

Sergey Subbotin  
*Editor*

# Computer Modeling and Intelligent Systems

Proceedings of The Fourth International Workshop  
CMIS-2021



Zaporizhzhia, Ukraine  
April 27, 2021

Subbotin S., (Ed.): The Third International Workshop on Computer Modeling and Intelligent Systems (CMIS-2021). Zaporizhzhia, Ukraine, April 27, 2021, CEUR-WS.org, online

This volume represents the proceedings of the Third International Workshop on Computer Modeling and Intelligent Systems (CMIS-2021), held in Zaporizhzhia, Ukraine, in April, 2021. It comprises 44 contributed papers that were carefully peer-reviewed and selected from 112 submissions.

Copyright © 2021 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 Fourth International Workshop on Computer Modeling and Intelligent Systems (CMIS-2021), held in Zaporizhzhia, Ukraine on 27 of April 2021.

The CMIS workshop was held in 2021 for the fourth 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.

The leading place in the training of IT-specialists at the National University "Zaporizhzhia Polytechnic" 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-2021 workshop.

In 2021, the University and the Department of Software Tools celebrate the 30th anniversary of the opening of the Software Engineering specialty at the university, as well as the 25th anniversary of the first graduation of the students on this specialty. The department has trained more than two thousand programmers of bachelor's, specialist's (engineer's), master's levels over the years of existence of this specialty. Thus, the Department of Software Tools has made a significant contribution to the development of IT education and industry in the Zaporizhzhia region. The Fourth International Workshop on Computer Modeling and Intelligent Systems (CMIS-2021) is timed to coincide with this significant dates.

The main purpose of the CMIS-2021 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:

- Artificial Intelligence (Pattern Recognition, Machine Learning, Computational Intelligence (Neural Networks, Fuzzy and Hybrid Systems, Intelligent Optimization and Search), Knowledge-based Systems);
- Computer Science Methods for Modeling;
- Info-Communication Technologies for Data Analysis (Data Mining, Decision Support, Signal and Image Processing) and Applications for Engineering, Medicine and Education.

The CMIS-2021 workshop joined scientists from Armenia, Azerbaijan, Belarus, Belgium, China, Czech Republic, Germany, Indonesia, Israel, Kyrgyzstan, Morocco, Norway, Poland, Portugal, Slovakia, Spain, Sweden, Turkey, Ukraine, United Kingdom of Great Britain and Northern Ireland, and Uzbekistan. The number of countries of CMIS-2021 authors and PC members is 21 (instead of 12 in 2020).

The language of CMIS-2021 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/cmis.workshop/>.

The CMIS-2021 Organizing Committee have received 122 paper submissions (instead of 177 submissions in 2020, and 130 submissions in 2019), out of which 44 were accepted for presentation as a regular papers (instead of 83 in 2020, and 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 2021 is 39% (instead of 47% in 2020, and 62% in 2019). These papers were published in this volume of CMIS-2021 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, 2021

Sergey Subbotin, Dr. Sc., Prof.,  
CMIS-2021 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. **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. **Anatoliy Sachenko**, Western-Ukrainean National University, Ukraine

Prof. **Natalya Shakhivska**, Lviv Polytechnic National University, Ukraine

Prof. **Galyna Tabunshchuk**, National University "Zaporizhzhia Polytechnic", 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

Prof. **José Otegi**, University of the Basque Country, Spain

Prof. **Gustavo Alves**, University of Porto, Portugal

Dr. **Peter Arras**, Katholieke Universiteit Leuven, Belgium

Dr. **Ivan Izonin**, Lviv Polytechnic National University, Ukraine

Dr. **Igor Korobiichuk**, Warsaw University of Technology, Poland

Dr. **Alexandr Nedzved**, Belarusian State University, Belarus

Dr. **Anzhelika Parkhomenko**, National University "Zaporizhzhia Polytechnic", Ukraine

Dr. **Andrii Oliinyk**, National University "Zaporizhzhia Polytechnic", Ukraine

Dr. **Thomas (Tom) Trigano**, Shamoon College of Engineering, Israel

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

Alexei Belotserkovsky

Anatoliy Melnyk

Andrii Chugay

Andrii Kopp

Andrii Tokar

Andrii Yarovyi

Andrii Zhuravka

Andriy Sverstiuk

Artyom Yegorov

Dmytro Fedasyuk

Dmytro Orlovskyi

Eduard Manziuk

Galina Kirichek

Georgiy Yaskov

Iaroslav Lytvynenko

Ievgeniia Kuzminykh

Igor Chimir

Iryna Piestova

Karina Melnyk

Kostiantyn Kasian

Kostiantyn Zashcholkin

Mariia Tiahunova

Mykhailo Poliakov

Mykola Ivanov

Natalia Borysova

Nataliia Kuznietsova

Nataliia Maksyshko

Nataliya Boyko

Nickolay Rudnichenko

Oksana Bahrii-Zaiats

Oksana Pichugina

Oleksandr Bezsonov

Oleksandr Drozd

Olga Nechyporenko

Olha Matsiy

Peter Bidyuk

Rykhard Bohush

Sergey Ivanov

Serhii Vladov

Serhiy Shtovba

Stepan Skrupsky

Svitlana Alyokhina

Tetiana Kapliienko

Tetiana Vakaliuk

Tetyana Marusenкова

Vasyl Martsenyuk

Viktor Melnyk

Vladimir Vichuzanin

Volodymyr Polishchuk

Vsevolod Bohaienko

Zoia Kochuieva

## 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



Co-funded by the  
Erasmus+ Programme  
of the European Union



**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).



**WORK4CE**

**International project "Cross-domain competences for healthy and safe work in the 21st century"(WORK4CE)** of Erasmus+ Programme co-funded by the European Union (Ref. No. 619034-EPP-1-2020-1-UA-EPPKA2-CBHE-JP)



**ViMaCs**

Virtual Master Cooperation  
Data Science

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

## Sponsor



<https://freshcodeit.com>  
<https://freshcodeit.jobs>

# Long-Term Forecasting Method in the Supply Chain Based on an Artificial Neural Network with Multi-Agent Metaheuristic Training

Eugene Fedorov<sup>a</sup> and Olga Nechyporenko<sup>a</sup>

<sup>a</sup>*Cherkasy State Technological University, Shevchenko blvd., 460, Cherkasy, 18006, Ukraine*

## Abstract

The problem of increasing the efficiency of long-term forecasting in the supply chain is examined. Neural network forecasting methods that are based on reservoir calculations, which increases the forecast accuracy, are proposed. Methods for identifying parameters of forecast models based on the metaheuristics are proposed for the methods mentioned above. These methods were researched on the basis of the data from the logistics company Ekol Ukraine and are intended for intelligent computer-based supply chain management systems.

## Keywords 1

long-term forecast, supply chain, metaheuristics, reservoir computing, forecast neural network model

## 1. Introduction

These days, domestic and foreign companies are striving to improve and optimize their business processes with the implementation of the Lean Production technology and principles, the uniqueness of which lies in the fact that it is effective for enterprises of various industries at any stage of the supply chain of products or services to the end consumer. [1-3]. The Lean Production concept is dominant in the formation of "perfect" supply chains, which, in the context of globalization and digitalization of the economy based on information and communication technologies, is the most important factor in competitiveness [4]. One of the most important problems in the field of supply chain management is the insufficiently high accuracy of the forecast. This leads to the fact that supply chain management can be ineffective. Therefore, the development of forecasting methods in the supply chain is an urgent task.

To date, many approaches are known as long-term forecasting tools, among which are:

- autoregressive forecasting methods [5];
- forecasting methods based on exponential smoothing [6];
- neural network forecasting methods [7-10].

Autoregressive methods have the complex determination of the functional dependencies type, the labor intense determination of the model parameters, low adaptability and the lack of the ability to model nonlinear processes.

Neural network forecasting methods provide a tangible advantage, consisting of: the relationships between factors are investigated on ready-made models; no assumptions about the distribution of factors are required; a priori information about factors may be missing; the original data may be highly correlated, incomplete or noisy; analysis of systems with a high degree of nonlinearity is possible; rapid model development; high adaptability; analysis of systems with a large number of

factors is possible; a complete enumeration of all possible models is not required; analysis of systems with heterogeneous factors is possible.

However, neural network methods have a lack of transparency, the complexity of the architecture definition, strict requirements for the training sample, the complexity of the training algorithm choice, and the resource-intensiveness of the training process. Therefore, the task of increasing the efficiency of neural network forecast is urgent.

The aim of the work is to develop a method for long-term forecasting in the supply chain. To achieve the goal, the following tasks were set and solved:

- analyze existing forecast methods;
- propose a neural network forecast model;
- choose a criterion for evaluating the effectiveness of a neural network forecast model;
- propose a method for determining the values of the neural network forecast model parameters based on multi-agent metaheuristics;
- perform numerical studies.

## 2. Problem statement

The problem of increasing the efficiency of long-term forecasting in the supply chain is reduced to the problem of finding such a vector of parameters  $W$ , which satisfies the forecast model adequacy

criterion  $F = \frac{1}{P} \sum_{\mu=1}^P (f(x_{\mu}, W) - d_{\mu})^2 \rightarrow \min_W$ , i.e. deliver the minimum of the mean squared error (the

difference between the model output and the desired output), where  $P$  – test set cardinality,  $x_{\mu}$  –  $\mu^{\text{th}}$  training input value,  $d_{\mu}$  –  $\mu^{\text{th}}$  training input value.

## 3. Literature review

The most commonly used forecast neural networks are:

1. Long short-term memory (LSTM) [11, 12];

This network is based on gates (FIR filters) and a multilayer perceptron. Instead of each hidden neuron, it uses a memory block that contains one or more cells, and is connected with input, output and forget gates. Gates determine how much information to pass through. If the input and output gates are close to 1 and the forget gate is close to 0, then the network turns into an Elman network. If the input gate is close to 0, then the short-term information from the input is ignored. If the forget gate is close to 0, then long-term information from the memory block is ignored. If the output gate is close to 0, then the output information is ignored. The advantage of this network is a higher forecast accuracy than in a conventional multilayer perceptron. The disadvantages are the complexity of the memory blocks implementation, insufficient forecast accuracy, the complexity of defining the architecture, insufficient learning rate.

2. Gated recurrent unit (GRU) [13-15];

This network is based on gates (FIR filters) and a multilayer perceptron. Instead of each hidden neuron, it uses a hidden block that is connected with reset and update gates. Gates determine how much information to pass through. If the reset gate is close to 1 and the update gate is close to 0, then the network turns into an Elman's network. If the reset gate and update gate are close to 0, then the long-term information from the hidden block is ignored and the network becomes a multilayer perceptron. If the update gate is close to 1, then the short-term information from the network input is ignored. The advantage of this network is a higher forecast accuracy than in a conventional multilayer perceptron. The disadvantages are the complexity of the hidden blocks implementation, insufficient forecast accuracy, the complexity of defining the architecture, insufficient learning rate.

3. Neural Turing machine (NTM) [16, 17];

This network is based on a Turing machine and a multilayer perceptron or LSTM and includes a controller and a memory matrix. At any given time, the controller receives input from the outside



world and sends the output to the outside world. The controller also reads from the memory matrix cells via the read heads and writes to the memory matrix cells via the write heads. The advantage of this network is a higher forecast accuracy than in a conventional multilayer perceptron. The disadvantages are the complexity of the controller implementation (in the case of LSTM) and the complexity of defining the architecture, insufficient forecast accuracy, insufficient learning rate.

#### 4. Echo state network (ESN) [18, 19];

This network is based on reservoir computing over sigmoid neurons and a multilayer perceptron. The hidden layer is called the reservoir. Each neuron in the reservoir may be unconnected or connected to other neurons in the reservoir. To train the network, the pseudoinverse matrix method is used. The advantages of this network are the highest forecast accuracy (due to the pseudoinverse matrix method) and the ease of implementation of sigmoid neurons in the reservoir. The disadvantages are the complexity of parallel learning and the complexity of defining the architecture.

#### 5. Long short-term memory (LSM) [20-23].

This network is based on reservoir computations over impulse neurons «Leaky Integrate and Fire» (LIF) and multilayer perceptron. Each neuron in the reservoir may be unconnected or connected to other neurons in the reservoir and is excitatory or inhibitory. A gradient learning method is used to train the network. The advantages of this network are a higher forecast accuracy than in a conventional multilayer perceptron and the possibility of parallel training for the part of the network corresponding to a multilayer perceptron. The disadvantages are the complexity of the implementation of impulse neurons, the complexity of defining the architecture and less high prediction accuracy, the complexity of parallel training for the part of the network corresponding to the reservoir.

Usually, the methods listed above either have a low forecast accuracy (due to falling into a local extremum) or a low learning rate (due to the high computational complexity of the hidden neuron or the complexity of parallelization of training) or the complexity of implementation (due to the complexity of the hidden neuron architecture) or the complexity of defining the architecture, which leads to a decrease in forecast efficiency.

Due to this, creation of a neural network with a training method and architecture that will eliminate the indicated disadvantages is an urgent task.

## 4. Block diagram of a neural network model for a long-term forecast

Figures 1-2 show a block diagram of a long-term forecast model based on a fully connected echo state network (FC-ESN), which is a recurrent two-layer neural network. Unlike traditional ESN, this network is fully connected, using cascades of unit delays. FC-ESN type 1 has a unit delay stage in the input layer. FC-ESN type 2 has a unit delay stage in the input and output layers. The number of input and output neurons is 1.

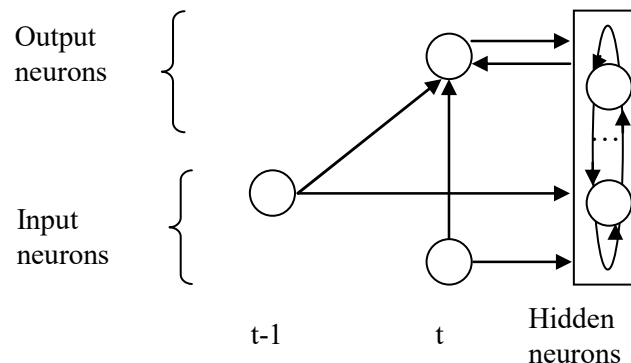


**Figure 1:** Block diagram of a long-term forecast model based on a fully connected echo state network with a cascade of unit delays for an input layer neuron (FC-ESN type 1)

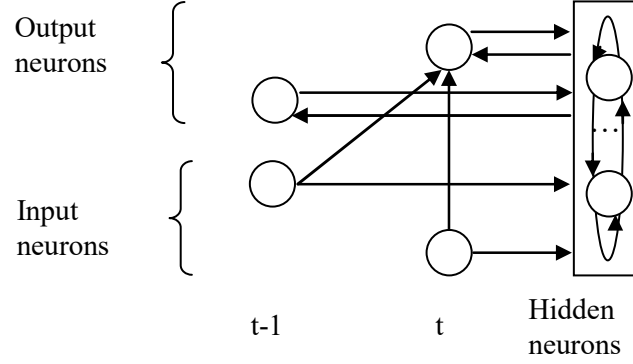


**Figure 2:** Block diagram of a long-term forecast model based on a fully connected echo-state network with a cascade of unit delays for a neuron of the input and output layers (FC-ESN type 2)

## 5. Neural network models for long-term forecast

### 5.1. Long-term forecast model FC-ESN type 1

#### 1. Initialization

$$n = 1.$$

$$y_i^{(1)}(n-1) = 0, \quad i \in \overline{1, N^{(1)}}.$$

#### 2. Forecast

##### 2.1. Initialization of the outputs of the neurons of the input layer

$$y_i^{(0)}(n) = x_i,$$

##### 2.2. Calculation of the outputs of the neurons of the hidden layer

$$y_j^{(1)}(n) = f^{(1)}(s_j^{(1)}(n)), \quad j \in \overline{1, N^{(1)}},$$

$$s_j^{(1)}(n) = b_j^{(1)}(n) + \sum_{i=0}^{M^{(0)}} w_{ij}^{(1)} y_i^{(0)}(n-i) + \sum_{i=M^{(0)}+1}^{M^{(0)}+N^{(1)}} w_{ij}^{(1)} y_{i-M^{(0)}}^{(1)}(n-1) + w_{M^{(0)}+N^{(1)}+1}^{(1)} y^{(2)}(n-1),$$

##### 2.3. Calculation of the outputs of the neurons of the output layer

$$y^{(2)}(n) = f^{(2)}(s^{(2)}(n)),$$

$$s^{(2)}(n) = b^{(2)}(n) + \sum_{i=0}^{M^{(0)}} w_i^{(2)} y_i^{(0)}(n-i) + \sum_{i=M^{(0)}+1}^{M^{(0)}+N^{(1)}} w_i^{(2)} y_{i-M^{(0)}}^{(1)}(n).$$

where  $N^{(1)}$  – the number of neurons in the first layer,

$M^{(k)}$  – the number of unit delays for the  $k^{\text{th}}$  layer,

$w_{ij}^{(k)}$  – the connection weight from the  $i^{\text{th}}$  neuron to the  $j^{\text{th}}$  neuron on the  $k^{\text{th}}$  layer,

$b_j^{(k)}$  – displacement (thresholds) on the  $k^{\text{th}}$  layer,

$y_j^{(k)}(n)$  – the output of the  $j^{\text{th}}$  neuron on the  $k^{\text{th}}$  layer at time  $n$ ,

$f^{(k)}$  – neurons activation function on the  $k^{\text{th}}$  layer (usually  $f^{(k)}(s) = \tanh(s)$ ).

### 5.2. Long-term forecast model FC-ESN type 2

#### 1. Initialization

$$n = 1.$$

$$y_i^{(1)}(n-1) = 0, \quad i \in \overline{1, N^{(1)}}.$$

## 2. Forecast

### 2.1. Initialization of the outputs of the neurons of the input layer

$$y_i^{(0)}(n) = x_i,$$

### 2.2. Calculation of the outputs of the neurons of the hidden layer

$$\begin{aligned} y_j^{(1)}(n) &= f^{(1)}(s_j^{(1)}(n)), \quad j \in \overline{1, N^{(1)}}, \\ s_j^{(1)}(n) &= b_j^{(1)}(n) + \sum_{i=0}^{M^{(0)}} w_{ij}^{(1)} y_i^{(0)}(n-i) + \sum_{i=M^{(0)+1}}^{M^{(0)}+N^{(1)}} w_{ij}^{(1)} y_{i-M^{(0)}}^{(1)}(n-1) + \\ &+ \sum_{i=M^{(0)}+N^{(1)}+1}^{M^{(0)}+N^{(1)}+M^{(2)}} w_{ij}^{(1)} y_i^{(2)}(n-(i-(M^{(0)}+N^{(1)}))), \end{aligned}$$

### 2.3. Calculation of the outputs of the neurons of the output layer

$$\begin{aligned} y^{(2)}(n) &= f^{(2)}(s^{(2)}(n)), \\ s^{(2)}(n) &= b^{(2)}(n) + \sum_{i=0}^{M^{(0)}} w_i^{(2)} y_i^{(0)}(n-i) + \sum_{i=M^{(0)+1}}^{M^{(0)}+N^{(1)}} w_i^{(2)} y_{i-M^{(0)}}^{(1)}(n). \end{aligned}$$

## 6. Criterion for evaluating the effectiveness of a neural network model for long-term forecast

In this work, to determine the parameters values of the FC-ESN model, the criterion of the model adequacy was chosen, which means the choice of such values of the parameters  $W = \{w_{ij}^{(1)}, w_i^{(2)}\}$ , which deliver the minimum of the mean squared error (the difference between the model output and the desired output):

$$F = \frac{1}{P} \sum_{\mu=1}^P (y_{\mu}^{(2)} - d_{\mu})^2 \rightarrow \min_W, \quad (1)$$

where  $P$  – the test set cardinality.

## 7. Method for determining the parameters values of the neural network model for long-term forecast

The method for determining the parameters values of the neural network model for long-term forecasting is reduced to calculating the weights of the hidden layer and the output layer of the FC-ESN model.

### 7.1. Calculating the weights of the hidden layer

The weights of the hidden layer are calculated as follows:

1. Initialize randomly biases (thresholds)  $b_j^{(1)}$  and weights  $w_{ij}^{(1)}$ .
2. Make up from weights  $w_{ij}^{(1)}$ ,  $i \in \overline{M^{(0)}+1, M^{(0)}+N^{(1)}}$ ,  $j \in \overline{1, N^{(1)}}$ , matrix  $W = [w_{ij}]$ ,  $i, j \in \overline{1, N^{(1)}}$ .
3. Determine the matrix  $\widehat{W}$  as  $\widehat{W} = \alpha \frac{W}{\max_{j \in \overline{1, N^{(1)}}} \{|\lambda_j|\}}$ ,

where  $\alpha$  – spectral radius of the matrix  $\widehat{W}$  (for large  $\alpha$  learning is faster, but long short-term memory decreases),  $0 < \alpha < 1$ ,

$\lambda_j$  – eigenvalues of matrix  $W$ .

4. Assign to the weights  $w_{ij}^{(l)}(n)$ ,  $i \in \overline{M^{(0)} + 1, M^{(0)} + N^{(l)}}$ ,  $j \in \overline{1, N^{(l)}}$ , the values of the corresponding elements of the matrix  $\widehat{W}$ .

## 7.2. The output layer weights calculation based on the multi-agent metaheuristic SAPSO method

The proposed SAPSO (simulated annealing and particle swarm optimization) method for numerical functions optimization consists of the following blocks (Figure 3).

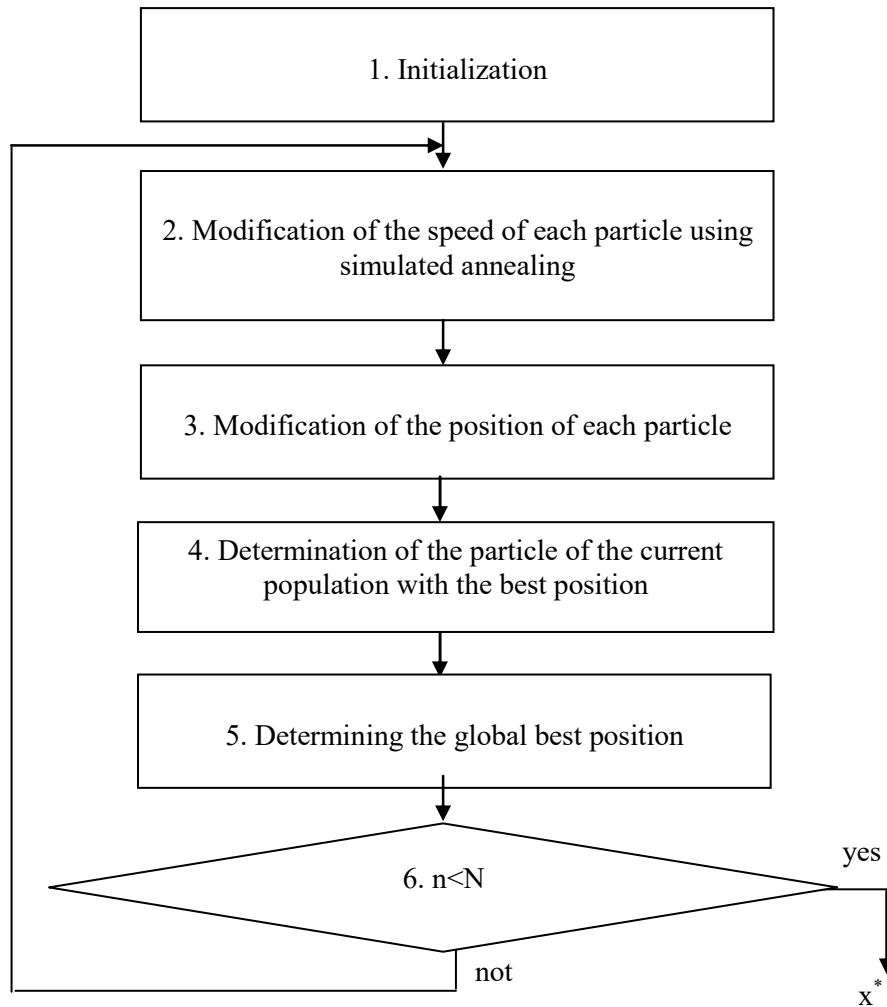


**Figure 3:** The sequence of procedures of the optimization method based on the multi-agent metaheuristic SAPSO method

Block 1 - Initialization:

- setting the maximum number of iterations  $N$ ;
- setting the size of the swarm  $K$  (usually no more than 40);
- setting the dimension of the particle position  $M$  (corresponds to the number of weights in the output layer);
- setting the number of the current iteration  $n$  to one;
- initialization of position  $x_k$  (corresponds to the solution, i.e. the vector of the weights of the output layer)

$$x_k = (x_{k1}, \dots, x_{kM}), x_{ij} = (x_j^{\max} - x_j^{\min})U(0,1) + x_j^{\min}, k \in \overline{1, K},$$

where  $U(0,1)$  – a function that returns a uniformly distributed random number in a range  $[0,1]$ ,  
 $x_j^{\min}$ ,  $x_j^{\max}$  – minimum and maximum value;

- initialization of personal (local) best position  $x_k^{best}$

$$x_k^{best} = x_k, k \in \overline{1, K};$$

- speed initialization  $v_k$

$$v_k = (v_{k1}, \dots, v_{kM}), v_{ij} = 0, k \in \overline{1, K};$$

- creating an initial particle swarm

$$Q = \{(x_k, x_k^{best}, v_k)\};$$

- determination of the particle of the current population with the best position

$$k^* = \arg \min_{k \in \overline{1, K}} F(x_k),$$

$$x^* = x_{k^*}.$$

Block 2 - Modification of the speed of each particle using simulated annealing

Block 2.1 – Calculating two vectors of random numbers for each particle

$$r1_k = (r1_{k1}, \dots, r1_{kM}), r1_{kj} \in \{U(0,1), C(0,1), N(0,1)\}, k \in \overline{1, K}, j \in \overline{1, M},$$

$$r2_k = (r2_{k1}, \dots, r2_{kM}), r2_{kj} \in \{U(0,1), C(0,1), N(0,1)\}, k \in \overline{1, K}, j \in \overline{1, M},$$

where  $N(0,1)$  – a function that returns a random number from a standard normal distribution,

$C(0,1)$  – a function that returns a random number from a standard Cauchy distribution,

Block 2.2 – Calculating annealing temperature

$$T(n) = \beta T(n-1), T(0) = T_0,$$

$$\beta = N^{-\frac{1}{N-1}}, T_0 = N^{\frac{N}{N-1}},$$

where  $T(n)$  – annealing temperature at iteration  $n$ ,

$T_0$  – initial annealing temperature,

$\beta$  – parameter controlling annealing temperature.

