available in these applications. Note that earlier in radar and radio communications to obtain phase data, special bulky devices were used – phase shifters, which could not be used in the field of voice signal processing. Currently, there are specialized microcircuits or digital signal processors that are also applicable in the field of digital processing of voice signals.

In addition, there are some features of the estimation, pre-processing and use of phase data. It should be noted that at present there is no experience and practice of using the signal phase with respect to voice authentication tasks.

This is confirmed by the fact that there are only a limited number of known works where phase data were used in the processing of speech signals. For example, in [6] the relevance of using phase information in the processing of speech data was pointed out, and in [7] a phase was used to clarify the frequency characteristics of the processed voice data. In [8], a comparative analysis of the procedures for estimating the

phase relationships between the vibrations of the pitch and harmonics of speech signals was performed, which the authors propose to use for solving problems of recognizing speech sounds and identifying speakers.

The above emphasizes the relevance of studies estimating the effect of phase data on the quality characteristics of voice authentication procedures. Phase data in voice authentication can be used in several ways that are practically important for digital processing of voice signals:

- increasing the signal-to-noise ratio of the registration materials (a known direction of using the phase in radar and radio communications);
- improving the quality of the formation of attributes for traditionally used templates, for example, the pitch frequency, formant information, etc.;
- development of new procedures for the formation of template elements based on phase data [9].

3 Work-related analysis

Let us analyze the latest scientific works in the field of speech signal processing when the issues of voice authentication in infocommunication systems have become particularly relevant. It is obvious that voice identification technologies have come to the user authentication systems from forensics. The scientific basis for the use of voice identification technology in forensics was investigated and discussed in detail in [10].

The general conclusion is that the voice identification differs from identification of fingerprints, where the variations are very small, and there is no absolutely reliable method for determining whether speech signals belong to the same person. In forensics, speaker recognition can only be probabilistic, i.e. indicating the likelihood that two speech signals belong to the same person. Under conditions of an analog telephone channel, even recognition of gender or age is sometimes complicated. Due to the small sample of speech signals, the confidence interval for evaluating the likelihood of two speech recordings belonging to the same speaker is so large that an unambiguous solution is impossible.

The task of segmentation of speakers is rather close. The segmentation of speakers in the conversation flow of different speakers (audio-indexing, diarization) is necessary when marking up sound transcripts, newsgroups, radio and television shows, interviews, etc. However, as in forensics, the quality of speaker extraction is low and unacceptable for solving user authentication problems [11].

As shown in [12], the individuality of the acoustic characteristics of the voice is determined by three factors: the mechanics of the vocal folds vibrations, the anatomy of the speech tract and the articulation control system. Naturally, the voice signal propagation channel can have some influence on acoustic characteristics (for example, the influence of external noise), the effect of which in modern systems is eliminated by digital processing procedures and organizational measures. Acoustically, the style is realized in the form of a contour of the pitch frequency, the duration of words and its segments, the rhythmic of the shock segments, the duration of pauses, and the volume level [12].

The attribute space, in which a decision is made on the identity of the speaker, should be formed taking into account all factors of the speech formation mechanism: the voice source, resonant frequencies of the speech path and their attenuation, as well as the dynamics of articulation control. In particular, in [12, 13], the following parameters of the voice source are considered: the average pitch frequency, the pitch frequency-period contour, fluctuations in the pitch frequency and the shape of the excitation pulse. The spectral characteristics of the vocal tract are described by the envelope of the spectrum and its average slope, formant frequencies and their bands, long-term spectrum or cepstrum [13].

It was shown in [14] that the most important factor in voice individuality is the fundamental frequency ($F0$), followed by formant frequencies, the size of fluctuations $F0$, and the slope of the spectrum. In [15], it was suggested that the attributes associated with $F0$ provide the best separability of voices, followed by the signal energy and the duration of the segments.

In some works, the formant frequencies are considered the most important factor [16, 17]. In particular, the fourth formant is practically independent of the type of phoneme and characterizes the tract [17].

The speaker recognition method is dominated by the cepstral method of transforming the spectrum of voice signals, which was first proposed in [18].

Cepstrum describes the shape of the envelope of the signal spectrum, which integrates the characteristics of the excitation sources (voice, turbulent and pulsed) and the shape of the speech tract. In experiments on subjective speaker recognition, it was found that the envelope of the spectrum strongly affects voice recognition [19]. Therefore, the use of a particular method of spectrum envelope analysis for speaker recognition is justified.

Instead of calculating the spectrum of the speech signal using the discrete Fourier transform over a short time interval, the amplitude-frequency characteristic of the signal found from the coefficients of linear speech prediction can also be used [20].

In [21], three informative areas were found: 100-300 Hz (the influence of a voice source), 4-5 kHz (pear-shaped cavities) and 6.5-7.8 kHz (possibly the effect of consonants). A small area is in the region of 1 kHz.

Due to the fact that the vast majority of speaker recognition systems use the same attribute space, for example, in the form of cepstral coefficients, their first and second differences, much attention is paid to the construction of decision rules, which were discussed above.

The development and application of the GMM method was considered in [22, 23]. The GMM method can be considered as an extension of the vector quantization method [23]. Vector quantization is the simplest model in speaker recognition systems, regardless of context.

The Support Vector Machines method (SVM) is actively used in various pattern recognition systems after the publication of the monograph [24]. This method allows to build a hyperplane in a multidimensional space that separates two classes, for example, the parameters of the target speaker and the parameters of speakers from the reference base. The hyperplane is calculated using not all parameter vectors, but only

specially selected ones. These vectors are called reference vectors. Since the dividing surface in the initial parameter space does not necessarily correspond to a hyperplane, a nonlinear transformation of the space of the measured parameters into a certain attribute space a higher dimension is performed. This nonlinear transformation must satisfy the requirement of linear separability in the new attribute space. If this condition is satisfied, then the dividing surface in the hyperplane is constructed using the support vector method. Obviously, the success of applying the support vector method depends on how well the non-linear transformation is selected in each specific case when recognizing speakers.

The Support Vector Machines method is used to verify speakers often in combination with the GMM or the HMM method.

The method of Hidden Markov Models (HMM) is also applied to speaker recognition, which has proven itself in problems of automatic speech recognition [25, 26]. In particular, it is assumed that for short phrases of a few seconds duration for a contest-dependent approach, it is best to use phoneme-dependent HMMs, rather than models based on the transition probabilities from frame to frame lasting 10 to 20 ms. The Hidden Markov models method can be used together with the GMM method.

The general conclusion from the analysis of well-known literature is that the templates for authentication (speaker recognition) are formed on the basis of digital processing of the amplitude-frequency spectrum of the user voice signal. At the same time, a more informative parameter of the user voice data is ignored, namely, the phase-frequency spectrum. This could be a promising area.

4 Methods and research results

We will analyze the experimental voice signal of the authentication system user, who pronounced the word “one”. The sampling frequency is 64 kHz and the signal-to-noise ratio is more than 20 dB. The analyzed voice signal is presented in Fig. 1.

Further, as in the well-known modern VASs, we calculate the amplitude-frequency spectrum from the experimental voice signal and perform its analysis. In this case, as indicated above [21], we will focus on the low-frequency region where the attributes of the authentication system user are located, focusing mainly on the pitch frequency and the associated formant frequencies.

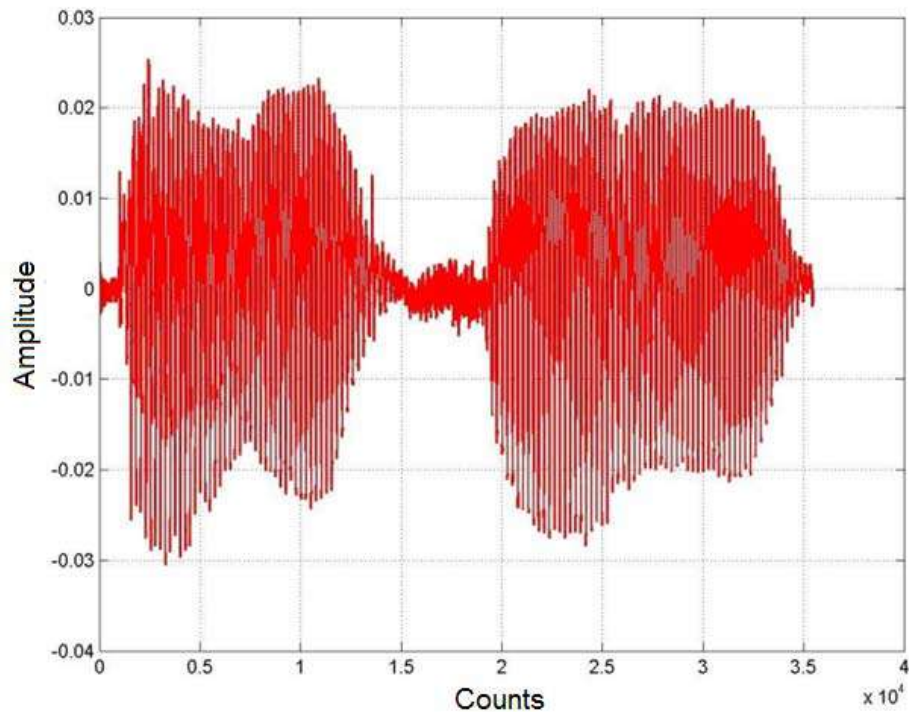


Fig. 1. Voice signal of the word “one”

It is known that the value of the pitch frequency is an individual characteristic of the speaker. It can vary depending on the emotional coloring of speech, but within fairly narrow limits. With parametric coding of speech, it is assumed that the pitch frequency of a person lies in the range of 80-400 Hz, and most formant frequencies are F_0 -fold.

The amplitude-frequency spectrum of the analyzed signal is presented in Fig. 2.

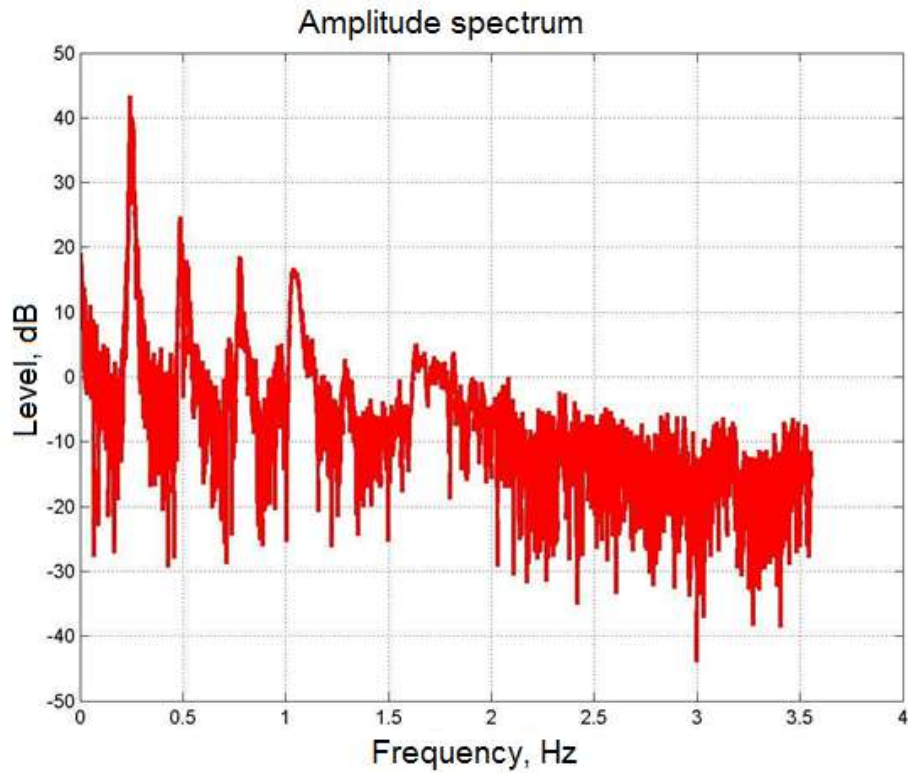


Fig. 2. A short range of voice signal “one”

Spectral analysis of the amplitude-frequency spectrum of the user's real voice signal made it possible to obtain an estimate of the pitch frequency in the region of 243 Hz. In this case, three formant frequencies are clearly pronounced (see Table 1), and the next ones have a low level of intensity.

Table 1. Characteristics of the amplitude spectrum formants

Level, dB	24.6	18.6	14.2
Frequency, Hz	243	486	776

Now we examine the characteristics considered with respect to the phase information of the voice signal of the authentication system user. For this, it is necessary to generate phase data that are not registered for a voice signal.

Therefore, phase data, as a rule, are calculated programmatically and algorithmically. To do this, it is necessary to restore the quadrature (imaginary) component of the voice signal from the registration materials. These procedures are associated with the application of the Hilbert transform [5].

$$y(t) = \frac{1}{2} \int_{-\infty}^{\infty} \frac{x(\tau)}{\pi(t - \tau)} d\tau,$$

where $x(t)$ is the recorded voice signal; $y(t)$ is the quadrature (imaginary) component of the analytical signal; t is an independent variable that has the physical meaning of a unit of time; τ is an integration variable. Next, we can calculate the phase of the voice signal using the following ratio

$$\varphi(t) = \arctg \frac{y(t)}{x(t)}.$$

Unfortunately, the function $\arctg$ gives angle values ranging from $-\pi/2$ to $\pi/2$. To determine the correct value of the phase angle, which for a voice signal varies from 0 to $2 \cdot \pi$, it is necessary to adjust the angle $\varphi(t)$ accordingly, taking into account the signs of the numerator and denominator in the ratio of the function $\arctg$. Otherwise, the phase spectrum will be incorrect. After correction, we obtain a phase angle, which has the form of a sawtooth signal of unknown duration.

As the results of previous studies [27] showed, after the formation of phase data, it is necessary to perform the procedures of their preliminary processing. This is due to some factors, among which we highlight the following:

- the polyharmonic nature of the voice signal, which is processed by the Hilbert transform. The latter is oriented to work with harmonic stationary data;
- incorrect data when the components $y(t)$ or $x(t)$ in the function $\arctg$ are equal to zero;
- for small values of the components $y(t)$ or $x(t)$, the latter can be lost in rounding noises.

These factors lead to the fact that both random errors and anomalous measurements can occur in sawtooth phase signals. This necessitates preliminary processing of both the voice signal and the phase data. Pre-processing can be based on a priori data about the nature of the phase change of the voice signal and will improve the quality of the characteristics formation for both existing and perspective components of templates.

Now we analyze the phase spectrum of the analyzed signal. Fig. 3 shows the phase spectrum of the corrected phase data, which we will consider below.

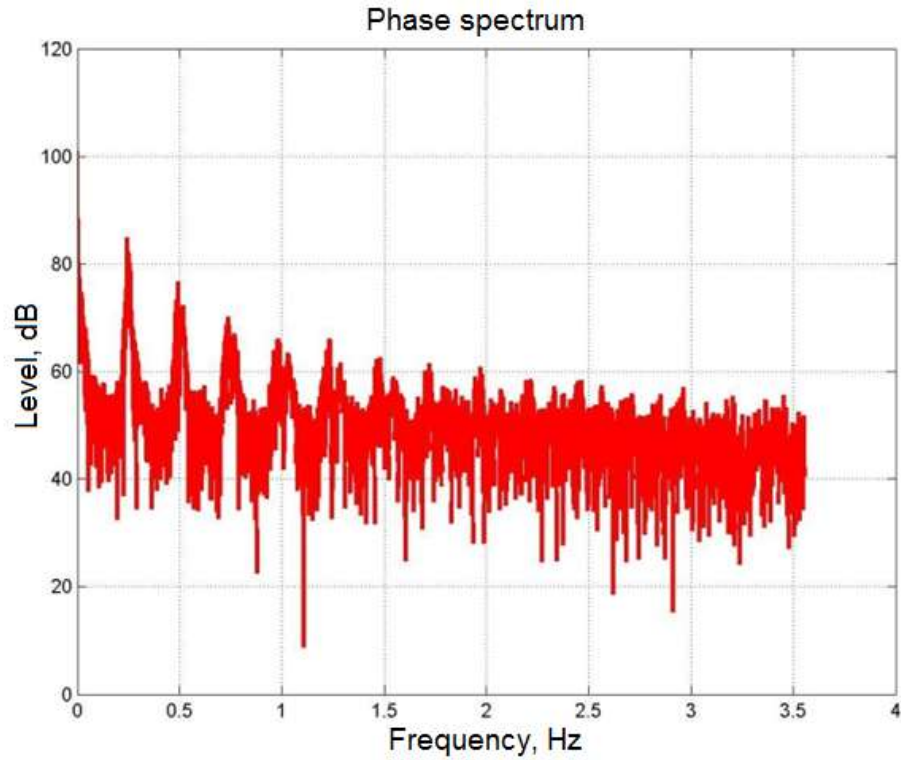


Fig. 3. A short phase spectrum of the voice signal

The results of processing the formant information of the phase spectrum are presented in Table. 2. In this spectrum, six formants can be distinguished, and the seventh and eighth have a slight energy difference. The pitch frequency, as in the amplitude spectrum, is 243 Hz.

The level of spectral density of the selected maxima is several times higher than the level of the maxima of the amplitude spectrum, which greatly simplifies the procedure for their selection. The number of selected formants in the phase spectrum is one and a half times greater. The aforementioned indicates a more informative phase spectrum of the voice signal.

Table 2. Characteristics of phase spectrum formants

Level, dB	84.9	76.7	70.3	65	64	62
Frequency, Hz	243	492	738	990	1217	1450

Another way to obtain an approximate estimate of the pitch frequency can be associated with the calculation of cepstral or mel-frequency cepstral coefficients (MFCC), which, as a rule, are included in the user template as attributes.

As it is known, cepstral coefficients are determined in accordance with the scheme presented in Fig. 4. The following notation is used in this Figure: FFT - Fast Fourier Transform block of a signal; LOG - block logarithmation spectrum; IFFT is the Inverse Fast Fourier Transform block.



Fig. 4. General scheme of cepstral signal analysis

Thus, the cepstral coefficients are the result of applying the inverse Fourier transform to the logarithmic power spectrum. The calculation of these coefficients is carried out on the samples of the signal, the duration of which are several tens of milliseconds. It is proposed to estimate the pitch frequency in each sample after performing the inverse Fourier transform.

In this case, the samples are selected with some overlap. The result of each inverse transformation allows to get an estimate of the maximum frequency in the sample. As a rule, approximately 40 coefficients are calculated, which means that we can get as many estimates of the pitch frequency. Averaging the results, we can form a more accurate estimate of the maximum pitch frequency.

The adequacy and reliability of the hypothesis put forward about a different method for estimating the pitch frequency is feasible in the process of a model experiment. To do this, we will perform digital processing according to the scheme shown in Fig. 4.

The amplitudes and phase data of the voice signal analyzed above were subjected to processing. When estimating the pitch frequency, the processed samples had an overlap coefficient of 0.75, the number of points of the discrete Fourier transform was 1024 and the Hamming smoothing width for the discrete Fourier transform.

The specifics of digital processing was the following: a sample may include vocalized or nonvocalized sounds. As we know, the pitch frequency is estimated from vocalized sounds that were extracted during the threshold processing of the spectral power level of the selected maxima.

As a result of processing the amplitude data, the following estimates were obtained: mathematical expectation is 247.5 Hz; the standard deviation is 15.7 Hz, and for phase data – 250.4 Hz and 17 Hz, respectively.

Thus, the proposed method for estimating the pitch frequency allows to get adequate and reliable results. The indicated method for estimating the pitch frequency can be useful in the presence of errors associated with gross errors. For example, taking the maximum frequencies of the first or second formants as the estimate of the pitch frequency. In this case, preference should be given to the estimate F_0 obtained in the process of calculating the cepstral coefficients.

5 Conclusion

The problem of improving the quality characteristics of voice authentication systems has been discussed in the article. As the main direction of solving this problem, it is proposed to use phase data of the analyzed voice signal in the process of digital processing. The reliability of the solution proposed for this problem and the analysis of the information content of the voice signal phase data are studied in the process of experimental evaluation of the pitch frequency and formant information, which are included in most user templates as required parameters. The cepstral or mel-frequency cepstral coefficients and a number of other attributes are additionally included in the template. The pitch frequency allows to solve the following problems: emotion recognition, gender determination, audio segmentation with multiple voices and speech separation into phrases.

In this regard, the current scientific task of studying new procedures to refine the estimates of the pitch frequency obtained on the basis of the amplitude-frequency spectrum analysis was considered in the work. The estimates were refined based on the use of phase data of the speech signal, as well as the estimates of the pitch frequency in the process of obtaining cepstral coefficients.

Therefore, the scientific novelty of the obtained results lies in the fact that for the first time a technique has been developed and experimental studies have been carried out to form an estimate of the pitch frequency (as well as formant frequencies) based on phase data of the voice signal. In addition, a new method has been developed for estimating the pitch frequency in the process of calculating cepstral coefficients. It can be performed during the analysis of both amplitude and phase data of the studied voice signal.

The results have been obtained in the process of statistical analysis of the simulation results using experimental voice data of the authentication system user.

Phase data of a voice signal allows obtaining adequate and reliable estimates in the process of spectral analysis. However, if there are errors associated with gross errors, for example, taking the maxima of the frequencies of the first or second formants as an estimate of the pitch frequency, preference should be given to the estimate obtained in the process of calculating the cepstral coefficients.

The practical significance of the results is as follows:

- a technique has been developed and features of phase information forming of the studied voice signal have been identified;
- an example of estimating the pitch frequency has shown higher informativity of the phase data, which allows selecting a larger number of formant frequencies;
- the developed method for estimating the pitch frequency eliminates gross errors in the formation of a template of the authentication system user.

It is advisable to carry out further studies in the direction of estimating the quality of the formation of attributes for traditionally used templates (for example, cepstral coefficients, mel-frequency cepstral coefficients, linear prediction coefficients, etc.) taking into account the phase of the voice signal, as well as the development of new procedures for the formation of template elements based on phase data.