Block 2.3 – Calculating parameter controlling the contribution of the component

$$\alpha_1(n) = \alpha_2(n) = \alpha(0) \exp(-1/T(n)), w(n) = w(0) \exp(-1/T(n)),$$

$$\alpha(0) = \alpha_0 = 0.5 + \ln 2, w(0) = w_0 = \frac{1}{2 \ln 2},$$

where  $\alpha_1(n)$  – parameter controlling the contribution of the component  $(x_k^{best} - x_k)(r_1)^T$  to the particle velocity at the iteration  $n$ ,

$\alpha_2(n)$  – parameter controlling the contribution of the component  $(x^* - x_k)(r_2)^T$  to the particle velocity at the iteration  $n$ ,

$w(n)$  – parameter controlling the contribution of the particle velocity at iteration  $n-1$  to the particle velocity at iteration  $n$ ,

$\alpha_0$  – initial value of parameters  $\alpha_1(n)$  and  $\alpha_2(n)$ ,

$w_0$  – initial value of parameter  $w(n)$ ,

The simulated annealing introduced in this work makes it possible to establish an inverse relationship between parameters  $\alpha_1(n)$ ,  $\alpha_2(n)$ ,  $w(n)$  and the iteration number, i.e. at the initial iterations, the entire search space is explored (in this case, the Cauchy distribution is used), and at the final iterations, the search becomes directional (in this case, the normal distribution is used). In addition, in this work, a direct relationship was established between parameters  $T_0$  and  $\beta$  and the iteration number, which makes it possible to automate the selection of these parameters.

The choice of initial values  $\alpha_0 = 0.5 + \ln 2$  and  $w_0 = \frac{1}{2\ln 2}$  is standard and satisfies the conditions for the particle swarm convergence  $w < 1$  and  $w_0 > \frac{1}{2}(\alpha_1 + \alpha_2) - 1$ .

Block 2.4 – Вычисление speed of each particle

$$v_k = w(n)v_k + \alpha_1(n)(x_k^{best} - x_k)(r_1)^T + \alpha_2(n)(x^* - x_k)(r_2)^T, \quad k \in \overline{1, K},$$

Block 3 – Modification of the position of each particle

Block 3.1 Limiting the speed of each particle

$$v_{kj} = \begin{cases} v_{kj} & v_{kj} \in (x_j^{\min}, x_j^{\max}) \\ 0 & v_{kj} \in \{x_j^{\min}, x_j^{\max}\} \end{cases}, \quad k \in \overline{1, K}, \quad j \in \overline{1, M}.$$

Block 3.2 – Calculating position of each particle

$$x_k = x_k + v_k, \quad k \in \overline{1, K},$$

$$x_{kj} = \begin{cases} x_j^{\min}, & x_{kj} \leq x_j^{\min} \\ x_{kj}, & x_{kj} \in (x_j^{\min}, x_j^{\max}) \\ x_j^{\max}, & x_{kj} \geq x_j^{\max} \end{cases}, \quad k \in \overline{1, K}, \quad j \in \overline{1, M},$$

Block 4 - Determination of the personal (local) best position of each particle

$$\text{If } F(x_k) \leq F(x_k^{best}), \text{ then } x_k^{best} = x_k, \quad k \in \overline{1, K}.$$

Block 5 - Determination of the particle of the current population with the best position

$$k^* = \arg \min_{k \in \overline{1, K}} F(x_k).$$

Block 6 - Determining the global best position

$$\text{If } F(x_{k^*}) < F(x^*), \text{ to } x^* = x_{k^*}.$$

Block 7 - Stop condition

If  $n < N$ , then increase the iteration number  $n$  by one and go to block 2.

## 8. Experiments and results

Modeling of the process of the neural network model values determination was carried out in the Matlab package using Parallel Computing Toolbox. Since the formation of each particle in block 1, the modification of the speed, position and local best position of each particle in blocks 2-4, respectively, occurs independently of other particles, and the order of formation and modification of particles is arbitrary, it is proposed to perform parallel processing of particles using a parallel parfor loop. Parfor is part of Parallel Computing Toolbox, replaces the sequential for loop and is based on OpenMP technology, but unlike it, it can be used not only on a local multicore machine, but also on a cluster. The advantage of this approach over the CUDA and MPI technologies (represented in the Parallel Computing Toolbox by the spmd block) is the simplicity and clarity of the technical implementation. Due to the small number of particles, it becomes possible to perform the formation and modification of each particle on the corresponding physical core of the machines processors united in a cluster.

Swarm size was selected as  $K=40$ .

To determine the type of distribution used in the SAPSO method, a number of experiments were carried out, the results of which are presented in Table 1.

**Table 1**  
Comparative characteristics of distribution types

| Distribution type    | U(0,1) | N(0,1) | C(0,1) |
|----------------------|--------|--------|--------|
| Criterion            |        |        |        |
| Number of iterations | 1000   | 10000  | 100000 |

According to Table 1, the distribution  $U(0,1)$  requires the least number of iterations while maintaining the required forecast accuracy.

To define the structure of a long-term forecast model based on FC-ESN, i.e. determining the number of hidden neurons, a number of experiments were carried out, the results of which are presented in Figure 4.

A sample of values based on data from the logistics company Ekol Ukraine was used as input data to determine the parameters values of the neural network model for the long-term forecast. The criterion for choosing the structure of the neural network model was the minimum mean squared forecast error. As can be seen from Figure 4, with an increase in the number of hidden neurons, the error value decreases. It is sufficient to use 16 neurons in the hidden layer for the forecast, since with a further increase in the number of neurons in the hidden layer, the change in the error value is insignificant.

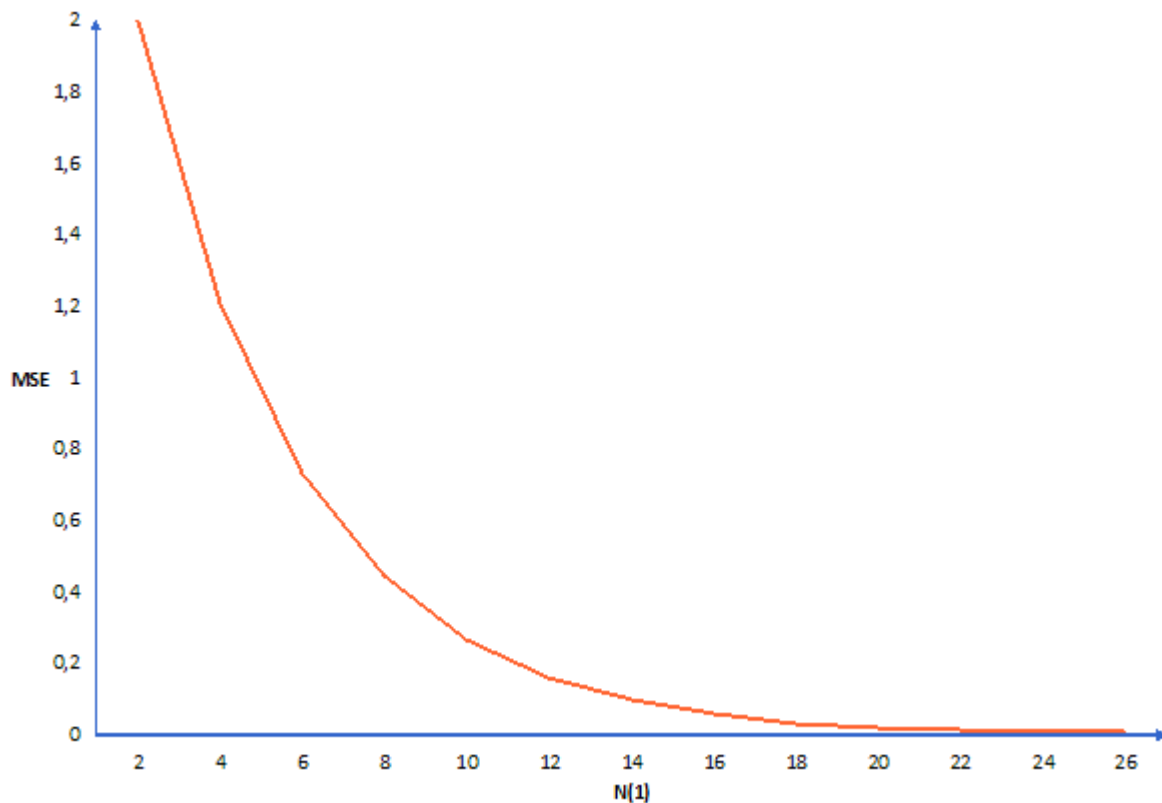


**Figure 4:** Graph of the dependence of the mean square error (MSE) of the forecast on the number of hidden neurons

The neural networks for long-term forecasting were investigated in the work according to the criterion of the minimum mean squared error (MSE) of the forecast and computational complexity (Table 2), where  $M^{(k)}$  – the number of unit delays for the  $k^{\text{th}}$  layer,  $S$  – the number of cell,  $N^{(1)}$  – the number of neurons in the first layer,  $P$  – training set cardinality,  $N$  – number of iterations of the multi-agent metaheuristic method SAPSO,  $N \ll P$ ,  $N^{(1)} \ll P$ .

According to Table 2, FC-ESN type 2 has the highest forecast accuracy, and Type 1 FC-ESN network has the lowest computational complexity.

Based on the performed experiments, the following conclusions can be drawn.

The LSTM network has average learning rates and forecast accuracy.

**Table 2**

Comparative characteristics of neural networks for long-term forecast

| Network<br>Criterion           | Full<br>LSTM                                      | GRU                                 | ESN                                                                   | FC-ESN type 1/2                                                   |
|--------------------------------|---------------------------------------------------|-------------------------------------|-----------------------------------------------------------------------|-------------------------------------------------------------------|
| Minimum MSE<br>of the forecast | 0.12                                              | 0.20                                | 0.08                                                                  | 0.06 / 0.02                                                       |
| Computational<br>complexity    | $\sim PN^{(1)}(5M^{(0)} + 3M^{(0)}S + 24S + S^2)$ | $\sim PN^{(1)}6(M^{(0)} + N^{(1)})$ | $\sim PN^{(1)}(M^{(0)} + N^{(1)}) + (\max\{P, M^{(0)} + N^{(1)}\})^2$ | $\sim N(M^{(0)} + N^{(1)}) / \sim N(M^{(0)} + M^{(2)} + N^{(1)})$ |

The GRU network is second only to the author's networks in learning speed (it uses a gradient learning method and less computational complexity than LSTM and ESN). But it has the least prediction accuracy (due to the gradient learning method and a simplified architecture compared to LSTM).

ESN networks are inferior in forecast accuracy only to the author's networks, since they are trained on the basis of the pseudoinverse matrix method. But it has the lowest learning rate (it has the highest computational complexity, and the pseudoinverse matrix method does not provide for parallelism).

The author's FC-ESN networks are trained on the basis of the proposed metaheuristic, which increases the forecast accuracy (low probability of hitting the local extremum) and the learning rate (provides parallel learning), and does not have the complex implementation.

## 9. Conclusions

The article discusses the problem of improving the efficiency of long-term forecasting in the supply chain. To solve this problem, the existing forecasting methods were investigated. These studies have shown that by far the most effective is the use of artificial neural networks. To improve the quality of the long-term forecast, an ESN neural network was chosen, modified (by introducing full connectivity and cascades of unit delays in the input and output layers), and in the course of a numerical study, the structure of its model was determined. The experiments have shown that with 16 hidden neurons, the value of the mean squared error does not change significantly, and the selected network gives forecast results with a minimum deviation. A method was proposed for determining the parameters values of the proposed neural network model for long-term forecast. This allowed to ensure high speed and accuracy of the forecast. The proposed methods are intended for software implementation in the Matlab package using Parallel Computing Toolbox, which speeds up the process of finding a solution. The software implementing the proposed methods was developed and researched on the database of the logistics company Ekol Ukraine. The conducted experiments have confirmed the efficiency of the developed software allowing to recommend it for practical use in solving problems of supply chain management. Prospects for further research are in applying the proposed methods on a wider set of benchmarks.

## 10. References

- [1] J. F. Cox, J. G. Schleher, Theory of constraints handbook, New York, NY, McGraw-Hill, 2010.
- [2] E. M. Goldratt, My saga to improve production, in: Selected Readings in Constraints Management, Falls Church, VA: APICS, 1996, pp. 43-48.
- [3] E. M. Goldratt, Production: The TOC way, including CD-ROM simulator and workbook, Revised edition, Great Barrington, MA: North River Press, 2003.
- [4] S. Smerichevska et al, Cluster Policy of Innovative Development of the National Economy: Integration and Infrastructure Aspects: monograph, S. Smerichevska (Eds.), Wydawnictwo naukowe WSPIA, 2020.
- [5] R. T. Baillie, G. Kapetanios, F. Papailias, Modified information criteria and selection of long memory time series models, in: Computational Statistics and Data Analysis, volume 76, 2014, pp. 116–131. doi: 10.1016/j.csda.2013.04.012.



- [6] P. Bidyuk, T. Prosyankina-Zharova, O. Terentiev, Modelling nonlinear nonsta-tionary processes in macroeconomy and finances, in: Z. Hu, S. Petoukhov, I. Dychka, M. He (Eds.), *Advances in Computer Science for Engineering and Education. Advances in Intelligent Systems and Computing*, volume 754, Springer, Cham, 2019, pp. 735–745. doi: 10.1007/978-3-319-91008-6\_72.
- [7] L. Lyubchik, E. Bodyansky, A. Rivtis, Adaptive harmonic components detection and forecasting in wave non-periodic time series using neural networks, in: *Proceedings of the ISCDMCI'2002*, Evpatoria, 2002, pp. 433-435.
- [8] K.-L. Du, K. M. S. Swamy, *Neural networks and statistical learning*, Springer-Verlag, London, 2014.
- [9] S. Haykin, *Neural networks*, Pearson Education, New York, NY, 1999.
- [10] S. N. Sivanandam, S. Sumathi, S. N. Deepa, *Introduction to neural networks using Matlab 6.0*, The McGraw-Hill Comp., Inc., New Delhi, 2006.
- [11] S. Hochreiter, J. Schmidhuber, Long short-term memory, in: *Neural Computation*, volume 9, 1997, pp. 1735-1780. doi: 10.1162/neco.1997.9.8.1735.
- [12] F. Gers, *Long Short-Term Memory in Recurrent Neural Networks*, PhD thesis, Ecole Polytechnique Federale de Lausanne.
- [13] K. Cho, B. van Merriënboer, C. Gulcehre, F. Bougares, H. Schwenk, Y. Bengio, Learning phrase representations using RNN encoder-decoder for statistical machine translation, in: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Doha, Qatar, 2014, pp. 1724–1734. doi: 10.3115/v1/D14-1179.
- [14] R. Dey, F. M. Salem, Gate-Variants of Gated Recurrent Unit (GRU) Neural Networks, arXiv:1701.05923, 2017. URL: <https://arxiv.org/ftp/arxiv/papers/1701/1701.05923.pdf>.
- [15] E. Fedorov, T. Utkina, O. Nechyporenko, Forecast method for natural language constructions based on a modified gated recursive block, in: *CEUR Workshop Proceedings*, vol. 2604, 2020, pp. 199-214.
- [16] A. Graves, G. Wayne, M. Reynolds et al., Hybrid computing using a neural network with dynamic external memory, *Nature* 538 (2016) 471–476. doi:10.1038/nature20101.
- [17] R. B. Greve, E. J. Jacobsen, S. Risi, Evolving neural turing machines for reward-based learning, in: *Proceedings of the 2016 Genetic and Evolutionary Computation Conference, GECCO'16*, ACM, 2016, pp. 117–124. doi: 10.1145/2908812.2908930.
- [18] H. Jaeger, Tutorial on Training Recurrent Neural Networks, Covering BPPT, RTRL, EKF and the Echo State Network Approach, GMD Report 159, German National Research Center for Information Technology, 2002.
- [19] H. Jaeger, M. Lukosevicius, D. Popovici, U. Siewert, Optimization and applications of echo state networks with leakyintegrator neurons, in: *Neural Networks* volume 20, 2007, pp. 335–352. doi:10.1016/j.neunet.2007.04.016.
- [20] T. Natshlager, W. Maas, H. Markram, The liquid computer: A novel strategy for real-time computing on time series, in: *Special Issue on Foundations of Information Processing of Telematik*, 2002, pp. 39–43.
- [21] Q. Wang, P. Li, D-lsm: Deep liquid state machine with unsupervised recurrent reservoir tuning, in *Pattern Recognition (ICPR)*, in: *23rd International Conference on Pattern Recognition (ICPR)* (Cancun: IEEE), 2016, pp. 2653–2658. doi: 10.1109/ICPR.2016.7900035.
- [22] W. Maass, Liquid state machines: motivation, theory, and applications, in: *Computability in context: computation and logic in the real world*, 2011, pp. 275–296. doi: 10.1142/9781848162778\_0008.
- [23] T. Neskorođieva, E. Fedorov, I. Izonin, Forecast method for audit data analysis by modified liquid state machine, in: *CEUR Workshop Proceedings*, 2020, volume 2631, pp. 145-158.

# Adaptive learning of evolving hyper basis function neural network

Yevgeniy Bodyanskiy<sup>a</sup>, Anastasiia Deineko<sup>b</sup>, Iryna Pliss<sup>a</sup> and Oleksandr Zeleniy<sup>c</sup>

<sup>a</sup> Kharkiv National University of Radio Electronics, Control Systems Research Laboratory, Nauky av., 14, Kharkiv, 61166, Ukraine

<sup>b</sup> Kharkiv National University of Radio Electronics, Artificial Intelligence Department, Nauky av., 14, Kharkiv, 61166, Ukraine

<sup>c</sup> Kharkiv National University of Radio Electronics, Department of Media Systems and Technologies, Nauky av., 14, Kharkiv, 61166, Ukraine

## Abstract

In the article architecture and learning method of the artificial evolving hyper basis neural network are proposed. The neural network under consideration tunes not only its synaptic weights, but automatically determines neurons number, coordinates of the kernel activation function centers and parameters of the receptive fields in online mode providing high speed data processing.

## Keywords

Artificial neural networks, adaptive learning, hyper basis function neural network, self-organizing T. Kohonen's map, V. Epanechnikov's activation kernel function

## 1. Introduction

To date, artificial neural networks (ANNs) are widely used to solve various problems of Data Mining and first of all for intelligent control, identification, pattern recognition, classification, clustering, forecasting, emulation in conditions of uncertainty and significant nonlinearity. If data should be processed in a sequential online mode, a convergence rate of a learning process comes to the forefront, that significantly limits the ANNs class suitable for work under these conditions. ANNs, which use kernel activation functions (radial-basis, bell-shaped, potential), are very effective from the speed optimization point of view in the learning process.

Radial-basis function neural networks (RBFN) are widely used because their output signal depends linearly on synaptic weights. The main idea of this ANNs is connected with potential function method [1], Parsen's estimations [2], kernel [3] and nonparametric [4] regressions. Universal approximation properties and ability to process data sequentially in online mode are its main benefits. However, the RBFN is exposed to the so-called "curse of dimensionality" which means that when the input space dimensionality increases, there's an exponential growth of the adjustable parameters' (weights') amount. To overcome this problem for solving practical tasks possible by using hyper basis function neural network (HBFN) [5], with have more advantages comparatively to traditional RBFN.

## 2. Radial basis and hyper basis function neural networks

In Figure 1 standard architecture of the radial-basis function network is shown whose hidden layer implements some nonlinear transformation of the input space  $R^n$  into higher dimension ( $h > n$ ) hidden

---

CMIS-2021: The Fourth International Workshop on Computer Modeling and Intelligent Systems, April 27, 2021, Zaporizhzhia, Ukraine

EMAIL: yevgeniy.bodyanskiy@nure.ua (Ye. Bodyanskiy); iryna.pliss@nure.ua (I. Pliss); anastasiia.deineko@nure.ua (A. Deineko);

oleksandr.zeleniy@nure.ua (O. Zeleniy)

ORCID: 0000-0001-5418-2143 (Ye. Bodyanskiy); 0000-0001-7918-7362 (I. Pliss); 0000-0002-3279-3135 (A. Deineko); 0000-0002-7583-7759 (O. Zeleniy)



© 2020 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

space  $R^h$  and its output layer is formed by adaptive linear associators that form the network response by performing a nonlinear transformation of the form

$$\hat{y}^R(k) = w_0 + \sum_{l=1}^h w_l \varphi_l(x(k)) = \sum_{l=0}^h w_l \varphi_l(x(k)) = w^T \tilde{\varphi}(x(k))$$

where  $x(k) = (x_1(k), x_2(k), \dots, x_n(k))^T$ ,  $\varphi_l(x(k)) = \varphi_l(\|x(k) - c_l\|, \sigma_l)$  – radial-basis function, depended of distance  $\|x(k) - c_l\|$  between input vector  $x(k)$  and activation function centers  $c_l$  and width parameter  $\sigma_l$ ,  $k$  – is current discrete time,  $\tilde{\varphi}(x(k)) = (1, \varphi^T(x(k)))^T$ ,  $\varphi(x(k)) = (\varphi_1(x(k)), \varphi_2(x(k)), \dots, \varphi_h(x(k)))^T$ .

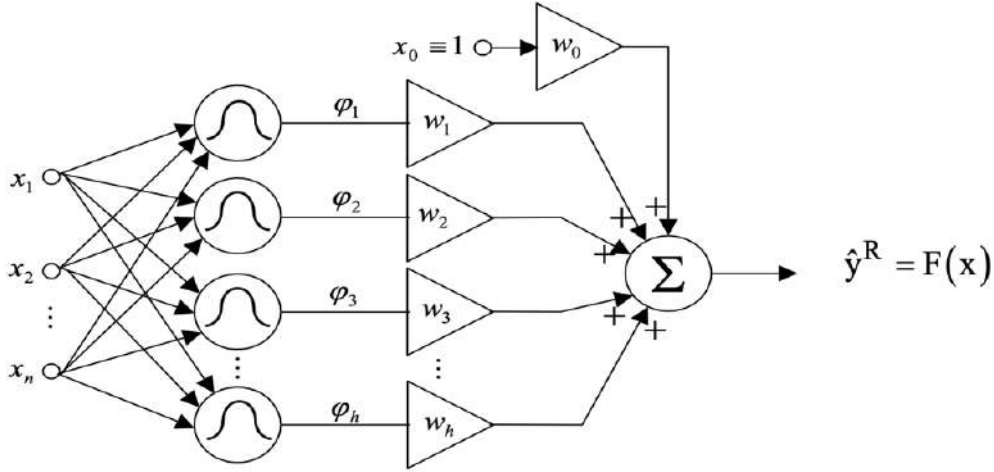


**Figure 1:** Standard radial-basis function network

The most commonly standard Gaussian function as activation function in the radial-basis ANNs is used in the form

$$\varphi_l(x(k)) = \exp\left(-\frac{\|x(k) - c_l\|^2}{\sigma_l^2}\right), l = 1, 2, \dots, h \quad (1)$$

where centers  $c_l$  and wight  $\sigma_l$  parameters are determined beforehand and do not tuned during learning process. The learning process itself is connected with adjusting of the synaptic weights vector  $w = (w_0, w_1, \dots, w_h)^T$ , for that different modifications of least-squares method or traditional gradient procedures are usually used.

Using the multidimensional construction instead of the Gaussian (1) it is possible to improve the approximating properties of the network

$$\varphi_l(x(k)) = \exp\left(-(x(k) - c_l)^T \Sigma_l^{-1} (x(k) - c_l)\right) = \exp\left(-\|x(k) - c_l\|_{\Sigma_l^{-1}}^2\right) \quad (2)$$

where  $\Sigma_l^{-1}$  covariance matrix that determines shape, size and receptive field orientation of  $l$  –  $th$  kernel activation function. This is the main difference between hyper basis function network and traditional RBFN.

If  $\Sigma_l = \sigma_l^2 I$  (here  $I = (n \times n)$  – identity matrix) the receptive field is a hypersphere with a center  $c_l$  and radius  $\sigma_l$ ; if  $\Sigma_l = \text{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2)$  – that is hyperelipsoid with axes coincide with input space axes and have length  $\sigma_i$ , at  $i^{th}$  axe, and finally, if  $\Sigma_l$  is arbitrary positive definite matrix

$$\Sigma_l = Q_l^T \Lambda_l Q_l$$

diagonal matrix of the eigenvectors  $\Lambda_l$  determines receptive field size and rotation orthogonal matrix  $Q_l$  – its orientation.

Speaking about hyper basis function ANNs learning it should be noted that here not only synaptic weights vector  $w$  can be adjusted but and centers  $c_l$ , and matrixes  $\Sigma_l$ . So, introducing into consideration the transformation implemented by the neural network in the form

$$\hat{y}^H(k) = \sum_{l=0}^h w_l \varphi_l\left(\|x(k) - c_l\|_{\Sigma_l^{-1}}^2\right),$$

learning criterion

$$E(k) = \frac{1}{2} e^2(k) = \frac{1}{2} \left( y(k) - \sum_{l=0}^h w_l \varphi_l \left( \|x(k) - c_l\|_{\Sigma_l^{-1}}^2 \right) \right)^2$$

(here  $y(k)$  – external reference signal) and derivations by all tuned parameters

$$\begin{cases} \frac{\partial E(k)}{\partial w_l} = -e(k) \varphi_l \left( \|x(k) - c_l\|_{\Sigma_l^{-1}}^2 \right), \\ \nabla_{c_l} E(k) = 2e(k) w_l \varphi_l' \left( \|x(k) - c_l\|_{\Sigma_l^{-1}}^2 \right) \Sigma_l^{-1} (x(k) - c_l), \\ \left\{ \frac{\partial E(k)}{\partial \Sigma_l^{-1}} \right\} = -e(k) w_l \varphi_l' \left( \|x(k) - c_l\|_{\Sigma_l^{-1}}^2 \right) (x(k) - c_l)(x(k) - c_l)^T, \end{cases} \quad (3)$$

the learning algorithm could be written [6]:

$$\begin{cases} w_l(k+1) = w_l(k) + \eta_w(k+1) e(k+1) \varphi_l \left( \|x(k+1) - c_l(k)\|_{\Sigma_l^{-1}(k)}^2 \right), \\ c_l(k+1) = c_l(k) - \eta_c(k+1) e(k+1) w_l(k+1) \varphi_l' \left( \|x(k+1) - c_l(k)\|_{\Sigma_l^{-1}(k)}^2 \right) \times \\ \quad \times \Sigma_l^{-1}(k) (x(k+1) - c_l(k)), \\ \Sigma_l^{-1}(k+1) = \Sigma_l^{-1}(k) + \eta_\Sigma(k+1) e(k+1) w_l(k+1) \varphi_l' \left( \|x(k+1) - c_l(k+1)\|_{\Sigma_l^{-1}(k)}^2 \right) \times \\ \quad \times (x(k+1) - c_l(k+1))(x(k+1) - c_l(k+1))^T \end{cases} \quad (4)$$

where  $\eta_w(k+1)$ ,  $\eta_c(k+1)$ ,  $\eta_\Sigma(k+1)$  – learning rate parameters for the corresponding variables.

Using as activation functions Gaussians (1) and (2) leads to learning procedure (4) becomes too cumbersome from computational point of view that naturally slows down the learning speed. In this regards we propose to introduce into consideration multidimensional modification of the V. Epanechnikov's function [7] in the form

$$\varphi_l \left( \|x(k+1) - c_l(k)\|_{\Sigma_l^{-1}}^2 \right) = 1 - \|x(k+1) - c_l(k)\|_{\Sigma_l^{-1}}^2,$$

whose derivations have the form

$$\begin{cases} \nabla_{c_l} \varphi_l \left( \|x(k) - c_l\|_{\Sigma_l^{-1}}^2 \right) = -2 \Sigma_l^{-1} (x(k) - c_l), \\ \left\{ \frac{\partial \varphi_l \left( \|x(k) - c_l\|_{\Sigma_l^{-1}}^2 \right)}{\partial \Sigma_l^{-1}} \right\} = (x(k) - c_l)(x(k) - c_l)^T. \end{cases} \quad (5)$$

Relations (5) avoid to rewrite the system (3) in the form

$$\begin{cases} \frac{\partial E(k)}{\partial w_l} = -e(k) \varphi_l \left( 1 - \|x(k) - c_l\|_{\Sigma_l^{-1}}^2 \right), \\ \nabla_{c_l} E(k) = 2e(k) w_l \Sigma_l^{-1} (x(k) - c_l), \\ \left\{ \frac{\partial E(k)}{\partial \Sigma_l^{-1}} \right\} = -e(k) (x(k) - c_l)(x(k) - c_l)^T, \end{cases}$$

and learning algorithm –

$$\begin{cases} w_l(k+1) = w_l(k) + \eta_w(k+1) e(k+1) \left( 1 - \|x(k+1) - c_l(k)\|_{\Sigma_l^{-1}(k)}^2 \right), \\ c_l(k+1) = c_l(k) - \eta_c(k+1) e(k+1) w_l(k+1) \Sigma_l^{-1}(k) (x(k+1) - c_l(k)), \\ \Sigma_l^{-1}(k+1) = \Sigma_l^{-1}(k) + \eta_\Sigma(k+1) e(k+1) w_l(k+1) \times \\ \quad \times (x(k+1) - c_l(k+1))(x(k+1) - c_l(k+1))^T. \end{cases} \quad (6)$$

It is easy to see that procedure (6) from the computational point of view is simpler that algorithm (4).

### 3. Hybrid neural network and evolution of its architecture

The question how to choose number of neurons in the network  $h$  and initial center locations  $c_l$  it is very important. The easiest way to overcome this problem is to use Subtractive Clustering algorithm [8], that is effective enough for processing data in the batch mode. But this algorithm requires to select

huge amount of the free parameters. If solving task is connected with processing of non-stationary data, then it is necessary to re-initialize the network from time to time.

Dynamic Decay Adjustment (DDA) also is one of the possible methods to tune neural network with kernel activation functions [9]. This method is belonging to constructive learning algorithms and works fast enough. But, in the mode of online processing of the non-stationary signals this method becomes noneffective.

Resource Allocation Network [10] uses hybrid learning based on both optimization and memory learning (the principle "Neurons in data points"), using of competition elements. At the same time, in the learning process using gradient procedures, both synaptic weights and the parameters of the neuron's centers closest to the received observation are adjusted. It can be noted that the standard Epanechnikov's functions are used instead of traditional Gaussians as activation functions in the network. The disadvantage of Resource Allocation Network is its high computational complexity.

In this regard, it is necessary to develop artificial evolving hyper basis function neural network, that tunes not only its all parameters but and automatically determines number of neurons in online mode with high speed of data processing.

In the figure 2 architecture of a hybrid evolving artificial neural network that is based on a hyper-basis neural network with a variable number of neurons and T. Kohonen's self-organizing map (SOM) [11], that controls their number and adjusts centers location in the self-learning mode is demonstrated.

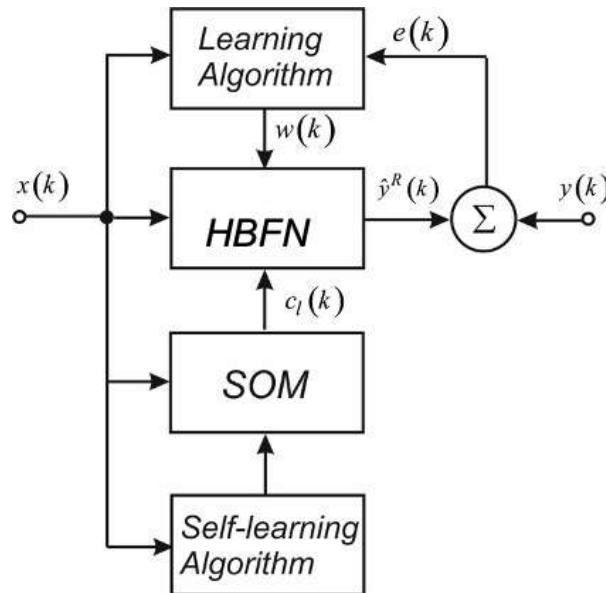


**Figure 2:** Structural schema of the hybrid evolving network

The functioning process of this system is as follows. When the first observation  $x(k)$  is fed to the neural network input, where the first neuron is formed according to the principle of "Neurons at data points", i. e. almost instantly. In the next incoming of the data they firstly fed to SOM, where they are compared with the already existing centroids and then if no coincidences are found, a new center of the kernel function is formed and, accordingly, a new neuron in HBFN.

According to the approach under consideration next method of the controlled numbers of neurons activation function is put into consideration:

Stage 1<sub>1</sub>: to encode all the values of the input variables into the interval  $-1 \leq x_i \leq 1$  and set the receptive field radius of the neighborhood function in the interval  $r \leq 0,33$ ;

Stage 2<sub>1</sub>: observation  $x(1)$  after this is fed set  $c_1 = x(1)$ ;

Stage 3<sub>1</sub>: observation  $x(2)$  is set:

- if  $\|x(2) - c_1\| \leq r$ , that  $c_1(1)$  is corrected by the rule

$$c_l(2) = \frac{c_1 + x(2)}{2};$$

- if  $r < \|x(2) - c_1(1)\| \leq 2r$ ,  $c_1(1)$  is corrected according to the self-learning of the self-organizing Kohonen's map by principle of "Winner takes more" (WTM) [11]

$$c_1(2) = c_1(1) + \eta(2)\psi_1(2)(x(2) - c_1(1))$$

with neighborhood function

$$\psi_l(2) = \max \left\{ 0, 1 - \left( \frac{\|x(2) - c_1(1)\|}{2r} \right)^2 \right\}$$

(Epanechnikov's function with receptive field with radius equal  $2r$ )

- if  $\|x(2) - c_1(1)\| > 2r$ ,

new kernel activation function is formed with center  $c_2(2) = x(2)$ .

This completes the first iteration of the activation functions forming of the hyper basis neural network. Let us for the  $k$ -th moment of time  $p \leq h$  activation functions  $\varphi_l(x(k))$  are formed with centers  $c_l(k)$  and  $x(k+1)$ -th observation is fed to processing. Next forming of the activation functions is performed as follows:

Stage  $1_{k+1}$ : to determine neuron-winner for each distance  $\|x(k+1) - c_l(k)\|$  minimal among all  $l = 1, 2, \dots, p$ ;

Stage  $2_{k+1}$ :

- if  $\|x(k+1) - c_l(k)\| \leq r$  then

$$c_l(k+1) = \frac{c_l(k) + x(k+1)}{2};$$

- if  $r < \|x(k+1) - c_l(k)\| \leq 2r$ ,
- then  $c_l(k+1) = c_l(k) + \eta(k+1)\psi_l(k+1)(x(k+1) - c_l(k))$ ;

$$\psi_l(k+1) = \max \left\{ 0, 1 - \left( \frac{\|x(k+1) - c_l(k)\|}{2r} \right)^2 \right\};$$

- if  $\|x(k+1) - c_l(k)\| > 2r$  then kernel activation function is formed with center  $c_{p+1}(k+1) = x(k+1)$ ;

- if in the process of activation functions formation arises a situation  $\|x(k+1) - c_l(k)\| > 2r$  and  $p = h$ , then it is necessary to increase the receptive field radius and return to stage  $2_{k+1}$  with an increased radius of the function  $\psi_l(k+1)$ .

As can be seen, this procedure is a hybrid of N. Kasabov's evolving algorithm [12] and T. Kohonen's self-organizing map. However, the proposed neural network is designed not only to solve clustering problems, but to control the number of neurons in the hyper basis neural network.

#### 4. Evolving cascaded hyper basic function neural network

If we have a short learning sample then it is possible to overcome the problem of overfitting by dividing an initial task in some way to subtasks of lower dimensionality and by grouping the obtained solutions to get a required result. From a computational point of view, the most convenient method in this case is the Group Method of Data Handling (GMDH) [14-17] that have demonstrated its efficiency while solving a number of practical tasks. In [17], a multilayer GMDH-neural network is considered. It contains two-input N-adalines as nodes and each node's output is a quadratic function of an input signal. Each neuron's synaptic weights are defined in a batch mode using the conventional least squares method. One way use some more hidden layers to provide a necessary approximation quality. That is why an online learning procedure becomes impossible. In this connection it's expedient to introduce a simplified architecture which is based on simple nodes and can be tuned under conditions of a short learning sample.

Let's introduce an architecture of the compartmental R-neuron that is shown in the Figure 4. Generally speaking, it's a simplified architecture of the conventional RBFN with two inputs  $x_i$  and  $x_j$ ,  $i, j = 1, 2, \dots, n$ , where  $n$  is a dimensionality of initial input space.

The compartmental R-neuron contains  $p$  activation functions  $\varphi_h^{ij}(x^{ij}, c_h^{ij}, \Sigma_h^{ij})$ ,  $p+1$  synaptic weights that are joined in vector  $w_l^{ij} = (w_{l0}^{ij}, w_{l1}^{ij}, \dots, w_{lp}^{ij})$ ,  $p$  two-dimensional centers' vectors  $c_h^{ij} = (c_h^i, c_h^j)^T$ ,  $p$   $(2 \times 2)$  -matrices of receptive fields' activation functions  $\Sigma_h^{ij}$ , a two-dimensional input

vector  $x^{ij} = (x_i, x_j)^T$ , output  $\hat{y}_l$ ;  $l = 1, 2, \dots, p$ ;  $k = 1, 2, \dots, N$  is a number of observations in a processing sample or an index of the current discrete time.

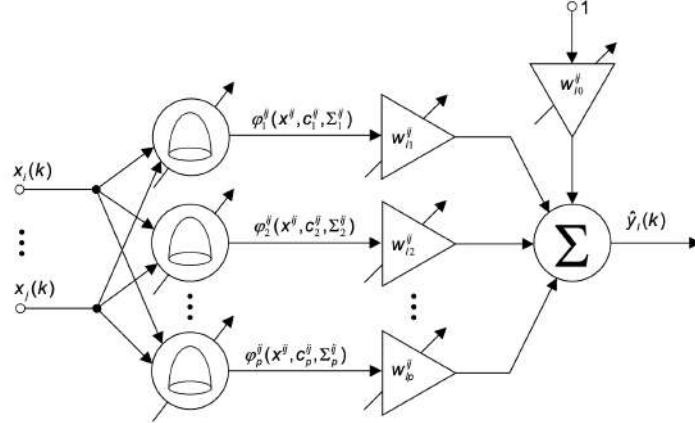


**Figure 3:** The compartmental R-neuron

Multidimensional Epanechnikov kernels are used as activation functions  $\varphi_h^{ij}(x^{ij}, c_h^{ij}, \Sigma_h^{ij})$

$$\varphi_h^{ij}(x^{ij}, c_h^{ij}, \Sigma_h^{ij}) = 1 - \|x^{ij} - c_h^{ij}\|_{(\Sigma_h^{ij})}^2 \quad (7)$$

which have a bell-shape form determined by a positive definite matrix of the receptive field  $\Sigma_h^{ij}$ . An advantage of the activation function (7) comparing to conventional ones is linearity of its derivatives with respect to all the parameters that makes it possible to adjust synaptic weights as well as both centers and receptive fields (which is very important when we have a very short learning sample). At the same time, the compartmental R-neuron implements a transformation in the form

$$\hat{y}_l = w_{l0}^{ij} + \sum_{h=1}^p w_{lh}^{ij} \varphi_h^{ij}(x^{ij}, c_h^{ij}, \Sigma_h^{ij}) = w_{l0}^{ij} + \sum_{h=1}^p w_{lh}^{ij} \left( 1 - \|x^{ij} - c_h^{ij}\|_{(\Sigma_h^{ij})}^2 \right).$$

It should be noticed that if we have a two-dimensional case it is easy to locate centers at the regular lattice's nodes and to define receptive fields as circles.

Introducing a  $(p+1) \times 1$ -vector of activation functions  $\varphi^{ij} = (1, \varphi_1^{ij}(x^{ij}(k), c_1^{ij}, \Sigma_1^{ij}), \dots, \varphi_p^{ij}(x^{ij}(k), c_p^{ij}, \Sigma_p^{ij}))^T$  and a learning criterion

$$E_l^N = \sum_{k=1}^N (y(k) - \hat{y}_l(k))^2 = \sum_{k=1}^N e_l^2(k) = \sum_{k=1}^N \left( y(k) - (w_l^{ij})^T \varphi^{ij}(k) \right)^2 \quad (8)$$

it's easy to obtain a required solution with the help of the traditional least squares method in the form

$$w_l^{ij} = \left( \sum_{k=1}^N \varphi^{ij}(k) (\varphi^{ij}(k))^T \right)^+ \sum_{k=1}^N \varphi^{ij}(k) y(k) \quad (9)$$

where  $(\cdot)^+$  is a symbol of the Moore-Penrose inversion. If data are fed sequentially in an online mode, recurrent form of the expression (8) can be used in the form

$$\begin{cases} w_l^{ij}(k) = w_l^{ij}(k-1) + \frac{P_{ij}(k-1) \left( y(k) - (w_l^{ij}(k-1))^T \varphi^{ij}(k) \right)}{1 + (\varphi^{ij}(k))^T P_{ij}(k-1) \varphi^{ij}(k)} \varphi^{ij}(k), \\ P_{ij}(k) = P_{ij}(k-1) - \frac{P_{ij}(k-1) \varphi^{ij}(k) (\varphi^{ij}(k))^T P_{ij}(k-1)}{1 + (\varphi^{ij}(k))^T P_{ij}(k-1) \varphi^{ij}(k)}, \quad P_{ij}(0) = \gamma I, \quad \gamma \gg 0. \end{cases} \quad (10)$$

The algorithms (9) and (10) are effective only in those cases when a required solution is stationary which means that optimal values of the synaptic weights aren't variable in time. But most of real-world practical tasks are characterized by an opposite situation. A high-performance adaptive learning

algorithm with both tracking and filtering properties [18] can be used for adaptive identification of non-stationary objects and non-stationary time series prediction in the form

$$\begin{cases} w_l^{ij}(k) = w_l^{ij}(k-1) + p_w^{-1}(k) \left( y(k) - \left( w_l^{ij}(k-1) \right)^T \varphi^{ij}(k) \right) \varphi^{ij}(k) = \\ \quad = w_l^{ij}(k-1) + p_w^{-1}(k) e_l(k) \varphi^{ij}(k), \\ p_w(k) = \beta p_w(k-1) + \|\varphi(x(k))\|^2, \quad 0 \leq \beta \leq 1. \end{cases} \quad (11)$$

To tune centers and covariance matrices, one could use a procedure

$$\begin{cases} w_l^{ij}(k) = w_l^{ij}(k-1) + \eta_w(k) e_l(k) \varphi^{ij}(k), \\ \eta_w^{-1}(k) = p_w(k) = \beta p_w(k-1) + \|\varphi(x(k))\|^2, \\ c_h^{ij}(k) = c_h^{ij}(k-1) + \eta_c(k) e_l(k) w_l^{ij}(k) \left( \Sigma_h^{ij}(k-1) \right)^{-1} \left( x^{ij}(k) - c_h^{ij}(k-1) \right) = \\ \quad = c_h^{ij}(k-1) + \eta_c(k) e_l(k) g_h(k), \\ \eta_c^{-1}(k) = p_c(k) = \beta p_c(k-1) + \|g_h(k)\|^2, \\ \left( \Sigma_h^{ij}(k) \right)^{-1} = \left( \Sigma_h^{ij}(k-1) \right)^{-1} - \eta_\Sigma(k) e_l(k) w_l^{ij}(k) \left( x^{ij}(k) - c_h^{ij}(k) \right) \left( x^{ij}(k) - c_h^{ij}(k) \right)^T = \\ \quad = \left( \Sigma_h^{ij}(k-1) \right)^{-1} - \eta_\Sigma(k) e_l(k) G_h(k), \\ \eta_\Sigma^{-1}(k) = P_\Sigma(k) = \beta P_\Sigma(k-1) + Tr \left( G_h(k) G_h^T(k) \right). \end{cases}$$

An architecture in Figure 4 combines the GMDH ideas and cascade neural networks.

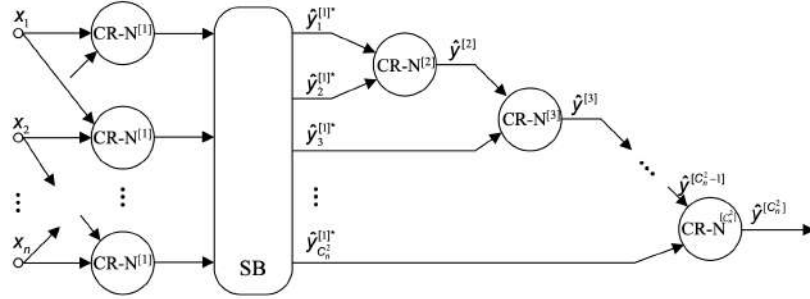


**Figure 4:** An evolving cascade neural network

The first hidden layer of the network is formed similarly to the first hidden layer of the GMDH neural network [17] and contains a number of neurons that equals to a number of combinations of  $n$  in 2, that is  $C_n^2$ . A selection block SB executes a sorting procedure by accuracy, for example, the most accurate signal among all output signal  $\hat{y}_l^{[1]}$  is  $\hat{y}_1^{[1]*}$  in the sense of variations, then comes  $\hat{y}_2^{[1]*}$  and the worst one is  $\hat{y}_{C_n^2}^{[1]*}$ . The SB outputs  $\hat{y}_1^{[1]*}$  and  $\hat{y}_2^{[1]*}$  are fed to the only neuron in the second cascade layer  $CR - N^{[2]}$  that computes a signal  $y^{[2]}$  which is later joined in the third cascade with the selection block's output  $\hat{y}_3^{[1]*}$ . A process of cascades' increasing goes on until the required accuracy is obtained. An overall neurons' number of this network is defined by a value  $2C_n^2 - 1$ . The neural network can process information that is fed in real time by adjusting both its parameters and its architecture in time [19].

## 5. Experimental results

In first series of experiments, we used synthetics generated data set that describes Normal Gauss bivariate distribution. In this data set number of clusters was chosen arbitrary and all data were generated in the interval  $[-1; 1]$ . Result of processing synthetic data set by evolving hyper basis function neural network (EHBFFNN) are shown at Figure 5.



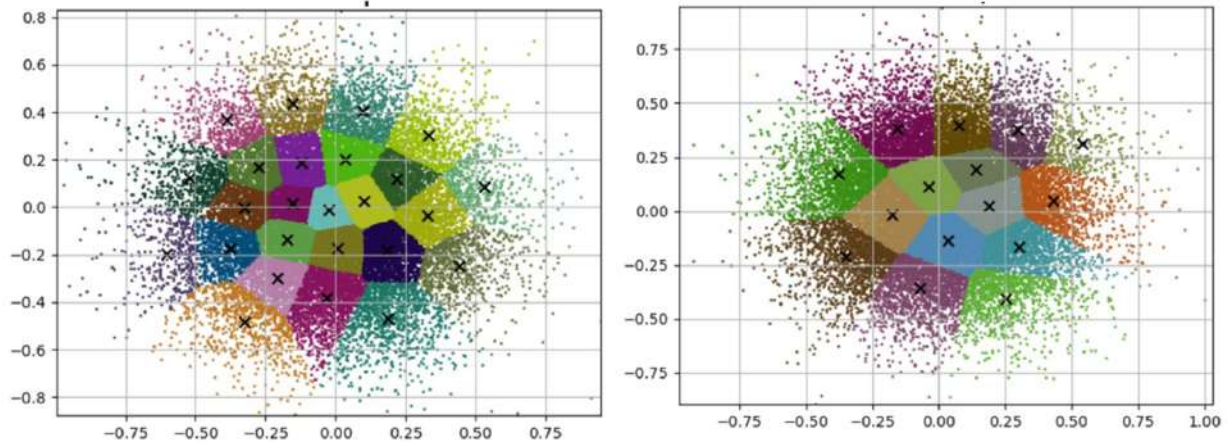


**Figure 5:** Result of processing synthetic data set by EHBFFN

For the next experiments series data set “Wine” and data set “Ionosphere” from UCI Repository [13] were taken. Efficiency of proposed evolving hyper basis function neural network (EHBFFN) has been investigated and compared with the standard radial-basis neural network and the standard T. Kohonen’s self-organizing map (SOM). Firstly, data preprocessing that include outliers and gaps analysis was made. Investigated data were coding into hypercube. For results visualization principal component analysis method was used.

The Figure 6 demonstrates evaluation of the centroid’s coordinates (as a visualization example data set “Wine” was taken).

Comparison of numerical results of proposed EHBFFN, standard RBFN and SOM are shown in the Table 1. Clustering quality were measured by Calinski-Harabash index [19] and in the Table 1 maximum, average and minimum values has been demonstrated.

This index in general formulation has a form

$$CH(m) = \frac{1}{m-1} Tr S_B^m \left( \frac{1}{N-m} Tr S_W^m \right)^{-1}$$

where  $S_B^m = \frac{1}{N} \sum_{j=1}^m N_j^m (w_j^m - w^{-m})(w_j^m - w^{-m})^T$ ,  $j = 1, 2, \dots, m$  – the inter-cluster distance matrix for  $m$  clusters;  $w^{-m} = \frac{1}{N} \sum_{j=1}^m N_j^m w_j^m$  – gravity data set center  $x$ ;  $N_j^m$  – number of observation belonged to the  $j$ -th clusters;  $S_W^m = \frac{1}{N} \sum_{j=1}^m \sum_{k=1}^N u_j(k) (x(k) - w_j^m)(x(k) - w_j^m)^T$  – intra-cluster distance matrix for  $m$  clusters,

$$u_j = \begin{cases} 1, & \text{if } x(k) \text{ belongs to } j\text{-th cluster,} \\ 0, & \text{otherwise} \end{cases}$$

where  $u_j$  – crisp membership function.