References

1. Ramishvili, G.S.: Avtomaticheskoe opoznavanie govoriashogo po golosu (Automatic speaker recognition over voice). Radio i svyaz, Moscow (1981).
2. Beigi, H.: Fundamentals of speaker recognition, Springer, New York (2011).
3. ISO/IEC 2382-37:2012 Information technology – Vocabulary – Part 37: Biometrics (2012).
4. Boll, R.M., Connel, J.Kh., Pankati, Sh., Ratkha, N.K., Senior E.U.: Handbook on biometry. Translated from English by Agapova N. Ye. Tekhnosfera, Moscow (2007).
5. Oppenheim, A.V., Lim, J.S.: The importance of phase in signals. In: Proceeding of the IEEE, Vol. 69(5), pp. 529 - 541 (1981) doi:10.1109/PROC.1981.12022
6. Paliwal, K.: Usefulness of phase in speech processing. In: Proc. IPSJ Spoken Language Processing Workshop, Gifu, Japan, pp. 1-6 (2003).
7. Paliwal, K., Atal, B.: Frequency-related representation of speech. In: Proceedings of the European Conference on Speech Communication and Technology (EUROSPEECH-2003), pp. 65–68 (2003).
8. Borisenko, S.Yu., Vorobiev, V.I., Davidov, A.G.: Sravnenie nekotorykh sposobov analiza fazovykh sootnoshenii mezhdu kvazigarmonicheskimi sostavlyayushimi rechevykh signalov (Comparison of some methods for analyzing phase relationships between the quasi-harmonic components of speech signals). In: Proceedings of the First All-Russian Acoustic Conference, pp. 2-7 (2004).
9. Wu, Z., Kinnunen, T., Chng, E., Li, H., Ambikairajah, E.: A study on spoofing attack in state-of-the-art speaker verification: the telephone speech case. In: Proc. Asia-Pacific Signal Information Processing Association Annual Summit and Conference (APSIPA ASC) (2012).
10. Broeders, Ton: Forensic Speech and Audio Analysis Forensic Linguistics. In: Proceedings 13th INTERPOL Forensic Science Symposium, Lyon, France, 16-19 October 2001, D2, pp. 54-84 (2001) <https://ssrn.com/abstract=2870568>
11. Fergani, B., Davy, M., Houacine, A.: Speaker diarization using one-class support vector machines. Speech Communication, vol.50, pp. 355–365 (2008) doi:10.1016/j.specom.2007.11.006
12. Kuwabara, H., Sagisaka, Y.: Acoustic characteristics of speaker individuality: Control and Conversion. Speech Communication, vol.16, pp. 165-173 (1995) doi:10.1016/0167-6393(94)00053-D
13. Sorokin, V.N., Tsyplikhin, A.I.: Speaker verification using the spectral and time parameters of voice signal. Journal of Communications Technology and Electronics, v.55, N12, pp. 1561-1574 (2010) doi:10.1134/S1064226910120302
14. Matsumoto, H., Hiki, S., Sone, T., Nimura, T.: Multidimensional representation of personal quality of vowels and its acoustical correlates. In: IEEE Trans. AU, vol. AU- 21, pp. 428-436 (1973) doi:10.1109/TAU.1973.1162507
15. Shriberg, E., Ferrer, L., Kajarekar, S., Venkataraman, A., Stolcke, A.: Modeling prosodic feature sequences for speaker recognition. Speech Communication, vol.46, N3–4, pp. 455–472 (2005) doi:10.1016/j.specom.2005.02.018
16. Lavner, Y., Gath, I., Rosenhouse, J.: The effects of acoustic modifications on the identification of familiar voices speaking isolated vowels. Speech Communication vol.30, 9-26 (2000) doi:10.1016/S0167-6393(99)00028-X
17. Takemoto, H., Adachi, S., Kitamura, T., Mokhtari, P., Honda, K.: Acoustic roles of the laryngeal cavity in vocal tract resonance. J. Acoust. Soc. Am., vol.120, pp. 2228–2239 (2006) doi:10.1121/1.2261270

18. Davis, S., Mermelstein, P.: Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. In: IEEE Trans. Acoustics, Speech, Signal Process., vol.28, N4, pp. 357–366 (1980) doi:10.1109/TASSP.1980.1163420
19. Itoh, K.: Perceptual analysis of speaker identity. In: Speech Science and Technology, Saito S. (Ed.), IOS Press, pp. 133–145 (1992).
20. Huang, X., Acero, A., Hon, H.-W.: Spoken Language Processing: a Guide to Theory, Algorithm, and System Development. Prentice-Hall, New Jersey (2001).
21. Lu, X., Dang, J.: An investigation of dependencies between frequency components and speaker characteristics for text-independent speaker identification. Speech Communication, vol.50, N4, pp. 312–322 (2007).
22. Reynolds, D.: Speaker identification and verification using Gaussian mixture speaker models. Speech Communication, vol.17, pp. 91–108 (1995).
23. Reynolds, D., Quatieri, T., Dunn, R.: Speaker verification using adapted gaussian mixture models. Digital Signal Process., vol.10, N1, pp. 19–41 (2000) doi:10.1006/dspr.1999.0361
24. Vapnik, V.N.: Statistical Learning Theory. Wiley, New York (1998).
25. BenZeghiba, M., Boulard, H.: On the combination of speech and speaker recognition. In: Proc. Eighth European Conf. on Speech Communication and Technology (Eurospeech), pp. 1361–1364 (2003).
26. Bimbot, F., Blomberg, M., Boves, L., Genoud, D., Hutter, H.-P., Jaboulet, C., Koolwaaij, J., Lindberg, J., Pierrot, J.-B.: An overview of the CAVE project research activities in speaker verification. Speech Communication, vol. 31, pp. 155–180 (2000).
27. Pastushenko, M., Pastushenko, V., Pastushenko, O.: Specifics of receiving and processing phase information in voice authentication systems. In: 2019 International Scientific-Practical Conference Problems of Infocommunications. Science and Technology (PIC S&T), Kyiv, Ukraine, pp. 621–624 (2019).

Speech Vocal Clustering Methods

Eugene Fedorov¹[0000-0003-3841-7373], Tetyana Utkina¹[0000-0002-6614-4133],
Kostiantyn Rudakov¹[0000-0003-0000-6077], Andriy Lukashenko²[0000-0002-6016-1899],
Ihor Zubko¹[0000-0002-3318-3347], Maryna Chychuzhko¹[0000-0001-5329-7897],
Valentyna Lukashenko¹[0000-0002-6749-9040]

¹ Cherkasy State Technological University, Cherkasy, Shevchenko blvd., 460, 18006, Ukraine
{t.utkina, ckc, k.rudakov, i.zubko, m.chychuzhko}@chdtu.edu.ua,
fedorovee75@ukr.net

² E. O. Paton Electric Welding Institute, Kyiv, Bozhenko str., 11, 03680, Ukraine
ineks-kiev@ukr.net

Abstract. The problem of increasing the clustering efficiency of vocal speech sounds is considered. Centroid and medoid clustering methods that use normalized distances which increases the efficiency of clustering, are proposed. Characteristics and quality criteria based on them are suggested for these methods. These methods have been investigated on TIMIT database and are intended for intelligent biometric identification systems.

Keywords: vocal speech sound, centroid, medoid, clustering, inter-cluster distance, intra-cluster distance, compression ratio.

1 Introduction

Automated biometric identification of a person means decision-making based on acoustic and visual information, which improves the quality of recognition of the person under investigation. Unlike the traditional approach [1], computer biometric identification speeds up and increases the likelihood of recognition, which is especially critical in conditions of limited time.

A special class of biometric identification of a person is formed by methods based on acoustic information [2].

To increase the efficiency of analysis, storage and transmission of a speech signal, knowledge of the structure of vocal speech sound is required, for which the following clustering methods are used.

Traditional methods (without the use of artificial neural networks and metaheuristics) [3-6] search for a solution faster than metaheuristic methods, some methods do not require setting of the number of clusters, but perform only a directed search for a solution and some methods require setting of the number of clusters or additional parameters.

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

Neural network methods [7-10] perform a search for a solution faster than meta-heuristic methods, but perform only a directed search for a solution and require exact setting of the number of clusters.

Metaheuristic methods [11-14] perform a random search for a solution, but the search for a solution is longer than that of other methods, and require exact setting of the number of clusters.

Thus, only some traditional methods do not require an exact setting of the number of clusters, but instead require setting of additional parameters.

The aim of the work is to increase the clustering efficiency of vocal speech sounds by providing an adaptive number of clusters, a minimum number of set parameters and parallel information processing.

To achieve this goal, it is necessary to solve the following tasks:

1. to develop a centroid clustering method;
2. to create medoid clustering methods;
3. to determine the characteristics and quality criterion of the clustering method;
4. to conduct a numerical study of the proposed clustering methods.

2 Problem statement

The problem of increasing the clustering efficiency of vocal speech sound comes down to the problem of finding such vector of parameters Θ^* that satisfies the criterion $F = SSWB + \frac{1}{C} \rightarrow \min_{\Theta}$, where $SSWB$ is the ratio of the sums of average intra-cluster and inter-cluster distances, C is the compression ratio of speech sound.

3 Literature review

Traditional clustering methods include:

1. Partition-based (partitioning-based) or center-based methods

In this case, a cluster is a set of objects, each of which is closer to the center of this cluster than to the center of any other cluster. The center of the cluster is usually a centroid (average value of the coordinates of all objects in the cluster) or a medoid (an object of the cluster, the average difference of which from other objects in the cluster is minimal). These methods consider intersecting areas with a high density of objects as different clusters, for example, K-means [15], PAM (k-medoids) [16], FCM [17], ISODATA [18] methods. The advantage of these methods consists in a quick search for a solution; some of them do not require setting of the number of clusters and have poor sensitivity to noise or random emissions. The disadvantage of these methods consists in the presence of only a directed search for a solution; a problem with clusters of different shapes, sizes and densities; some of them require setting of the number of clusters and are sensitive to noise or random emissions or require setting of additional parameters.

2. Model mixture or distribution-based or model-based methods

In this case, a cluster is described by the probability density function. These methods should be used when clusters are of different sizes, and the set of objects of all clusters can be described by a mixture of distribution densities, for example, EM method [19]. The advantage of these methods consists in a quick search for a solution; determination of the shape and size of each cluster by the corresponding distribution density. The disadvantage of these methods consists in the presence of only a directed search for a solution; requirement for precise setting of the number of clusters; sensitivity to noise or random emissions; the difficulty of choosing the appropriate distribution density.

3. Density-based methods

In this case, a cluster is an area with a high density of objects, which is separated by areas with a low density of objects from areas with a high density of objects. These methods consider intersecting areas with a high density of objects as a single cluster. These methods should be used when noise and random emissions are present, clusters have different shapes and sizes, and the number of clusters is unknown. Methods that define clusters (for example, DBSCAN method [20]) and those that provide visualization of clusters (for example, OPTICS method [21]) are distinguished. The advantage of these methods consists in the absence of a requirement to set the number of clusters; the presence of various shapes and sizes of a cluster; poor sensitivity to noise or random emissions. The disadvantage of these methods consists in the presence of only a directed search for a solution; a problem with clusters of heterogeneous density; slow search for a solution.

4. Hierarchical methods

These methods provide visualization of a cluster tree called a dendrogram. On a dendrogram, pairs of clusters that are joined (in the case of agglomerative methods) or obtained as a result of separation (in the case of divisive methods) are connected by a U-shaped arc, the height of which corresponds to the distance between clusters. By the method of constructing a dendrogram, these methods are divided into agglomerative or bottom up ones – each object is considered as a singleton cluster, after which a stepwise combination of pairs of the closest clusters is performed (for example, centroid communication, Ward, single communication, full communication, group average methods [22, 23]) and divisive or top down ones – all objects are considered as one cluster, and at each step one of the constructed clusters is divided into a couple of clusters (for example, DIANA, DISMEA methods [22, 23]). The advantage of these methods consists in the absence of a requirement to set the number of clusters; visibility; some of them don't strongly tend to clusters of a certain shape and size and have poor sensitivity to noise or random emissions. The disadvantage of these methods consists in the presence of only a directed search for a solution; slow search for a solution; some of them strongly tend to clusters of a certain shape and size, are sensitive to noise or random emissions, tend to discard clusters of high power.

Usually, the methods listed above either require setting of the number of clusters or have a slow search for a solution or require setting of additional parameters, which leads to a decrease in the clustering efficiency.

Therefore, the urgent task is to increase the clustering efficiency of vocal speech sounds by providing an adaptive number of clusters, a minimum number of set parameters and parallel information processing.

4 Method clustering of vocal speech sound

4.1 Centroid clustering of vocal speech sound based on minimum distance method

Centroid clustering of vocal speech sound based on the author's minimum distance method includes the following steps:

1. Set a lot of samples of vocal speech sound $S = \{s_i(n)\}$, $i \in \overline{1, I}, n \in \overline{1, N}$, which are in a single amplitude-time window, where I is the number of samples, N is the length of the sample. Set the number of quantization levels of a speech signal L (for an 8-bit sound sample $L = 256$). Set normalized threshold ε , $0 < \varepsilon < 1$. Set the number of clusters $K = 0$.
2. Calculate normalized squared distance between each pair of sound samples

$$D_{ij} = \frac{\|s_i - s_j\|^2}{NL^2}, i \in \overline{1, I}, j \in \overline{1, I}.$$

3. Calculate the distance between each sound sample and a lot of sound samples

$$d_i = \sum_{j=1}^I D_{ij}, i \in \overline{1, I}.$$

4. Determine the number of sound sample with a minimum distance

$$i^* = \arg \min_i d_i, i \in \overline{1, I}.$$

5. Set a sound sample with a minimum distance as the new cluster center, i.e. $m_{K+1} = s_{i^*}$, set the number of sound samples in the new cluster per unit, i.e. $a_{K+1} = 1$, increase the number of clusters, i.e. $K = K + 1$.
6. Set the sound sample number $i = 1$.
7. If $i^* = i$, then go to step 13.
8. Calculate normalized squared distance between the i^{th} sound sample and cluster centers

$$D_k = \frac{\|s_i - m_k\|^2}{NL^2}, k \in \overline{1, K}.$$

9. Calculate the smallest normalized squared distance between the i^{th} sound sample and cluster centers

$$d^* = \min_k D_k, \quad k \in \overline{1, K}.$$

10. Determine the number of the cluster with a minimum distance

$$k^* = \arg \min_k D_k, \quad k \in \overline{1, K}.$$

11. If $d^* \leq \varepsilon$, then calculate the new center of the k^{th} cluster, i.e. $m_{k^*} = \frac{a_{k^*} m_{k^*} + s_i}{a_{k^*} + 1}$,

increase the number of sound samples in the k^{th} cluster, i.e. $a_{k^*} = a_{k^*} + 1$.

12. If $d^* > \varepsilon$, then set the i^{th} sample as the new cluster center, i.e. $m_{K+1} = s_i$, set the number of sound samples in the new cluster per unit $a_{K+1} = 1$, increase the number of clusters, i.e. $K = K + 1$.

13. If $i < I$, then go to the new sample, i.e. $i = i + 1$, go to step 7.

The method results in an adaptive set of centroids.

4.2 Medoid clustering of vocal speech sound based on minimax distance method

Medoid clustering of vocal speech sound based on minimax distance method, unlike the traditional version, preliminarily determines the number of a sound sample with a minimum distance for a non-random choice of the first cluster center and includes the following steps:

1. Set a lot of samples of vocal speech sound $S = \{s_i(n)\}$, $i \in \overline{1, I}, n \in \overline{1, N}$, which are in a single amplitude-time window, where I is the number of samples, N is the length of the sample. Set the number of clusters $c = 0$.
2. Calculate the squared distance between each pair of sound samples

$$D_{ij} = \|s_i - s_j\|^2, \quad i \in \overline{1, I}, \quad j \in \overline{1, I}.$$

3. Calculate the distance between each sound sample and a lot of sound samples

$$d_i = \sum_{j=1}^I D_{ij}, \quad i \in \overline{1, I}.$$

4. Determine the number of sound sample with a minimum distance

$$i^* = \arg \min_i d_i, \quad i \in \overline{1, I}.$$

5. Set a sound sample with a minimum distance as the new cluster center, i.e. $\phi_{c+1}(n) = s_{i^*}(n)$, $n \in \overline{1, N}$, increase the number of clusters, i.e. $c = c + 1$.
6. Calculate the squared distance between each sound sample and each cluster center

$$D_{ik} = \|s_i - \phi_k\|^2, \quad i \in \overline{1, I}, \quad k \in \overline{1, c}.$$

7. Calculate a minimax squared distance between sound samples and cluster centers

$$d^* = \max_i \min_k D_{ik}, \quad i \in \overline{1, I}, \quad k \in \overline{1, c}.$$

8. Determine the number of sound sample with a minimax squared distance

$$i^* = \arg \max_i \min_k D_{ik}, \quad i \in \overline{1, I}, \quad k \in \overline{1, c}.$$

9. If $c = 1$, then set a sound sample with a minimax squared distance as the new cluster center, i.e. $\phi_{c+1}(n) = s_{i^*}(n)$, $n \in \overline{1, N}$, increase the number of clusters, i.e. $c = c + 1$, go to step 6.
10. Calculate the average squared distance between cluster centers

$$\varepsilon = \frac{2}{c^2 - c} \sum_{i=1}^c \sum_{j=i}^c \|\phi_i - \phi_j\|^2.$$

11. Verification of the termination condition.

If $d^* > \frac{\varepsilon}{2}$, then set a sound sample with a minimax squared distance as the new cluster center, i.e. $\phi_{c+1}(n) = s_{i^*}(n)$, $n \in \overline{1, N}$, increase the number of clusters, i.e. $c = c + 1$, go to step 6.

12. Calculate the distance between each sound sample and each cluster center

$$D_{ik} = \|s_i - \phi_k\|, \quad i \in \overline{1, I}, \quad k \in \overline{1, c}.$$

13. Determine for each sound sample the cluster center closest to it

$$u_i = \arg \min_k D_{ik}, \quad i \in \overline{1, I}, \quad k \in \overline{1, c}.$$

The method results in an adaptive set of medoids.

4.3 Medoid clustering of vocal speech sound based on subtractive clustering method

Medoid clustering of vocal speech sound based on subtractive clustering method, unlike the traditional version, uses the normalization of squared distances for calculat-

ing the potentials of sound samples to specify normalized standard deviations and includes the following steps:

1. Set a lot of samples of vocal speech sound $S = \{s_i(n)\}$, $i \in \overline{1, I}, n \in \overline{1, N}$, which are in a single amplitude-time window, where I is the number of samples, N is the length of the sample. Set the number of quantization levels of a speech signal L (for an 8-bit sound sample $L = 256$). Set the threshold to stop the method δ , $0 < \delta < 1$. Set standard deviations σ_a and σ_b , $\sigma_b > \sigma_a$, $0 < \sigma_a < 1$, $0 < \sigma_b < 1$. Set the number of clusters $c = 0$.
2. Calculate the potential of each sound sample

$$P(i) = \sum_{j=1}^I e^{-\frac{4\|s_i - s_j\|^2}{\sigma_a^2 NL^2}}, \quad i \in \overline{1, I}.$$

3. Determine the number of sound sample with a maximum potential

$$i^* = \arg \max_i P(i), \quad i \in \overline{1, I}.$$

4. Set the potential $P(i^*)$ as the new cluster potential, i.e. $\bar{P}(c+1) = P(i^*)$.
5. Calculate the new potential of each sound sample

$$P(i) = P(i) - \bar{P}(c+1) e^{-\frac{4\|s_i - \phi_{c+1}\|^2}{\sigma_b^2 NL^2}}, \quad i \in \overline{1, I}.$$

6. Verification of the termination condition.

If $\frac{\bar{P}(c+1)}{\bar{P}(1)} \geq \delta$, then set a sound sample with the highest potential as the new cluster center, i.e. $\phi_{c+1}(n) = s_{i^*}(n)$, $n \in \overline{1, N}$, increase the number of clusters, i.e. $c = c + 1$, go to step 3.

7. Calculate the distance between each sound sample and each cluster center

$$D_{ik} = \|s_i - \phi_k\|, \quad i \in \overline{1, I}, \quad k \in \overline{1, c}.$$

8. Determine for each normalized sound sample the cluster center closest to it

$$u_i = \arg \min_k D_{ik}, \quad i \in \overline{1, I}, \quad k \in \overline{1, c}.$$

The method results in an adaptive set of medoids.

4.4 Medoid clustering of vocal speech sound based on minimum average distance method

Medoid clustering of vocal speech sound based on the author's method of minimum average distance includes the following steps:

1. Set a lot of samples of vocal speech sound $S = \{s_i(n)\}$, $i \in \overline{1, I}, n \in \overline{1, N}$, which are in a single amplitude-time window, where I is the number of samples, N is the length of the sample. Set the number of quantization levels of a speech signal L (for an 8-bit sound sample $L = 256$). Set the radius of the neighborhood of sound samples ε , $0 < \varepsilon < 1$. Set the number of clusters $c = 0$. Set a lot of samples of speech sound that haven't fallen into existing clusters, $\tilde{S} = \{s_i(n)\}$.
2. Calculate normalized squared distance between each pair of sound samples

$$D_{ij} = \frac{\|s_i - s_j\|^2}{NL^2}, \quad i \in \overline{1, I}, \quad j \in \overline{1, I}.$$

3. Calculate the distance between each sound sample and a lot of sound samples

$$d_i = \sum_{j=1}^I D_{ij}, \quad i \in \overline{1, I}.$$

4. Determine the neighborhood of each sound sample

$$U_{i,\varepsilon} = \{j \mid D_{ij} < \varepsilon, j \in \overline{1, I}\}, \quad i \in \overline{1, I}.$$

5. Determine the neighborhood of the sound sample with a minimum average distance

$$i^* = \arg \min_{i \in \tilde{S}} \frac{d_i}{|\tilde{S}|},$$

$$U^* = U_{i^*, \varepsilon}.$$

6. Calculate the new distance between each sound sample and a lot of sound samples that haven't fallen into existing clusters

$$d_i = \begin{cases} 0, & i \in U^* \\ d_i - \sum_k D_{ik}, & i \notin U^*, \quad k \in U_{i,\varepsilon} \cap U^*, \quad i \in \tilde{S}. \end{cases}$$

7. Determine the new neighborhood for each sound sample that hasn't fallen into existing clusters.

$$U_{i,\varepsilon} = \begin{cases} \emptyset, & i \in U^* \\ U_{i,\varepsilon} \setminus U^*, & i \notin U^*, i \in \tilde{S} \end{cases}.$$

8. Set a sound sample with a minimum average distance as the new cluster center, i.e.

$$\phi_{c+1}(n) = s_{i^*}(n), \quad n \in \overline{1, N}, \text{ increase the number of clusters, i.e. } c = c + 1.$$

9. Verification of the termination condition.

If $\tilde{S} \setminus U^* \neq \emptyset$, then $\tilde{S} = \tilde{S} \setminus U^*$, go to step 5.

10. Calculate the distance between each sound sample and each cluster center (code-book vector)

$$D_{ik} = \|s_i - \phi_k\|, \quad i \in \overline{1, I}, \quad k \in \overline{1, c}.$$

11. Determine for each sound sample the cluster center closest to it

$$u_i = \arg \min_k D_{ik}, \quad i \in \overline{1, I}, \quad k \in \overline{1, c}.$$

The method results in an adaptive set of medoids.

5 Determination of characteristics and quality criterion for clustering methods of vocal speech sound

To evaluate clustering methods, the following characteristics are used in the work:

1. The sum of average intra-cluster distances:

— in the case of centroid clustering methods

$$SSW = \sum_{k=1}^K \left(\frac{\sum_{i=1}^I \chi_{A_k}(s_i) \|s_i - m_k\|^2}{\sum_{i=1}^I \chi_{A_k}(s_i)} \right),$$

— in the case of medoid clustering methods

$$SSW = \sum_{k=1}^c \left(\frac{\sum_{i=1}^I \chi_{A_k}(s_i) \|s_i - \phi_k\|^2}{\sum_{i=1}^I \chi_{A_k}(s_i)} \right),$$

$$\chi_{A_k}(s_i) = \begin{cases} 1, & s_i \in A_k \\ 0, & s_i \notin A_k \end{cases},$$

where A_k is a cluster.

2. The sum of inter-cluster distances:

— in the case of centroid clustering methods

$$SSB = \sum_{k=1}^K \|m_k - m\|^2 ,$$

$$m = \frac{\sum_{i=1}^I s_i}{I}$$

— in the case of medoid clustering methods

$$SSB = \sum_{k=1}^c \|\phi_k - \phi_1\|^2 .$$

3. The ratio of the sums of average intra-cluster and inter-cluster distances

$$SSWB = \frac{SSW}{SSB} .$$

In addition, to assess the quality of clustering methods, the following characteristics – compression ratio for speech sound – is proposed in the work:

$$C = \frac{I}{c} ,$$

where I is the number of samples of vocal speech sound, c – the number of clusters.

The following criterion for the quality of clustering, which means choosing such a value Θ that minimizes the sum of the inverse of compression ratio and the ratio of the sums of average intra-cluster and inter-cluster distances, is formulated in the paper:

$$F = SSWB + \frac{1}{C} \rightarrow \min_{\Theta} . \quad (1)$$

For centroid clustering of vocal speech sound based on minimum distance method $\Theta = \varepsilon$. For medoid clustering of vocal speech sound based on subtractive clustering method $\Theta = (\sigma_a, \sigma_b, \delta)$. For medoid clustering of vocal speech sound based on minimum average distance method $\Theta = \varepsilon$.

6 Experiments and results

Numerical experiments were carried out using notebook Intel Core i5 8th Gen, MATLAB package, CUDA technology of parallel information processing on the Ge-

Force 920M graphics card with the number of threads in the block $N^s = 1024$. In this case, the most time-consuming (computational complexity $O(I^2)$, where I is the number of samples of vocal speech sound) step 2 of all four proposed clustering methods was parallelized. This made it possible to speed up the search for a solution.

For speech signals containing vocal sounds, the sampling frequency $f_d = 8$ kHz and the number of quantization levels $L = 256$ were set. The length of the sample of vocal speech sound $N = 256$.

For modified method of subtractive clustering, the following fixed parameter values $\delta = 0.01$, $\sigma_b = 1.25\sigma_a$ were set.

The results of a numerical study of the proposed clustering methods for vocal sounds of people from TIMIT database are presented in Table 1.

The result presented in Table 1 shows that the author's method of minimum average distance provides the smallest F value, calculated according to (1). The compression ratio is approximately 7.5, i.e. the number of stored samples is reduced by about 7.5 times.

Based on the experiments, the following conclusions can be drawn.

The author's method of minimum distance should be used only when centroids are required, and not medoids, since it has the largest F and the lowest compression ratio.

Table 1. Results of a numerical study of vector quantization methods

Characteristics of clustering methods	Author's method of minimum distance	Modified minimax distance method	Modified subtractive clustering method	Author's method of minimum average distance
Compression ratio	4.5429	10.7500	7.5834	7.5834
F	0.3157	0.2712	0.2420	0.2400
Number of parameters	1	0	3	1

The modified method of minimax distance performs a coarser adjustment of the number of clusters than the modified method of subtractive clustering and the author's method of minimum average distance, because it does not use parameters. On the other hand, an operator does not need to set any parameters, the values of which are established empirically.

The modified method of subtractive clustering performs finer adjustment of the number of clusters than the modified method of minimax distance, because it uses parameters. It requires more complex setting than the author's method of minimum average distance, because it uses three parameters, the values of which are established empirically.

The author's method of minimum average distance performs finer adjustment of the number of clusters than the modified method of minimax distance, because it uses

parameters. It requires simpler adjustment than the modified method of subtractive clustering, because it uses only one parameter, the value of which is established empirically.

7 Conclusions

The article considers the problem of increasing the clustering efficiency of vocal speech sounds. The following clustering methods are proposed – the author’s method of minimum distance, the author’s method of minimum average distance, the modified method of minimax distance (unlike the traditional version, it preliminarily determines the number of a sound sample with a minimum distance for non-random choice of the first cluster center), the modified method of subtractive clustering (unlike the traditional version, it uses squared distances normalization to set normal standard deviations in calculating the potentials of sound samples). Characteristics and quality criteria based on them are proposed for these methods. The proposed methods allow to increase the clustering efficiency of vocal speech sounds by providing an adaptive number of clusters, a minimum number of specified parameters and parallel information processing. The proposed methods are intended for software implementation on GPU using CUDA technology, which speeds up the process of finding a solution. Software that implements the proposed methods has been developed and researched on TIMIT database. The conducted experiments have confirmed the operability of the developed software and allow to recommend it for use in practice when solving problems of biometric identification of a person. Prospects for further research are to test the proposed methods on a wider set of test databases.

References

1. Singh, N., Khan, R.A., Shree, R.: Applications of Speaker Recognition. *Procedia Engineering*. 38, 3122–3126 (2012).
2. Weychan, R., Marciniak, T., Dabrowski, A.: Implementation aspects of speaker recognition using Python language and Raspberry Pi platform. 2015 *Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA)*. (2015). doi: 10.1109/SPA.2015.7365153
3. Brusco, M.J., Shireman, E., Steinley, D.: A comparison of latent class, K-means, and K-median methods for clustering dichotomous data. *Psychological Methods*. 22, 563–580 (2017). doi: 10.1037/met0000095
4. Lenarczyk, P., Piotrowski, Z.: Speaker recognition system based on GMM multivariate probability distributions built-in a digital watermarking token. *Przegląd Elektrotechniczny*. 89(2), 59–63 (2013).
5. Ram, A., Jalal, S., Jalal, A.S., Kumar, M.: A Density Based Algorithm for Discovering Density Varied Clusters in Large Spatial Databases. *International Journal of Computer Applications*. 3, 1–4 (2010). doi 10.5120/739-1038
6. Krishnamurthy, A., Balakrishnan, S., Xu, M., Singh, A.: Efficient Active Algorithms for Hierarchical Clustering. *CoRR*. abs/1206.4672, (2012).
7. Haykin, S.S.: *Neural networks and learning machines*. Pearson, Delhi (2016).

8. Du, K.-L., Swamy, M.N.S.: *Neural Networks and Statistical Learning*. (2014). doi: 10.1007/978-1-4471-5571-3
9. Sivanandam, S.N., Sumathi, S., Deepa, S.N.: *Introduction to neural networks using MATLAB 6.0*. Tata McGraw-Hill Education, New Delhi (2010).
10. Larin, V.J., Fedorov, E.E.: Combination of PNN network and DTW method for identification of reserved words, used in aviation during radio negotiation. *Radioelectronics and Communications Systems*. 57, 362–368 (2014). doi: 10.3103/s0735272714080044
11. Das, S., Abraham, A., Konar, A.: *Metaheuristic clustering*. Springer Berlin, Berlin (2009).
12. Saemi, B., Hosseinabadi, A.A.R., Kardgar, M., Balas, V.E., Ebadi, H.: Nature Inspired Partitioning Clustering Algorithms: A Review and Analysis. *Soft Computing Applications Advances in Intelligent Systems and Computing*. 96–116 (2017). doi: 10.1007/978-3-319-62524-9_9
13. Brownlee, J.: *Clever algorithms: nature-inspired programming recipes*. Creative Commons, Melbourne (2011).
14. Subbotin, S., Oliinyk, A., Levashenko, V., Zaitseva, E. Diagnostic rule mining based on artificial immune system for a case of uneven distribution of classes in sample. *Communications*. 3. 3-11. (2016).
15. Steinley, D., Brusco, M.J.: Initializing K-means Batch Clustering: A Critical Evaluation of Several Techniques. *Journal of Classification*. 24, 99–121 (2007). doi: 10.1007/s00357-007-0003-0
16. Kaufman, L., Rousseeuw, P.J.: *Finding groups in data: an introduction to cluster analysis*. Wiley-Interscience, Hoboken, NJ (2005).
17. Wu, K.-L., Yu, J., Yang, M.-S.: A novel fuzzy clustering algorithm based on a fuzzy scatter matrix with optimality tests. *Pattern Recognition Letters*. 26, 639–652 (2005). doi: 10.1016/j.patrec.2004.09.016
18. Memarsadeghi, N., Mount, D.M., Netanyahu, N.S., Moigne, J.L.: A Fast Implementation Of The Isodata Clustering Algorithm. *International Journal of Computational Geometry & Applications*. 17, 71–103 (2007). doi: 10.1142/S0218195907002252.
19. Hastie, T., Friedman, J., Tibshirani, R.: *The Elements of statistical learning: data mining, inference, and prediction*. Springer, New York (2017).
20. Borah, B., Bhattacharyya, D.: DDSC : A Density Differentiated Spatial Clustering Technique. *Journal of Computers*. 3, (2008). doi: 10.4304/jcp.3.2.72-79
21. Liu, P., Zhou, D., Wu, N.: VDBSCAN: Varied Density Based Spatial Clustering of Applications with Noise. 2007 International Conference on Service Systems and Service Management. (2007). doi: 10.1109/ICSSSM.2007.4280175
22. Aggarwal, C.C., Reddy, C.K.: *Data Clustering: Algorithms and Applications*. Chapman and Hall/CRC, Boca Raton, FL (2018).
23. Gan, G., Ma, C., Wu, J.: *Data clustering: theory, algorithms, and applications*. Siam, Philadelphia (2007).

Online Video Summarization with the Kohonen SOM in Real Time

Oleksii Gorokhovatskyi^[0000-0003-3477-2132], Oleh Teslenko^[0000-0003-3105-9323], Volodymyr Zatkhei^[0000-0003-4426-7789]

Simon Kuznets Kharkiv National University of Economics, 9a Nauka Ave., Kharkiv 61166
Ukraine
oleksii.gorokhovatskyi@gmail.com

Abstract. In this paper, we propose an algorithm for the creation of an automatic video summary. It is based on the Kohonen's Self-Organizing Map as a method for training and clustering of frame features in online mode. The decision about whether the frame should be in summary depends on the stability of the last sequential clustering results. Three-way matching of images between automatic summary and corresponding user one is proposed and tested. Open Video and SumMe datasets were used for accuracy and performance comparison. It is shown, that the proposed approach can achieve real time summarization combining with its online properties without the requirement to see the whole video. The accuracy (measured by F1 scores) of the proposed approach can compete with batch processing methods. We also compared the performance to the state-of-the-art existing methods of online real time processing.

Keywords: video summarization, keyframe, self-organizing map, clustering, image features, matching, summary

1 Introduction and the related work

In recent years tremendous development of the information, computer and communication technologies made humans impossible to process all available and appearing data themselves. Especially, this is true for video content, e.g. every minute 300-500 hours of video (by different sources) are uploaded to YouTube, additionally, users watch over a billion hours every day.

Video summarization is a process of selecting some specific subset of keyframes (still images) or keyshots (small sequences of frames) from a video stream which preserves the main idea of video [1]. A summary should keep important frames of the initial video that creates the core summarization challenge – the same frames may be important and unimportant at the same time for different users, making in such a way the summary of video to be a quite subjective term.

A lot of researches [2–5] distinguish all summarization methods into two classes: unsupervised [6–8] and supervised [9–12]. A brief description of some known methods is below.

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

Some researchers [13–15] used self-organizing maps or close approaches. Hierarchical Growing Cell Structures (GCS) methods are proposed in [13, 14] as an extension of the Kohonen's Self-Organizing Maps (SOM) that allows to build a flexible structure without knowing the number of classes a priori. A graphical user interface to investigate the construction of two-dimensional SOM is proposed in [15]. Unfortunately, these researches don't contain significant modeling results using at least a dataset of medium size.

One of the most popular summarization approach is VSUMM [7, 16]. The authors proposed the method based on the extraction of color features from video frames and unsupervised classification. Additionally, a new measure to compare automatic and user-defined summaries (Comparison of User Summaries, CUS) was presented.

The variety of other video summarizing methods, based on generative adversarial networks [17], long short-term memory networks [11, 12], attention-based [18, 19] and deep learning approaches [20], using of text annotations both with visual features [21], fuzzy-based incremental clustering [22] were proposed before.

Paper [8] describes the idea to generate video summaries in online mode immediately without seeing the entire video in quasi real time. This method includes the building of the dictionary via group sparse coding for some initial video frames, following by the reconstruction attempt of unseen frames. If reconstruction error is significant enough, a dictionary is updated and the current frame is added to summary.

Ideas proposed in [23] extend dictionary learning with the prediction of interestingness using global camera motion analysis and colorfulness. This approach seems not to be designed for online processing but focused on real time processing mainly.

The implementations of the abovementioned approaches [8] and [23] seem not to be available, as well as test videos for [8].

The contributions of the paper include:

- video summarization method based on the self-organizing maps, that can work in online mode (without seeing the whole video) in real time;
- new three-stage matching of two sets of frames, that includes both keypoint and raw image pixels comparison;
- selection of the keyframes based on Kohonen's SOM clustering stability;
- we performed the quality assessment of the proposed online summary generation method using the dataset, which was created by volunteers in online fashion also.

2 Problem statement

Formally the problem is to create the summary of the video as the set of still keyframes. We want to generate summary frames on the fly in online mode without seeing the whole video. The summarization method should be fast enough to fit real time processing requirement. We want to generate such a summary that is close to the corresponding one created by the human in the same online conditions.

3 Self-organizing maps

Kohonen's self-organizing map (SOM) [24] is one of the most popular neural network unsupervised clustering approaches. This type of network preserves the topology between input and output values and allows to map multidimensional input into low-dimensional (typically 1D or 2D) outputs.

The important property of the Kohonen network is that it is capable of online data processing when input samples come one by one followed by immediate clustering (classification) decision.

Training of Kohonen SOM network implies the update of all weights after the processing of each training sample one by one (online training) and may be done according to stages below. Let n be the quantity of outputs (known a priori) and we denote the quantity of features as m .

1. Initializing of weights for each neuron with the small random weights.
2. Selection for input features vector x .
3. Train the features vector while error for it is bigger than training epsilon (0.0001):
 - 3.1 Find the closest (Best Matching Unit, BMU) g neuron in terms of some distance, e.g. Euclidean.
 - 3.2 Update weights w_{ij} for all n neurons according to (1):

$$w_{ij} = w_{ij} + \alpha(t)\theta(x_i, g, t)(x_i - w_{ij}) \quad (1)$$

where $\alpha(t) = 0.1e^{-0.001t}$ is the learning rate, t – training step, $\theta(x_i, g, t) = e^{-d^2/\sigma(t)}$ – the value of the neighborhood function between best matching unit vector g and current neuron number i at the training step t , where $d = \|r_g - r_{x_i}\|$ is the distance between the coordinates of best matching unit r_g and current one r_{x_i} , $\sigma(t) = n^{2-0.002t}$ is the radius of Gauss function, x – current input vector.

4. Repeat step 2 for all features vector increasing t each time.

One of the most important parameters to be set up for the usage of the SOM is the size of a one-dimensional or two-dimensional map. In this work, we used a one-dimensional map with 20 possible clusters to reach the required real time performance. Our experiments showed that the bigger quantity of clusters requires more training time, the lesser quantity doesn't catch the difference between frames well enough. The successful choice of this parameter allows to balance the quality and the performance of the method being proposed.

The training stage of the SOM continues until all clusters are trained and the quantity of already processed feature vectors is less than 1000.

4 Features and keyframes selection

4.1 Image features

The selection of features is an important step to represent the specific properties of each frame. The best choice from our point of view is such features, which can be calculated quickly and/or in parallel to preserve real-time processing.

We selected the common color features, obtained from floating non-intersection windowing of an image. Averaged R, G and B color components in range [0;1] are used as features from the single window.

We used square windows with size $w=16$ or $w=32$ in experiments and rescaled images accordingly to preserve the same length of features vector. The processing of building a feature vector for the entire image was performed in two parallel independent threads and merged at the end.

4.2 The selection of keyframes

We will define the keyframe as a frame that varies slightly during some quantity T of previously examined frames in a video stream. We define the quantitative measure of the variability in the two-step procedure.

The first step includes the clustering of a frame by SOM. So, we cluster the video stream frame by frame, counting the quantity of frames Q , belonging to the same cluster in a row. When this quantity becomes bigger than the predefined threshold T_0 , we consider this part of a video stream as stable enough and select keyframe candidate from the middle with index $k^* = i - Q/2$, where i – is the number of the frame being processed.

At the next step we compare keyframe candidate k^* with previously added frame in order to avoid the addition of the very similar, belonging to the same cluster but having some frame with another cluster in between accidentally. If frame candidate k^* and previously added frame k_{prev}^* are similar, we replace previous frame with a frame from an updated index: $k^* = (i - Q/2 - k_{prev}^*)/2 + k_{prev}^*$. If they are not similar, candidate frame k^* is confirmed as keyframe.

Measuring the difference between frames k_1 and k_2 is based on the average difference between corresponding feature vectors f^1 and f^2 in LAB color space (c^1 and c^2):

$$d = \frac{1}{m} \sum_{i=1}^m \Delta E_{00}(c_i^1, c_i^2) \quad (2)$$

where $\Delta E_{00}(c_i^1, c_i^2)$ – is the CIEDE2000 difference between two colors [25]. We claim images to be similar if $d \leq 1.5$, where 1.5 is the just noticeable difference (JND) for this metric according to [26].

So, we present the entire scheme of the suggested summarization method in Fig. 1. We denote as $T_0 = 42$ the quantity of frames, required to be classified as the same class in a row with SOM.

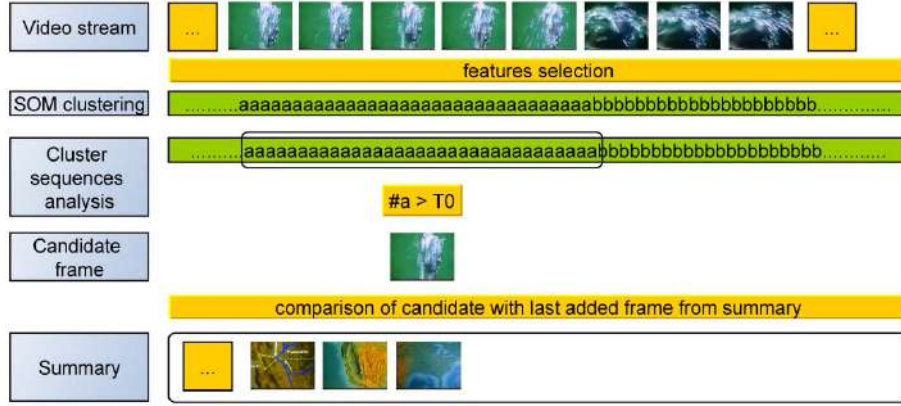


Fig. 1. The whole scheme of the proposed summarization method

5 Image matching

In order to check the quality of summarizing with the suggested approach we need to compare two sets of images, the first one contains images from our automatic summary, the second (ground truth or etalon) one contains images, proposed by humans.

We implemented the match between two sets in three consecutive stages.

The first stage includes the rough estimate of match candidate images with Fast Retina Keypoint (FREAK, [27]) descriptor. FREAK is built on FAST keypoint detector [28] and requires initial threshold t to be set up as a required difference between a central pixel and surrounding to identify central pixel as a corner. The bigger value t leads to the less quantity of keypoint being detected. We found $t = 20$ to be a good default value for experiments.

FREAK descriptors contain 512 bits, we compare them in cascades 128 bits each as proposed in [27]. A comparison of chains requires another threshold F_0 , which allows some differences to appear, as the same bit arrays even for close images is the rare case. If corresponding bit chains in descriptors have more differences compared to the threshold, descriptors are considered being different and comparison stops. In our experiments we found threshold $F_0 = 32$ to be the good choice (25% bit values may differ between descriptors).

We suggest that two images probably match if the overall quantity of matched descriptors between them is greater than the quantity of non-matched.

When the list of matching candidates is ready, we compare each pair of candidates using comparison (2). We perform these comparisons on feature vectors, built on full size images with window size $w = 32$.

The impact of different F_0 values on the matching results is shown in Fig. 2. The first row contains some frames from the user summary. Three rows below contain some frames from an automatic summary and a corresponding F_0 value (24, 32 or 40). As one can see from Fig. 2, we have one successful match for $F_0 = 24$, three correct matches for $F_0 = 32$. In the last case when $F_0 = 40$ we have five matches, one of which is absolutely false (third frame) and one is quite subjective to make the decision about the match (last frame).

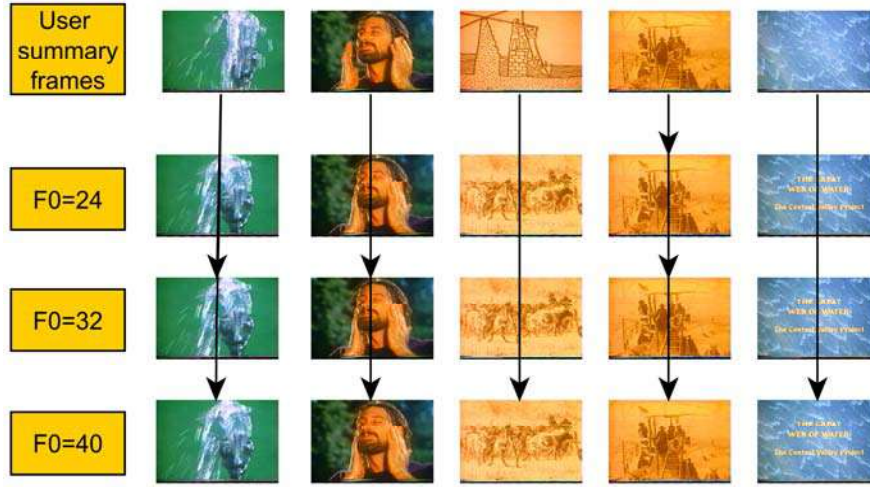


Fig. 2. Frames matching with different F_0 values

The last stage is independent of the previous ones and is used for the frames, for which FREAK descriptor is not effective enough. We process the closest (in the scope of Manhattan distance between frame numbers) frames from automatic and user summaries, which were not matched before. Some frames may be skipped earlier because they contain not enough FREAK descriptors. We analyze them with searching the difference between average R, G and B values of the entire full size image. Images are assumed to be similar if the average difference of all three R, G and B is less than 20.