In situation when observations came to process sequentially in online mode is it necessary to organize calculation of the Calinski-Harabash index on the sliding window of dimensional  $s$  ( $s = 1, 2, \dots, N$ ) and we could rewrite this index in the form

$$CH(m, k) = \frac{\frac{1}{m-1} \sum_{j=1}^m N_j^m(\tau) \|w_j^m(\tau) - w^{-m}(\tau)\|^2}{\frac{1}{N-m} \sum_{j=1}^m u_j(\tau) \|x(\tau) - w_j^m(\tau)\|^2}$$

where

$$w^{-m}(\tau) = \frac{1}{s} \sum_{\tau=k-s+1}^k x(\tau).$$

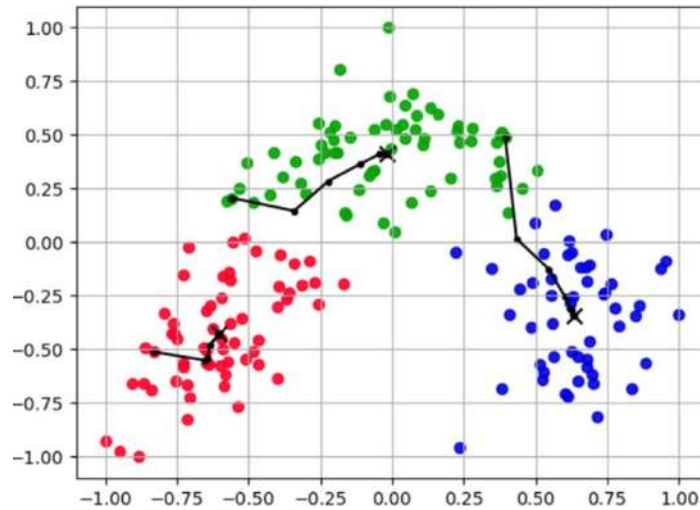


**Figure 6:** Evaluation of the centroid's coordinates

Visualizations of clustering results by proposed evolving hyper basis function neural network are presented at the Figure 7 (Figure 7a – data set “Wine”; Figure 7b – data set “Ionosphere”)

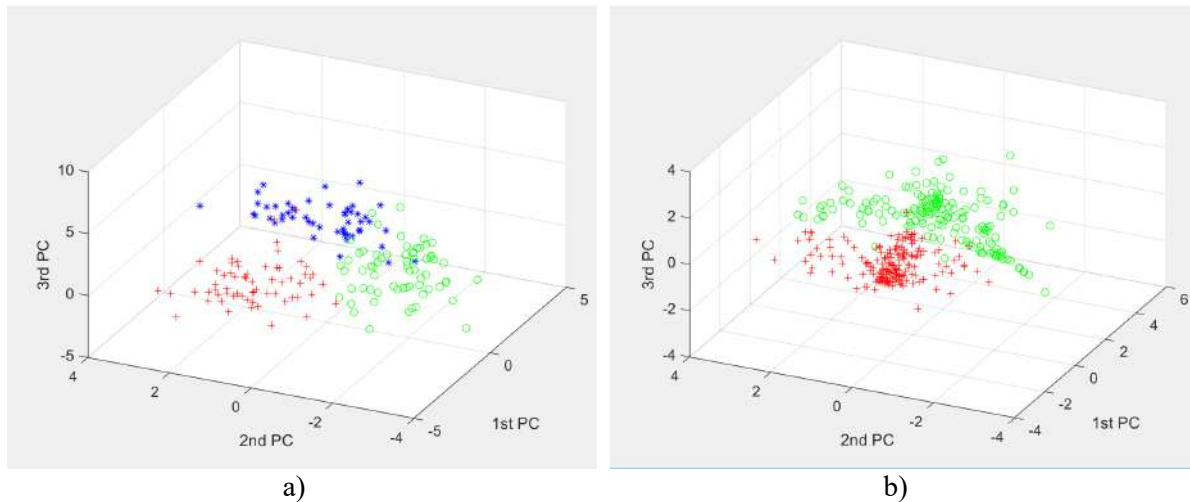


**Figure 7:** Visualizations of clustering results by proposed evolving hyper basis function neural network:  
a) data set “Wine”,  
b) data set “Ionosphere”

**Table 1**

Comparison of numerical results of investigated ANNs

| Investigated ANNs     | Calinski-Harabash index |        |       |
|-----------------------|-------------------------|--------|-------|
|                       | avg                     | max    | min   |
| “Wine” data set       |                         |        |       |
| RBFN                  | 65.34                   | 69.98  | 40.93 |
| SOM                   | 56.20                   | 70.67  | 47.62 |
| EHBFFN                | 69.52                   | 70.94  | 41.90 |
| “Ionosphere” data set |                         |        |       |
| RBFN                  | 62.78                   | 90.52  | 55.35 |
| SOM                   | 70.56                   | 86.16  | 64.82 |
| EHBFFN                | 74.70                   | 122.64 | 62.78 |

## 6. Conclusions

The approach, which combines training of both synaptic weights and activation functions' centers, and which is based on both supervised learning and self-learning, is proposed in this paper. The main advantage of the proposed approach is that it can be used in an online mode, when a training set is fed to a system's input sequentially, and its volume is not fixed beforehand. The results can be used for solving a wide class of Dynamic Data Mining and Data Stream Mining problems.

## 7. References

- [1] M.A. Aizerman, E.M. Braverman, L.I. Rozonoer, Method of potential functions in the theory of machine learning. Moscow: Nauka, 1970.
- [2] E. Parzen, "On the estimation of a probability density function and the mode", Ann. Math. Statist. 3, (1962), 1065-1076.
- [3] S. Haykin, Neural Networks. A Comprehensive Foundation. Upper Saddle River, N.J.: Prentice Hall, Inc., (1999).
- [4] T.V. Varyadchenko, V.Ya. Katkovnik, "Nonparametric method of inversion of regression functions", Stochastic control systems, Novosibirsk: Nauka, (1979), pp. 4-14.
- [5] Y. Zhon, T. Mu, Zh-H. Pang, Ch. Zheng, "A survey on hyper basis function neural network", System Science & Control Engineering, 7(1), (2019), 495-507.
- [6] Ye. Bodyanskiy, O. Tychenko, A. Deineko, "An evolving radial basis neural network with adaptive learning of its parameters and architecture", Automatic Control and Computer Science, 49(5), 2015, 255-260.
- [7] V. A. Epanechnikov, "Nonparametric estimation of multivariate probability density", Probability Theory and its Application, 14 (2), (1968), 156-161.
- [8] S. Chiu, "Fuzzy model identification based on cluster estimation", Journal of Intelligent & Fuzzy Systems, 2(3), (1994).
- [9] N. Karimi, S. Kazem, D. Ahmadian, H. Adibi, L. V. Ballestra, "On a generalized Gaussian radial basis function: Analysis and applications", Eng. Anal. with Boundary Elements. (2020), vol. 112, 46-57,
- [10] M. Dehghan, V. Mohammadi, "The numerical solution of Fokker–Planck equation with radial basis functions (RBFs) based on the meshless technique of Kansa's approach and Galerkin method", Eng. Anal. Boundary Elements, (2014), vol. 47, 38-63.
- [11] T. Kohonen, Self-Organizing Maps. Berlin: Springer-Verlag. 1995.
- [12] N. Kasabov. Evolving Connectionist Systems. London: Springer-Verlag. 2003.
- [13] A. Frank, and A. Asuncion, UCI Machine Learning Repository, [<http://archive.ics.uci.edu/ml>]. Irvine, CA: University of California, School of Information and Computer Science, 2013.
- [14] A.G. Ivakhnenko, Self-learning systems for recognition and automatic control. Technika, Kyev, 1969.
- [15] A.G. Ivakhnenko, Long-term prediction and complex systems control. Technika, Kyev, 1975.
- [16] A.G. Ivakhnenko, H.R. Madala, Inductive Learning Algorithms for Complex Systems Modeling. CRC Press, London-Tokio, 1994.
- [17] Z. Zhao, Y. Lou, Y. Chen, H. Lin, R. Li, G. Yu, "Prediction of interfacial interactions related with membrane fouling in a membrane bioreactor based on radial basis function artificial neural network (ANN)", Bioresource Techno, (2019), vol. 282, 262-268.
- [18] Ye .V. Bodyanskiy, O.K. Tyshchenko, A.O. Deineko, Evolving neuro-fuzzy systems with kernel activation function. Saarbruecken, Germany: LAP Lambert Academic Publishing. 2015.
- [19] R. Xu, D.C. Wunsch. Clustering, IEEE Press Series on Computational Intelligence. Hoboken, NJ: John Wiley & Sons, Inc., 2009.

# DCGAN for Enhancing Eye Diseases Classification

Mahmoud Smaida<sup>a</sup>, Serhii Yaroshchak<sup>a</sup>, Youness El Barg<sup>c</sup>

<sup>a</sup> The national university of water and Environmental Engineering, Rivne, Ukraine

<sup>b</sup> Multi-disciplinary Faculty Larache, Larache, Morocco

## Abstract

One of the most important factors in image classification is the amount of data, especially in medical images. However, In the medical field, obtaining these datasets remains a challenge. In this work, we present Deep Convolutional Generative Adversarial Network (DCGAN) method that generate synthetic medical images. In addition, using GMD (Glaucoma, Myopia and Diabetic retinopathy) model to improve eye diseases classification with and without synthetic medical images. Four types of eye diseases, Glaucoma, Myopia, Diabetic retinopathy and Normal are our dataset. Our method is demonstrated on a limited dataset of eye disease (300 Glaucoma, 300 Myopia, 300 Diabetic retinopathy and 300 Normal). Firstly, we exploit DCGAN to generate synthetic medical images. then we utilize GMD method for eye diseases classification. finally, training our method using original data and synthetic medical images, then compare performance.

The accuracy of the model had improved significantly from 76.58% in training set and 76.42% in validation set, to 80.45% in training set and 83.74% in validation set. We suggest that this work can be applied to other image classification models such as Vgg16, Inception v3, ResNet to enhance the accuracy.

## Keywords 1

Eye diseases, Deep learning, Myopia, DCGAN, Glaucoma, Diabetic retonapthy

## 1. Introduction

One of the main challenges in the field of medical imaging is how to deal with small data sets, especially when we use supervised learning to classify images. In medical imaging assignments, explanatory reports are made by specialists with experience in his or her specialty. Although, some of medical data sets are able to be used online, and major challenges have been declared, most data sets remain limited and only apply to specific medical issues. Collecting medical data is a complex and costly procedure that requires collaboration between researchers and clinics [1].

Many researchers try to solve this problem with augmentation diagrams, including making some adjustments to the images of the data set such as rotation, cropping and size. Using such redundant data to improve network training has become a standard procedure for computer vision tasks, recently one of the most important method used is GANs.

Generative Adversarial Networks (GANs), are a type of machine learning network that Ian Goodfellow and colleagues invented in 2014. Two neural networks compete with each other in a game (meaning game theory, often but not always in the form of a zero-sum game). The aim of it is to creating fabricated data similar to real data, which are difficult for a human or mechanical observer to distinguish between them. This technique learns to generate new data with the same statistical properties of a training set. For example, a GAN trained in photographs can create new images that

---

CMIS-2021: The Fourth International Workshop on Computer Modeling and Intelligent Systems, April 27, 2021, Zaporizhzhia, Ukraine

EMAIL: m.e.smaida@nuwm.edu.ua (M. Smaida); s.v.yaroshchak@nuwm.edu.ua (S. Yaroshchak); youness.elbarg@etu.uae.ac.ma (Y. El Barg)

ORCID: 0000-0002-5552-2768 (M. Smaida); 0000-0001-9576-2929 (S. Yaroshchak); 0000-0002-6386-5402 (Y. El Barg)



© 2020 Copyright for this paper by its authors.  
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).  
CEUR Workshop Proceedings (CEUR-WS.org)

look real to human observers, and have many photorealistic characteristics. Although it was originally proposed as a form of the generative model of unsupervised learning, GANs have also proven useful for semi-supervised learning, fully supervised learning, and reinforcement learning. At the 2016 symposium, AI expert Yann LeCun described GANs as "the coolest idea in the field of machine learning in the past 20 years" [2].

The Generative Adversarial Network consists of two separate neural networks, Generator (G) and Discriminator (D) constantly competing against each other, the generator, which tries to generate examples that seem the real data, it works to deceive the discriminator, while the role of discriminative network is to distinguish between the real and fake data. a vector of random number will be feeds to the generator, and the outputs are unreal synthetic data denoted by  $G(z)$ , the discriminator feeds real data and the output of generator. The output of the discriminator is the possibility whether the entered data is real or fake, the output denoted by  $D(x)$ . The output  $D(x)$  represents the probability that  $x$  will be a real image if it is 1, which means 100% is a real image, and the output is 0 which means that the image cannot be real.

During training, the goal of creating the Generator is to try to create real images to deceive the Discriminator. The goal of Discriminator is to separate the Generated images from the real images as closely as possible.

What is the outcome of the final match? In the most perfect case, Generator can create a real and fake  $G(z)$  images. For D it is difficult to determine whether the image generated by G is real so  $D(G(z)) = 0.5$ .

In this paper, we investigate the problem of lack of medical images such as eye diseases for image classification. We propose Generative Adversarial networks (GANs) to increase our dataset in order to improve the accuracy of classification using our model (Glaucoma, Myopia and Diabetic retinopathy model). The model based on CNN for image classification.

In this way, our goal is achieved and we obtain the generative model G, which can be used to create images in order to increase our dataset [4].

### 1.1. How GAN is work:

The above is just a rough overview of the basic principles of GAN, how can it be described in mathematical language? Here is a straightforward excerpt from the formula from Ian Goodfellow's paper, which called objective function of GAN [3]:

$$\min_G \max_D V(D, G) = E_{x \sim p_{data}(x)} [\log(D(x))] + E_{z \sim p_Z(z)} [\log(1 - D(G(z)))] \quad (1)$$

You only need to analyze this formula:

- The complete formula consists of two periods.  $X$  represents the real image;  $Z$  represents the noise input to the G network.
- $G(z)$  is a fake image generated by the generator G.
- $D(x)$  is the probability of that  $X$  came from the real data (0~1), the discriminator should classify a real image as real. And this value should be close to 1.
- Purpose of G as mentioned above.  $D(G(z))$  is the probability that a D network governs whether the image generated by G is real. In other words, G wants  $D(G(z))$  to be as large as possible (close to 1), and  $V(D, G)$  will be smaller at this time. So, we see that the forward tag of the formula is  $\min_G$ .
- Purpose of D: is simply a classifier. It tries to distinguish real data from the data created by the generator. train D to classify real images as real,  $D(x)$  should be close to 1. Train D to classify fake images as fake,  $D(G(z))$  should be close to 0.

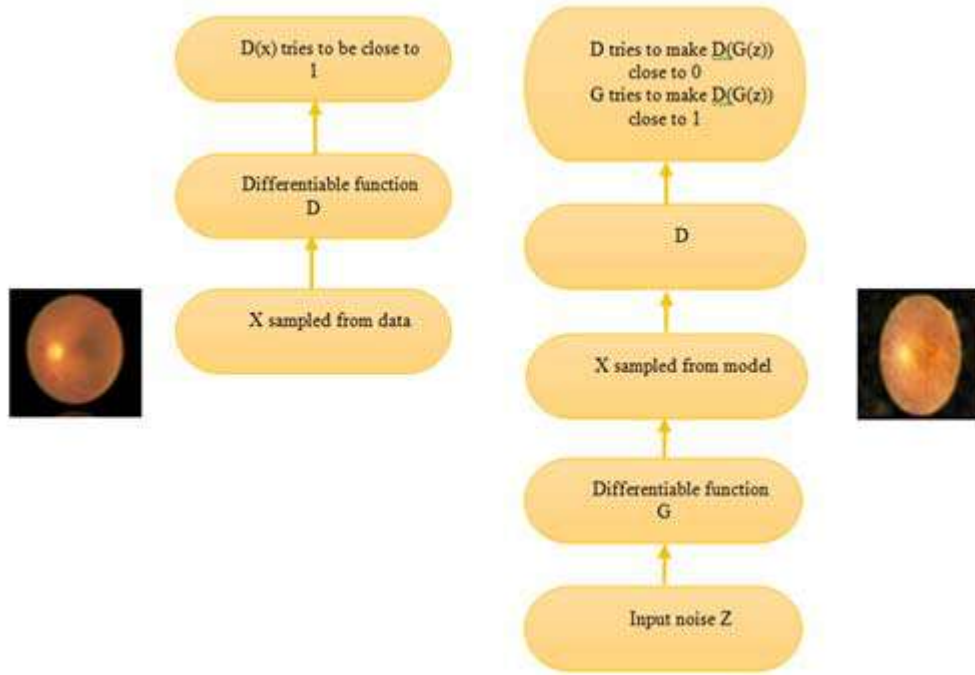


**Figure 1:** Adversarial Nets Framework

## 1.2. Training process in GAN:

Since the generative model and the discriminant model in the generative adversarial network are completely independent models and cannot be trained at the same time, separate repetitive training is used. For the discriminant, the label for the true sample is marked with the number 1, and the label for the generated sample is marked as 0, regardless of the quality of the sample generated.

Updating the discriminator by ascending its stochastic gradient:

$$\nabla_{\theta_d} \frac{1}{m} \sum_{i=1}^m \log D(x^{(i)}) + \log(1 - D(G(z^{(i)}))) \quad (2)$$

Updating the generator by descending its stochastic gradient:

$$\nabla_{\theta_g} \frac{1}{m} \sum_{i=1}^m \log(1 - D(G(z^{(i)}))) \quad (3)$$

We train D with the hope that  $V(G, D)$  is close to 1, so we add a gradient (ascending). When training G in the second step,  $V(G, D)$  is as small as possible close to 0, so the gradient is subtracted. And then the entire training process rotates.

## 1.3. Different types of GANs

There have been many types of GAN since its first appearance in 2014, including Convolutional GAN (CGAN), Deep Convolutional GANs (DCGANs), Least Squares GANs(LSGAN), Boundary Equilibrium GANs(BEGAN), StackGAN, InfoGANs, Wasserstein GANs (WGAN) and DiscoGANs. The dcgan that has been used in this paper will be explained briefly [4], [5], [6], [7], [8].

### 1.3.1. Concept of DCGAN

DCGAN is Deep Convolutional Generative Adversarial Network, the same as GAN, so we won't repeat it here. It only replaces the G and D above with two convolutional neuronal networks (CNN). But the change is not directly enough, DCGAN made some changes to the structure of the



convolutional neural network to improve sample quality and convergence velocity. These changes include:

- Instead of using pooling layers use stride convolutions (discriminator) and fractional-stride convolutions (generator).
- Use batch normalization in both D and G.
- Remove the FC layer and make the network a fully convolutional network.
- The G ReLU network is used as the activation function, and the last layer uses tanh.
- Use LeakyReLU as the activation function in the D network.

Figure 2 shows G-Network in DCGAN [3].

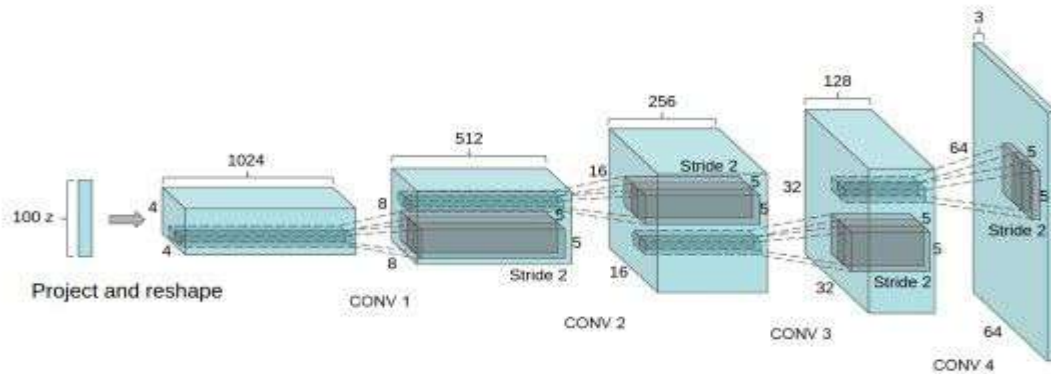


**Figure 2:** G-Network in DCGAN

## 2. Related work

several researches have been published in the study of GAN. Most of these studies done recently. Emphasis will be placed on increasing the sample size of the database, a few reviews are as follows:

Frid-Adar, Maayan, et al. [9] In this article, the authors proposed an enlargement scheme based on a combination of standard imaging and artificial liver lesions using GANs to improve the classification of liver lesions. Using the GAN, they pooled high-quality focal liver lesions from CT images and developed CNN for the liver lesion classification task and augmented the CNN training set using synthetic data created to improve classification outcomes.

The classification efficiency gave a sensitivity of 78.6% and a specificity of 88.4%. By adding the synthetic data, they obtained a sensitivity of 85.7% and a specificity of 92.4%.

Iqbal, Talha, and Hazrat Ali. [10] in this work authors propose a new GAN for Medical Images (MIGAN). The MIGAN generates synthetic retinal image with their segmented masks, which will be used to apply a controlled analysis of medical images. One of the features of the proposed model is the ability to learn features from a few dataset.

Shmelkov, Konstantin, Cordelia Schmid, and Karteek Alahari. [11] The authors suggest two measures based on the classification of images, GANTrain and GANTest, which come close to recall and precision. They introduce new assessment metrics to compare class conditional GAN structure with GANTrain and GANTest scores. They used NN architecture for the classification of images for these two measurements. Authors train their networks with generated images to calculate the GANTrain. Then they obtained performance on a test set which collected from real images.

Bowles, Christopher, et al. [12] this article explains the feasibility of injecting synthetic data obtained from the GAN into the original datasets in two brain segmentation problems, resulting in an improvement in the dice similarity rate by 1 to 5 percentage points under various conditions. They used a Progressive Growing of GANs (PGGAN) network to generate synthetic data. PGGAN was chosen, because the ability of training at large image sizes. They added the synthetic data with the real data to increase their dataset. The best result obtained when the authors used GAN + Rotation augmentation.

Zeid Baker, Mousa. [13] the researcher in this paper used Generative Adversarial Networks to generate synthetic images similar to the real dataset in order to expand a dataset. they used two types

of experiments were achieved, the first is using fine-tune a Deep Convolutional Generative Adversarial Network for a specific dataset, while the second experiment was used to analyze how synthetic data affect the accuracy in a classification task. authors used three types of datasets MNIST, Fashion-MNIST and Flower photos. The authors conclusion that DCGAN leads to an increase the model accuracy depends on the type of the dataset and the data preprocessing plays a big roll on a DCGANs performance, as for almost any ML algorithm.

Wu, Qiufeng, Yiping Chen, and Jun Meng. [14] in this work, authors used Generated images augmented by deep convolutional generative adversarial networks (DCGAN) and original images to Identify Tomato Leaf Disease. They used GoogLeNet classifier to training and testing 5 classes of tomato leaf images, this model achieved accuracy of 94.33%. the authors result that images generated by DCGAN not only enlarge the size of the data set, but also have the characteristics of diversity, which makes the model have a good generalization effect.

### 3. Material and method

#### 3.1. Datasets

The dataset was analyzed and preprocessed of four different classes which contain 1692 of Glaucoma, Myopia, Diabetic retinopathy and Normal images, were selected for this study to aim for the highest variance among classes. where 1200 images used for training, 246 images for testing and 246 images used for validation purpose, which mean 15% of the total images were used for testing and 15% for validation purpose. All the images were collected in total from Kaggle dataset and iChallenge-GON Comprehension, in high resolution images.



**Figure 3:** Images of Normal Fundus, Glaucoma, Diabetic retinopathy and Myopia

The size of all images should be the same, colored and png format. In machine learning and deep learning, accuracy of model will be affected if the samples are not equally distributed [15].

**Table 1**

Eye diseases samples in original dataset

| Diseases             | Training | Validation |
|----------------------|----------|------------|
| Glaucoma             | 300      | 48         |
| Diabetic retinopathy | 300      | 66         |
| Myopia               | 300      | 66         |
| Normal               | 300      | 66         |
| Total                | 1200     | 246        |

In addition, 300 images have been used of each class to train our model. There are two issue should be considered: The number of parameters when the number of network layers increases, and the small number of eye diseases which we have collected. That will lead to the overfit under the influence of many parameters and few datasets. Increasing of the dataset is an effective way to solve this issue. In this paper, authors proposed DCGAN network to increase eye diseases dataset by generate synthetic dataset. Therefore, the training data set, and validation set will be increased (data enhancement).



## 3.2. Model building

The main goal in this work is to generate images that look like realistic images in each class in order to solve the problem of insufficient medical images. In addition, using GMD model to classify eye diseases with and without DCGAN.

### 3.2.1. GMD model

As shown in Figure 4, the size of input image is set to 64×64 pixels with 3 RGB channels. To extract the features from the image, two convolution layers were used: the first is 256 filters of size 3×3 pixels and the second is 128 filters of size 3×3 pixels. For the pooling layer, a window of size 2×2 pixels were used, which compresses the original image size for further processing. After that, another three convolution layers were used of 128, 64, and 32 filters with size 3×3 pixels with a maximum pooling size 2×2 pixels. Then, fully connection is used (Dense 100 units and 60 units) and output layer (4 units) to predict the eye cases. CNNs adjust their filter weights through backpropagation, which means that after the forward pass, the network is able to look at the loss function and make a backward pass to update the weights.

The GMD model performance was evaluated against training, testing, and validation datasets using the accuracy measure, which reflects the ability of the model to classify the whole database into four categories with almost no error. The following represents the accuracy measure equation [16]:

$$\text{Accuracy} = \frac{TP}{TP + FP + FN + TN} \quad (4)$$

Where, TP is True Positive: Your predicted positive and it is true, TN is True Negative: Your predicted negative and it is true, FP is False Positive: your predicted positive but it is false and False Negative FN: Your predicted negative but it is false.