6 Measuring quality of the automatic summary

We will estimate the quality of the suggested approach with measuring the CUS values like it was suggested in [7] and F1 scores, like it is described in detail in [18] and applied in [11, 12] and others.

Two summary quality CUS_A (accuracy) and CUS_E (error) metrics are proposed in [7]:

$$CUS_A = \frac{n_{mAS}}{n_{US}}, \quad CUS_E = \frac{\bar{n}_{mAS}}{n_{US}}, \quad (3)$$

where n_{mAS} is the quantity of matching keyframes from automatic summary, $\bar{n}_{mAS}$ is the quantity of non-matching keyframes from automatic summary, n_{US} is the total quantity of keyframes from user summary.

$F1$ score is calculated as $F1 = 2PR/(P + R)$, where P is the precision, F is the recall, calculated on the true positives (quantity of matched frames), false positives (quantity of frames which are present if automatic summary but absent in user one), false negatives (quantity of frames which are present in user summary but missing in automatic).

7 Experiments

SumMe [10] and TVSum [6] are the most popular datasets in video summarizing, but not suitable for us. They have importance assigned to each frame as a result of the user summary performed on the entire video. This is convenient to build etalon summaries from this user summaries just solving 0/1 knapsack problem for a fixed length of the summary, typically 15%. In our case we make the decision in online mode, so we don't know the final length naturally. Despite we could extend out an automatic summary to make the required length, criteria of frame selection in batch and online modes are very different.

So, we used dataset, gathered from Open Video dataset by [7, 16], that contains 50 color videos in MPEG-1 format (30 fps, 352x240 pixels) approximately 75 minutes in total.

User summaries were generated by 2 volunteers in online mode. We asked them to select important (from their point of view, no explanations required) frames while looking video-sequences without sound. The modification of previously selected summary frames was not allowed. Users made the decision about the importance of the frame without knowing the content of the entire video.

The specific of human behavior, called chronological bias, is described in [6]. The authors say that humans sometimes claim frames that appear earlier to be more important just because of chronologically, regardless of frame content. This effect leads to the selection of very close frames as important ones. To avoid the influence of chronological bias, we apply the elimination of duplicates in user summaries applying matching of frames, described above.

After we compared automatic summaries using the proposed approach with user summaries, created by volunteers in online mode and cleaned from duplicates, we got such average values: $CUS_A = 0,58$, $CUS_E = 0,38$, $F1 = 0,61$. Feature vector of length 210 was built on the size of images that two times less than original ones.

We have recalculated these scores for OV [16, 29], DT [30] and VSUMM [7] methods using our matching approach, results are presented in Table 1 and they are close to ones, shown in [7]. As one can see, our approach has much bigger CUS_E value compared to the same value for our user summaries, created in online mode ($CUS_E = 0,38$). That means, that the rules for creation of user summaries (online without seeing the entire video or the selection of keyframes after watching entire video) matters.

Table 1. CUS_A , CUS_E and $F1$ scores for the known approaches, suggested matching and known user summaries

Method	CUS_A	CUS_E	$F1$
OV	0,63	0,61	0,56
DT	0,44	0,37	0,48
VSUMM1	0,76	0,47	0,71
VSUMM2	0,62	0,35	0,65
Our (compared to existing summaries)	0,58	0,53	0,57
Our (compared to our user summaries, created in online fashion)	0,58	0,38	0,61

The total duration of the 50 videos [16] from Open Video dataset is about 75 minutes, we built the summary for all videos in 17 minutes. The average length of the summary is 0.3% of the entire video. One-dimensional map with 20 clusters was used.

We used SumMe [10] dataset to evaluate the processing speed. The duration of 25 videos from the SumMe dataset is approximately 70 minutes, time of processing, the quantity of keyframes in summary and length of the features vector are shown in Table 2.

First column contains the name of video and its duration in minutes and seconds. Values in the second column were calculated for the frames with size two times less, than original, in the third column – three times less. Values in the last column were calculated with specific downscaling factor for each video that limits the quantity of features (maximum allowed width of the frame is 200, height is 140). The empty value in the last column means that the result of processing corresponds to one of the values from previous columns.

Values in the last column in Table 2 show the ratio between the best processing time and the total time of a video. The processing of four videos (Air_Force_One, Eiffel Tower, Notre_Dame and Scuba) didn't satisfy real time requirement. The length of the summary decreases significantly with decreasing of feature vector length for one video (Bearpark_climbing).

Table 2. The performance of summarization for videos of SumMe dataset

Video (duration, m:s)	Time / Summary length / features vector length (0.5)	Time / Summary length / features vector length (0.33)	Time / Summary length / features vector length (200x140)	Ratio
Air_Force_One (3:00)	12:10 / 2 / 5841	7:47 / 2 / 2574	4:04 / 2 / 273	1.35
Base jumping (2:39)	2:38 / 19 / 1092	1:51 / 21 / 459	1:28 / 19 / 273	0.55
Bearpark_climbing (2:14)	4:11 / 8 / 2574	2:47 / 8 / 1092	1:46 / 3 / 273	0.79
Bike Polo (1:43)	3:58 / 11 / 2574	2:42 / 11 / 1092	1:42 / 11 / 273	0.99
Bus_in_Rock_Tunnel (2:51)	1:29 / 8 / 627	1:04 / 9 / 273	-	0.37
Car_over_camera (2:26)	1:46 / 4 / 798	1:08 / 4 / 251	1:08 / 4 / 351	0.46
Car_railcrossing (2:49)	6:06 / 8 / 2574	4:03 / 12 / 1092	2:37 / 13 / 273	0.93
Cockpit_Landing (5:02)	9:30 / 19 / 2574	6:52 / 16 / 1092	4:28 / 13 / 273	0.89
Cooking (1:27)	0:09 / 6 / 189	0:06 / 4 / 72	0:25 / 6 / 798	0.07
Eiffel Tower (3:20)	13:57 / 26 / 5841	9:35 / 25 / 2574	5:04 / 23 / 273	1.52
Excavators river crossing (6:29)	3:06 / 37 / 627	2:11 / 34 / 273	-	0.34
Fire Domino (0:55)	0:25 / 7 / 462	0:17 / 5 / 189	-	0.31
Jumps (0:39)	0:12 / 4 / 336	0:08 / 4 / 135	-	0.21
Kids_playing_in_leav es (1:46)	4:01 / 5 / 2574	2:42 / 5 / 1092	1:38 / 6 / 273	0.92
Notre_Dame (3:12)	13:09 / 18 / 5841	8:49 / 18 / 2574	4:40 / 15 / 273	1.46
Paintball (4:16)	7:13 / 5 / 2574	4:43 / 2 / 1092	2:48 / 5 / 273	0.66
Paluma_jump (1:26)	1:06 / 1 / 798	0:46 / 1 / 351	-	0.53
Playing_ball (1:44)	0:48 / 12 / 462	0:34 / 13 / 189	-	0.33

Playing_on_water_slide (1:42)	0:46 / 12 / 462	0:31 / 12 / 189	-	0.30
Saving dolphins (3:43)	6:06 / 2 / 1914	4:08 / 2 / 798	2:44 / 2 / 264	0.74
Scuba (1:14)	2:58 / 10 / 2574	1:59 / 8 / 1092	1:15 / 8 / 273	1.01
St Maarten Landing (1:10)	1:02 / 3 / 1092	0:42 / 2 / 459	0:34 / 3 / 273	0.49
Statue of Liberty (2:36)	1:36 / 7 / 798	1:04 / 7 / 351	-	0.41
Uncut_Evening_Flight (5:23)	10:32 / 50 / 2574	7:46 / 48 / 1092	5:04 / 47 / 273	0.94
Valparaiso_Downhill (2:53)	6:27 / 13 / 2574	4:10 / 13 / 1092	2: 28 / 12 / 273	0.86

We tested also the performance of the proposed online summarization method on the two long movies. First one contains 260222 frames and lasts 3 hrs. 00 min. 53 sec. (10853 seconds in total), the second one has 203882 frames and lasts 2 hrs. 21 min. and 43 sec. (8503 seconds in total).

The online summarization of the first movie took 3800 seconds for 336 features per frame, the length of the summary was 705 frames. Second video required 2209 seconds to build summary containing 585 frames using 252 features per frame.

8 Conclusion

We proposed an approach to generate the video summary in the form of keyframes, which are still images. It is based on the clustering of separate frames in online mode without the analysis of the whole video using Kohonen's self-organized maps. The frame from the video stream is selected as a summary frame if some quantity of frames $T_0 / 2$ after it and before were classified as the same cluster.

The proposed method was tested on Open Video dataset and the performance is tested on SumMe dataset. We showed that the quality is comparable to some batch video summarization methods, and the performance combined with the flexibility in the selection of the quantity of frame features allows achieving real time processing in most cases. The quality of the suggested method is better compared to the user summary created in online mode.

The investigation of the dependency of quality and performance on the size of SOM map may be the topic of future research.

References

1. Otani, M., Nakashima, Y., Rahtu, E., Heikkilä, J.: Rethinking the Evaluation of Video Summaries. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 15-20 June 2019, pp. 7588-7596 (2019) doi: 10.1109/CVPR.2019.00778
2. Panda, R., Das, A., Wu, Z., Ernst, J., Roy-Chowdhury, A.K.: Weakly Supervised Summarization of Web Videos. In: 2017 IEEE International Conference on Computer Vision (ICCV), 22-29 October 2017, pp. 3677-3686 (2017)
3. Pan, G., Zheng, Y., Zhang, R., Han, D., Sun, D., Qu, X.: A bottom-up summarization algorithm for videos in the wild. *EURASIP Journal on Advances in Signal Processing*, vol. 2019, #15 (2019) doi: 10.1186/s13634-019-0611-y
4. Rochan, M., Wang, Y.: Video Summarization by Learning from Unpaired Data. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 15-20 June 2019, pp. 7894-7903 (2019) doi: 10.1109/CVPR.2019.00809
5. Cai, S., Zuo, W., Davis, L.S., Zhang, L.: Weakly-supervised Video Summarization using Variational Encoder-Decoder and Web Prior. In: Ferrari V., Hebert M., Sminchisescu C., Weiss Y. (eds) *Computer Vision – ECCV 2018*. ECCV 2018. Lecture Notes in Computer Science, vol. 11218, pp. 193-210 (2018) doi: 10.1007/978-3-030-01264-9_12
6. Song, Y., Vallmitjana, J., Stent, A., Jaimes A.: TVSum: Summarizing Web Videos Using Titles. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 7-12 June 2015, pp. 5179-5187 (2015) doi: 10.1109/CVPR.2015.7299154
7. De Avila, S. E. F., Lopes, A. P. B., da Luz, A. Jr., Araujo, A.D.L.: Vsumm: A mechanism designed to produce static video summaries and a novel evaluation method. *Pattern Recognition Letters*, vol. 32, no. 1, pp. 56-68 (2011)
8. Zhao, B., Xing E.P.: Quasi Real-Time Summarization for Consumer Videos. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition, 23-28 June 2014, pp. 2513-2520 (2014) doi: 10.1109/CVPR.2014.322
9. Gygli, M., Grabner, H., Van Gool L.: Video summarization by learning submodular mixtures of objectives. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 7-12 June 2015, pp. 3090–3098 (2015) doi: 10.1109/CVPR.2015.7298928
10. Gygli, M., Grabner, H., Riemenschneider, H., Van Gool L.: Creating summaries from user videos. In: Fleet D., Pajdla T., Schiele B., Tuytelaars T. (eds) *Computer Vision – ECCV 2014*. ECCV 2014. Lecture Notes in Computer Science, vol 8695, pp. 505–520 (2014) doi: 10.1007/978-3-319-10584-0_33
11. Zhang, K., Chao, W.L., Sha, F., Grauman, K.: Video Summarization with Long Short-Term Memory. In: Leibe B., Matas J., Sebe N., Welling M. (eds) *Computer Vision – ECCV 2016*. ECCV 2016. Lecture Notes in Computer Science, vol. 9911, pp. 766–782 (2016) doi: 10.1007/978-3-319-46478-7_47
12. Mahasseni, B., Lam, M., Todorovic, S.: Unsupervised video summarization with adversarial LSTM networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 21-26 July 2017, pp. 1–10 (2017) doi: 10.1109/CVPR.2017.318
13. Koprinska, I., Clark, J.: Video Summarization and Browsing Using Growing Cell Structures. In: *International Joint Conference on Neural Networks (IJCNN)*, 25-29 July 2004, pp. 2601-2606 (2004) doi: 10.1109/IJCNN.2004.1381056
14. Koprinska, I., Clark, J., Carrato, S.: VideoGCS - A Clustering-Based System for Video Summarization and Browsing. In: *The Proceedings of the 6th COST 276 Workshop, Thessaloniki, Greece, May 2004*, pp. 34-40 (2014)

15. Baerecke, T., Kijak, E., Nuernberger, A., Detyniecki, M.: Summarizing Video Information Using Self-Organizing Maps . In: 2006 IEEE International Conference on Fuzzy Systems, 16-21 July 2006 (2006) doi: 10.1109/FUZZY.2006.1681764
16. VSUMM (Video SUMMarization). URL: <https://sites.google.com/site/vsummsite/home>. Accessed: 01 March 2020.
17. Zhang, Y., Kampffmeyer, M., Liang, X., Zhang, D., Tan, M., Xing, E.: Dilated Temporal Relational Adversarial Network for Generic Video Summarization. *Multimedia Tools and Applications* , vol. 78, pp. 35237-35261 (2019) doi:10.1007/s11042-019-08175-y
18. Fajtl, J., Sokeh, H.S., Argyriou, V., Monekosso, D., Remagnino, P.: Summarizing Videos with Attention. In: Carneiro G., You S. (eds) *Computer Vision – ACCV 2018 Workshops. ACCV 2018. Lecture Notes in Computer Science*, vol 11367, pp. 39-54 (2019) doi: 10.1007/978-3-030-21074-8_4
19. Ji, Z., Xiong, K., Pang, Y., Li, X.: Video Summarization with Attention-Based Encoder-Decoder Networks. URL: <https://arxiv.org/pdf/1708.09545.pdf>. Accessed: 01 March 2020.
20. Otani, M., Nakashima, Y., Rahtu, E., Heikkilä, J., Yokoya, N.: Video Summarization Using Deep Semantic Features. In: Lai SH., Lepetit V., Nishino K., Sato Y. (eds) *Computer Vision – ACCV 2016. ACCV 2016. Lecture Notes in Computer Science*, vol. 10115, pp. 361-377 (2017) doi: 10.1007/978-3-319-54193-8_23
21. Plummer, B.A., Brown, M., Lazebnik, S.: Enhancing Video Summarization via Vision-Language Embedding. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 21-26 July 2017, pp. 1052-1060 (2017) doi: 10.1109/CVPR.2017.118
22. Pournazari, M., Mahmoudi, F., Moghadam, A. M. E.: Video Summarization Based on a Fuzzy Based Incremental Clustering. *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 4, No. 4, pp. 593-602 (2014)
23. Marvaniya, S., Damoder, M., Gopalakrishnan, V., Iyer, K. N., Soni, K.: Real-time video summarization on mobile. In: 2016 IEEE International Conference on Image Processing (ICIP), 25-28 September 2016, pp. 176-180 (2016) doi: 10.1109/ICIP.2016.7532342
24. Kohonen, T.: *Self-Organizing Maps*. Berlin: Springer-Verlag (1995)
25. Sharma, G., Wu, W., Dalal, E. N.: The CIEDE2000 Color-Difference Formula: Implementation Notes, Supplementary Test Data, and Mathematical Observations. *Color Research & Application*, vol. 30(1), pp. 21-30 (2004) doi:10.1002/col.20070
26. Yang, Y., Ming, J., Yu, N.: Color Image Quality Assessment Based on CIEDE2000. *Advances in Multimedia*, vol. 2012 (ID of paper 273723) (2012) doi: 10.1155/2012/273723
27. Alahi, A., Ortiz, R., Vandergheynst, P.: FREAK: Fast Retina Keypoint. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition, 16-21 June 2012, pp. 510-517 (2012) doi: 10.1109/CVPR.2012.6247715
28. Rosten, E., Drummond, T.: Machine Learning for High-Speed Corner Detection. In: Leonardis A., Bischof H., Pinz A. (eds) *Computer Vision – ECCV 2006. ECCV 2006. Lecture Notes in Computer Science*, vol. 3951, pp. 430-443 (2006) doi: 10.1007/11744023_34
29. DeMenthon, D., Kobla, V., Doermann, D.: Video summarization by curve simplification. In: *Proceedings of the sixth ACM international conference on Multimedia*, September 1998, pp. 211-218 (1998) doi: 10.1145/290747.290773
30. Mundur, P., Rao, Y., Yesha, Y.: Keyframe-based video summarization using Delaunay clustering. *International Journal on Digital Libraries*, vol. 6, no. 1, pp. 219-232 (2006) doi: 10.1007/s00799-005-0129-9

Application of the Fuzzy Clustering Technique for Processing and Analysis of Medical Images

Alexander Stepanenko¹[0000-0002-3668-5828], Andrii Oliinyk²[0000-0002-6740-6078], Ievgen Fedorchenko³[0000-0003-1605-8066]

Department of Software Tools, National University “Zaporizhzhia Polytechnic”,
64 Zhukovskoho str., Zaporizhzhia, Ukraine, 69063
¹alex@zntu.edu.ua, ²olejnikaa@gmail.com,
³evg.fedorchenko@gmail.com

Abstract. The problem of improving the efficiency of medical diagnostics is continuously present in the medical practice. The feasibility of using the mathematical tools for fuzzy logic as applied to the tasks of medical radiological images analysis is under investigation. The paper considers a new approach to the application of the fuzzy clustering technique for analysis and processing of medical images having various structures. The results of practical application of the new approach with the use of real medical images are shown.

Keywords: medical images, image recognition, segmentation, noise, fuzzy clustering, MCCPAO

1 Introduction

The ultrasonic imaging is one of the main and common methods of medical diagnostics [1], which places greater focus on the enhancement of sensitivity and reliability of the procedure for visualization of ultrasonic pulse-echo image. The paper [2] considers the opportunities for the enhancement of sensitivity of detection of the boundaries of low-contrast structures and neutralization of the impact of multiple reflection of ultrasonic pulses hindering the visual analysis of pulse-echo images. However, from the point of view of determination of homogeneity, roughness and regularity, the procedure for segmentation of the ultrasonic images is of the greatest interest. The issues of segmentation of different types of images fall within the domain of computer vision [3], but the presence of modulation speckle noises is the distinctive feature of ultrasound diagnostics. Therefore, the application of any method for segmentation of ultrasonic images should take into account the presence of this factor and come with the use of modulation noise filtering procedure.

Another distinctive feature of ultrasound diagnostics is the fact that the boundaries of the anatomical regions are “blurred” on the ultrasonic images, since the acoustical impedances of adjacent biological structures may be close or continuous variable. This fact necessitates the application of the fuzzy clustering techniques. The fuzzy clustering introduces the notion of fuzzy clusters and the function of membership of

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

the objects thereto, which varies in the interval $[0,1]$ and allows to estimate the degree of membership of the objects to the different classes. The fuzzy clustering is based on the FCM (Fuzzy C-Means) technique, which has many modifications [4]. The FCM technique is based on the use of ideas and mathematical tools for fuzzy logic, and it is widely used as applied to the tasks of medical radiological imaging analysis [5]. However, the information capabilities of the method as applied to the analysis of ultrasonic images have not been previously investigated.

2 Analysis of published data and problem definition

Recently, one of the urgent directions of the development of computer technology in medicine is the processing of digital images: improving image quality, restoring damaged images, recognizing individual elements. Ultrasonic (ultrasound) technique is the most common and allows you to diagnose a large number of dangerous conditions of human health.

To date, many leading research centers and laboratories have developed many image processing algorithms, including medical ones. When processing and analyzing images, the following main stages are distinguished: filtering, pre-processing, segmentation, recognition and diagnostics. The effectiveness of the subsequent stages of image processing directly depends on the results of filtering and pre-processing.

In [2], a clustering algorithm for medical images was proposed — a clustering algorithm using the dendrogram method. The proposed clustering method is based on the statistical pattern recognition method - the dendrogram method. This method consists in the fact that successively a pair of the closest points in the space of images is replaced by one point (center of gravity). The procedure is repeated as many times as needed until several points remain — the centers of the clusters. The dendrogram is a specialized type of diagram. Objects on the dendrogram are arranged in hierarchical levels to emphasize their mutual affinity. However, since this method was not previously used for clustering images, the authors made its substantial modification to solve the problem of distinguishing various image objects. To analyze the effectiveness of the developed algorithm, the authors of the article compare the proposed algorithm with the most common algorithm - the k-means clustering algorithm, which is used in most modern systems for searching images in databases using visual content. The advantages of the proposed method of dendrograms are: less processor time, higher quality clustering. However, along with the advantages of the dendrogram method, in the present work, an obvious drawback of this method is noted - the need to use RAM for storing the basic structure of the algorithm (matrix of compositional similarity coefficients).

In [3], the results of segmentation of ultrasound medical images based on the Canny contouring method are presented. The Canny algorithm consists of four stages: image blur (the dispersion of additive noise in the image is reduced); differentiating the blurry image and calculating the gradient values in the x direction and y direction; not maximum suppression; threshold processing. In fact, this is a set of sequentially applied algorithms. This approach is resistant to noise and usually gives better results

compared to other methods. But since this is just a set of algorithms, the speed of this method is inferior to simpler operators.

The advantage of the proposed algorithm is the minimization of the level of errors, ensuring the finding of most boundaries; maximum accuracy of selection, that is, ensuring a minimum distance between the detected and actual boundaries; the only response in a place where there is only one border.

The Canny algorithm has a two-level threshold for trimming redundant information. This drawback does not allow the Canny algorithm to be used in automatic mode, since the required participation of the user in setting the upper and lower thresholds is required.