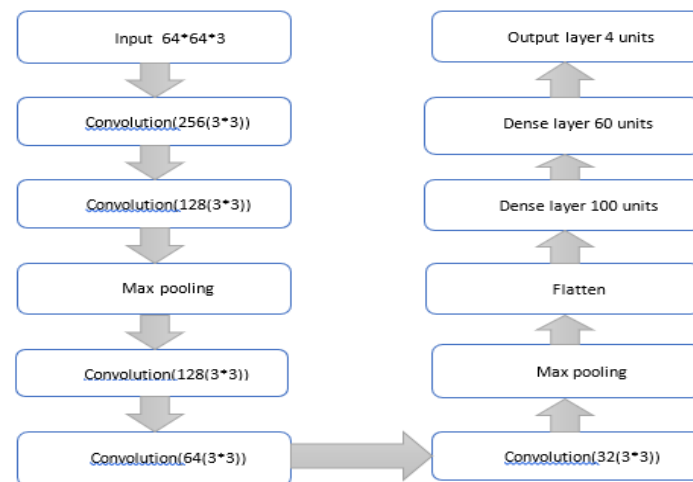


Figure 4: Block diagram of GMD model

### 3.2.2. DCGAN architecture

GAN consists of two separate neural networks, Generator (G) and Discriminator (D) constantly competing against each other, the generator, which tries to generate examples that seem the real data, it works to deceive the discriminator, while the role of discriminative network is to distinguish between the real and fake data. A vector of random number will be fed to the generator, and the outputs are unreal synthetic data denoted by G (z), the discriminator feeds real data and the output of generator. The output of the discriminator is the possibility whether the entered data is real or fake, the output denoted by D (x). The output D (x) represents the probability that x will be a real image if it is 1, which means 100% is a real image, and the output is 0 which means that the image cannot be real [17].

Objective function and how GAN is work, we mentioned above in introduction in section 1. DCGAN is Deep Convolutional GAN, it only replaces the G and D with two convolutional neuronal networks (CNN). Compared with GAN, DCGAN made some changes to the CNN architecture to improve sample quality and convergence speed. These changes are mentioned above in section 1.3.1. In this paper, the block diagram of eye diseases identification is shown in Figure 5. We also will interduce the proposed generator and discriminator network used by referring to the DCGAN architectures as shown in Figure 6:

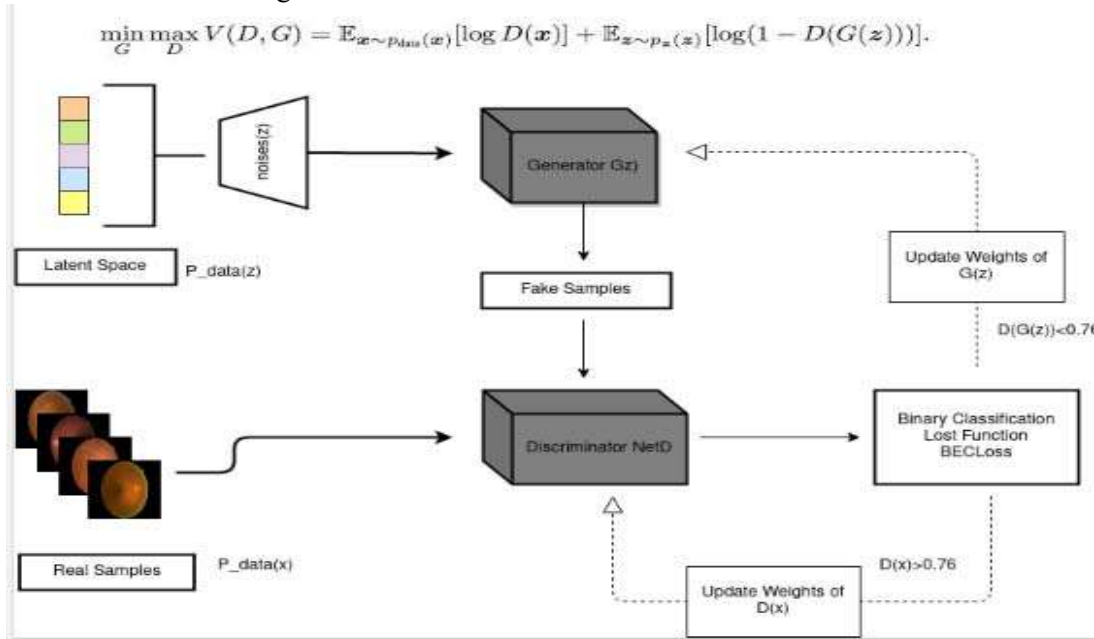


**Figure 5:** Block diagram of eye diseases identification

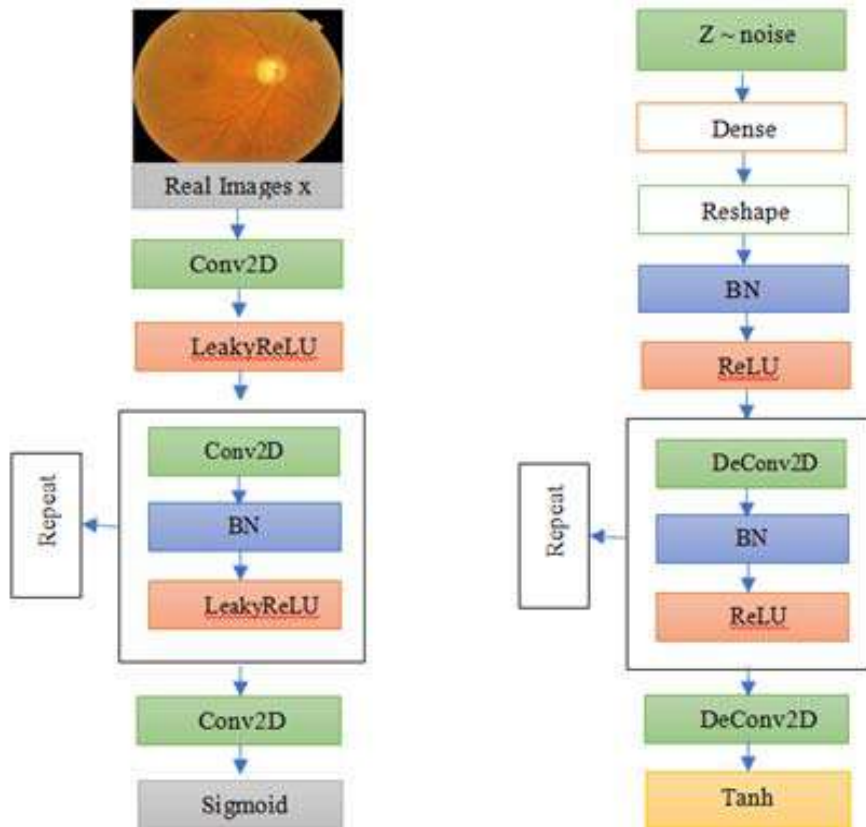


**Figure 6:** discriminator (left) and generator architecture (right)

## 4. Experiments and Results

According to the models explained above, all these models were implemented by Python language (Google Colab) using dual Graphics Processing Unit (GPU). The main objective of this paper is to increase our dataset using DCGAN. These data will be fed to GMD model in order to improve eye diseases classification. Accuracy will be measured before and after adding synthetic dataset. Dataset has been explained in section 3.1. it consists of four types of eye diseases, Glaucoma, Diabetic retinopathy, Myopia and Normal eyes. Table 1 shows the number of samples of each class in training set and validation set.

### 4.1. DCGAN results

DCGAN is to increase our dataset in order to improve the diagnosis of eye diseases. The schema of our experiment as shown in Figure 5, has been applied. We utilized our dataset which consists of 1200 training samples, 246 validation samples of eye diseases size 64x64 coloured images for four general classes. We used the same network structure for the generator and discriminator as shown in figure 6, and training our medical images, the parameters of the discriminator and generator are updated in the same number of iterations. During this process BCElose will be calculated on all real and fake batch, and calculate gradients for D and G in backward propagation.

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <pre> # (1) Update D network: maximize log(D(x)) + log(1 - D(G(z))) ## Train with all-real batch netD.zero_grad() # Format batch real_cpu = data[0].to(device)[:1] b_size = real_cpu.size(0) label = torch.full((b_size,), real_label, dtype=torch.float, device=device) # Forward pass real batch through D output = netD(real_cpu).view(-1) # Calculate loss on all-real batch errD_real = criterion(output, label) # Calculate gradients for D in backward pass errD_real.backward() D_x = output.mean().item() ## Train with all-fake batch # Generate batch of latent vectors noise = torch.randn(b_size, nz, 1, 1, device=device) # Generate fake image batch with G fake = netG(noise) label.fill_(fake_label) # Classify all fake batch with D output = netD(fake.detach().to(device)[:1]).view(-1) # Calculate D's loss on the all-fake batch errD_fake = criterion(output, label) # Calculate the gradients for this batch </pre> | <pre> errD_fake.backward() D_G_z1 = output.mean().item() # Add the gradients from the all-real and all-fake batches errD = errD_real + errD_fake # Update D optimizerD.step() # (2) Update G network: maximize log(D(G(z))) netG.zero_grad() label.fill_(real_label) # fake labels are real for generator # Since we just updated D, perform another forward pass of all-fake batch through D output = netD(fake.to(device)[:1]).view(-1) # Calculate G's loss based on this output errG = criterion(output, label) # Calculate gradients for G errG.backward() D_G_z2 = output.mean().item() # Update G optimizerG.step() </pre> |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

For the discriminant, the label for the true sample is marked with the number 1, and the label for the generated sample is marked as 0, regardless of the quality of the sample generated. So,  $D(x)$

should be close to 1 and  $D(G(z))$  should be close to 0. In our experiment we save just the images that  $D(G(z)) \geq 0.76$  which close to 1.

$x=0.76$

```
if errD_fake >= float(x) or errD_fake == errD_real :
```

```
vutils.save_image(vutils.make_grid(fake[:1].to(device)[:1], padding=2, normalize=True), "%s/fake_samples_epoch_%03d.png" % ("/content/drive/MyDrive/GAN2/FakeImages", epoch), normalize = False)
```

Adam was used as the gradient method for learning parameters of model. Its initial learning rate is 0002.

```
optimizerD = optim.Adam(netD.parameters(), lr=lr, betas=(beta1, 0.999))
```

```
optimizerG = optim.Adam(netG.parameters(), lr=lr, betas=(beta1, 0.999))
```

The number of epochs we have used is 4000, that is not mean 4000 images will be generated, because we save the images that  $D(G(z)) \geq 0.76$ . Table 2 shows the synthetic images and total images which we obtained. Figure 7 shows samples of the fake and real images.

**Table 2**

Eye diseases samples after using DCGAN

| Diseases             | synthetic data using DCGAN | Original data + synthetic data |            |
|----------------------|----------------------------|--------------------------------|------------|
|                      |                            | Training                       | Validation |
| Glaucoma             | 318                        | 618                            | 48         |
| Diabetic retinopathy | 294                        | 594                            | 66         |
| Myopia               | 428                        | 728                            | 66         |
| Normal               | 306                        | 606                            | 66         |
| Total                | 1346                       | 2546                           | 246        |

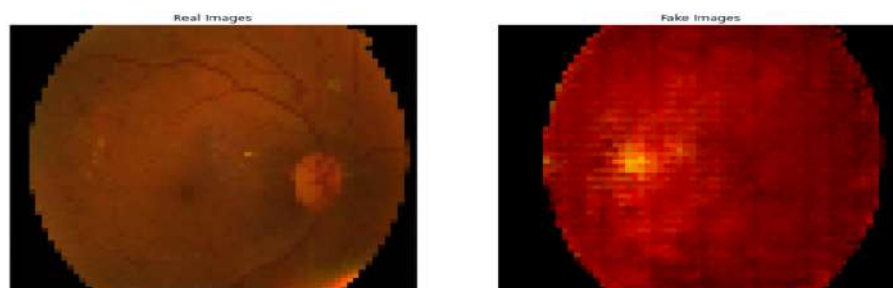


**Figure 7:** sample of real and fake image

The figures below show the loss during training in both Generator and Discriminator for all classes:

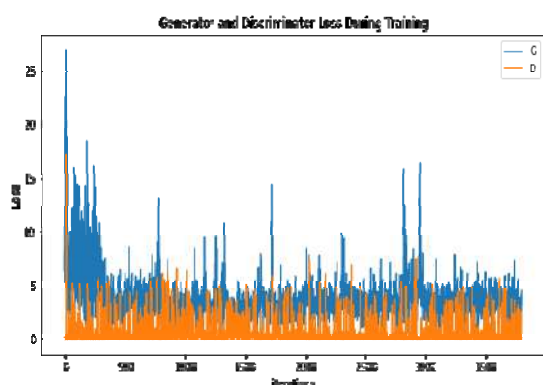


**Figure 8:** Loss in Glaucoma

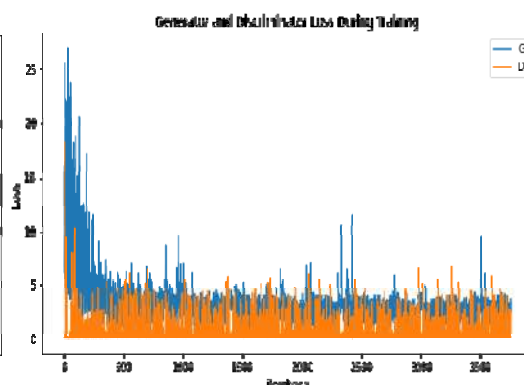


**Figure 9:** Loss in Diabetic Retonaphthy

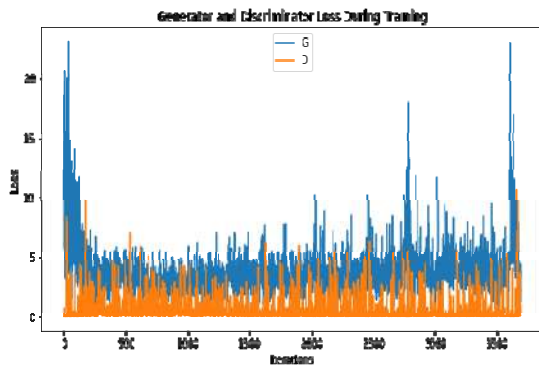


Figure 10: Loss in Myopia

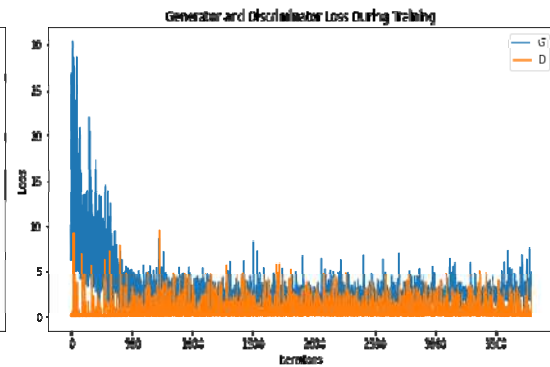


Figure 11: Loss in Normal

## 4.2. GMD model Results with Original data

GMD model has been applied, Constant learning rate is the default learning rate schedule in SGD. Both momentum and decay rate are set to zero. Learning rate was set to 0.001 which showed a good performance to start. In constant learning rate schedule, the model was defined and the learning rate value was set to constant as follows.

```
# define CNN_model
model1 = CNN_model()
epochs = 100
# define SGD optimizer and set to default except lr
learning_rate = 0.001
sgd = SGD(lr=learning_rate, momentum=0.0, decay=0.0, nesterov=False).
```

Then, compiling, fitting GMD model and plotting the model accuracy as shown in Figure.12. We obtained Accuracy on training set 76.58% and Accuracy on validation set 76.42%.

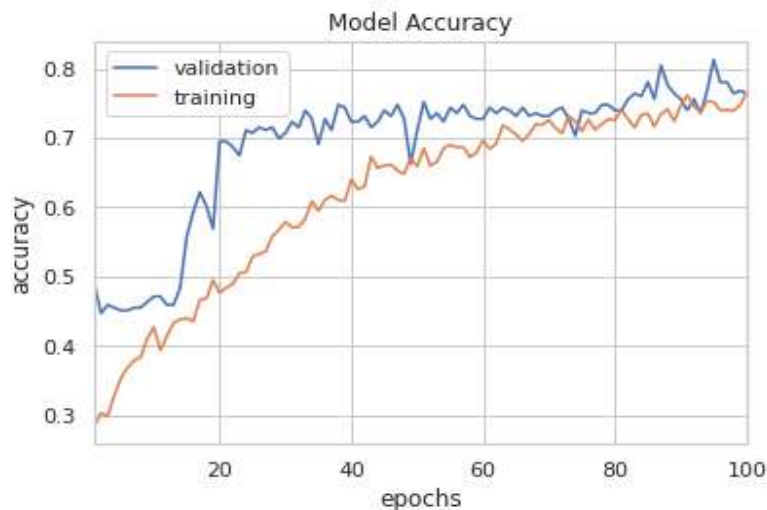


Figure 12: model accuracy using constant learning rate without data generated

### 4.3. GMD model Results with synthetic data

The model has been applied and constant learning rate is the default learning rate schedule in SGD as mentioned in paragraph 4.1. Then, compiling, fitting our model and plotting the model accuracy as shown in Figure.13. The result obtained, accuracy on training set 80.45% and accuracy on validation set 83.74%.

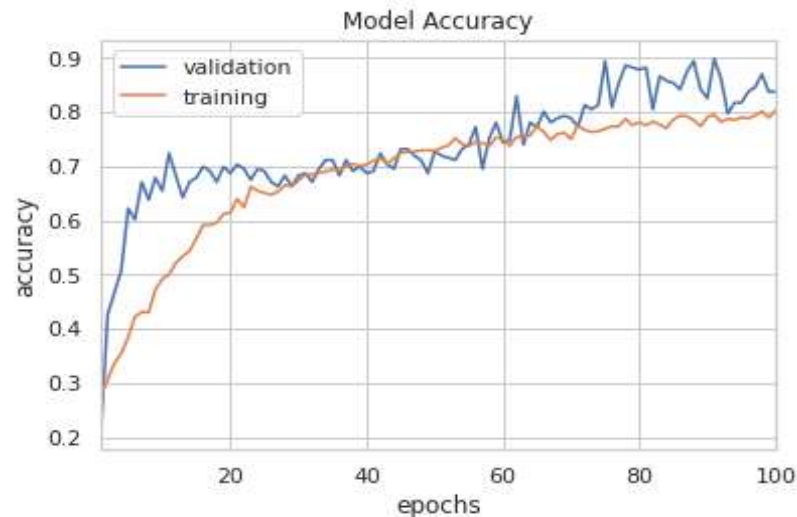


**Figure 13:** model accuracy using constant learning rate with data generated

## 5. Conclusion

In this research, a Deep Convolutional Generative Adversarial Network, and GMD model based on Keras, TensorFlow and pycharm has been deployed using Python on a database of four types of ophthalmic cases: Glaucoma, Myopia, Diabetic retinopathy, and Normal eyes. The implemented DCGAN and GMD model have been trained using GPU in Google Colab, in order to increase our dataset to improve eye diseases classification.

In conclusion, DCGAN strongly can be used for generating high quality of medical images and increasing training data. Furthermore, the performance of GMD model with original data and with synthetic images data were compared. We confirmed that with the help of DCGAN, GMD model (CNN model) can yield a better *accuracy* compared to traditional methods. Moreover, the *accuracy* of the model had improved significantly from 76.58% in training set and 76.42% in validation set, to 80.45% in training set and 83.74% in validation set. Based on this work, we suggest that DCGAN can be optimized for multi-class generator. We also suggest that this work can be applied to other image classification models such as Vgg16, Inception v3, ResNet to enhance the accuracy.

We noticed that the images we have obtained from DCGAN had some noise. Therefore, in the next paper we will use denoising Autoencoder to enhance these images to improve the model accuracy.

## 6. References

- [1] H. Greenspan, B. van Ginneken, and R. M. Summers, "Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1153–1159, May 2016.
- [2] Mustaffa Hussain, GANs-What and Where?, 2020. URL: <https://medium.com/thecyphy/gans-what-and-where-b377672283c5>.
- [3] Goodfellow, Ian J., et al. "Generative adversarial networks." *arXiv preprint arXiv:1406.2661* (2014).
- [4] Radford, Alec, Luke Metz, and Soumith Chintala. "Unsupervised representation learning with deep convolutional generative adversarial networks." *arXiv preprint arXiv:1511.06434* (2015).

- [5] Mirza, Mehdi, and Simon Osindero. "Conditional generative adversarial nets." arXiv preprint arXiv:1411.1784 (2014).
- [6] Mao, Xudong, et al. "Least squares generative adversarial networks." Proceedings of the IEEE international conference on computer vision. 2017.
- [7] Berthelot, David, Thomas Schumm, and Luke Metz. "Began: Boundary equilibrium generative adversarial networks." arXiv preprint arXiv:1703.10717 (2017).
- [8] Martín Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein Generative Adversarial Networks. International Conference on Machine Learning, ICML, 2017.
- [9] Frid-Adar, Maayan, et al. "Synthetic data augmentation using GAN for improved liver lesion classification." *2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018)*. IEEE, 2018.
- [10] Iqbal, Talha, and Hazrat Ali. "Generative adversarial network for medical images (MI-GAN)." *Journal of medical systems* 42.11 (2018): 1-11.
- [11] Shmelkov, Konstantin, Cordelia Schmid, and Karteek Alahari. "How good is my GAN?." Proceedings of the European Conference on Computer Vision (ECCV). 2018.
- [12] Bowles, Christopher, et al. "Gan augmentation: Augmenting training data using generative adversarial networks." arXiv preprint arXiv:1810.10863 (2018).
- [13] Zeid Baker, Mousa. "Generation of Synthetic Images with Generative Adversarial Networks." (2018).
- [14] Wu, Qiufeng, Yiping Chen, and Jun Meng. "DCGAN-based data augmentation for tomato leaf disease identification." *IEEE Access* 8 (2020): 98716-98728.
- [15] F. J. Pulgar, A. J. Rivera, F. Charte, and M. J. del Jesus, "On the impact of imbalanced data in convolutional neural networks performance," in Proc. Int. Conf. Hybrid Artif. Intell. Syst., vol. 10334. Cham, Switzerland: Springer, 2017, pp. 220–232.
- [16] Visa, Sofia, et al. "Confusion Matrix-based Feature Selection." *MAICS* 710 (2011): 120-127.
- [17] M. Mirza and S. Osindero, "Conditional generative adversarial nets," 2014, arXiv:1411.1784. [Online]. Available: <http://arxiv.org/abs/1411.1784>
- [18] Kong, Quan, et al. "Active generative adversarial network for image classification." Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 33. No. 01. 2019.
- [19] Khalifa, Nour Eldeen M., et al. "Detection of coronavirus (covid-19) associated pneumonia based on generative adversarial networks and a fine-tuned deep transfer learning model using chest x-ray dataset." arXiv preprint arXiv:2004.01184 (2020).



# Intelligent Big Data system based on scientific machine learning of cyber-physical systems of medical and biological processes

Vasyl Martsenyuk<sup>a</sup>, Aleksandra Klos-Witkowska<sup>a</sup>, Andriy Sverstiuk<sup>b</sup>, Oksana Bahrii-Zaiats<sup>b</sup>, Marcin Bernas<sup>a</sup>, Krzysztof Witos<sup>a</sup>

<sup>a</sup> University of Bielsko-Biala, Willowa St., 2, Bielsko-Biala, 43-300, Poland

<sup>b</sup> I. Horbachevsky Ternopil National Medical University, Maidan Voli St., 1, Ternopil, 46002, Ukraine

## Abstract

The work focuses on developing a Big Data system allowing us to process medical data intelligently, which are ingested from different sources and of various types. It aims to design biosensor devices with desired qualitative characteristics, namely their operational stability. The article suggests the mathematical representation of the discrete population dynamics and the dynamic logic of the studied models of the most advanced biosensors as biopixels arrays which are used at the analytics stage of scientific machine learning. Application software for the intelligent Big Data system for investigating the stability of cyber-physical systems for medical purposes using the R package has been developed. The software package for the intelligent Big Data system for investigating the stability of cyber-physical systems for medical applications consists of the following main software modules: unit of ingesting and processing Big Data used to identify input parameters of the model of cyber-physical systems; the software module of research of dynamic behavior of cyber-physical systems; the software module of research of dynamic logic of cyber-physical systems; the block of decision-making on the stability of cyber-physical systems; the block of visualization, lattice images of macrophage/monoclonal antibody data source, a database that receives lattice images of macrophage binding to monoclonal antibodies and a database that receives images of fluorescent biopixels. Using the software package, the results of computer simulation of mathematical models of cyber-physical systems of medical and biological processes in the form of images of macrophage, monoclonal antibodies, connections of macrophages with monoclonal antibodies, fluorescent pixels, and an electrical signal from the converter are obtained.

## Keywords 1

cyber-physical system, Big Data, stability of the model, differential equations, difference equations, rectangular lattice, hexagonal lattice

## 1. Introduction

The most popular areas of use of Big Data in medicine are diagnostics using medical images in ophthalmology [1-3], oncology [4, 5], dermatology [6], cardiology [7]. These are the branches of modern medicine where most images are used for diagnosis.

Over the last decade, we have seen a rapid growth of digital developments, including a variety of sensors, microcontrollers, communication systems to meet societal needs. Thanks to the Internet of Things (IoT) and the use of biosensors, digital electronic health services are increasingly used every year. Biosensors play a critical role in the Internet of Things when it comes to eHealth. There are a

---

CMIS-2021: The Fourth International Workshop on Computer Modeling and Intelligent Systems, April 27, 2021, Zaporizhzhia, Ukraine  
EMAIL: vmartsenyuk@ath.bielsko.pl (V. Martsenyuk); awitkowska@ath.bielsko.pl (A. Klos-Witkowska); sverstyuk@tdmu.edu.ua (A. Sverstiuk); bagrijayats@tdmu.edu.ua (O. Bahrii-Zaiats); mbernas@ath.bielsko.pl (M. Bernas); kwitos@ath.bielsko.pl (K. Witos)  
ORCID: 0000-0001-5622-1038 (V. Martsenyuk); 0000-0003-2319-5974 (A. Klos-Witkowska); 0000-0001-8644-0776 (A. Sverstiuk); 0000-0002-5533-3561 (O. Bahrii-Zaiats); 0000-0002-0099-1647 (M. Bernas); 0000-0001-7959-7329 (K. Witos)



© 2021 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)



number of sensors available on the market that help people monitor their daily fitness, blood glucose levels, and many sensible applications of primary diagnostics at home. In [8] the problems of safe and confidential use of the most modern biosensors for the Internet of Medical Things are considered.

The architecture of intelligent monitoring of electronic health of patients with chronic diseases is presented in [9]. Elements of the architecture include portable devices, biosensors, and smartphones for collecting medical and biological indicators [10]. The intelligent system uses scientific machine learning (SciML) as a tool for analytics in the context of complex applications across science, engineering, and medicine [11, 12]. It focuses on Big Data, obtained from various hospitals, and data obtained from the patient, to diagnose and generate warnings of critical and crisis situations. The system of health monitoring using prognostic calculations using Big Data analysis is presented in [13]. The use of Big Data in endoprosthesis is considered in [14]. Google works with US clinics to track searches for the COVID-19 pandemic and to predict in which region the outbreak will occur [15].

In 2013, thanks to an electronic medical card, which stored all the information about the client, scientists established a relationship between brain degradation and diabetes. Scientists from the University of Cape Town, thanks to a correct analysis using Big Data, analyzed the most common types of cancer and found that malignant tumors of the intestine, lung and ovary have clear genetic markers. Medical Big Data helps prevent the development of the disease at an early stage through the analysis of cardiovascular pressure, pulse, respiration and blood sugar [16, 17].

A few years ago, Apple and IBM joined forces to give IBM Watson even more structured information from the iPhone and Apple Watch. One of the most promising vectors of Big Data development in medicine is surgery [18, 19], where the relevant systems analyze all available protocols of operations in order to assess the possible risks of surgical interventions in the form of visualization in different colors. At home, Americans simply connect a stethoscope to the iPhone and send the data to their doctor. At the same time, a large cluster collects information on all patients. It is sorted by criteria and then looks for dependencies, confirm or refute myths.

Nowadays, the integration between physical resources and computing leads to the creation of complicated computing systems with distributed parameters. Such systems are controlled or managed by integrated into the Internet computing resources [20, 21]. In recent years, we have seen an interest in cyber-physical systems (CPS), which are tools for controlling and monitoring the studied indicators using computer technology, in which software is closely related to physical objects. CPS for medical applications are used to measure and analyze biochemical parameters in biological fluids, detection of cancer, markers of heart failure, pathogenic bacteria, to determine the level of contamination of food and the environment. In recent years, CPSs have been used effectively, in which the function of selective elements is performed by biosensors. Such systems are called cyber-physical biosensor systems (CPBSS), which have a number of advantages: high selectivity, stability, performance, affordable cost, the ability to use a wide range of users.

As a result of the use of CPBSS, Big Data is obtained, which requires processing and interrelation with the measured medical and biological parameters. The analytical analysis of the obtained data in the form of useful knowledge for use by software services that control the studied systems is carried out.

For Big Data analysis, SciML is used as a data-driven method for studying computer algorithms that are automatically improved through experience [22, 23].

In CPS, “experience” is usually presented in the form of data, and SciML allows you to create mathematical models based on sample data for forecasting and decision making. SciML has made significant strides in a variety of fields, including computer vision, language recognition, and control systems, as well as related scientific fields in physics, chemistry, and biology [24].

Processing data ingested from CPS of medical and biological processes is one of the important solutions which fits the “5 V’s of Big Data” characterization of Big Data [25]. The volume of such type of data grows drastically. So, the amount of healthcare data gathered from 2013 to 2020 has grown 15 times to 2314 exabytes [26]. The functioning of most CPS in medical applications is in real-time mode, corresponding to the requirements of velocity for BigData. Data sources used for medical CPS varies from numerical and literal signals (ECG, EEG, EMG, etc.) to medical images (CT, MRT, USG, etc.), which fits a variety of data sources. The veracity of medical CPS data is assured by the quality of standardized measurements used. Finally, the Value of Big Data solutions within medical

data-driven-CPS is supported by new results on diagnostics, treatment, and prophylaxis, which can be exceptionally obtained using SciML.

Algorithms and methods for processing measurement parameters in the CPS of medical and biological processes using Big Data are created on the basis of their mathematical models. Mathematical models of biosensors for CPBSS on rectangular lattice using lattice differential equations with delay have been developed [27, 28]. The mathematical models of biosensors for CPBSS on rectangular lattice using lattice difference equations with time delay are developed in the article [29]. A detailed description of the considered model is presented in [30, 31].

When developing the modern CPS with the desired qualitative characteristics we need to process the medical information of the different kinds, which corresponds to the notion of Big Data. Namely, the medical data fit the main features of Big Data such as variety, value, velocity, and veracity.

Medical data in CPBSS come from different sources and types, so in their operation it is necessary to ensure independence and operational stability. Therefore, there is a need to develop a software package for the intelligent system BigData in the study of the sustainability of medical CPS. These studies are particularly relevant for autonomy, which is defined as the improvement of enzyme activity, protein when maintained under certain conditions, as well as operational stability to ensure the performance of CPS when measuring biomedical parameters. The usefulness of the studied systems is determined by the stability of the sensitive element, which is located in the receptor, and the functioning of the corresponding components of the matrix of the system during its use.

The purpose of the study: to develop a software package for the intelligent system BigData in the study of the stability of medical CPS.

## 1. Mathematical models of biosensors of the intelligent Big Data system for medical applications

### 1.1. Mathematical model of biosensor on rectangular lattice using differential equations with time delay

In models of lattice differential equations, the spatial structure has a discrete character and lattice dynamics is widely used in the examples considered in [30, 31] and the mathematical model of the immunosensor, which characterizes the change in the concentration of macrophages  $M_{i,j}(t)$  and monoclonal antibodies  $A_{i,j}(t)$  was substantiated:

$$\begin{aligned} \frac{dM_{i,j}(t)}{dt} &= (\beta - \gamma A_{i,j}(t - \tau) - \delta_v M_{i,j}(t - \tau)) M_{i,j}(t) + \hat{S}\{M_{i,j}\}, \\ \frac{dA_{i,j}(t)}{dt} &= (-\mu_f + \eta \gamma M_{i,j}(t - \tau) - \delta_f A_{i,j}(t)) A_{i,j}(t), \quad t > 0, \end{aligned} \quad (1)$$

where  $\gamma$  - the speed of detection of macrophages;

$\beta$  - birth rate constant of macrophages;

$\tau$  - time delays in the formation of the macrophage-monoclonal antibody complex;

$\delta_M$  - the average rate of decrease in the concentration of macrophages;

$\delta_A$  - the average rate of decrease in the concentration of monoclonal antibodies;

$\eta$  - probabilistic rate of formation of the macrophage-monoclonal antibody complex;

$\Delta$  - distance between immunopixels;

$D$  - macrophage diffusion coefficient;

$\hat{S}\{M_{i,j}\}$  - operator of diffusion processes.

The model of the immunosensor (1) using lattice differential equations most fully takes into account all parameters of immunosensory systems, namely: to take into account when measuring the concentration of macrophages  $M_{i,j}(t)$  and monoclonal antibodies  $A_{i,j}(t)$  and their binding properties, constant fertility of macrophages, time delay formation of a macrophage-monoclonal antibody complex; the average rate of decrease in the concentration of macrophages and monoclonal

antibodies; probabilistic rate of formation of the macrophage-monoclonal antibody complex; the distance between the immunopixels; macrophage diffusion coefficient; diffusion process operator.

These parameters are used by immunosensory systems.

Model (1) is given by initial functions (2)

$$\begin{aligned} M_{i,j}(t) = M_{i,j}^0(t) \geq 0, \quad A_{i,j}(t) = A_{i,j}^0(t) \geq 0, \\ t \in [-\tau, 0), \quad M_{i,j}(0), A_{i,j}(0) > 0. \end{aligned} \quad (2)$$

For a square array  $N \times N$ , the discrete diffusion is used for the spatial operator, taking into account the imbalance constant  $n_{dsbn}$

$$\hat{S}\{M_{i,j}\} = \begin{cases} D\Delta^2[M_{1,2} + M_{2,1} + M_{i,j-1} - 2n_{dsbn}M_{1,1}] & i, j = 1 \\ D\Delta^2[M_{2,j} + M_{1,j-1} + M_{1,j+1} + M_{i,j+1} - 3n_{dsbn}M_{i,j}] & i = 1, j \in \overline{2, N-1} \\ D\Delta^2[M_{1,N-1} + M_{2,N} - 2n_{dsbn}M_{1,N}] & i, j \in \overline{2, N-1} \\ D\Delta^2[M_{i-1,N} + M_{i+1,N} + M_{i,N-1} - 3n_{dsbn}M_{i,N}] & i \in \overline{2, N-1}, j = N \\ D\Delta^2[M_{N-1,N} + M_{N,N-1} - 2n_{dsbn}M_{N,N}] & i = N, j = N \\ D\Delta^2[M_{00,N-1,j} + M_{N,j-1} + M_{N,j+1} + M_{i,j+1} - 3n_{dsbn}M_{N,j}] & i = N, j \in \overline{2, N-1} \\ D\Delta^2[M_{N-1,1} + M_{N,2} - 2n_{dsbn}M_{N,1}] & i = N, j = 1 \\ D\Delta^2[M_{i-1,1} + M_{i+1,1} + M_{i,2} - 3n_{dsbn}M_{i,1}] & i \in \overline{2, N-1}, j = 1 \\ D\Delta^2[M_{i-1,j} + M_{i+1,j} + M_{i,j-1} + M_{i,j+1} - 4n_{dsbn}M_{i,j}] & i, j \in \overline{2, N-1}. \end{cases} \quad (3)$$

Each pixel is exposed to macrophages from four adjacent pixels that are separated by equal distances  $\Delta$ .

The boundary condition  $M_{i,j} = 0$  for the edges of the array  $i, j = 0, N + 1$  is used.

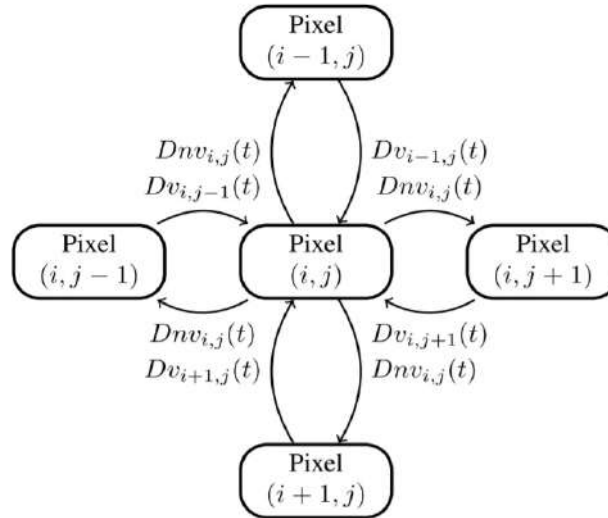


**Figure 1:** A rectangular grid that connects four adjacent pixels in the biosensor model using a Cartesian coordinate system

## 1.2. Mathematical model of biosensor on hexagonal lattice using lattice difference equations with time delay

For the intelligent Big Data system of medical and biological processes dynamics, we use the mathematical description with the help of nonlinear difference equations with delay.

The model of the biosensor on the basis of a hexagonal lattice is considered. In this case, for the numbering of biopixels  $(i, j, k)$ ,  $i, j, k = \overline{-N, N}$ ,  $i + j + k = 0$  the cubic coordinate system is used [32, 33].

Let  $M_{i,j,k}(t)$  is the concentration of macrophages,  $A_{i,j,k}(t)$  is the concentration of monoclonal antibodies in the biopixel  $(i, j, k)$ ;  $i, j, k = \overline{-N, N}$ ,  $i + j + k = 0$ .

In the model it is used next parameters:

$\gamma > 0$  - the speed of detection of macrophages;

$\beta$  - birth rate constant of macrophages;

$\tau$  - time delays in the formation of the macrophage-monoclonal antibody complex;

$\delta_M$  - the average rate of decrease in the concentration of macrophages;

$\delta_A$  - the average rate of decrease in the concentration of monoclonal antibodies;

$\eta$  - probabilistic rate of formation of the macrophage-monoclonal antibody complex;

$\Delta > 0$  - distance between immunopixels;

$D$  - macrophage diffusion coefficient;

$\hat{S}\{M_{i,j}\}$  - operator of diffusion processes.

When colonies of monoclonal antibodies are absent, colonies of macrophages are regulated by a logistic equation with time delay:

$$M_{i,j,k}(n+1) = (1 + \beta - \delta_v M_{i,j,k}(n-r)) M_{i,j,k}(n), \quad (4)$$

where  $\beta$  and  $\delta_v$  – positive numbers, and  $r \geq 0$  mean latency of the negative response of the macrophages' colonies.

The diffusion of macrophages from six adjacent pixels is considered  $(i+1, j, k-1)$ ,  $(i+1, j-1, k)$ ,  $(i, j-1, k+1)$ ,  $(i-1, j, k+1)$ ,  $(i-1, j+1, k)$  i  $(i, j+1, k-1)$  (Fig. 2) with diffusion speed  $D\Delta^{-2}$ .

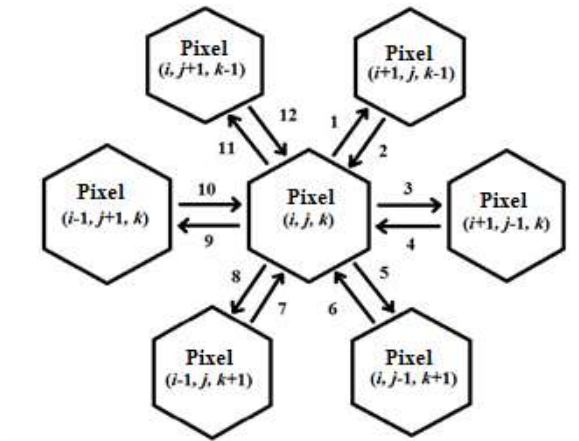


**Fig. 2.** Hexagonal lattice, which binds six neighboring pixels in the model of the biopixels using the cubic coordinates:

$$1, 3, 5, 8, 9, 11 - \left( \frac{D}{\Delta^2} M_{i,j,k}(t) \right); 2 - \left( \frac{D}{\Delta^2} M_{i+1,j,k-1}(t) \right);$$

$$4 - \left( \frac{D}{\Delta^2} M_{i+1,j-1,k}(t) \right); 6 - \left( \frac{D}{\Delta^2} M_{i-1,j,k+1}(t) \right); 7 - \left( \frac{D}{\Delta^2} M_{i-1,j+1,k}(t) \right); 10 - \left( \frac{D}{\Delta^2} M_{i-1,j+1,k}(t) \right); 12 - \left( \frac{D}{\Delta^2} M_{i,j+1,k-1}(t) \right).$$

The mathematical model of late- macrophage - monoclonal antibody interaction for a hexagonal array of biopixels using based on the well-known Marchuk model [34-36] and uses the spatial operator  $\hat{S}$  proposed in [37].

$$M_{i,j,k}(n+1) = M_{i,j,k}(n) \exp \left\{ \beta - \gamma A_{i,j,k}(n-r) - \delta_v M_{i,j,k}(n-r) \right\} + \hat{S}\{M_{i,j,k}(n)\}, \quad (5)$$

$$A_{i,j,k}(n+1) = A_{i,j,k}(n) \exp \left\{ -\mu_f + \eta \gamma M_{i,j,k}(n-r) - \delta_f A_{i,j,k}(n) \right\}, \quad n > 0$$

where  $\hat{S}\{M_{i,j,k}\}$  is a discrete diffusion for a spatial operator  $\hat{S}$ .

$$\begin{aligned} M_{i,j,k}(n) &= M_{i,j,k}^0(n) \geq 0, \quad A_{i,j,k}(n) = A_{i,j,k}^0(n) \geq 0, \\ n &\in [-r, 0), \quad M_{i,j,k}(0), A_{i,j,k}(0) > 0. \end{aligned} \quad (6)$$

## 2. Dynamic logical simulation of intelligent Big Data system for investigation the stability of CPS of medical and biological processes.

In order to simulate the dynamic logic of an intelligent Big Data system for investigation the stability of CPS of medical and biological processes, the syntax which is proposed by A. Platzer for the general CPS [38, 39] it is used. The CPS uses the Hybrid Programming Language (HP), which has more features than differential equations. Consider the dynamic logical simulation of intelligent Big Data system for investigation the stability of CPS for medical applications on the example of the mathematical model of biosensor on hexagonal lattice using lattice difference equations with time delay. Dynamic program is the first level of HP and it is defined by the following grammar:

$$\begin{aligned} a &::= M_{i,j,k}(n+1) = M_{i,j,k}(n) \exp\{\beta - \gamma A_{i,j,k}(n-r) - \delta_x A_{i,j,k}(n-r)\} + \hat{S}\{M_{i,j,k}(n)\}, \\ A_{i,j,k}(n+1) &= A_{i,j,k}(n) \exp\{-\mu_f + \eta \gamma M_{i,j,k}(n-r) - \delta_y A_{i,j,k}(n)\} \& \Phi_t. \end{aligned} \quad (7)$$

where  $\Phi_t$  is an evolutionary domain constraint in the form of a formula for the logic of the first order of real arithmetic

$$\begin{aligned} \Phi_t &\stackrel{\text{def}}{=} M^{\min} \leq V_{i,j,k}(n) \leq M^{\max} \\ &\wedge A^{\min} \leq A_{i,j,k}(n) \leq A^{\max} \wedge i, j, k = \overline{-N, N} \wedge n > 0, i + j + k = 0 \end{aligned} \quad (8)$$

The functioning of the biopixel  $(i, j, k)$  is determined by two states, with respect to fluorescence. Namely,  $S_{fl}$  is a state of fluorescence and  $S_{nonfl}$  is one of the non-fluorescence states. The use of the first order of semantics of logic and the satisfaction ratio  $s|L$  for the first-order formula  $L$  of real arithmetic and state  $s$  can be determined for some pixels  $(i, j, k); i, j, k = \overline{-N, N}, i + j + k = 0$  states  $S_{fl}$  and  $S_{nonfl}$  as

$$\begin{aligned} s_{fl}| &= k_{fl} M_{i,j,k}(n) A_{i,j,k}(n) \geq \Theta_{fl}, \\ s_{nonfl}| &= k_{fl} M_{i,j,k}(n) A_{i,j,k}(n) < \Theta_{fl} \end{aligned} \quad (9)$$

Discrete changes occur in computer programs when they accept new values for variables. This situation occurs when a fluorescence phenomenon occurs in a pixel  $(i, j, k); i, j, k = \overline{-N, N}, i + j + k = 0$ . The state  $s_{fl,i,j,k} := 1$  is assigned a value of 1 to the variable  $s_{fl,i,j,k}$ . This leads to a discrete, jump-like change, as the value  $s_{fl,i,j,k}$  does not change smoothly, but rapidly when it suddenly changes from 1 to  $s_{fl,i,j,k}$ , causing a discrete jump of values  $s_{fl,i,j,k}$ . In this way, we obtain a discrete model of change  $s_{fl,i,j,k} := 1$ , except for the model of change (9).

## 3. Software complex of the intelligent Big Data system for investigation the stability of CPS of medical and biological processes

The software package for the intelligent Big Data system for investigation the stability of CPS of medical and biological processes consists of the following main software modules: unit of ingesting and processing Big Data used to identify input parameters of the model of CPS, the software module of research of dynamic behavior of CPS, the software module of research of dynamic logic of CPS,

the block of decision-making on stability of CPS, the block of visualization, lattice images of macrophage/ monoclonal antibody data source, a data source that receives lattice images of macrophage binding to monoclonal antibodies and a database that receives images of fluorescent biopixels.

A software package for the study of CPS phase diagrams using the R package (<http://www.r-project.org/>) has been developed [40]. The block diagram of the software package of the intelligent Big Data system for investigation the stability of CPS of medical and biological processes is shown in Fig. 4.

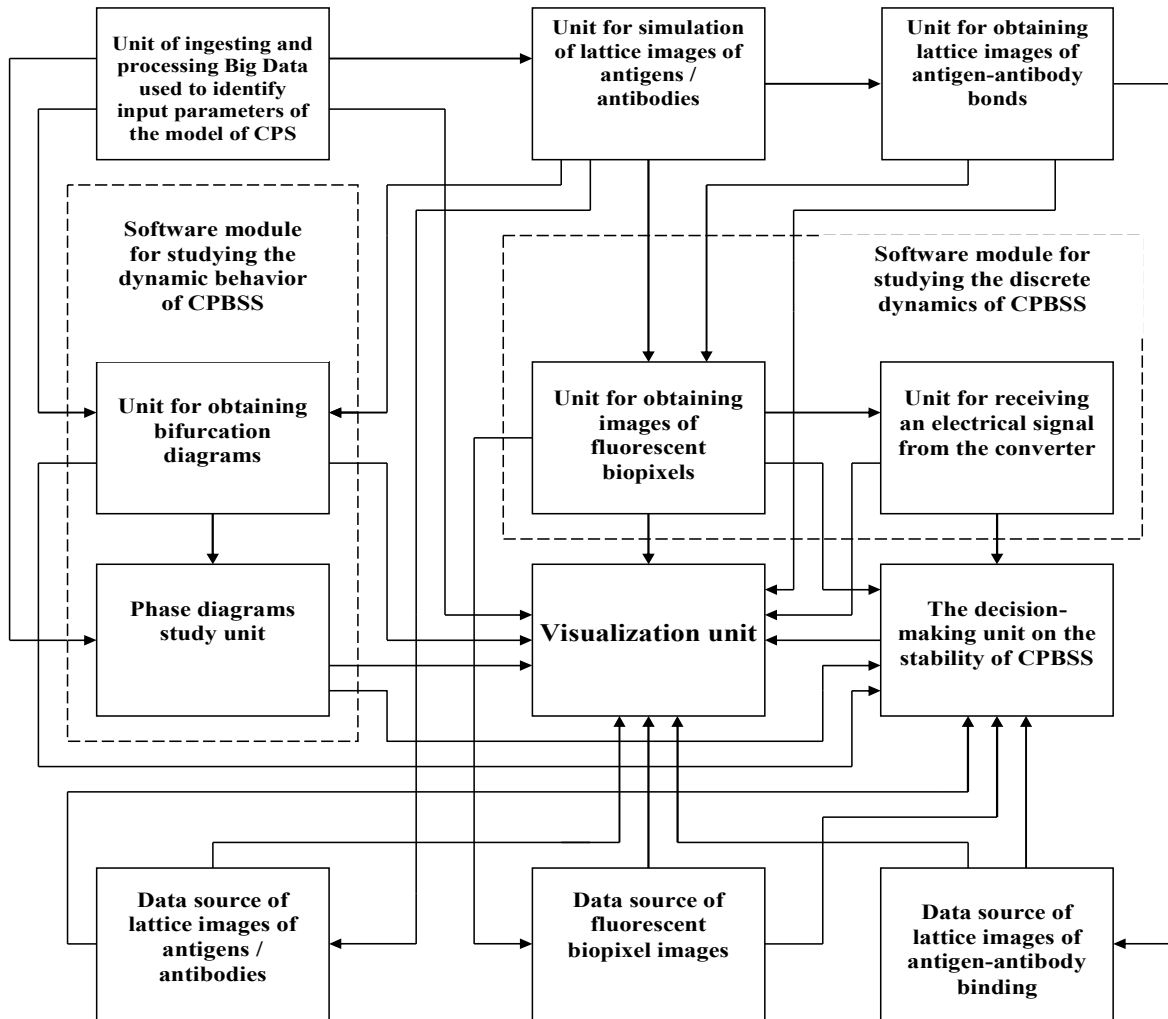


**Figure 4:** Block diagram of the software package of the intelligent Big Data system for investigation the stability of CPS of medical and biological processes

The software module for studying the dynamic behavior of CPS consists of blocks for obtaining bifurcation and phase diagrams. Using software to study the stability of intelligent Big Data system for investigation the stability of CPS of medical and biological processes, obtained the results of computer simulation in the form of bifurcation and phase diagrams of macrophage populations relative to monoclonal antibodies, lattice images of macrophages, monoclonal antibodies, probabilities of macrophage binding to monoclonal antibodies and monoclonal antibodies, the signal from the converter. The software module for studying the dynamic logic of CPS consists of blocks for modeling lattice images of macrophages/ monoclonal antibodies, block for obtaining lattice images of macrophage binding to monoclonal antibodies, obtaining images of fluorescent biopixels and the electrical signal from the converter. The study of CPS stability of medical and biological processes based on the Big Data intelligent system is an important data source: lattice images of macrophage/

monoclonal antibody data source, a data source that receives lattice images of macrophage binding to monoclonal antibodies, and a database that receives images of fluorescent biopixels.