In [4], a method for analyzing medical images based on the Laplacian contouring was proposed. 1st and 2nd order Laplacian - the method is focused on sharpening graphic data. The main goal of sharpening is to emphasize small details or to improve those details that were out of focus due to errors or imperfections in the method of digitizing medical data. The method is based on the application of the first or second derivatives.

Proposed by the authors, the method has high speed image processing and relatively low computational complexity. In addition, for this method, parallelization of calculation processes is easily applicable. High speed, ease of calculation, parallelization of processes gives a great advantage in the application of this method for processing data in layers.

The disadvantage of this method is that the value of the relative number of erroneously determined contour pixels of this method decreases nonlinearly. A mathematical model of noise is proposed in [5], and then the result of removing artificially added noise using various filters is compared. The proposed noise model is essentially a multiplicative noise that changes not 1 pixel, but a group of adjacent pixels in the range from 0 to 9 along each axis. The authors compare the developed model with standard multiplicative noise. Real ultrasound image is processed by averaging filter. Multiplicative noise is applied to one copy of the resulting image filtering, and multiplicative noise proposed by the authors is applied to another. Artificially noisy copies of the filtered image are compared with the original image by the standard deviation criterion (MSE).

Based on the table of reduced MSE values for 5 different images, it is concluded that the model proposed by the authors is more accurate than standard multiplicative noise.

It should be noted that the proposed model does not have a normal distribution, by technical means, at the moment, it is impossible to completely suppress it, therefore, the approach to the development of methods based on artificially created images is theoretically poorly justified.

In [6], the authors solve the problems of modern diagnostics based on ultrasound images, referring to their earlier studies: the human factor, specific noise and the general low definition of the edges of objects in the image. It is noted in the materials of the article that, based on research by Shrimadi and others, the authors of the paper draw conclusions about the percentage error of all ultrasonic measurements, setting it from 9% to 36%. To solve this problem, the authors developed a system. The system

presented by the authors, on the basis of the proposed method of anisotropic diffusion, filters the noise of the ultrasound image and then, based on the method of analysis of global and local histograms, segments the ultrasound image. However, the paper does not provide a description of the experimental testing of the system, nor does it provide numerical test results.

In [7], a search is made for points on the edge of an object based on polynomial interpolation of its boundaries on the image that is filtered and filtered by a Gaussian filter. The boundary point is the point at which both interpolation curves intersect. The authors do not define the general principle of searching for the points of the curves used to construct the polynomial, the sizes of the "inter-pixel interval", as well as cases when the selected curves do not intersect, or intersect at several points. From this we can conclude that the points should be selected so that in the selected interval there are no sharp fluctuations in brightness caused by noise. The paper gives numerical examples using Gaussian filters of various sizes on curves with a length of 5 to 20 pixels. But the main disadvantage of this method is the restriction on the exact definition of the upper and lower levels. For a signal outside the edge that is not smooth and also contains some noise, the signal level is calculated as the average of several points. These points should be symmetrical around the reference level (input, initial) of the intersection point. The distance and number of pixels from which the average value is determined depending on the task.

In [8], a method for isolating contours on ultrasound images of the heart was proposed. The proposed processing algorithm begins with smoothing the image with a 9×9 Gaussian filter, and normalizing the result. Then a new image is generated as follows. For each point of the normalized image, the maximum "polar difference" is calculated, which is defined as the difference in brightness of diametrically opposite pixels in a variety of directions (8 neighborhoods are probably used). For the resulting values, the maximum "polar difference" is again calculated, and it is subtracted from the values obtained in the previous step. The result is scaled logarithmically for visual perception. The effectiveness of the algorithm proposed by the authors is determined on the basis of visual comparison with the result of the expert.

The main advantage is the short processing time, since this algorithm does not require re-smoothing the image to obtain a pronounced peak. The disadvantage of this algorithm is that it does not give a single, precisely defined contour of the object, but makes the boundaries of strong brightness differences more noticeable.

Approaches to binarization of images are no less diverse than to cleaning from noise. There are a huge number of histogram conversion algorithms, adaptive local and global threshold filters based on various statistics and parameters. In [9], the authors used a morphological dilatation filter to eliminate "narrow troughs" of brightness values in the image. The authors proceed from the fact that the brightness of the lumen pixels is characterized by a low average intensity and a low standard deviation, the adventitious pixel brightness is characterized by a high average intensity and a low standard deviation, and the remaining pixels must have a high average intensity and a high standard deviation. As a result, the authors obtained a two-dimensional histogram of the image of the carotid artery. For each pixel, we considered a 10×10 environment by which the mean value and standard deviation were calculated. The

mean and standard deviation were normalized from 0 to 1 and grouped into 50 classes in the 0.02 interval. The authors in the classification according to the values of 2 parameters attributed: the arterial lumen pixels by the average brightness value to the first class, and by the standard deviation value to the first to seventh class. Thus, the work performs segmentation by the threshold of statistical quantities. An important advantage of this approach is its efficiency in processing low-contrast images, as well as images for which it is necessary to apply a variable binarization threshold.

The disadvantage of the algorithm is the limited class of application, since the main requirement for its use is the continuity and smoothness of the contour of the desired object.

In [10], a method for distinguishing the contours of objects in low-contrast blurry images was proposed. Generalized method is as follows. Two images of one scene with varying degrees of blur are subtracted from one another. Points obtained by subtracting the "image" at which the sign will change are considered the contours of the objects of the original image. A comparative evaluation was carried out by the standard deviation and peak signal-to-noise ratio. Noise and blurring of test images was carried out in the Matlab program, using a Gaussian filter and white noise (according to the Gaussian distribution).

The advantage of the proposed method is that the method allows to reduce the width of the contour line and increase the accuracy of its localization.

The disadvantage of the proposed method is an increase in processing time and an increase in memory, since the algorithm works with two images (input and blurry). In [11], segmentation of ultrasound medical images based on global and local histograms. Initially, it is proposed to establish 4 brightness ranges, each of which will correspond to certain objects in the image: 1 - cavities filled with water; 2,3 - areas of medium echogenicity (mainly muscle tissue, which can be both hypo- and hyperechoic); solid inclusions (bones, calculi). The image is preliminarily filtered, then its histogram is approximated. It does not specify what kind of low-pass filtering yields an image, nor does it refine the method of approximating the histogram. Then, according to a principle that is not specified, local minima are determined and threshold values are set for the above 4 brightness ranges. Then the image is analyzed by a window of 7x7 or 9x9 pixels, for each of which local histograms are built. Pixels are excluded from the obtained histograms, the number of which in the window does not exceed 5% of all pixels in the window. Then, the number of local maxima in the obtained histograms is calculated. If more than 1 local maximum is found, the point-center of the window is recognized by the contour; if the local maximum is 1, the point-center of the window refers to the type of object to which the local maximum of the histogram of its window belongs.

The main advantage is that the algorithm provides high segmentation accuracy based on the analysis of local histogram statistics, since the obtained limit line has a less complex shape, does not contain gaps, and there are no random "holes".

The disadvantage of this method is the low visual information content, which does not make it possible to widely apply it.

In [12], the question of choosing the dimension of the mask of weight coefficients was solved in the method of sharpening low-contrast images. The authors proposed a

method for sharpening low-contrast two-dimensional images. Differences in brightness of one pixel wide are obtained by processing the image with a sliding window with weights that are calculated separately for each pixel in the image. To analyze the dependences of the accuracy of the sharpening method on the size of the weight mask used in it, an artificial generated, noisy and blurry image was used. As a result of experiments with masks of various sizes, it was concluded in the work that the size of the mask must necessarily coincide with the width of the differences in image intensities. When choosing a larger mask, the selected contours are more rounded than the true ones and cover large areas. The advantage of the proposed approach is that it allows you to work with low-contrast images, sharpen them, while providing a width of the intensity difference of one pixel, and also to highlight the contours that provide the determination of geometric parameters with high accuracy.

The disadvantage is that when you select a smaller mask, the image becomes grainy, erroneous contours appear.

Another algorithmic approach to segmentation was proposed in [13], namely, an algorithm for threshold image processing was proposed that calculates the threshold using a special type of histogram called the “three-peak histogram of the intersection points of the original and the smoothed image”. A smoothed image is obtained from the original using Gaussian filters.

An important advantage of the algorithm is that it allows you to automatically determine the threshold value as a result of approximating a portion of a three-peak histogram, this indicates its adaptive nature. That is, this allows you to apply it not only for threshold processing, but also for highlighting the image outline.

The disadvantage is that the principle of selecting the size of the Gaussian filter mask is not specified, nor is the formula for obtaining the set of intersection points of the original and the smoothed image based on which a three-peak histogram constructed. It is proposed to set the brightness threshold value equal to the brightness corresponding to the maximum value of the average peak of the three-peak histogram.

An adaptive filter based on the anisotropic diffusion method was proposed in [14], which changes the threshold value of the diffusion coefficient and automatically sets the number of iterations depending on the image noise level. The results presented by the authors show the effective operation of the filter for images of the noise level up to 90% of the signal level compared to the anisotropic diffusion filter, which uses a randomly selected threshold value and the number of iterations, which allows optimizing the filter parameters.

The advantage of an adaptive filter based on anisotropic diffusion using an adaptive threshold value and a function of the number of iterations allows you to obtain an image with preserved edges of small objects even at a noise level of 90% of the signal amplitude. The disadvantage of the proposed method is the occurrence of computational complexity in the context of lowering the speed of the algorithm depending on the number of iterations.

In [15], a genetic algorithm (GA) was proposed that automates the process of constructing ultrasound image segmentation methods. GA finds solutions with a predetermined processing result. A parallel genetic algorithm for finding correspondences between the values of texture parameters of images and processing methods suitable

for them is also proposed. The result of the two aforementioned GAs is a set of image processing sequences and the corresponding vectors of texture parameter ranges.

The advantage of the algorithm under consideration is that it allows achieving a segmentation quality close to expert for a given ultrasound image (90% match).

The disadvantage of the parallel genetic algorithm is that it allows with a high degree of certainty (probability ≥ 0.8) to obtain the desired processing result, provided that the parameters of the given image are close enough to the parameters of any image from the training sample.

In [16], a method of secondary processing of various types of images is proposed, which makes it possible to distinguish contours characterized by a maximum value of the brightness difference. When visualizing ultrasound images, the bulk of the information is contained in the differences in brightness. In this case, the psychophysical properties of vision are such that the contrast sensitivity depends on the intensity of the surrounding background and, therefore, the reaction of the eye to a change in illumination is non-linear [1]. The above leads to the fact that when highlighting the boundaries of objects and conducting geometric measurements, a systematic error arises related to the peculiarities of human vision. This error can be eliminated using the secondary processing of ultrasound images.

This method can significantly reduce the systematic errors of the geometric measurements of the heart associated with the psychophysical features of a person's vision.

However, the application of this method in medical practice leads to the fact that for different types of images characterized by texture features and saturation, the efficiency of border extraction can be different. In [17], the results of experimental studies of the influence of features of human vision on the measurement error of the geometric dimensions of objects with a blurred outline are presented. The study was conducted in a group of 11 operators. Each participant was asked to choose the point at which, in their opinion, the maximum value of the brightness difference is observed in several test images. Images were offered to operators in a loop in order to average the error value on the same image. The results of the study confirm the need to create supporting software designed to reduce the subjectivity of expert analysis of ultrasound images.

In [18], segmentation is performed on an ultrasound image of layers of artery tissue in order to measure their thickness. The experiments were carried out on 100 ultrasound longitudinal images of the coronary artery, the results of computer calculations were compared with the measurements of two experts, and no significant differences were found in the computer and expert assessment. Testing of the method was carried out as follows. On each of the images, a region containing the artery wall was manually selected. The selected area was filtered to eliminate noise using the filters described in, then it was binarized, the binarized image was dilated over a 3x3 set, the borders at the bottom of the selected area were eliminated, and a contour was drawn using B-splines. As a result, selected tissue boundaries were obtained. Statistical calculations of deviations in computer and expert assessment of tissue thicknesses in various ultrasound images of the artery are also presented. However, the results of the proposed method were not presented for less "good" cases of ultrasound of the arter-

ies, on which there is no such clear difference in the brightness of the layers of the artery. As a result, selected tissue boundaries were obtained.

In [19], a GA was proposed for obtaining effective threshold detectors for signal identification, which is also applicable in image processing and machine vision problems. GA is used to study compound operators used in the task of identifying objects. The proposed results show that GA is one of the methods for synthesizing complex operators that allow recognition of objects. In order to generate noise / target classification rules and object identification rules, the GA algorithm is also used. The algorithm is used to extract dependent features from a plurality of predefined sets of image features. The algorithm is used to obtain automatic object recognizers. In [20], the statement of the problem of segmentation of medical magnetic resonance images is presented. A hybrid ant algorithm for its solution is proposed, which allows to improve the quality and speed of image processing. In solving the problem, the methodology of swarm intelligence, cluster analysis, the theory of evolutionary calculations, mathematical statistics, computer modeling and programming are used. The authors present the results of experiments obtained on the basis of data from a library of medical MRI images. The optimal values of the parameters determining the behavior and efficiency of the algorithm are established.

Advantages of the algorithm: the comparative simplicity of the actions performed, the possibility of a highly efficient implementation for multiprocessor architectures, guaranteed convergence (although the convergence time is not defined).

The disadvantages are mainly associated with not always optimal settings. Meanwhile, the objects on MRI images have a high degree of complexity and multifactoriality, which imposes high requirements on the reliability and accuracy of their research.

In [21], the implementation and analysis of a mixed algorithm for segmentation of K-medium and ant colonies was performed, and a software system for visualizing and testing the developed algorithm was implemented. The developed algorithm was tested on public benchmarks (Berkeley benchmark was used). The output processed images are obtained, as well as the values of the heuristic coefficients of the developed algorithm.

As can be seen from the results presented by the authors, the proposed method coped quite well with the task. The advantage of the algorithm is the correct separation of parts of the image from the background. However, as can be seen from the presented results, in order to achieve the optimum, the method still needs the correct selection of heuristic parameters. The study of the functions and features of modern specialized systems for the analysis and processing of medical images for various purposes has shown that these systems have several disadvantages. The main drawback seems to be that most of the systems contain only a wide range of image analysis and processing methods available to the researcher, without indicating which method should be applied to achieve the conversion goal. In this regard, the following problems were identified: it is impossible to guarantee the optimal (in the sense of achieving the conversion goal) the choice of a method (or combination of methods) for image processing, since this choice is based only on the knowledge and experience of the user; it is impossible to search all the methods available to the researcher (and

their combinations) to achieve the best processing result, since it will be too time-consuming.

3 Research method and results

The result of the Fuzzy C-Means algorithm is that each pixel in an image is assigned a vector of the function of membership to each class, with reference to which we can make a conclusion about the nature of the object, but subject to the previous neutralization of the influence of the noise factors. Currently, there is no single approach to the task solution. In [22] it is suggested that Wiener filtering is used, and in [23] the use of adaptive anisotropic filtration is proposed. The main problem is the fact that a-priori knowledge of spectral characteristics of both noise and noise-free (i.e. ideal) images is required for optimal filtering. In this regard, the method of selective singular decomposition of automorphic visualization (AVSSDM) [24-26] of noisy images is of interest, because this method does not require a-priori knowledge of spectral characteristics, but only approximate knowledge of the noise correlation interval.

AVSSDM is relating to the image analysis framework methods with the size of the square frame corresponding to the correlation interval. A multidimensional array is formed on the basis of AVSSDM, to which a procedure for decorrelation transformation is applied with the formation of a new array of “own” images. The first “own” image corresponding to the highest singular value is taken as the filtered image.

The experimental findings given in [27] were applicable to the analysis of geophysical fields images. The informational capabilities of the method as applied to the analysis of ultrasonic pulse-echo images remained unclear.

Therefore, for the further improvement of efficiency of medical diagnostics, it is necessary to investigate the information capabilities of AVSSDM and fuzzy FCM clustering as applied to the tasks of medical ultrasonic pulse-echo image segmentation.

The fuzzy clustering is a generalized version of the traditional C-means algorithm overcoming the constraints of group membership of the samplings. The fuzzy clustering uses a fuzzy set in order to specify the degree of membership of the samplings to only one group. The numerical value of the degree of membership lies in the range [28-30], so that the sum of all values of membership of one element (pixel) to all groups is equal to unity.

The fuzzy clustering algorithm is based on the use of procedure for iterative minimization of the objective function

$$J(U, V) = \sum_{i=1}^C \times \sum_{n=1}^N u_{in}^m |x_n - v_i|^2, \quad (1)$$

where $V=\{v_1, \dots, v_c\}$ – cluster centers; $U=[u_{in}]$ – matrix of order $C \times N$, where u_{in} is the membership function i of n entry x_n , $m \in [1, \infty)$ – fuzzification parameter, C – assigned cluster count. The values of the membership functions should satisfy the following constraints:

$$0 \leq u_{in} \leq 1, \quad i = 1, 2, \dots, c, \quad n = 1, 2, \dots, N; \quad (2)$$

$$\sum_{i=1}^c u_{in} = 1, \quad n = 1, 2, \dots, N; \quad (3)$$

$$0 < \sum_{n=1}^N u_{in} < N, \quad i = 1, 2, \dots, c. \quad (4)$$

This iterative algorithm is based on the sequential computation of the following equations:

$$v_i = \frac{1}{\sum_{n=1}^N u_{in}^m} \sum_{n=1}^N u_{in}^m x_{in}, \quad i = 1, 2, \dots, c; \quad (5)$$

$$u_{in} = \frac{\left[\frac{1}{|x_n - v_i|^2} \right]^{1/(m-1)}}{\sum_{j=1}^c \left[\frac{1}{|x_n - v_j|^2} \right]^{1/(m-1)}}; \quad i = 1, 2, \dots, c, \quad n = 1, 2, \dots, N. \quad (6)$$

For each pixel in the ensemble of images, the value of its membership function for each cluster depends on the distance to the corresponding center of the cluster. The fuzzification parameter m reduces the influence of low values of the membership function. The advantages of the Fuzzy C-Means algorithm can include its flexibility, which allows to work with values of membership functions. The disadvantages include the need to assign a-priori cluster count and the theoretical uncertainty of the choice of initial values and fuzzification parameter, resulting in the variant results of clusterization, as well as noise sensitivity.

The Figure 1 a-d shows the ultrasonic pulse-echo image of the child's brain (a) and the first three "own" AVSSDM images (b-d) in "space-time" coordinates. Such echo-impulse images are usually considered as a set of signal paths (image columns). It is apparent that the original image on the Figure 1a is noisy, whereas on the first "own" AVSSDM image (Fig. 1b) the influence of modulation distortion is virtually neutralized. When considering the amplitude-time slices (Fig. 2 a and b) of the 50th impulse line (50th column of the Figures 1a and 1b), one can see removing excessive noise from a signal (Fig. 2b) (rapidly changing amplitudes in peaks and declines (Fig. 2a)).

The Figure 2 c and d shows the amplitude-time slices of the second and third "own" AVSSDM images (Fig. 1c, d) corresponding to the noise components, and their rejection is actually equivalent to the adaptive filtration effect, since no assump-

tions were made about the spectral and statistical characteristics of the modulation noise [31-33].

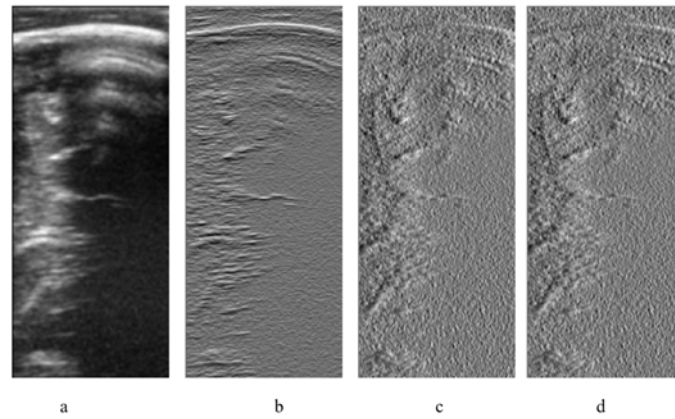


Fig. 1. Ultrasonic image of the brain: a – original image; (b – d) – first three “own” images synthesized based on the application of AVSSDM (horizontal axis is a space axis; vertical axis is a time axis)

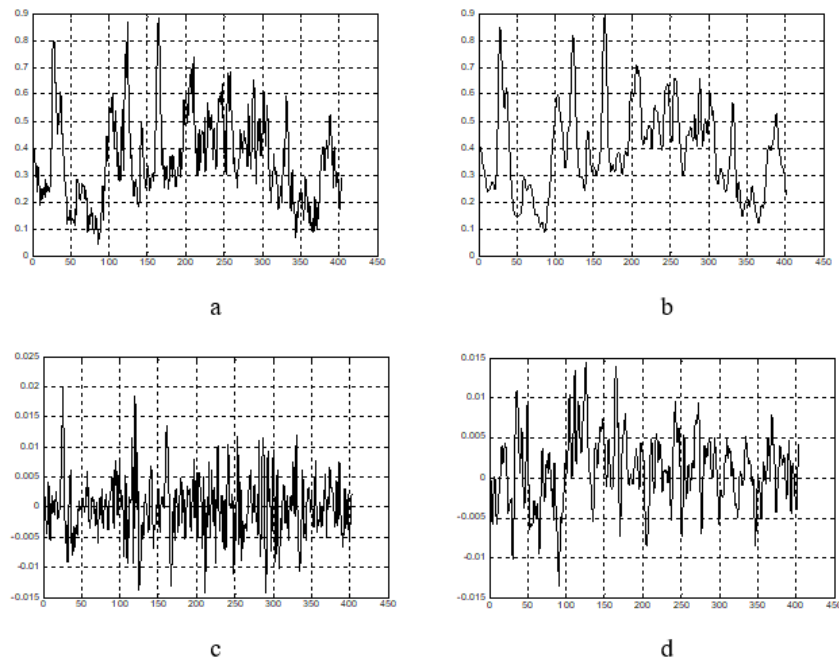


Fig. 2. Amplitude-time slices of the 50th impulse line of the images on the Figure 1 a-g, respectively (horizontal axis is a space axis; vertical axis is a relative amplitude)

The Figure 3 depicts the results of the application of the fuzzy clustering technique for segmentation of Fig. 1a and Fig. 1b into seven classes

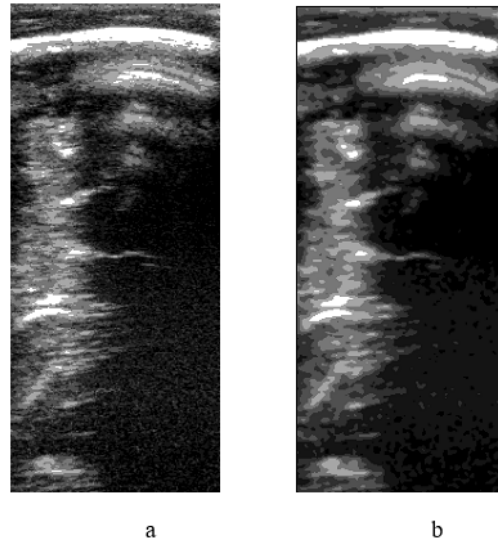


Fig. 3. Results of the application of the fuzzy clustering technique: a – Fig. 1a; b – Fig. 1b

It follows from the comparison of Fig. 1a and Fig. 3a that the application of the fuzzy clustering technique for segmentation of the noisy image actually accomplishes nothing, whereas its application to the filtered image (Fig. 3 b) allows to facilitate the procedure for visual analysis and interpretation of diagnostic results e.g. by varying the number of classes [34 - 39].