#### 4. Computer simulation of mathematical models of biosensors using an intelligent Big Data system for investigation the stability of CPS for medical applications

##### 4.1. Parameters of mathematical models of biosensors for investigation the stability of CPS for medical applications

To study the occurrence of bifurcation and deterministic chaos in compartmental mathematical models of lattice type on a rectangular lattice using differential equations, models (1), (4) at  $N = 4$  and values of parameters  $\beta = 2 \text{ min}^{-1}$ ,  $\gamma = 2 \frac{\text{ml}}{\text{min} \cdot \text{mkg}}$ ,  $\mu_f = 1 \text{ min}^{-1}$ ,  $\eta = 0.8 / \gamma$ ,  $\delta_v = 0.5 \frac{\text{ml}}{\text{min} \cdot \text{mkg}}$ ,  $\delta_f = 0.5 \frac{\text{ml}}{\text{min} \cdot \text{mkg}}$ ,  $D = 0.2 \frac{\text{nm}^2}{\text{min}}$ ,  $\Delta = 0.3 \text{ nm}$  are considered. The concentrations of macrophages populations  $V_{i,j}(t)$ ,  $V_{i,j,k}(t)$  and monoclonal antibodies populations  $F_{i,j}(t)$ ,  $F_{i,j,k}(t)$  are measured in  $\frac{\text{mkg}}{\text{ml}}$ .

##### 4.2. Results of computer simulation of mathematical model of biosensor on rectangular lattice using lattice differential equations with delay

The computer simulation was implemented for different values  $\tau$ . The long-term behavior of the model (1) – (3) for  $\tau = 0.05$ ,  $\tau = 0.22$ ,  $\tau = 0.2865$  with a set of parameter values, which are presented above is analyzed. Qualitative changes in the behavior of biopixels and the intelligent Big Data system model for the investigation of the stability of the CPS of medical and biological processes are observed.

Figures 5 and 6 show the result of computer simulation of the discrete dynamics of the intelligent Big Data system for the investigation of the stability of CPS of medical and biological processes in the form of lattice images of macrophages and monoclonal antibodies in pixels of the studied system.

Figures 5 (a) and 6 (a) show the results of computer simulation of lattice images of macrophages and monoclonal antibodies in the pixels of the system (1) at  $\tau = 0.05$ , which corresponds to a stable focus. For  $\tau = 0.22$  there is a less pronounced (Fig. 5 (b) and Fig. 6 (b)), and for  $\tau = 0.2865$  more pronounced traveling wave of monoclonal antibodies, which is presented in Figures 5 (c) and 6 (c).

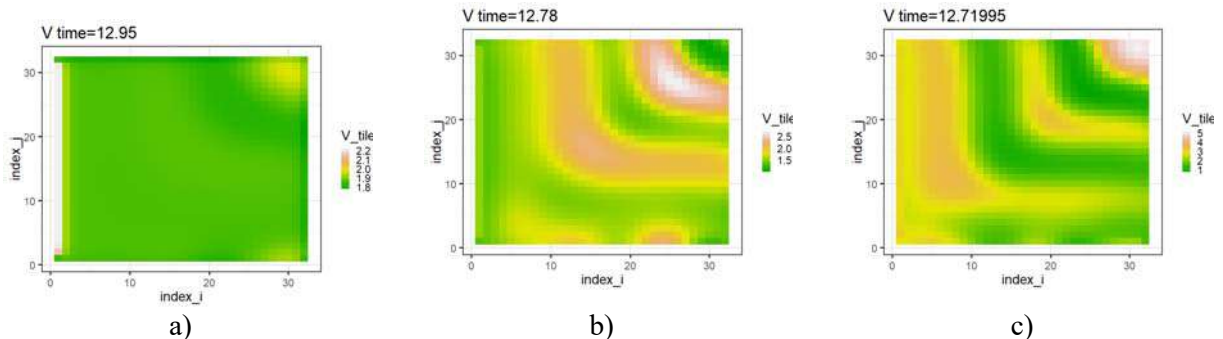


**Figure 5:** Lattice images of macrophages in pixels of system (1) at  $\tau = 0.05$  (a),  $\tau = 0.22$  (b),  $\tau = 0.2865$  (c)

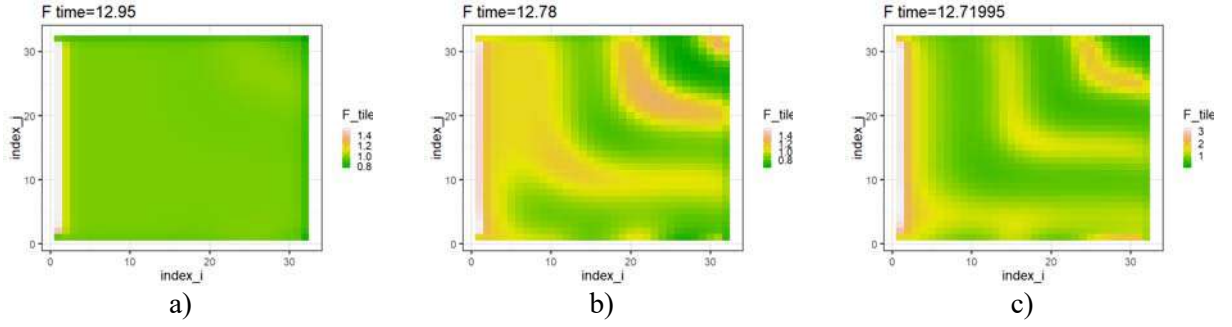


**Figure 6:** Lattice images of monoclonal antibodies in pixels of system (1) at  $\tau = 0.05$  (a),  $\tau = 0.22$  (b),  $\tau = 0.2865$  (c)

In the second stage of computer modeling of intelligent Big Data system for investigation of the stability of CPS of medical and biological processes lattice graphs are used on which for each pixel the probability of contact of macrophages with monoclonal antibodies, as  $V_{i,j} \times F_{i,j}$  at  $\tau = 0.05$ ,  $\tau = 0.22$ ,  $\tau = 0.287$  is presented. They are shown in Figures 7 (a - c).

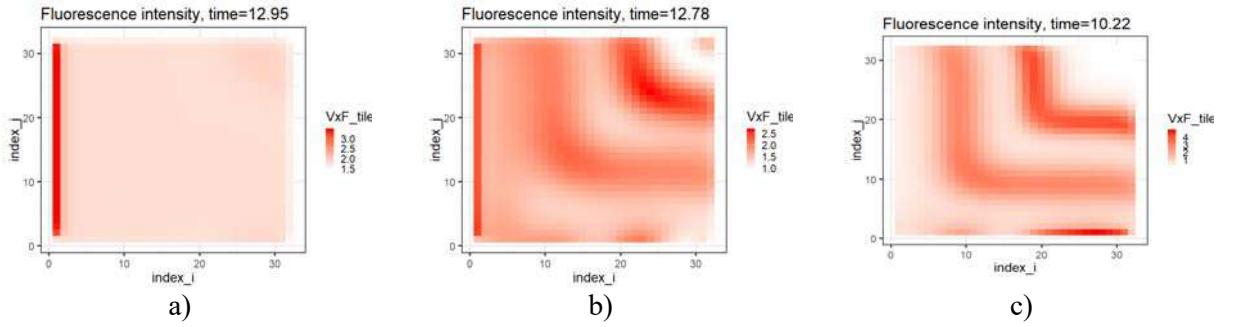


**Figure 7:** Lattice images of connections of macrophages with monoclonal antibodies in pixels of system (1) at  $\tau = 0.05$  (a),  $\tau = 0.22$  (b),  $\tau = 0.2865$  (c)

Graphs of fluorescent pixels are shown in Figures 8 (a - c).

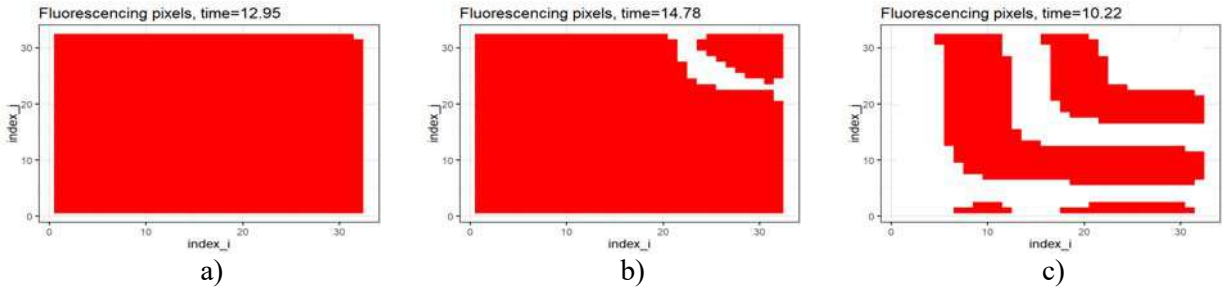


**Figure 8:** Fluorescence image of system (1) as a result of numerical simulation at  $\tau = 0.05$  (a),  $\tau = 0.22$  (b),  $\tau = 0.2865$  (c)

Figure 8 (a) shows the result of numerical simulation of system (1) at  $\tau = 0.05$ , which corresponds to a stable focus. For  $\tau = 0.22$  there is a less pronounced (Fig. 8 (b)), and for  $\tau = 0.2865$  it is more pronounced traveling wave of fluorescent pixels, which is shown in Figure 8 (c). Figure 8 (c) shows the result of numerical simulation of system (1), which corresponds to the approach to the limit cycle (there is a traveling wave of fluorescent pixels). In the case  $\tau = 0.2865$ , chaotic behavior is observed, which begins with wave-like changes in fluorescent pixels (Fig. 8 (c)) and quickly progresses to chaotic changes. Figure 8 (c) shows the result of numerical simulation of system (1), in which a chaotic wave of fluorescent pixels is observed.



To control the measurement process in the intelligent Big Data system for investigation of the stability of the CPS for medical applications, the result of computer simulation of the electrical signal from the converter (Fig. 9 (a, b)), which characterizes the number of fluorescent pixels at values of  $\tau = 0.05$ ,  $\tau = 0.2865$  plays an extremely important role. Analyzing the type of electrical signal in Figure 9, we see that when the value  $\tau$  changes qualitatively changes the behavior of the pixels and the whole intelligent Big Data system to investigate the stability of the CPS for medical and biological applications. In Figure 9 (a) there is a steady state at  $\tau = 0.05$ . In Figure 9 (b) ( $\tau = 0.2865$ ) a traveling wave of non-fluorescent pixels is clearly visible.

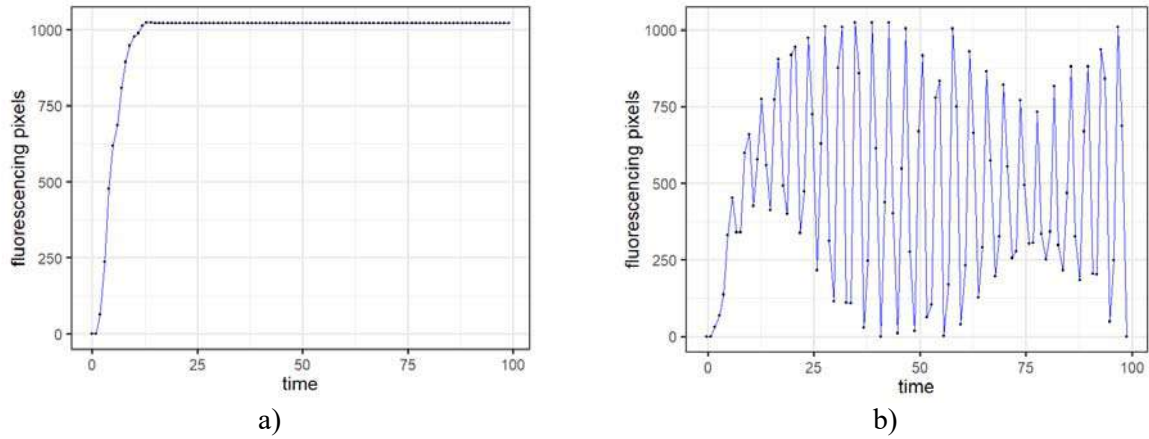


**Figure 9:** Electrical signal from the converter, which characterizes the number of fluorescent pixels at  $\tau = 0.05$  (a),  $\tau = 0.2865$  (b)

As shown by numerical analysis, fluorescent states in biopixels change according to the laws of discrete dynamics. The threshold value for fluorescence is  $\Theta_f = 1,5$ . Taking into account the continuous dynamics of the immunological response, each biopixel is considered as a CPS.

#### 4.3. Results of computer simulation of mathematical model of biosensor on hexagonal lattice using lattice difference equations with delay

The long - term behavior of model (4) at  $r = 5$ ,  $r = 17$ ,  $r = 22$  with a set of parameter values, which are presented above (Fig. 10 - 14), is analyzed. We observe qualitative changes in the behavior of biopixels and the biosensor model on hexagonal lattice using lattice difference equations with delay in general. Figures 10 and 11 show the first stage of computer modeling of discrete dynamics of an intelligent Big Data system for investigation of the stability of CPS of medical and biological processes in the form of lattice images of macrophages and monoclonal antibodies in pixels of the studied system. Figure 10 (a) shows the result of numerical simulation of system (4) at  $r = 5$ , which corresponds to a stable focus. For  $r = 17$  there is less pronounced (Fig. 10 (b)), and for  $r = 22$  more pronounced wavy changes in the lattice images of macrophages and monoclonal antibodies in the pixels of the system (4), as shown in Figures 10 (c).

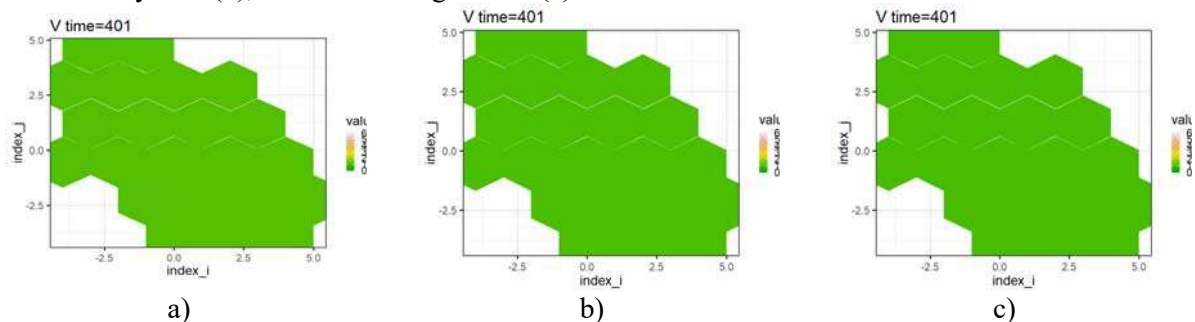


**Figure 10:** Lattice images of macrophages in pixels of system (4) at  $r = 5$  (a),  $r = 17$  (b),  $r = 22$  (c)

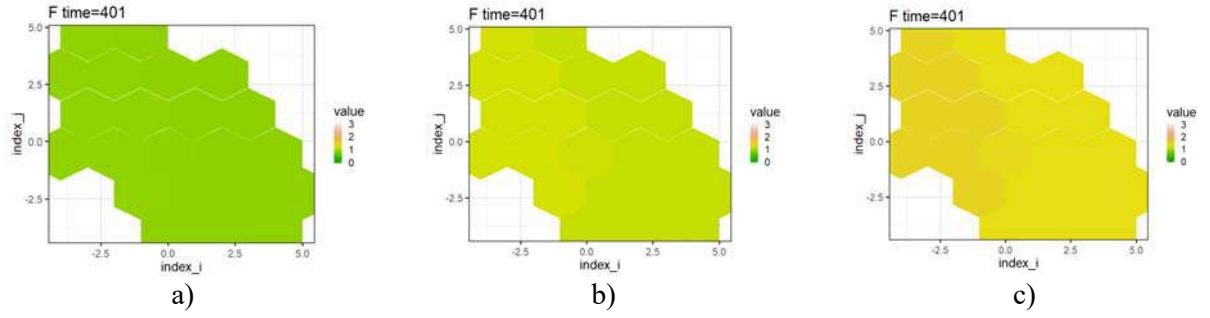


**Figure 11:** Lattice images of monoclonal antibodies in pixels of the system (4) at  $r = 5$  (a),  $r = 17$  (b),  $r = 22$  (c)

In the second stage of computer modeling of the intelligent Big Data system, lattice graphs were used to investigate the stability of the CPS of medical and biological processes. Firstly, the corresponding graphs are constructed, on which for each pixel the image of probability of contact of macrophages with monoclonal antibodies, as  $V_{i,j,k} \times F_{i,j,k}$  at  $r = 5$ ,  $r = 17$ ,  $r = 22$  is presented in Figure 12 (a - c).

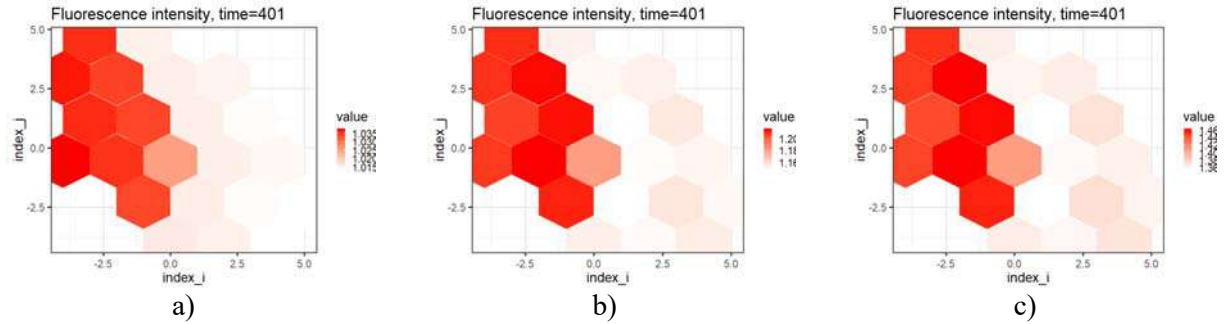


**Figure 12:** Lattice images of macrophage- monoclonal antibody binding in pixels of system (4) at  $r = 5$  (a),  $r = 17$  (b),  $r = 22$  (c)

Figure 12 (a) shows the result of numerical simulation of system (4) at  $r = 5$ , which corresponds to a stable focus. For  $r = 17$  there is less pronounced (Fig. 12 (b)), and for  $r = 22$  more pronounced wavy changes in the images of the probability of contact of macrophages with monoclonal antibodies of system (4), as shown in Figure 12 (c).

In the third stage of computer modeling of the intelligent Big Data system for investigation of the stability of CPS of medical and biological processes, lattice graphs of fluorescent pixels were obtained based on the fulfillment of condition (6), which are shown in Figures 13 (a - c).

Figure 13 (a) shows the result of numerical simulation of system (4) at  $r = 5$ , which corresponds to a stable focus. For  $r = 17$  there is a less pronounced (Fig. 13 (b)), and for  $r = 22$  - it is more pronounced traveling wave of fluorescent pixels of the system (4), which is shown in Figure 13 (c).

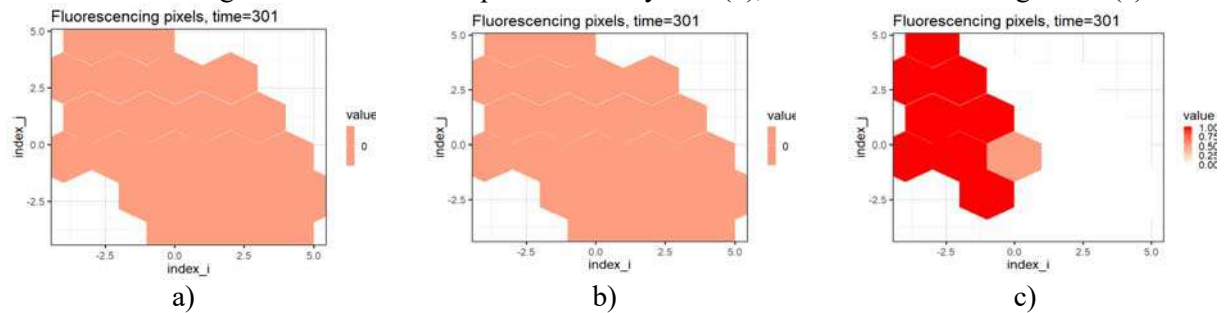


**Figure 13:** Fluorescence image of the system (4) as a result of numerical simulation at  $r = 5$  (a),  $r = 17$  (b),  $r = 22$  (c)

Hopf bifurcation is observed with increasing time delay [41]. Figure 13 (c) shows the result of numerical simulation of the system (4) at  $r = 17$ , which corresponds to the limit cycle (there is a traveling wave of fluorescent pixels). In the case  $r = 22$ , chaotic behavior is observed, which begins with wave-like changes in fluorescent pixels (Fig. 13 (c)) and quickly turns into chaotic changes.

Figure 14 shows the electric signal from the converter (Fig. 14), which characterizes the number of fluorescent pixels at  $r = 17$ ,  $r = 22$ .

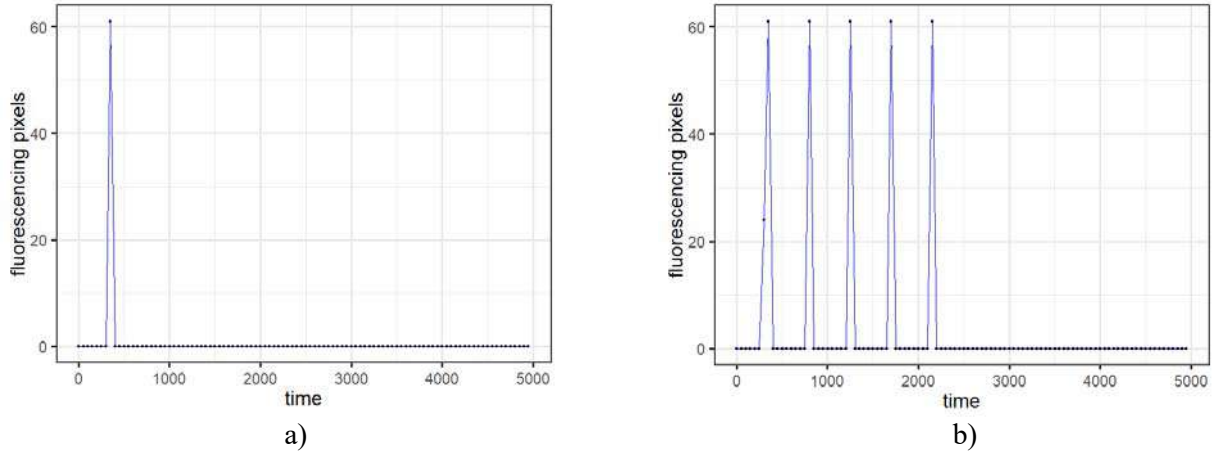


**Figure 14:** Electrical signal from the converter, which characterizes the number of fluorescent pixels at  $r = 17$  (a),  $r = 22$  (b)

To control the measurement process using an electrical signal to an intelligent big data system for investigation of the stability of the CPS of medical and biological processes is extremely important the result of computer simulation.

Analyzing the type of electrical signal in Figure 14, we can conclude that changing the value qualitatively changes the behavior of the pixels and the entire biosensor model based on lattice difference equations. In Figure 14 (a) ( $r = 17$ ) occurs, and in Figure 14 (b) ( $r = 22$ ) there is a traveling wave of non-fluorescent pixels. The threshold value for fluorescence is  $\Theta_{fl} = 1,5$ .

A two-dimensional array of biopixels was used to develop the dynamic logic of an intelligent Big Data system to investigate the stability of the CPS of medical and biological processes on a hexagonal lattice using delay equations. As shown by the results of numerical analysis, fluorescent states in biopixels change according to the laws of discrete dynamics.

## 5. Conclusion

In the work, the general scheme of the cyber-physical sensor system proposed in [42] was extended to the usage of Big Data. Peculiarities of immunosensors functioning are taken into account in CPS mathematical models. According to the laws of discrete dynamics, lattice images in immunopixels are modified. The proposed mathematical models take into account the interaction of immunopixels through the diffusion of macrophages.

The intelligent Big Data system uses the discrete dynamics of macrophages and antibodies, which is based on dynamic logic, to study the stability of CPS in medicine. The classes of rectangular lattice using differential equations or hexagonal lattice using difference equations are used in the work as modeling of the interaction of macrophages and monoclonal antibodies in immunopixels. This also takes into account the presence of macrophage and monoclonal antibodies colonies, that are localized in pixels, as well as the diffusion of macrophage colonies between pixels using the R package.

According to a series of experiments of intelligent Big Data system for investigation of the stability of CPS of medical and biological processes using lattice differential equations with delay it is established that in the case of using a rectangular lattice at constant time delay  $\tau \in [0, 0.22]$  the solution of the studied system is a stable focus. When  $\tau = 0.23$  (in the case of a rectangular lattice) there is a Hopf bifurcation and all subsequent trajectories correspond to stable boundary cycles for all

imunopixels. With a further increase in the delay constant  $\tau$ , chaotic behavior of the CPS of medical processes occurs on the basis of lattice differential equations with time delay.

According to the results of a series of experiments of intelligent Big Data system for investigation of the stability of CPS of medical processes using lattice difference equations with delay it is established that in case of using hexagonal lattice at  $r \in [0, 16]$  the solution of the investigated system tend to endemic states, which are a stable focus. For  $r = 17$  (in the case of a hexagonal lattice) there is a Hopf bifurcation and all subsequent trajectories correspond to stable boundary cycles for all imunopixels. With a further increase in the delay constant  $r$ , chaotic behavior of the CPS occurs on the basis of lattice difference equations.