4 Conclusion

The fuzzy clustering technique is a fairly simple and convenient method for improving the efficiency of the procedure for visual analysis of medical ultrasonic pulse-echo images. The neutralization of the influence of modulation noises is an indispensable prerequisite for the successful application of the fuzzy clustering technique. The method of selective singular decomposition of automorphic visualization of noisy images is an effective method of modulation noise filtering, because this method does not require a-priori knowledge of spectral and statistical characteristics

Acknowledgement

The work was performed within the framework of the research topic named Methods and Intelligent methods and software for diagnostics and non-destructive quality con-

trol of military and civilian applications (State Registration No. 0119U100360), Department of Software Tools, National University “Zaporizhzhia Polytechnic”.

References

1. Yang, X.: Ultrasound Common Carotid Artery Segmentation Based on Active Shape Model. Hindawi Publishing Corporation Computational and Mathematical Methods in Medicine, vol. 2013, pp. 1–11(2013).
2. Wu, Y.H: Evolutionary Feature Construction for Ultrasound Image Processing and Its Application to Automatic Liver Disease Diagnosis. International Conference on Complex, Intelligent, and Software Intensive Systems, pp. 565-570(2011).
3. Molinari, F.: Completely automated robust edge snapper for carotid ultrasound IMT measurement on a multi-institutional database of 300 images. Medical & Biological Engineering & Computing, vol. 49(8), pp. 935-945.
4. Sassi, O.B.: Improved Spatial Gray Level Dependence Matrices For Texture Analysis. International Journal of Computer Science & Information Technology (IJCSIT), vol.4(6), pp. 209-219 (2012).
5. Pattichis, C., Loizou, C.P.: Despeckle Filtering for Ultrasound Imaging and Video, Volume I: Algorithms and Software, Second. Synthesis Lectures on Algorithms and Software in Engineering, pp. 166 (2015). ISBN 9781627057424.
6. Pattichis C., Loizou, C.P.: Despeckle Filtering for Ultrasound Imaging and Video: Volume II: Selected Applications. Synthesis Lectures on Algorithms and Software in Engineering, pp. 180 (2015). ISBN 9781627058155.
7. Nillesen, M. M.: Automated Assessment of Right Ventricular Volumes and Function Using Three-dimensional Transesophageal Echocardiography. Ultrasound in Medicine and Biology, pp. 1-11 (2015).
8. Harvard Medical School Health Topics A-Z. Harvard Medical School faculty members, (2013).
9. Prema, T.Badiger, S.: Medical Imaging Techniques in Clinical Medicine. International Journal of Medical, Health, Pharmaceutical and Biomedical Engineering, vol. 8(11), pp. 797-800 (2014).
10. Johnston, K. W., Cronenwett, J.L.: Rutherford's Vascular Surgery. Elsevier Saunders, vol. 2, pp. 2688 (2014).
11. Loizou, C. P.: Despeckle filtering software toolbox for ultrasound dimaging of the common carotid artery. Computer methods and programs in biomedicine, vol. 114, pp. 109–124(2014).
12. Silva, J., Campilho, A., Rocha, R.: Automatic detection of the carotid lumen axis in B-mode ultrasound images. Comput Methods Programs Biomed, vol.115(3), pp. 110–118 (2014).
13. Mamou, J.: Quantitative Ultrasound in Soft Tissues. Springer Science, pp. 444 (2013).
14. Cong, Z., Fei, B., Qin, X.: Automatic segmentation of right ventricular ultrasound images using sparse matrix transform and a level set. Phys Med Biol, vol. 58(21), pp. 7609–7624 (2013).
15. Santos, M.A.: A novel automatic algorithm for the segmentation of the lumen of the carotid artery in ultrasound B-mode images. Expert Systems with Applications, vol. 40, pp. 6570–6579 (2013).

16. Lee, Y., Lee, H., Wang, W. Review: The physiological and computational approaches for atherosclerosis treatment. *International Journal of Cardiology*, vol. 167(5), pp. 1664-1676 (2013).
17. El-Khatib, S., Rodzin, S., Skobtcov, Y.: Investigation of Optimal Heuristical Parameters for Mixed ACO-k-means Segmentation Algorithm for MRI Images. *Proc. of the Conf. on Information Technologies in Science, Management, Social Sphere and Medicine*, vol. 51, pp. 216–221 (2016).
18. Kanung, G.K., Singh, N., Dash, J., Mishra, A.: Mammogram Image Segmentation Using Hybridization of Fuzzy Clustering and Optimization Algorithms. *Intelligent Computing, Communication and Devices*. Springer India, vol. 309, pp. 403–413 (2015).
19. Kanungo, G. K., Singh, N., Dash, J., Mishra, A. Mammogram image segmentation using hybridization of fuzzy clustering and optimization algorithms. *Springer India* 2015. *Intelligent Computing, Communication and Devices, Advances in Intelligent Systems and Computing*, pp. 309 (2015).
20. Jebathangam, J., Purushothaman, S.: Analysis of segmentation methods for locating microcalcification in mammogram image. *International Journal of Emerging Technologies in Engineering Research (IJETER)*, vol. 4, no. 9 (2016).
21. Gopi Raju, N., Nageswara Rao, P. A.: Particle swarm optimization methods for image segmentation applied in mammography. *Int. Journal of Engineering Research and Application*, vol. 3, no. 6, pp. 1572–1579 (2013). ISSN: 2248-9622.
22. Oliinyk, A.A., Subbotin, S.A.: The decision tree construction based on a stochastic search for the neuro-fuzzy network synthesis. *Optical Memory and Neural Networks (Information Optics)*, vol. 24(1), 18-27 (2015). doi: 10.1007/s10559-012-9405-z.
23. Fedorchenko, I., Oliinyk, A., Stepanenko, A., Zaiko, T., Shylo, S., Svyrydenko, A.: Development of the modified methods to train a neural network to solve the task on recognition of road users. *Eastern European Journal of Enterprise Technologies*, issue 9/98, pp. 46–55 (2019). DOI: 10.15587/1729-4061.2019.164789
24. Oliinyk A., Fedorchenko I., Stepanenko A., Rud M., Goncharenko D.: Evolutionary method for solving the traveling salesman problem. *Problems of Infocommunications. Science and Technology: 5th International Scientific-Practical Conference PICST2018, Kharkiv, 9–12 October 2018, Kharkiv, Kharkiv National University of Radioelectronics*, pp. 331-339 (2018). Doi: 10.1109/INFOCOMMST.2018.8632033.
25. Kotsur, M., Andrienko, P., Kotsur, I., Blyzniakov, O.: Converter for frequency-current slip-power recovery scheme. *Naukovyi Visnyk Natsionalnoho Hirnychoho Universyte-tu*, vol. 4, pp. 49-54 (2017).
26. Oliinyk, A., Fedorchenko, I., Stepanenko, A., Rud, M., Goncharenko, D.: Combinatorial optimization problems solving based on evolutionary approach. In: 2019 15th International Conference on the Experience of Designing and Application of CAD Systems (CADSM), pp. 41-45 (2019). DOI: 10.1109/CADSM.2019.8779290.
27. Kotsur, M., Kotsur, I., Bezverkhnia, Y., Andrienko, D.: Increasing of thermal reliability of a regulated induction motor in non-standard cycle time conditions. In: 2017 International Conference on Modern Electrical and Energy Systems (MEES), vol. 2018, pp. 88-91(2017). Doi: 10.1109/MEES.2017.8248960.
28. Bruijne, M.: Machine learning approaches in medical image analysis: From detection to diagnosis. *Medical Image Analysis*, vol. 33, pp.94-97 (2016). Doi: 10.1016/j.media.2016.06.032.
29. Stathopoulos, S., Kalamboukis, T.: Applying latent semantic analysis to large-scale medical image databases. *Computerized Medical Imaging and Graphics*, vol. 39, pp.27-34 (2015). Doi: 10.1016/j.compmedimag.2014.05.009.

30. Oliinyk, A., Zaiko, T., Subbotin, S.: Training sample reduction based on association rules for neuro-fuzzy networks synthesis. *Optical Memory and Neural Networks (Information Optics)*, vol. 23(2), pp. 89-95 (2014). doi: 10.3103/S1060992X14020039.
31. Criminisi, A.: Machine learning for medical images analysis. *Medical Image Analysis*, vol. 33, pp.91-93 (2016). Doi: 10.1016/j.media.2016.06.002.
32. Weese, J., Lorenz, C.: Four challenges in medical image analysis from an industrial perspective. *Medical Image Analysis*, vol. 33, pp.44-49 (2016). Doi: 10.1016/j.media.2016.06.023.
33. Khatami, A., Khosravi, A., Nguyen, T., Lim, C., Nahavandi, S.: Medical image analysis using wavelet transform and deep belief networks. *Expert Systems with Applications*, vol. 86, pp.190-198 (2017). Doi: 10.1016/j.eswa.2017.05.073.
34. Oliinyk, A.O., Oliinyk, O.O., Subbotin, S.A.: Software-hardware systems: Agent technologies for feature selection. *Cybernetics and Systems Analysis*, vol. 48(2), 257-267 (2012). doi: 10.1007/s10559-012-9405-z.
35. Fedorchenko, I., Oliinyk, A., Stepanenko, A., Zaiko, T., Korniienko S., Burtsev N.: Development of a genetic algorithm for placing power supply sources in a distributed electric network. *Eastern European Journal of Enterprise Technologies*, issue 5/3 (101), pp. 6–16. (2019). doi: 10.15587/1729-4061.2019.180897.
36. Oliinyk, A., Fedorchenko, I., Stepanenko, A., Katschan, A., Fedorchenko, Y., Kharchenko, A., Goncharenko, D.: Development of genetic methods for predicting the incidence of volumes of emissions of pollutants in air. In: *2nd International Workshop on Informatics and Data-Driven Medicine*, pp. 340-353 (2019). ISSN: 16130073.
37. Shkarupylo, V., Skrupsky, S., Oliinyk, A., Kolpakova, T., 2017. Development of stratified approach to software defined networks simulation. *EasternEuropean Journal of Enterprise Technologies*, Vol. 89, Issue 5/9, pp. 67–73. DOI: 10.15587/1729-4061.2017.110142.
38. A. Oliinyk, T. Zayko and S. Subbotin, "Synthesis of Neuro-Fuzzy Networks on the Basis of Association Rules", *Cybernetics and Systems Analysis*, vol. 50, no. 3, pp. 348-357, 2014. DOI: 10.1007/s10559-014-9623-7.
39. Stepanenko, A., Oliinyk, A., Deineha, L., Zaiko, T., "Development of the method for decomposition of superpositions of unknown pulsed signals using the secondorder adaptive spectral analysis", *EasternEuropean Journal of Enterprise Technologies* 2(9-92), pp. 48-54, 2018. DOI: 10.15587/1729-4061.2018.126578.

Bit Speed Control Method in Compression of Predicted Frames in Video Sequence

Vladimir Barannik¹ [0000-0002-2848-4524], Oleg Kulitsa² [0000-0003-2589-6520], Anna Hahanova³ [0000-0002-1318-7973], Valerii Barannik³ [0000-0003-3516-5553]

¹Kharkiv national Air Force University, Kharkiv, Ukraine

vvbar.off@gmail.com

²National University of Civil Defence of Ukraine, Cherkasy, Ukraine

valera462000@gmail.com

³Kharkiv National University of Radio Electronics, Kharkiv, Ukraine

vika.iv55@gmail.com

Abstract. The method of bit rate control in compressing videodata is proposed, taking into account achieving the maximum image quality. In the process of finding optimal compression parameters it was decided to use the dichotomy method. It allows you to find the optimal values with a given accuracy, without having to resort to a complete search of options. This leads to the reduction of video group frame processing time, which is especially important in case of media content transmission in real time mode. The simplicity of the method is also one of the necessary requirements for control algorithms, when used in devices with limited computing power. A quality factor was selected as the main compression parameter, with which the quantization matrix is formed. At this stage, the greatest loss of information occurs. By choosing the method of block processing in the frame, it is possible to reduce the bit rate to the necessary limit for transmission over the communication channels.

Keywords. predicted frames, bit rate, brightness component, quantization, quality factor, the bit rate control.

1 Introduction

MPEG and H.264 standards require that the video frames processed by separate units - the macroblock. If the controlled encoding parameters are held constant (e.g., the size of the search field motion compensation step quantizer, and the like), the number of code bits of each macroblock will vary from macroblock to macroblock depending on the picture content, leading to variations in the output bit stream rate (measured in bits / frame or bit / s). Typically, encoder with fixed parameters pro-

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

duces more bits for the reference frame in which the rapid motion sealed or removed small parts, for frames with a slow change and without parts he will need fewer bits. Such jumps bit rate can produce a big problem for many transport and storage protocols. For instance, channel with constant speed (channel switching) is unable to transmit data streams at variable speed. Network based on packet switching can support variable bit rate, but the average throughput at any given time is limited by certain factors, depending on the connection speed and congestion.

In these cases, monitoring and bit rate adaptation, produced by the codec for its compliance with the transport speed. Variable coding rate may be adapted for transporting through the channels at a constant rate by the encoder and decoder buffers. However, for such adaptation have to pay amount of buffer memory and decoding delay. The higher the bit rate variation, the greater the required volume of buffer and the decoding delay is longer. In addition, this method cannot cope with arbitrary irregular bit rate, if not prevent unreasonably large amounts of buffers and not to make too long delays. For this it is necessary to implement a feedback mechanism for controlling the encoder output bit rate in order to prevent overflow or underload of buffers.

A common approach to monitoring the bit rate is to change the compression parameters during the encoding time in order to maintain the speed limit in the communication channel. One of the choices of parameters optimal performance compression method is the interval bisection or dichotomy.

2 Construction of basic bit rate control method during compression of predicted frames in a video sequence

2.1 Problem statement

Currently, solutions of the bit rate adapting problems, when transmitting the video data through the channels at a constant rate buffers used in the encoder and decoder. Thus, the higher the bit rate variation, the bigger volume of buffer required and the decoding delay is longer. Furthermore, this method cannot cope with arbitrary irregular bit rate, since the volumes of buffers and delays in the transmission and decoding have limits [4, 5]. Therefore, a bit rate control technique must be implemented using feedback that at lower bit rate also takes into account the quality of the reconstruction and processing of the video sequence frame time.

Since I-frames are the most critical to image quality deterioration, their compression ratio will be low. Therefore, to achieve the required values of the video sequence at the encoder rate in output is proposed to design the control method for the frames of P-type. Since their requirement for image quality is significantly lower, this allows to use techniques of compression with high ratios. And also, the number of the video sequence in order to support larger, which has a significant impact on the overall bit rate of the video stream.

For the encoder output stream, which corresponds to the capacity of the data channel, the control algorithms used bit rate. The desired output speed is achieved by con-

trolling the quantization parameter. Depending on the application, compression may be used as the constant output rate (CBR - constant bit rate), and a variable (VBR - variable bit rate) [4]. To compress with constant rate, depending on the purpose of the algorithm, may vary the size of the buffer for the encoder and decoder. If necessary, a more detailed output rate control algorithm used by encoder multi-pass encoding. Widespread received two-pass coding scheme: the first pass - preliminary analysis step and the establishment of the required encoding parameters; a second passage - coding the previously specified parameters. In this case the selected parameters may be used as a separate block in the frame, and the entire frame.

2.2 Construction of method

During the frame processing occurs in the dimension of its partition blocks $m \times n$. Thus, the whole frame represents a plurality of blocks $\{b_1, b_2, \dots, b_k\}$. These blocks are encoded separately from each other [43]. We denote by $d(t)_i$ and $\sigma(t)_i$ bit costs and standard error for the block b_i current frame t . meaning $d(t)_i$ and $\sigma(t)_i$ depend on making vector Ψ^i , which is used for encoding each block. A vector solution is part of the solution set, which is denoted by the letter Ψ . A vector solution includes k components, i.e. $\Psi^i = \{\psi_1^i, \psi_2^i, \dots, \psi_k^i\}$. Thus, the expression for $d(t)_i$ and $\sigma(t)_i$ It can be represented as follows:

$$\begin{aligned} d(t)_i &= d_i(t, \Psi^i) = d(t)_i(\psi_1^i, \psi_2^i, \dots, \psi_k^i); \\ \sigma(t)_i &= \sigma_i(t, \Psi^i) = \sigma(t)_i(\psi_1^i, \psi_2^i, \dots, \psi_k^i). \end{aligned} \quad (1)$$

The article [1] proposed an algorithm of processing of blocks in P-frames. In accordance with what is proposed blocks brightness component processed by two different methods: using the DCT as an I-type or using the DICM as P-type, and the color difference components will be processed only by using a DICM previous frame.

When compression decisions of P-frames, vector for bit rate control method consists of two components: informativeness measure unit and the parameter quality $\Psi^i = \{\psi_1^i, \psi_2^i\}$. Accordingly, for each i -th block of the bit rate and the mean square error will be a function of two variables:

$$\begin{aligned} d(t)_i &= d_i(t, \Psi^i) = d(t)_i(\psi_1^i, \psi_2^i, \dots, \psi_k^i); \\ \sigma(t)_i &= \sigma_i(t, \Psi^i) = \sigma(t)_i(\psi_1^i, \psi_2^i, \dots, \psi_k^i). \end{aligned} \quad (2)$$

The standard error for the entire frame is given by:

$$\sigma(t, \Psi) = \sum_{i=1}^k \sigma(t, \Psi^i).$$

Similarly, the bit cost per frame are defined as:

$$d(t, \Psi) = \sum_{i=1}^k d(t, \Psi^i).$$

To optimize the compression parameters necessary to find such values of the vector solutions that will meet the following conditions:

$$\begin{cases} \sigma(t, \Psi^*) = \min_{\Psi^i \in \Psi} \sigma(t, \Psi^i); \\ d(t, \Psi^*) \leq d_{\text{req}}. \end{cases} \quad (3)$$

Where d_{req} - the required bit cost per frame; Ψ^* - optimal vector solutions.

In processing blocks as a control parameter is proposed to use only quality factor, which is used for generating the quantization matrix. Since at this stage conducts a correction of component transformant under psychovisual characteristics of visual perception. Informativeness measure unit determines by transformant forming method and its impact on the bit rate is not significant.

In the formation of matrices of quantization in the JPEG standard [7] used two approaches. One is that the standard JPEG recommended to include two quantization tables, one for the brightness (Table 1), or the second for color.

Table 1. The base matrix quantization brightness component JPEG standard

16	11	10	16	24	40	51	61
12	12	14	19	26	58	60	55
14	13	16	24	40	57	69	56
14	17	22	29	51	87	80	62
18	22	37	56	68	109	103	77
24	35	55	64	81	104	113	92
49	64	78	87	103	121	120	101
72	92	95	98	112	100	103	99

The second approach is to compute the quantization table values in real time. In this case, for each element of the transformed image matrix there is a corresponding element of the quantization matrix.

The resulting matrix is obtained by dividing each element of the matrix element corresponding to transformant quantization matrices and subsequent rounding off the result to the nearest integer. As a rule, the value of the quantization matrix elements are growing in the direction from left to right and top to bottom.

Quantization matrix calculation is as follows: the quality factor is given by one value (Quality Factor - QF) is usually in the range from 1 to 25, and calculates the formula of the matrix of values:

$$q(t)_{i,j} = 1 + (1 + i + j) \cdot QF. \quad (4)$$

The quality factor sets the interval between adjacent quantization levels matrix located at its diagonals.

This paper proposes to manage not only the number of bits to the compressed block, but also its visual quality. It is necessary to find the value of the quality factor QF_{opt} , in which the mean square error is minimal $\sigma(t, \Psi^*) = \min_{\Psi^i \in \Psi} \sigma(t, \Psi^i)$, given the fact that the speed, and therefore the bit cost will not exceed the desired value $d(t, \Psi^*) \leq d_{req}$.

To find the optimum ratio QF using known half segment division method (dichotomy) which, relative to other methods is faster, easier and provides the desired accuracy (ε).

QF denote as a parameter β and correspondingly QF_{opt} as a parameter β_{opt} . The method of implementing the search for the problem can be illustrated by the following sequence of steps:

Prepare stage

To determine such values $\beta_1 = \min$ and $\beta_2 = \max$ that certainly is true $d(t, \Psi_{\beta_2}) < d_{req} < d(t, \Psi_{\beta_1})$.

Iteration

Find the average value β in accordance with the method of the dichotomy in the interval $[\beta_1; \beta_2]$: $\beta = \left\lfloor \frac{\beta_1 + \beta_2}{2} \right\rfloor$ to move closer to β_{opt} .

To check compliance $\beta \approx \beta_{opt}$ You need to calculate: $d(t, \Psi_\beta) \cdot \sigma(t, \Psi_\beta)$.

If $d(t, \Psi_\beta) > d_{req}$ and $\sigma(t, \Psi_\beta) \leq \sigma_{req}$ the lower limit of the interval is displaced ($\beta_1 := \beta$);

If $d(t, \Psi_\beta) > d_{req}$ and $\sigma(t, \Psi_\beta) > \sigma_{req}$ then these parameters there is no solution and the control unit initiates a change in the value QF or ΔD .

If $d(t, \Psi_\beta) \leq d_{req}$ the upper limit of the interval is displaced ($\beta_2 := \beta$).

The verification of algorithm complete condition.

If $|\beta_1 - \beta_2| < \varepsilon$, the search for solutions is completed and the result is optimal $\Psi^* = \Psi_\beta$.

If not, it proceeds to the next iteration.

A block diagram of the implementation of the method for finding the optimal parameter control bit rate is shown in Fig.1.

Consider an example of operation of this method.

Assign values $\beta_1 = QF_{\min}$ and $\beta_2 = QF_{\max}$ such that the conditions $d(t, \Psi_{\beta_2}) < d_{\text{req}} < d(t, \Psi_{\beta_1})$.

We spend the first iteration. We find $\beta^1 = \left\lfloor \frac{\beta_1 + \beta_2}{2} \right\rfloor$ and calculate values for a given $d(t, \Psi_{\beta})$ and $\sigma(t, \Psi_{\beta})$. We make a comparison $d(t, \Psi_{\beta}) > d_{\text{req}}$. If the condition is not satisfied, the bit rate should be increased to improve the image quality.

In this case, assign $\beta_2 := \beta$ and then examines the range of $[\beta_1; \beta]$. If the condition $d(t, \Psi_{\beta}) > d_{\text{req}}$ is performed, to review by the standard error: $\sigma(t, \Psi_{\beta}) \leq \sigma_{\text{req}}$.

If the error is less than the required $\sigma(t, \Psi_{\beta}) \leq \sigma_{\text{req}}$ assign $\beta_1 := \beta$ and further finding the optimum value will be in the range $[\beta; \beta_2]$. Thus, after the first iteration in the example shift will be performed: $\beta_{\max} \rightarrow \beta$.

In case if the speed and the error exceeds the required value any of the conditions does not satisfy: $d(t, \Psi_{\beta}) > d_{\text{req}}$ or $\sigma(t, \Psi_{\beta}) > \sigma_{\text{req}}$, We see that the optimization of the data compression method is not possible and necessary changes of the original parameters. In this case it is proposed to increase the quality factor in the quantization in increments of 1, if the bit cost will not be reduced to the desired level, the next step will be to increase ΔD threshold.

Thus, the control unit selects the quality factor for each P-frame. Also, when deciding the control unit comprises a memory in which stored optimum parameters QF and ΔD , which further reduces the time of the decision and selection of the desired values in the quantization.

The last step is carried out verification of optimality. We check the condition $|\beta_1 - \beta_2| < \varepsilon$, where the parameter ε - shows the set accuracy of calculations. If the required accuracy was achieved, the optimal value is considered $\sigma(t, \Psi_{\beta})$. Solution found. Otherwise the search will continue for implementation of the next iteration.

Thus, the elaborated method does not require complete (or nearly complete) sorting a plurality of solutions such as, for example, by dynamic programming. This reduces processing time and frame transmission, which is necessary when working in real time.

The simplicity of the method also makes it possible to reduce the load on the computing device encoder, so there is the possibility of its use in systems in limited computing capabilities.

In view of this method represent the overall control algorithm to work under compression P frames.

At the initial stage of the control set initial parameters of blocks compression for brightness and chrominance components, respectively: thresholds $\Delta D(Y)_{in}$, $\Delta D(C_a)_{in}$ and quality of factors $QF(Y)_{in}$, $QF(C_a)_{in}$, then assesses the total bit rate $V(t)_{comp}$ and the mean square error $\sigma(t)$ video frame being processed.

Since the brightness component when restoring the image information has a great load than color-difference components C_r, C_b , the threshold value of [3,4] in the processing of the brightness component blocks $\Delta D(Y)$ We will choose less than the processing of the color difference components $\Delta D(Y) < \Delta D(C_a)$. $\Delta D(C_r) = \Delta D(C_b) = \Delta D(C_a)$, where C_a - represents one of the components color difference planes: $C_a = C_r$ or $C_a = C_b$.

Similarly, the quality factors for brightness and chrominance components will also be chosen with different values: $QF(Y) < QF(C_a)$.

The initial parameters selected for reasons best image quality, but they are set in the ranges:

$$\Delta D_{min} \leq \Delta D(Y)_{in}, \Delta D(C_a)_{in} \leq \Delta D_{max}$$

$$QF_{min} \leq QF(Y)_{in}, QF(C_a)_{in} \leq QF_{max}.$$

If the standard error $\sigma(t)$ will exceed the required value $\sigma(t) \geq \sigma_{req}$, the quality of the reconstructed image is less than a predetermined, carry out reduction value of quality factor brightness component $QF(Y)$.

After that assesses the total bit rate $V(t)_{comp}$ and the mean square error $\sigma(t)$ stream [5]. Further, if necessary, $QF(Y)$ decreases again, until a minimum value QF_{min} . If $QF(Y) = QF_{min}$ and the condition for the image quality is not yet achieved $\sigma(t) \geq \sigma_{req}$, the decision to change the next parameter - $QF(C_a)_{in}$ for the components of color difference. Reducing procedure $QF(C_a)_{in}$ repeated similarly.

If the parameters of the quality factors achieved the minimum values $QF(Y) = QF_{min}$ and $QF(C_a) = QF_{min}$ and root mean square error $\sigma(t)$ still higher than the desired value $\sigma(t) \geq \sigma_{req}$ it occurs threshold increase first for brightness $\Delta D(Y)_{in}$ and then for the color difference components $\sigma(t) \geq \sigma_{req}$. This increases the I-type blocks in the frame, which in turn will improve the quality of the compressed P frame. At each change thresholds also assesses total bit rate $V(t)_{comp}$ and the mean square error $\sigma(t)$.

Thresholds $\Delta D(Y)$ and $\Delta D(C_a)$ can only increase to a predetermined maximum values: $\Delta D(Y) \leq \Delta D_{\max}$, $\Delta D(C_a) \leq \Delta D_{\max}$.

Selection is carried out as long as the predetermined image quality at a bit rate desired value is reached: $\sigma(t) < \sigma_{\text{req}}$.

If too much compression possible for this parameter is not allowed to reach a given quality: $\sigma(t) \geq \sigma_{\text{req}}$ In accordance with the method developed by the management decision to change the color subsampling format (for example format of 4: 2: 2 to 4: 4: 4).

This approach will greatly improve the quality of the reconstructed image due to the transmission of the complete information from all color planes model YC_rC_b , however this can lead to a sharp increase in the final compressed bit-rate frame.

Consider the case when the total bit rate $V(t)_{\text{comp}}$ higher than necessary: $V(t)_{\text{comp}} \geq V_{\text{req}}$.

Here, the characteristics of the intensity of the control mechanism of the video stream are utilized in the following order:

1. increase $QF(C_a)$ for the color difference components;
2. increase $QF(Y)$ for the brightness component;
3. decrease $\Delta D(C_a)$ for the color difference components;
4. decrease $\Delta D(Y)$ for brightness component.

The values of the quality factors $QF(Y)$, $QF(C_a)$ and thresholds $\Delta D(Y)$, $\Delta D(C_a)$ can only increase or decrease to predetermined values:

$$QF(C_a) \leq QF_{\max} \cdot \Delta D(Y) \geq \Delta D_{\min} ; \Delta D(C_a) \geq \Delta D_{\min} \cdot \Delta D(C_a) \geq \Delta D_{\min} .$$

If too much compression possible for this parameter is not allowed to reduce the bit rate to a desired value: $V(t)_{\text{comp}} \geq V_{\text{req}}$ In accordance with the control method is decided to change the color subsampling format (for example format of 4: 2: 2 to 4: 1: 1), which will greatly reduce the amount of data per color-difference components.

Evaluation of the final bit rate $V(t)_{\text{comp}}$ it is performed after change of each parameter, as long as it is less than the desired value $V(t)_{\text{comp}} < V_{\text{req}}$. If, after all the changes did not reach the set speed during the processing time that has been allotted to one frame compression, it may be decided to permit the passage of the current P-frame.

2.3 Results and discussion

The designed total bit rate control method allowing to make an adjustment of the intensity of the video stream in accordance with the parameters of the telecommunications network.

For comparison, the developed method bit rate control existing source video stream primarily been compressed codec standard with characteristics of transmitted streams: a frame size of 1280x720 pixels; video codec H.264 / AVC variable bit rate (VBR); frame rate - 30 frames / s.

From the data obtained, were taken averaged maximum values of bit rate, as they play a crucial during accelerations and loss (the average maximum value of the intensity of the video stream for MPEG was 1,153 Mbit / s, the average value at which no overload is defined as required $V_{tr} = 733, 6$ Kbyte / s). Similar operations are also performed using the developed method and coding control. As a result, the dependencies bit rate were calculated ($V(t)_{comp}$) And peak signal / noise ratio (PSNR) of the threshold (ΔD) For various values of the quality factor (QF), when subsampling formats 4: 1: 1, 4: 2: 2 and 4: 4: 4 (Figure 1-4).

As the minimum limit for the peak signal / noise ratio is taken equal to the standard value of 26 dB MPEG.

From the results obtained it can be concluded that the best treatment for the P-frame is a color sub-sampling format 4: 2: 2, wherein the quality factor (quantization step) = 0 and the decision threshold = 33 out of 255 possible is provided:

- gain in the degree of reduction of the bit rate for a given parameter PSNR on average 36.36% with respect to the MPEG standard;
- predetermined level of bit rate control process for its higher quality indicators determined parameter PSNR, at the level of 30.18 dB, which is on average 16.07% better MPEG standard.

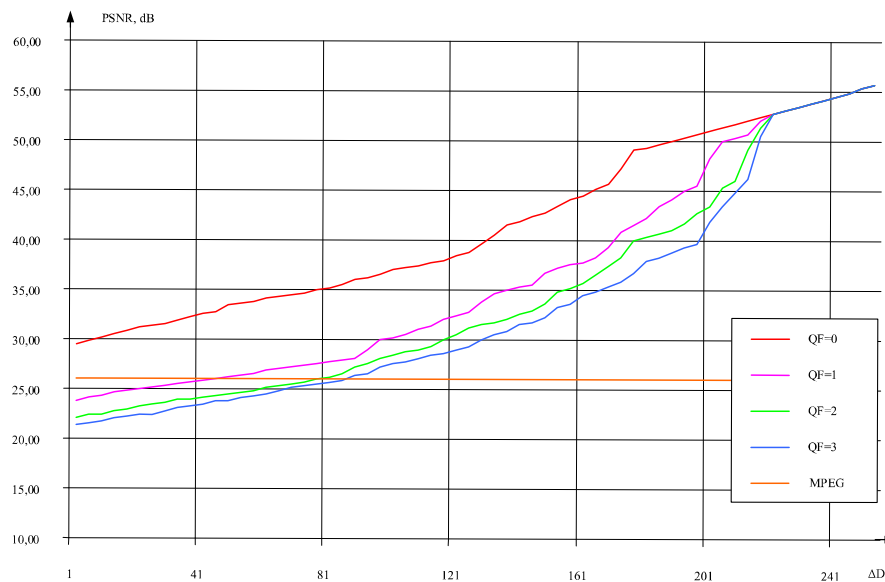


Fig. 1. The dependence of the peak signal / noise ratio (PSNR) from the threshold (ΔD) for different values of the quality factor (QF)

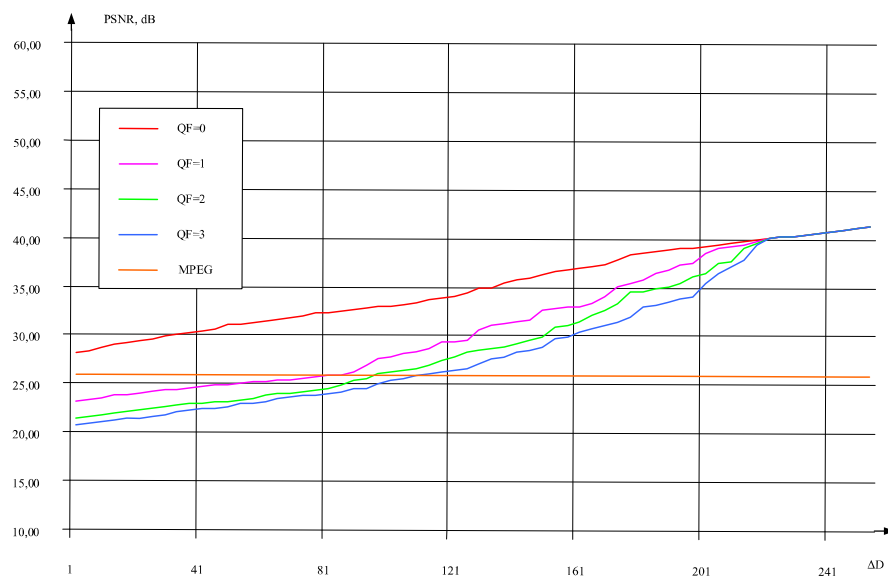


Fig. 2. The dependence of the peak signal / noise ratio (PSNR) from the threshold (ΔD) for different values of the quality factor (QF)

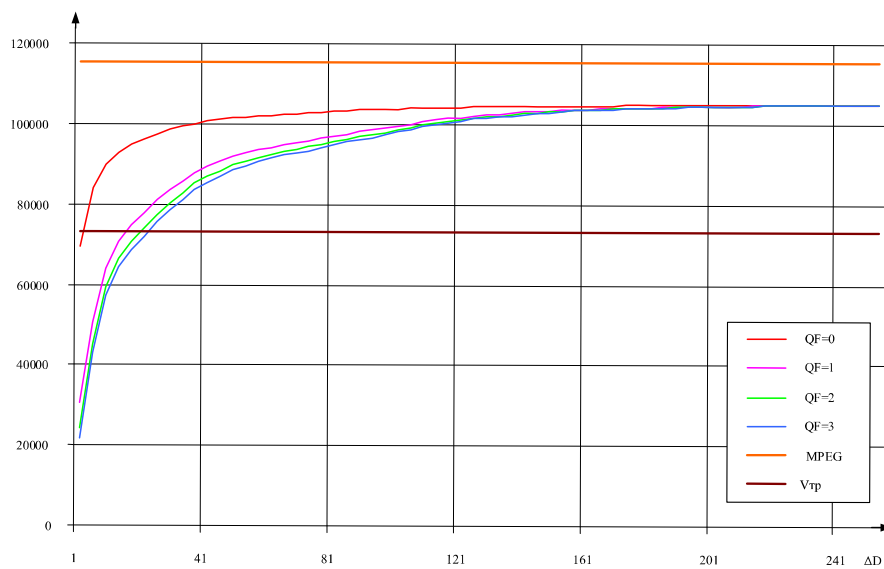


Fig. 3. The dependence of bit rate from the threshold (ΔD) for different values of the quality factor (QF)

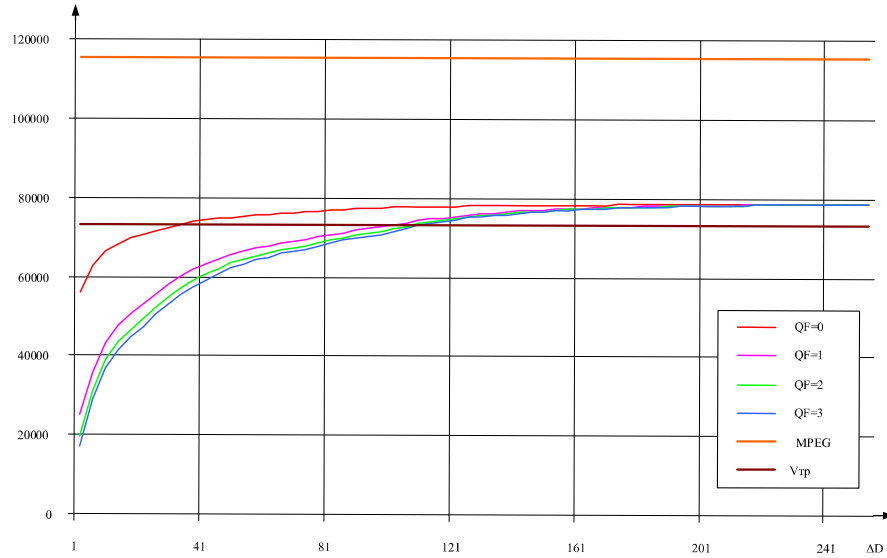


Fig. 4. The dependence of bit rate from the threshold (ΔD) for different values of the quality factor (QF)

3 Conclusion

Thus, the designed total bit rate control method allowing making an adjustment of the video intensity stream in accordance with the parameters of the telecommunications network. The developed method takes into account the following mechanisms:

1. used differential processing blocks selectable by a predetermined limit, which indicates a measure of the current block information content, its algorithm described in [4,5];
2. in the process of compression algorithm produced the following parameters: color sub-sampling, threshold information content, as well as the quality factor. Combining mechanism of the selected parameters is made in such way that the decision was made about the optimality with minimal latency. This allows you to adapt quickly to the throughput of the communication channel and to select the quality of the transmitted image.
3. for processing brightness and chrominance components of a frame in accordance with the method developed by the threshold value and the quality factor, will be chosen to lower brightness as compared with the chrominance components. This leads to the fact that the brightness component of the frame is compressed with better quality than chrominance components. Depending on the conditions of desired velocity or a predetermined quality parameters, quality factors and thresholds will

change so to select optimum values for the compression in the shortest possible time span.

References

1. End-user multimedia QoS categories; G.1010 (11/01), ITU-T STANDARD, INTERNATIONAL TELECOMMUNICATION UNION, GENEVA ; CH, no. G.1010 (11/01), 29 November 2001 (2001-11-29), pages 1 - 18, XP017463369
2. Barannik, V.V., Ryabukha, Yu.N. and Kulitsa, O.S. : The method for improving security of the remote video information resource on the basis of intellectual processing of video frames in the telecommunication systems. In: Telecommunications and Radio Engineering. Vol. 76. No 9. pp. 785-797. (2017). doi: 10.1615/TelecomRadEng.v76.i9.40
3. Barannik, V., Krasnorutskiy, A., Ryabukha, Y., Okladnoy, D.: Model intelligent processing of aerial photographs with a dedicated key features interpretation. In: Modern Problems of Radio Engineering, Telecommunications and Computer Science (TCSET), pp. 736-739. Lviv-Slavske (2016). doi: 10.1109/TCSET.2016.7452167
4. Barannik V., Podlesny S., Krasnorutskiy A., Musienko A., Himenko V.: The ensuring the integrity of information streams under the cyberattacks action. In: East-West Design & Test Symposium (EWDTS), pp. 1-5. IEEE Press (2016) doi: 10.1109/EWDTS.2016.7807752
5. Barannik V., Podlesny S., Tarasenko D., Barannik D., Kulitsa O.: The video stream encoding method in infocommunication systems. In: Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering (TCSET), pp. 538-541. Lviv 2018. doi: 10.1109/TCSET.2018.8336259
6. Gonzales R.C., Woods R.E.: Digital image processing. Prentice Hall, New Jersey, edition. II, 1072-1073. (2002)
7. Richardson I.: H.264 and MPEG-4 Video Compression: Video Coding for Next-Generation Multimedia. Ian Richardson (2005)
8. Zhang Y., Negahdaripour S., Li Q.: Error-resilient coding for underwater video transmission. In: OCEANS MTS/IEEE, pp. 1-7. Monterey, CA (2016)
9. Barannik V.V., Ryabukha Yu. N., Tverdokhle V.V., Barannik D.V.: Methodological basis for constructing a method for compressing of transformants bit representation, based on non-equilibrium positional encoding. In: Advanced Information and Communication Technologies (AICT), 2017 2nd International Conference, pp.188-192. (2017). doi: 10.1109 / AIACT.2017.8020096
10. Baccouch H., Ageneau P. L., Tizon N., Boukhatem N.: Prioritized network coding scheme for multilayer video streaming. In: 14th IEEE Annual Consumer Communications & Networking Conference (CCNC), pp. 802-809, IEEE Press (2017)
11. Tsai W. J., Sun Y. C.: Error-resilient video coding using multiple reference frames. In: IEEE International Conference on Image Processing, pp. 1875-1879, IEEE Press (2013).
12. Chigorin A., Krivoviyaz G., Velizhev A., Konushin A.: A method for traffic sign detection in an image with learning from synthetic data. In: 14th International Conference Digital Signal Processing and its Applications, pp. 316-335. IEEE Press (2012)
13. Salomon D.: Data Compression: The Complete Reference. Fourth Edition. Springer-Verlag London Limited, pp. 899 (2007)
14. Vatolin D., Ratushnyak A., Smirnov M. , Yukin V.: Methods of data compression. The device archiver, compression of images and videos. DIALOG MIFI, Moskow (2013)

15. Barannik, V., Krasnoruckiy, A. and Hahanova, A.: The positional structural-weight coding of the binary view of transformants. In: East-West Design & Test Symposium (EWDTS), pp. 1-4. Rostov-on-Don (2013) doi: 10.1109/EWDTS.2013.6673178
16. Ding Z., Chen H., Gua Y., Peng Q.: GPU accelerated interactive space-time video matting. In: Computer Graphics International, pp. 163-168. (2010)
17. Ablamejko S.V., Lagunovskij D.M. :Obrabotka izobrazhenij: tehnologija, metody, prime-nenie. Amalfeja, Minsk (2000)
18. Rao K. R., Hwang J. J.: Techniques and Standards for Image, Video and Audio Coding. Englewood Cliffs, NJ: Prentice-Hall (1996)
19. Shi, Yun Q.: Image and video compression for multimedia engineering: fundamentals, algorithms, and standards. Huifang Sun, NY, CRC Press (2008)
20. Bai X., Wang J.: Towards temporally-coherent video matting. In: Proceedings of the 5th international conference on Computer vision/computer graphics collaboration techniques, pp. 63-74. MIRAGE'11, Springer-Verlag (2011)
21. Musienko A., Ganjaric J. : Technology of coding of digital aerial photographs taking into account classes of a semantic saturation of blocks in system of air monitoring. In: VII Inter University Conference of Students, PhD Students and Young Scientists ["Engineer of XXI Century, pp. 215-220. Bielsko-Biala, Poland (2016)
22. JPEG 2000 image coding system: Core coding system. ITU-T recommendation. t. 800 and iso/iec 15444-1. ITU-T and ISO/IEC JTC 1 (2000)
23. In J., Shirani S., Kossentini F.: Jpeg compliant efficient progressive image coding. De-partment of Electrical and Computer Engineering, University of British Columbia, Van-couver BC V6T 1Z4, Canada
24. Zheng B., S. Gao: A soft-output error control method for wireless video transmission. In: 8th IEEE International Conference on Communication Software and Networks (ICCSN), pp. 561-564. IEEE Press, Beijing (2016)
25. Kubasov D.V., Vatolin S.: Review of methods of motion compensation. In: Computer graphics and multimedia, vol. 3, pp. 33-43. KPI, Kharkiv (2010)
26. Kudryashov B.D.: Information theory. Peter, St. Petersburg (2009)
27. Akimov D., Shestov A., Voronov A., Vatolin D.: Occlusion Refinement for Stereo Video Using Optical Flow. In: International Conference on 3D Imaging, pp. 115-138. (2012)
28. Pratt W. K., Chen W. H., Welch L. R.: Slant transform image coding. Proc. Computer Processing in communications. New York: Polytechnic Press, pp. 63-84. (1969)
29. Stankiewicz O., Wegner K., Karwowski D., Stankowski J., Klimaszewski K. Grajek T.: Encoding mode selection in HEVC with the use of noise reduction. In: International Con-ference on Systems, Signals and Image Processing (IWSSIP), pp. 1-6. Poznan (2017)

Robust object detection in images corrupted by impulse noise

Rykhard Bohush¹ [0000-0002-6609-5810], Sergey Ablameyko² [0000-0001-9404-1206] and
Yahor Adamovskiy¹ [0000-0003-1044-8741]

¹Polotsk State University, Blokhina st., 29, Novopolotsk, Republic of Belarus, 211440
{ r.bogush, adamovskiy.y }@pdu.by

²Belarusian State University, Nezavisimosti avenue 4, Minsk, Republic of Belarus
ablameyko@bsu.by

Abstract. This paper proposes two effective normalized similarity functions for robust object detection in very high density impulse noisy images. These functions form an integral similarity estimate based on relations of minimum by maximum values for all pairs of analyzed image features. To provide invariance under the constant brightness changes, zero-mean additive modification is used. We explore properties of our functions and compare them with other commonly used for object detection in images corrupted by impulse noise. The efficiency of our approach is illustrated and confirmed by experimental results.

Keywords: similarity functions, object detection, impulse noise

1 Introduction

Impulse noise is frequently encountered in digital image and video due to transmission errors, defective pixels in camera sensors, faulty memory locations, and timing errors during the conversion [1, 2]. An important property for this noise type is that the corrupted pixels can take only the maximum and minimum values from dynamic range and that only part of the pixels is corrupted, and the rest are noise-free. Noise reduction increases the computational cost and it does not allow to get clear image for high noise level. Therefore, it is important to use image processing noise reducing methods for solving the mentioned tasks.

Object detection and precise definition of object location in images and videos are widely used for many applied tasks: industrial inspection and medical diagnostics, stereo vision, reference marks detection and localization in space images of the earth's surface, tracking targets of airborne radar stations, etc [3, 4]. Absolute and relative object position in image or video can be determined after their localization. Camera calibration, image georeferencing, formation of stereo images by computer methods and other applications contain similar tasks. Therefore, many scientific works are devoted to the development of methods for detecting and localizing objects in images and video. Generally, many approaches calculate the similarity of reference object and an input image and compared with a threshold value.

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

Unluckily, there is no a single similarity function that works very well for various images and for all tasks. Different tasks require different measure properties. Therefore, traditional similarity metrics are improved [5, 6] or new functions are provided [7, 8, 9] considering expansion of applications of image processing, and different similarity functions are selected for various applications [10, 11]. For example, image detection in large databases requires high speed and low resistance to deformation of the image, but people tracking in video desires robustness to various types of noise and high accuracy of comparing selected features.

Even in solving only one task uncertainty may also occur. For example, to correctly identity tracked people in a current frame, a maximum accuracy of the estimation for people feature similarity with previous frames is required [12]. However, high accuracy for high density noise or overlapping parts of people by objects will lead to loss their index. It means that a person from previous frames will be identified as another person who has entered into the surveillance area and displayed on the current frame. This problem is also related to the fact that the number of possible grayscale and color images of the same size is huge, their correlation characteristics in practice are not perfect. This leads to false identification or inaccurate object localization when detection is performed in presence of various types of noise components in video or image.

In this paper, we present a new effective function similarity for image and video corrupted by impulsive noise. These functions form an integral normalized similarity estimate based on sequential division of minimum by maximum values for all pairs of features. In Section 2, comprehensive analysis of main similarity functions used in image processing is presented. Section 3 presents our normalized similarity functions and computational costs for them. Section 4 presents experimental results. Finally, the conclusion and feature work details are provided in Section 5.

2 Main similarity functions for image processing

If two images O and B are identical, then value of the normalized metric $M(O, B) = 0$, and the value of the normalized similarity function $S(O, B) = 1$. For another extreme case, when the differences between the images are maximized, the normalized metric $M(O, B) = 1$, and the similarity function usually $S(O, B) = 0$ or $S(O, B) = -1$, if the mean value of the brightness levels is not taken into account.

To compare two images $O = \{o_{ij}\}$ and $B = \{b_{ij}\}$, $N \times N$ size, several similarity functions are often used [13]. It is known that linear relation between brightness of images allows to efficiently apply the correlation coefficient for image corrupted by additive noise [14]. The value of the normalized cross-correlation function (NCC) varies from 0 to 1 and is calculated as:

$$S^{NCC} = \frac{\sum_{i=1}^N \sum_{j=1}^N o_{ij} b_{ij}}{\sqrt{\sum_{i=1}^N \sum_{j=1}^N (o_{ij})^2} \sqrt{\sum_{i=1}^N \sum_{j=1}^N (b_{ij})^2}} \quad (1)$$

The normalization embodied into the zero-mean normalized cross-correlation function (ZNCC) allows one to tolerate linear brightness variations. Also, due to the subtraction of the local mean, the ZNCC provides better robustness than the NCC since it tolerates uniform brightness variations as well. In this case, ZNCC takes a value of (-1) to (+1) and is calculated by formula:

$$S^{ZNCC} = \frac{\sum_{i=1}^N \sum_{j=1}^N (o_{ij} - \bar{o})(b_{ij} - \bar{b})}{\sqrt{\sum_{i=1}^N \sum_{j=1}^N (o_{ij} - \bar{o})^2} \sqrt{\sum_{i=1}^N \sum_{j=1}^N (b_{ij} - \bar{b})^2}} \quad (2)$$

where $\bar{o}$ and $\bar{b}$ – mean values for images O and B (3):

$$\bar{o} = \frac{1}{N^2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} o_{ij}, \quad \bar{b} = \frac{1}{N^2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} b_{ij} \quad (3)$$

Similarity function based on the Euclidean distance (ED) is characterized by a lower computational cost and is defined as:

$$S^{ED} = 1 - \sqrt{\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (o_{ij} - b_{ij})^2} \quad (4)$$

Sum of squared differences (SSD) function is robust to Gaussian noise [15] and is calculated as:

$$S^{SSD} = 1 - \frac{1}{LN^2} \sum_{i=1}^N \sum_{j=1}^N (o_{ij} - b_{ij})^2 \quad (5)$$

where L – of valid brightness values range.

The function based on the weighted sum of squared differences (SSDW) is more stable to linear distortion of levels of analyzed characteristics compared to previous functions:

$$S^{SSDW} = 1 - \frac{1}{L} \frac{\sum_{i=1}^N \sum_{j=1}^N (o_{ij} - b_{ij})^2}{\sqrt{\sum_{i=1}^N \sum_{j=1}^N (o_{ij})^2} \sqrt{\sum_{i=1}^N \sum_{j=1}^N (b_{ij})^2}} \quad (6)$$

The Hausdorff metric-based function (Hd) can only be used for low noise, as increasing noise level quickly reduces the estimation accuracy. This function is determined by the formula:

$$S^{Hd} = 1 - \frac{1}{L} \max_{ij} |o_{ij} - b_{ij}| \quad (7)$$

In general case, to detect an object in an image, the similarity value is calculated for features of all subimages and object. The decision on object presence is made based on the comparison of obtained values with a threshold. If the condition is satisfied, decision about correspondence on subimage and object is made. For accurate localization similarity function value must exceed threshold only if the object is correctly positioned. However, ensuring the minimum probability of missing object (false-negative error) requires a lower threshold level, but this leads to an increase in the number of subimages, including near the correct location of the object in the image that does not correspond to it, known as false-positive error.

3 New normalized similarity functions

The proposed two similarity functions use relationship calculation between the minimum and maximum values for all pairs of analyzed features. Summation of the calculated values is used to obtain an integral normalized value that characterizes similarity of two images. Relationship calculation between descriptors will better emphasize local differences compared to subtraction. High resistance to noise is achieved through the use of summation when obtaining a complex normalized value. Minimum or maximum attribute is necessary to determine when searching a relationship between them, therefore, the proposed functions are called normalized minimax similarity functions.

To estimate image similarity of object $O = \{o_{ij}\}$, $N \times N$ size, and image $B = \{b_{ij}\}$, $N \times N$ size, functions are described as:

- normalized minimax additive similarity p-function ($MMADD^p$):

$$S^{MMADD^p} = \frac{1}{NN} \sum_{i=1}^N \sum_{j=1}^N \frac{\min(o_{ij}^p, b_{ij}^p)}{\max(o_{ij}^p, b_{ij}^p)}, p \in \mathbb{Z}, p \geq 1 \quad (8)$$

- normalized zero-mean minimax additive similarity p-function ($ZMMADD^p$):

$$S^{ZMMADD^p} = \begin{cases} \frac{1}{NN} \sum_{i=1}^N \sum_{j=1}^N \left(\frac{b_{ij} - \bar{b}}{o_{ij} - \bar{o}} \right)^2, & \text{if } |o_{ij} - \bar{o}| \geq |b_{ij} - \bar{b}| \\ \frac{1}{NN} \sum_{i=1}^N \sum_{j=1}^N \left(\frac{o_{ij} - \bar{o}}{b_{ij} - \bar{b}} \right)^2, & \text{if } |o_{ij} - \bar{o}| < |b_{ij} - \bar{b}| \end{cases} \quad (9)$$

For both proposed minimax functions, the following basic properties are obvious:

$$\begin{aligned} S(O, B) &= 1, \Leftrightarrow O = B; \\ S(O, B) &= S(B, O); \\ S(O, B) &= 0, \Leftrightarrow o_{ij} = 0, b_{ij} = L \end{aligned} \quad (10)$$

Presented functions are universal, as MMADD^p returns a normalized value from (0) to (1) and ZMMADD^p returns a value from (-1) to (1) for randomly selected features of a pair of images. Table 1 shows a comparison of computational costs for similarity functions for images with size $N \times N$ pixels.

Table 1. Computational costs for similarity functions

Function name	Number of addition / subtraction operations	Number of multiplication / division operations	Number of comparison operations
NCC	$3N(N-1)$	$3N^2+3$	–
ZNCC	$2N^2+5(N-1)^2$	$3N^2+7$	–
ED	$N^2+N(N-1)+1$	N^2+1	–
SSD	$N^2+N(N-1)+1$	N^2+3	–
SSDV	$N^2+3N(N-1)+1$	$3N^2+4$	–
Hd	N^2+1	1	N^2-1
MMADD ^p (p=2)	$N(N-1)$	$3N^2+2$	N^2
ZMMADD ^p (p=1)	$2N^2+3N(N-1)$	N^2+4	N^2

Thus, in comparison with function of normalized correlation, the offered minimax function provides reduction of calculation complexity, as minimum twice. As result, minimax similarity functions can be used for search of objects in a static image and for detecting moving objects in video sequences.

4 Experimental results

4.1 Parameter determination

It is proposed to use the following parameters for the similarity functions analytical assessment when detecting objects in image:

- function value calculated for object and subimage (A – main peak). Value should aim to 1;
- main peak variance (D_A). The value should aim to zero, meaning smaller value deviations A from expected value;
- function maximum value from all side peak (S_L) can be used to determine threshold T , which should be less than T .
- maximum side peak variance (D_{S_L}) should also be as less as possible;
- side peak number at levels higher than 0.95 (N_{S_L}). The parameter can be used to evaluate the possible false-positive detection results number if the threshold value is incorrectly selected.

Research software was developed using the MatLab package for experimental determination of normalized minimax similarity functions qualitative characteristics. The software is based on an object detection algorithm in a grayscale image using pixel brightness as features. It includes the following steps:

1. Selecting a subimage B_{kl} ($k \in 0..M-m, l \in 0..N-n$), $m \times n$ size, from the upper left of the bitmap image;
2. Similarity function calculation for a reference object O of size $m \times n$ and the selected subimage B_{kl} :

$$S = F(O, B_{kl}) \quad (11)$$

where F is the mathematical transformation operator;

3. Deciding on object availability by the rule:

$$\begin{cases} \text{if } S > T & \text{then } B_{kl} = O \\ \text{else } B_{kl} \neq O \end{cases} \quad (12)$$

where T is the threshold;

4. If the shift number is less than $(M-m) \times (N-n)$, move right or down one pixel and go to step 1, otherwise the search is finished. The analyzed total number of fragments is defined as $(M-m+1) \times (N-n+1)$.

All parameters were calculated for 20 different images without noise and distortion of size 150×150 . For each of them 20 reference objects of 15×15 size were used with obtained values averaging. Values $A=1$ and $D_A=1$ are obtained for all similarity functions.

The resulting values of the remaining parameters are given in Table 2. Analysis of the Table 2 shows that the proposed $ZMMADD^p$ function has the best characteristics for accurately determining object coordinates in images without noise.

Table 2. Function characteristics for noise-free images

Similarity Function	S_L	D_{S_L}	N_{S_L}
NCC	0,97786	0,00008	8568
ZNCC	0,69987	0,01455	3
ED	0,91882	0	481
SSD	0,9918	0	11971
SSDV	0,99948	0	20263
Hd	0,74588	0,00054	43
MMADD ^p (p=2)	0,64973	0,00496	263
ZMMADD ^p (p=1)	0,47403	0,01489	6

4.2 Noise resistance research

Impulse noise is one of the interference types and it appears as random white or black dots in an image, i.e. corrupted pixels take a maximum or minimum valid value, for example, 255 and 0 for 8-bit images. Interference with an amplitude value greater than the useful signal dynamic range occurs, for example, during a fast transient, and is the cause of the appearance of impulse noise in the image. The impulse noise probability density function of a random variable z is given by:

$$p(z) = \begin{cases} P_a & \text{for } z = a \\ P_c & \text{for } z = c \\ \text{else} & 0 \end{cases} \quad (13)$$

A pixel with a brightness value of $c > a$ looks like a white point when c (?), and a pixel with a brightness value of c looks like a black point in image. Почитай. Здесь совсем непонятно.

Impulse noise is a combination of white and black points (“salt and pepper”). Impulse noise can be characterized by the noise density ρ , which determines the percentage of the corrupted pixels n to a total number, i.e. the probability of distortion for each pixel. For impulse noise, ρ values [0.1, 0.3, 0.5, 0.8, 0.95] was chosen. Objects for detection in Fig. 1. are presented.

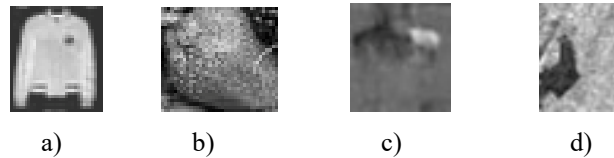


Fig. 1. Object templates: (a) for Fig. 2a, e, i; (b) for Fig. 2b, f, j; (c) for Fig. 2c, g, k; (d) for Fig. 2d, h, l;

Fig. 2 shows some examples of images with the various noise levels.

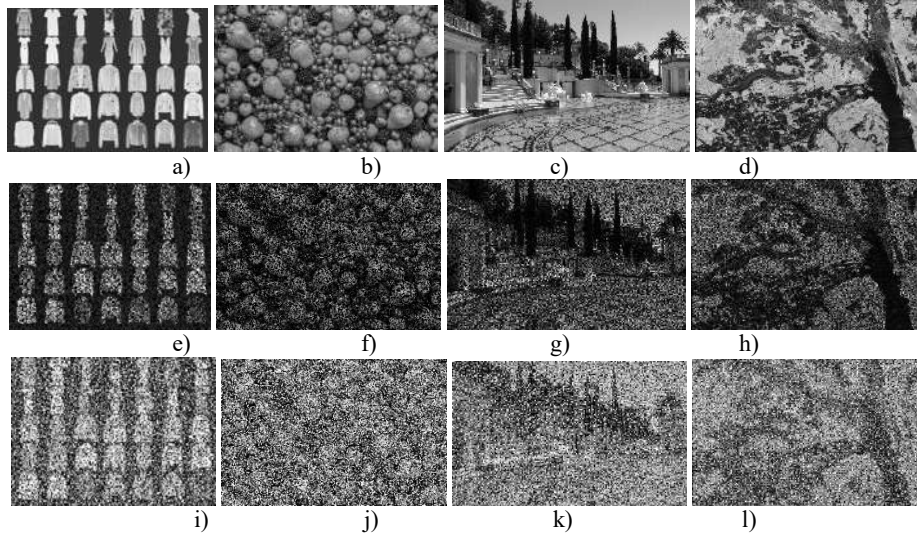


Fig. 2. Some test images: (a-e) noise-free images; (f-j) images corrupted by single-level impulse noise with $\rho = 0.5$; (k-n) images corrupted by «salt and pepper» noise with $\rho = 0.5$

The difference values between similarity function main and maximum side peaks show an assessment completeness and comparison of similarity functions in terms of resistance to various noise types (Table 3-4).

Table 3. Experimental results for similarity functions stability to single-level impulse noise

Similarity Function	$(A - S_L)$				
	$\rho=0.1$	$\rho=0.3$	$\rho=0.5$	$\rho=0.8$	$\rho=0.95$
NCC	-0.00365	-0.03498	-0.06312	-0.11406	-0.2196
ZNCC	0.06579	-0.00573	-0.07419	-0.18602	-0.21083
ED	0.01497	-0.02903	-0.02102	-0.02727	-0.0044
SSD	-0.0009	-0.00857	-0.01165	-0.01096	-0.01039
SSDV	-0.00001	-0.00037	-0.00083	-0.0038	-0.01349
Hd	-0.13882	-0.05725	-0.05569	-0.04392	-0.03216
MMADD ^p (p=2)	0.20776	0.14718	0.1144	0.02539	-0.01688
ZMMADD ^p (p=1)	0.07336	0.08461	-0.1426	-0.14419	-0.10195

Correct object detection is possible when the side peak amplitude is below the similarity function main peak level:

$$(A - S_L) > 0 \quad (14)$$

Object detection is possible for $MMADD^p$ function ($p=2$) at $\rho = 0.8$ when the image is distorted by single-level impulse noise and salt and pepper noise. For other functions at this noise level, condition $(A - S_L) > 0$ is not observed.

Table 4. Experimental results for similarity functions stability to salt and pepper noise

Similarity Function	$(A - S_L)$				
	$\rho = 0.1$	$\rho = 0.3$	$\rho = 0.5$	$\rho = 0.8$	$\rho = 0.95$
NCC	-0.01046	-0.05543	-0.09423	-0.08755	-0.09563
ZNCC	0.12058	0.01313	-0.12529	-0.19709	-0.34173
ED	0.01589	0.00848	-0.0138	-0.03309	-0.03137
SSD	0.00663	-0.00103	-0.02537	-0.02423	-0.05641
SSDV	0.00006	-0.00009	-0.00053	-0.00045	-0.00062
Hd	-0.20314	-0.09176	-0.11529	-0.07843	-0.07843
$MMADD^p$ ($p=2$)	0.22005	0.16443	0.10312	0.01452	-0.0042
$ZMMADD^p$ ($p=1$)	0.38812	0.16848	0.04697	-0.03497	-0.07924

The $ZMMADD^p$ ($\rho = 1$) provides invariance under the constant change in brightness. Object detection is possible for this function at $\rho = 0.3$ when the image is distorted by single-level impulse noise and $\rho = 0.5$ for salt and pepper noise.

5 Conclusion

Normalized similarity functions for image evaluation are proposed. They are characterized by satisfactory computational costs and can improve the detection and object localization in image and video corrupted by impulse noise. Proposed functions are universal, they return a normalized value from (0) to (1) for randomly selected features. Similarity function resistant to a linear change of analyzed feature values of the compared images is proposed that returns a normalized value from (-1) to (1). Experiments were performed to compare these functions with the known ones for images corrupted by single-level impulse and salt and pepper noise. Pixel brightness values are used as image features for experiments. The obtained results have confirmed efficiency of presented similarity functions.

6 References

1. Gonzalez R., Woods R.: Digital Image Processing, 3rd Edition. Prentice Hall (2008)
2. Pitas I., Venetsanopoulos A.: Order statistics in digital image processing Proceedings of the IEEE 80 (12) pp 1893-1921 (1992)
3. Rezaei M., Fränti P.: Set Matching Measures for External Cluster Validity Transactions on Knowledge and Data Engineering 28 (8) pp 2173-86 (2016)
4. Ges V., Starovoitov V.: Distance-based functions for image comparison Pattern Recogn. Lett. 20 207-14 (1999)

5. Wang L., Zhang Y., Feng J. On the Euclidean Distance of Images IEEE Transactions on Pattern Analysis and Machine Intelligence 27 (8) pp 1334–39 (2005)
6. Nakhmani A., Tannenbaum A.: New Distance Measure Based on Generalized Image Normalized Cross-Correlation for Robust Video Tracking and Image Recognition Pattern Recogn. Lett. 34 (3) pp 315–21 (2013)
7. Zagoruiko N. G., Borisova I. A., Dyubanov V. V., Kutnenko O. A.: A quantitative measure of compactness and similarity in a competitive space J. Appl. Industr. Math. 5 (1) pp 144–54 (2011)
8. Lv G.: A novel similarity measure for matching local image descriptors IEEE Access 6 pp 55315–25 (2018)
9. Kim J., Hyun Ch., Han H., Kim H.: Evaluation of Matching Costs for High-Quality Sea-Ice Surface Reconstruction from Aerial Images Remote Sensing 11 (9) pp 1055-72 (2019)
10. Zitova B., Flusser J.: Image registration methods: a survey Image and vision computing 21 (11) pp 977–1000 (2003)
11. Cha S. H.: Comprehensive survey on distance/similarity measures between probability density functions Int. J. of Mathematical Models and Methods in Applied Sciences 1 (4) p 300 (2007)
12. Bohush R., Zakharava I.: Robust Person Tracking Algorithm Based on Convolutional Neural Network for Indoor Video Surveillance Communications in Computer and Information Science 1055 pp 289–300 (2019)
13. Goshtasby A. A.: Similarity and dissimilarity measures in Image Registration Principles, Tools and Methods. Advances in Computer Vision and Pattern Recognition (London) pp 7–66 (2012)
14. Voronov S. V., Tashlinskii A. G.: Efficiency analysis of information theoretic measures in image registration Pattern Recognit. Image Anal. 26 502–5 (2016)
15. Voronov S. V., Tashlinskii A. G.: Analysis of objective functions in the problem of estimating mutual geometric deformations of images Pattern Recognit. Image Anal. 24 (4) 575–582 (2014)