The developed computer programs for the study stability of intelligent Big Data system for investigation the stability of CPS of medical processes in the form of phase diagrams, lattice images of macrophages / monoclonal antibodies, lattice images of macrophage binding to monoclonal antibodies, images of fluorescent imunopixels and the electrical signal from the converter should be used in research, design organizations, medical and laboratory centers in the development and testing of cyber-physical systems of medical processes.

In the following researches is necessary to offer mathematical models of intelligent systems of Big Data on the basis of other geometrical structures and to carry out research of their stability.

## 6. Acknowledgements

The work was co-funded by the European Union's Erasmus + Programme for Education under KA2 grant (project no. 2020-1-PL01-KA203-082197 "Innovations for Big Data in a Real World")

## 7. References

- [1] S. M. Khan, X. Liu, S. Nath, E. Korot, L. Faes, S. K. Wagner, P. A. Keane, N. J. Sebire, M. J. Burton, A. K. Denniston, A global review of publicly available datasets for ophthalmological imaging: barriers to access, usability, and generalisability, in: *The Lancet Digital Health*, Volume 3, Issue 1, 2021, pp. e51-e66. doi:10.1016/S2589-7500(20)30240-5.
- [2] V. Sujatha, S. Prasanna Devi, S. Vinu Kiran, S. Manivannan, Big Data Analytics on Diabetic Retinopathy Study (DRS) on Real-time Data Set Identifying Survival Time and Length of Stay, in: *Procedia Computer Science*, Volume 87, 2016, pp. 227-232. doi: 10.1016/j.procs.2016.05.153.
- [3] T. Inomata, J. Sung, M. Nakamura, K. Fujisawa, K. Muto, N. Ebihara, M. Iwagami, M. Nakamura, K. Fujio, Y. Okumura, M. Okano, A. Murakami, New medical big data for P4 medicine on allergic conjunctivitis, in: *Allergology International*, Volume 69, Issue 4, 2020, pp. 510-518. doi: 10.1016/j.alit.2020.06.001.
- [4] D. C. Elbers, N. R. Fillmore, F.-Chi Sung, S. S. Ganas, A. Prokhorenkov, Ch. Meyer, R. B. Hall, S. J. Ajjarapu, D. C. Chen, F. Meng, R. L. Grossman, M. T. Brophy, N. V. Do, The Veterans Affairs Precision Oncology Data Repository, a Clinical, Genomic, and Imaging Research Database, in: *Patterns*, Volume 1, Issue 6, 2020, p. 100083. doi: 10.1016/j.patter.2020.100083.
- [5] S. M. Willems, S. Abeln, K. A. Feenstra, Remco de Bree, E. F. van der Poel, R. J. Baatenburg de Jong, J. Heringa, M. W.M. van den Brekel, The potential use of big data in oncology, in: *Oral Oncology*, Volume 98, 2019, pp. 8-12. doi: 10.1016/j.oraloncology.2019.09.003.
- [6] R. Tripathi, R. S. Mazmudar, K. D. Knusel, J. S. Bordeaux, J. F. Scott, Big data in dermatology: Publicly available health care databases for population health research, in: *Journal of the American Academy of Dermatology*, Volume 83, Issue 5, 2020, pp. 1546-1556. doi: 10.1016/j.jaad.2020.04.145.
- [7] A. L. P. Ribeiro, G. M. M. Paixão, P. R. Gomes, M. H. Ribeiro, A. H. Ribeiro, J. A. Canazart, D. M. Oliveira, M. P. Ferreira, E. M. Lima, J. Lopes de Moraes, N. Castro, L. B. Ribeiro, P. W. Macfarlane, Tele-electrocardiography and Big Data: The CODE (Clinical Outcomes in Digital Electrocardiography) study, in: *Journal of Electrocardiology*, Volume 57, Supplement, 2019, pp. S75-S78. doi: 10.1016/j.jelectrocard.2019.09.008.

- [8] P. P. Ray, D. Dash, N. Kumar, Sensors for internet of medical things: State-of-the-art, security and privacy issues, challenges and future directions, in: *Computer Communications*, Volume 160, 2020, pp. 111-131. doi: 10.1016/j.comcom.2020.05.029.
- [9] J. Lloret, L. Parra, M. Taha, J. Tomás, An architecture and protocol for smart continuous eHealth monitoring using 5G, in: *Computer Networks*, Volume 129, Part 2, 2017, pp. 340-351. doi: 10.1016/j.comnet.2017.05.018.
- [10] I. Jayatilleka, M. N. Halgamuge, Chapter 1 - Internet of Things in healthcare: Smart devices, sensors, and systems related to diseases and health conditions, Editor(s): Himansu Das, Nilanjan Dey, Valentina Emilia Balas, In *Advances in Ubiquitous Sensing Applications for Healthcare, Real-Time Data Analytics for Large Scale Sensor Data*, Academic Press, Volume 6, 2020, pp. 1-35.
- [11] Center for Scientific Machine Learning. URL: <https://www.odn.utexas.edu/research/centers-groups/scientific-machine-learning/>.
- [12] B. Yuan, C. Shen, A. Luna, A. Korkut, D. S. Marks, J. Ingraham, Ch. Sander, CellBox: Interpretable Machine Learning for Perturbation Biology with Application to the Design of Cancer Combination Therapy, in: *Cell Systems*, Volume 12, Issue 2, 2021, pp. 128-140. doi: 10.1016/j.cels.2020.11.013.
- [13] M. Srivathsan, K. Yogesh Arjun, Health Monitoring System by Prognostic Computing Using Big Data Analytics, in: *Procedia Computer Science*, Volume 50, 2015, pp. 602-609. doi:10.1016/j.procs.2015.04.092.
- [14] P. N. Ramkumar, H. S. Haeberle, M. R. Bloomfield, J. L. Schaffer, A. F. Kamath, B. M. Patterson, V. E. Krebs, Artificial Intelligence and Arthroplasty at a Single Institution: Real-World Applications of Machine Learning to Big Data, Value-Based Care, Mobile Health, and Remote Patient Monitoring, in: *The Journal of Arthroplasty*, Volume 34, Issue 10, 2019, pp. 2204-2209. doi: 10.1016/j.arth.2019.06.018.
- [15] S. J. Kurian, A. R. Bhatti, M. A. Alvi, H. H. Ting, C. Storlie, P. M. Wilson, N. D. Shah, H. Liu, M. Bydon, Correlations Between COVID-19 Cases and Google Trends Data in the United States: A State-by-State Analysis, in: *Mayo Clinic Proceedings*, Volume 95, Issue 11, 2020, pp. 2370-2381. doi: 10.1016/j.mayocp.2020.08.022.
- [16] C. Baladrón, J. J. Gómez de Diego, I. J. Amat-Santos, Big data and new information technology: what cardiologists need to know, in: *Revista Española de Cardiología (English Edition)*, Volume 74, Issue 1, 2021, pp. 81-89. doi: 10.1016/j.rec.2020.06.036.
- [17] A. Zamanifar, 7 - Data analytics in IOT-based health care, Editor(s): K. Ch. Lee, S. S. Roy, P. Samui, V. Kumar, in: *Data Analytics in Biomedical Engineering and Healthcare*, Academic Press, 2021, pp. 119-130. doi: 10.1016/B978-0-12-819314-3.00007-0.
- [18] A. Perry, P. Kerezoudis, Ch. S. Graffeo, L. P. Carlstrom, M. Peris-Celda, F. B. Meyer, M. Bydon, M. J. Link, Little Insights from Big Data: Cerebrospinal Fluid Leak After Skull Base Surgery and the Limitations of Database Research, in: *World Neurosurgery*, Volume 127, 2019, pp. e561-e569. doi: 10.1016/j.wneu.2019.03.207.
- [19] J. Wall, T. Krummel, The digital surgeon: How big data, automation, and artificial intelligence will change surgical practice, in: *Journal of Pediatric Surgery*, Volume 55, Supplement, 2020, pp. 47-50. doi: 10.1016/j.jpedsurg.2019.09.008.
- [20] Y. Krainyk, Y. Davydenko and V. Starchenko, Message-level Decoding of Error Patterns for Turbo-Product Codes, in: *Proceedings of the 39th International Conference on Electronics and Nanotechnology (ELNANO)*, Kyiv, Ukraine, 2019, pp. 660-663. doi: 10.1109/ELNANO.2019.8783849.
- [21] Y. Krainyk, Y. Davydenko and V. Tomas, Configurable Control Node for Wireless Sensor Network, in: *Proceedings of the 3rd International Conference on Advanced Information and Communications Technologies (AICT)*, Lviv, Ukraine, 2019, pp. 258-262. doi: 10.1109/AICT.2019.8847732.
- [22] D. Michie, D.J. Spiegelhalter, C. Taylor Machine Learning. Neural and Statistical Classification (1994), pp. 1-298.
- [23] M. Rezaei-Ravari, M. Eftekhari, F. Saberi-Movahed, Regularizing extreme learning machine by dual locally linear embedding manifold learning for training multi-label neural network classifiers, in: *Engineering Applications of Artificial Intelligence*, Volume 97, 2021, p. 104062. doi: 10.1016/j.engappai.2020.104062.

- [24] L. Huang, H. Leng, X. Li, K. Ren, J. Song, D. Wang, A Data-Driven Method for Hybrid Data Assimilation with Multilayer Perceptron, in: *Big Data Research*, Volume 23, 2021, p. 100179. doi: 10.1016/j.bdr.2020.100179.
- [25] The seven Vs of Big Data. URL: <https://www.bbva.com/en/five-vs-big-data/>.
- [26] Total amount of global healthcare data generated in 2013 and a projection for 2020. URL: <https://www.statista.com/statistics/1037970/global-healthcare-data-volume/>.
- [27] V.P. Martsenyuk, I.Ye. Andrushchak, P.M. Zinko, A.S. Sverstiuk, On Application of Latticed Differential Equations with a Delay for Immunosensor Modeling, in: *Journal of Automation and Information Sciences*, Volume 50, Issue 6, 2018, pp. 55–65. doi: 10.1615/JAutomatInfScien.v50.i6.50.
- [28] V.P. Martsenyuk, A.S. Sverstiuk, I.Ye. Andrushchak, Approach to the Study of Global Asymptotic Stability of Lattice Differential Equations with Delay for Modeling of Immunosensors, in: *Journal of Automation and Information Science*, Volume 48, Issue 8, 2019, pp. 58–71. doi: 10.1615/JAutomatInfScien.v51.i2.70.
- [29] V.P. Martsenyuk, A. Klos-Witkowska, A.S. Sverstiuk, Stability investigation of biosensor model based on lattice difference equations. *Difference Equations and Discrete Dynamical Systems with Applications*, in: *Springer Proceedings in Mathematics and Statistics*. 24th ICDEA, Dresden, Germany. Volume 312, 2020, pp. 297–322. doi: 10.1007/978-3-030-35502-9\_13.
- [30] V.P. Martsenyuk, A. Klos-Witkowska, A.S. Sverstiuk, Stability, bifurcation and transition to chaos in a model of immunosensor based on lattice differential equations with delay, in: *Electronic Journal of Qualitative Theory of Differential Equations*: No. 27 (2018), pp. 1–31. doi: 10.14232/ejqtde.2018.1.27.
- [31] V.P. Martsenyuk, A.S. Sverstiuk, A. Klos-Witkowska, N.V. Kozodii, O.A. Bagriy-Zayats, I.R. Zubenko, Numerical Analysis of Results Simulation of Cyber-physical Biosensor Systems, in: 1st International Workshop Information-Communication Technologies & Embedded Systems, 14–15 November, Mykolaiv, Volume 1, 2019, pp. 149–164.
- [32] A. Patel, Hexagonal Grids. URL: <https://www.redblobgames.com/grids/hexagons/>.
- [33] L. Čomić, B. Nagy, A combinatorial coordinate system for the body-centered cubic grid, in: *Graphical Models*, Volume 87, 2016, pp. 11–22. doi: 10.1016/j.gmod.2016.08.001.
- [34] U. Forys, Marchuk's model of immune system dynamics with application to tumour growth, in: *Journal of Theoretical Medicine*, Volume 4, no. 1, 2002, pp. 85–93. doi: 10.1080/10273660290052151.
- [35] U. Forys, Stability and bifurcations for the chronic state in Marchuk's model of an immune system, in: *Journal of Mathematical Analysis and Applications*, Volume 352, Issue 2, 2009, pp. 922–942. doi: 10.1016/j.jmaa.2008.11.055.
- [36] Z. Huang, Qigui Yang, Junfei Cao, Stochastic stability and bifurcation for the chronic state in Marchuk's model with noise, in: *Applied Mathematical Modelling*, Volume 35, Issue 12, 2011, pp. 5842–5855. doi: 10.1016/j.apm.2011.05.027.
- [37] A. Prindle, P. Samayoa, I. Razinkov, T. Danino, LS Tsimring, and J. Hasty, A sensing array of radically coupled genetic 'biopixels', in: *Nature*, Volume 481, no. 7379, Dec. 2011, pp. 39–44.
- [38] A. Platzer, Logics of Dynamical Systems. 27th Annual IEEE Symposium on Logic in Computer Science, Dubrovnik, 2012, pp. 13–24. doi: 10.1038/nature10722.
- [39] A. Platzer, The complete proof theory of hybrid systems, in: 27th Annual ACM/IEEE Symposium on Logic in Computer Science, 2012, pp. 541–550. doi: 10.1109/lics.2012.64.
- [40] V.P. Martsenyuk, A.S. Sverstiuk, O.A. Bahrii-Zaiats, Yu.A. Rudyak, B. Shelestovsky, Software complex in the study of the mathematical model of cyber-physical systems, in: *ICT&ES 2nd International Workshop Information-Communication Technologies & Embedded Systems*, 12 November, 2020 Mykolaiv, Ukraine. Volume 1, 2020, pp. 87–97.
- [41] D. Duan, B. Niu, J. Wei, Hopf-Hopf bifurcation and chaotic attractors in a delayed diffusive predator-prey model with fear effect, in: *Chaos, Solitons & Fractals*, Volume 123, 2019, pp. 206–216. doi: 10.1016/j.chaos.2019.04.012.
- [42] C. Berger, A. Hees, S. Braunreuther, G. Reinhart, Characterization of Cyber-Physical Sensor Systems, in: *Procedia CIRP*, Volume 41, 2016, pp. 638–643. doi: 10.1016/j.procir.2015.12.019.



# Homogeneity hypothesis in discriminant analysis

Dmitriy Klyushin<sup>a</sup>

<sup>a</sup>Taras Shevchenko National University of Kyiv, Ukraine, 03680, Kyiv, Akademika Glushkova Avenue 4D

## Abstract

One of the most important properties of a machine learning algorithm is its ability to generalize results of learning on finite training sets. This property is based on the compactness hypothesis stating that objects of the same class in the feature space, as a rule, are located closer to each other than to objects of other classes. The compactness hypothesis has a geometric nature and uses the concept of proximity in the feature space, which is most often expressed in terms of a metric. Meanwhile, this hypothesis does not fully take into account the probabilistic nature of the features. It is quite suitable for data with unimodal distributions that have a compact support, but in a general case it may not hold leading to incorrect generalizations. In the paper, an alternative approach is described in which the homogeneity hypothesis is used instead of the compactness hypothesis. Within the framework of this approach, objects are called homogeneous, if their features follow identical distributions. We propose as measures of homogeneity the Petunin's  $p$ -statistics and its versions, which is highly efficient in recognizing both disjoint and significantly overlapping samples that violate the compactness assumption. This approach has a rigorous mathematical foundation and high efficiency in practical applications.

## Keywords <sup>1</sup>

Discriminant analysis, relational analysis, featureless pattern recognition, compactness hypothesis, homogeneity hypothesis

## 1. Introduction

The complexity of machine learning significantly depends on the compactness hypothesis, which allows generalizations based on finite training samples. Intuitively, the hypothesis states that in the feature space, similar objects should be closer to each other than to dissimilar ones. This definition appeals to the geometric concept of proximity and implicitly uses a metric. A typical example of a method based on such principles is the nearest neighbor method, which recognizes test objects by their closeness to training objects.

The compactness hypothesis ignores the probabilistic nature of the random training data. More precisely, it is only acceptable for classifying random data with unimodal distributions having compact support. In practice, such a condition is too burdensome. Therefore, it is necessary to develop a method that would estimate the proximity between random samples on different principles. To solve the problem, let us introduce the concept of object homogeneity, which means that objects are drawn from the same population, i.e. their features obey the same distribution. This allows objects to be classified using criteria to test static hypotheses of homogeneity.

The aim of the article is to describe a new approach to machine learning based on the homogeneity hypothesis as an alternative to the compactness hypothesis. Using a measure of homogeneity, not just a metric, allows generalizing relational discriminant analysis and increasing its efficiency.



## 2. Distance-based machine learning techniques

Duin, Pekalska et al. [1–4], Mottle, Seredin et al. [5–8] and others proposed the concept of featureless discriminant analysis. They suggested replacing the feature vector of an object with an estimate of its proximity to some training set using a metrics. Unfortunately, this approach is poorly suited to solving problems often encountered in biomedical research. Let us say a researcher studies the parameters of a set of cells. In this case, it gets samples of real numbers, not an ordered vector. In such cases, the metric is not applicable and the only useful tool is a homogeneity (similarity) measure. Among the numerous statistical tests for the homogeneity of samples, only the Kolmogorov-Smirnov test, the Wilcoxon test and the Klyushin-Petunin test allow us to assess robustly the homogeneity of samples in the form of the probability of belonging to the same general population [9].

As it was noted above, in the featureless discriminant analysis, objects are represented not as vectors in a feature space, but as a measure of proximity to a training set. As a result, the starting point of featureless analysis is a distance matrix filled with distances or labels characterizing the similarity between objects in the training set (reference points). As a basic distance, usually Euclidean and pseudo-Euclidean distances are used coupled with the kernel trick. Obviously, this approach leads to problems with generalization power and strong dependence from a training set. In addition, it is not valid for samples containing of independent identically distributed (i.i.d.) random values.

Recently, this approach was renewed as machine learning techniques such as Minimal Learning Machine [10] and the Extreme Minimal Learning Machine [11]. The main tool in these methods is the nonlinear distance regression, which estimates the dissimilarity between observations. Nowadays, various metrics and learning techniques are used in this field [12–21]. Excellent surveys of these methods may be found in [22, 23]. These methods have some useful advantages, but they use Euclidean distance-constructed probability distributions. Thus, they fail in situations when reference points are not vectors in some vector space but samples of i.i.d. random values.

The problem we try to solve is to extend the application field of the distance-based machine learning techniques using not metrics to estimate distances between objects but the homogeneity (similarity) measures described below, and propose an alternative way of similarity-based classification. These homogeneity measures do not dependent on underlying distributions of training samples and have useful properties of generalization. For example, in opposite to standard counterparts (the Kolmogorov–Smirnov statistics and the Wilcoxon statistics), they successfully work both with samples following distributions with different means and identical variance and with identical means and different variances.

## 3. Two-sample homogeneity measure

Consider training samples  $x = (x_1, x_2, \dots, x_n) \in G_1$  and  $y = (y_1, y_2, \dots, y_n) \in G_2$  from populations  $G_1$  and  $G_2$  following absolutely continuous distribution functions  $F_1$  and  $F_2$ . We reduce the classification of a test sample  $z = (z_1, z_2, \dots, z_n)$  to testing of homogeneity  $z$  and  $x$  from the one side, and  $z$  and  $y$  from the other side. There are many nonparametric tests for two samples homogeneity: Kolmogorov–Smirnov test, Wilcoxon sign rank test etc. (see, for example, [24]). However, as it will be shown, the most effective tool for testing homogeneity of two samples is the Petunin's  $p$ -statistics [25]. This is explained by the fact that the  $p$ -statistics has similar high significance and sensitivity independently in both cases when samples are disjoint or almost overlapped.

### 3.1. Original Klyushin–Petunin test

The Klyushin–Petunin test [25] is non-parametric one and use only assumption that distribution functions are absolutely continuous. This test uses the Hill's assumption  $A_{(n)}$  [26] stating that for



exchangeable random values  $x_1, x_2, \dots, x_n \in G$  following to an absolutely continuous distribution function we have:

$$P\left(x \in (x_{(i)}, x_{(j)})\right) = \frac{j-i}{n+1}, \quad j < i, \quad (1)$$

where  $x_{(i)}$  and  $x_{(j)}$  are the  $i$ -th and  $j$ -th order statistics. The Hill's assumption was proved as for i.i.d. random values [27] and for exchangeable i.d. random values [28]. Finding the relative frequency  $h_{ij}$  of the event  $z_m \in (x_{(i)}, x_{(j)})$  for the elements of  $z$ , we can estimate a proximity between  $h_{ij}$  and  $\frac{j-i}{n+1}$ . This may be made using numerous confidence intervals for binomial proportion. For definiteness, we use the Wilson confidence interval  $I_{ij}^{(n)} = (p_{ij}^{(1)}, p_{ij}^{(2)})$  where

$$\begin{aligned} p_{ij}^{(1)} &= \frac{h_{ij}n + g^2/2 - g\sqrt{h_{ij}(1-h_{ij})n + g^2/4}}{n + g^2}, \\ p_{ij}^{(2)} &= \frac{h_{ij}n + g^2/2 + g\sqrt{h_{ij}(1-h_{ij})n + g^2/4}}{n + g^2}. \end{aligned} \quad (2)$$

The significance level of this interval is the function of  $g$ . When  $g = 3$  the significance level of  $I_{ij}^{(n)}$  does not exceed 0.05 [25].  $P$ -statistics, estimating the homogeneity of samples  $x$  and  $z$ , is defined by the equation

$$h = \# \left\{ p_{ij} = \frac{j-i}{n+1} \in I_{ij}^{(n)} \right\} / \left( \frac{n(n-1)}{2} \right), \quad (3)$$

As we see, the  $p$ -statistics is the estimation of the probability that the samples are homogeneous, therefore, using (2) we may formulate the following test: the null hypothesis is accepted if  $h$  is greater than 0.95, else it is rejected.

When the null hypothesis is true, the events  $\left\{ p_{ij} = \frac{j-i}{n+1} \in I_{ij}^{(n)} \right\}$  generate a generalized Bernoulli scheme [29, 30]. When the alternative hypothesis is true, they generate a modified Bernoulli scheme. When the null hypothesis can be either true or false, they generate the Matveichuk–Petunin scheme [31]. When the null hypothesis holds,  $\lim_{n \rightarrow \infty} \frac{j-i}{n+1} \in (0,1)$ , and  $\lim_{n \rightarrow \infty} \frac{i}{n+1} \in (0,1)$ , then the asymptotic significance level  $\beta$  of a sequence of confidence intervals  $I_{ij}^{(n)}$  is less than 0.05 [25].

### 3.2. Modified Klyushin–Petunin test

In practice, samples, as a rule, contain rounded numbers and duplicates (ties). Thus, we must distinguish a hypothetical sample drawn from a hypothetical population  $G$  of precise measurements and an empirical sample drawn from an empirical population  $\tilde{G}$  of rounded measurements. Let us introduce a sample  $\tilde{x} = (\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_n)$  approximating a hypothetical sample  $x = (x_1, x_2, \dots, x_n)$  and let the variational series  $x_{(1)} < x_{(2)} < \dots < x_{(n)}$  and  $\tilde{x}_{(1)} < \tilde{x}_{(2)} < \dots < \tilde{x}_{(m)}$  be variational series of hypothetical and empirical samples.

If a number  $x^*$  is drawn from  $G$  independently from  $x$  then

$$p\left(x^* \in [x_{(k)}, x_{(k+1)}]\right) = \frac{1}{n+1}, \quad (4)$$

where  $k = 0, 1, \dots, n$ ,  $x_{(0)} = -\infty$ , and  $x_{(n+1)} = \infty$ . Thus,

$$p\left(\tilde{x}^* \in [\tilde{x}_{(i)}, \tilde{x}_{(j)}]\right) \approx \gamma_i + \gamma_{i+1} + \dots + \gamma_{j-1} + \frac{j-i}{n+1}, \quad (5)$$

where  $t_i = t(\tilde{x}_{(i)})$  is the multiplicity of  $\tilde{x}_{(i)}$ . If  $\tilde{x}$  does not contain ties then  $\gamma_i = 0$ .

Suppose, that the hypothetical population  $G$  follows a hypothetical absolutely continuous distribution function  $F$ . Then, (4) holds. Consider empirical samples  $\tilde{x} = (\tilde{x}_1, \dots, \tilde{x}_n)$  and  $\tilde{z} = (\tilde{z}_1, \dots, \tilde{z}_n)$ . Using the Wilson confidence interval  $I_{ij} = (p_{ij}^{(1)}, p_{ij}^{(2)})$  for the probability (5) of the event  $\{\tilde{z}_k \in (\tilde{x}_{(i)}, \tilde{x}_{(j)})\}$  we find an observed relative frequency. Let us denote  $N = \#I_{ij} = \frac{n(n-1)}{2}$  and compute the empirical  $p$ -statistics  $h = \frac{1}{N} \# \left\{ \frac{j-i}{n+1} \in I_{ij} \right\}$ . Then, we can formulate the following test: the null hypothesis is accepted if the  $h$  (the probability that the samples are homogeneous) is greater than 0.95, else the null hypothesis is rejected.

### 3.3. Exact Klyushin–Petunin test

As we see, the versions of the Klyushin–Petunin test based on the Wald confidence interval depend on the parameter  $g$ , that varies from 1.96 for the normal distribution to 3 for an general unimodal distribution. To avoid this uncertainty, we propose to use the exact confidence interval for the unknown probability  $p$  on the basis of the proportion  $h$  in the Bernoulli model consisting of  $n$  trials [32]. To do this, consider two functions depending on  $p \in [0, 1]$ :

$$\varphi(p) = |h - p|$$

and

$$\psi(p) = \frac{1}{2n} + \frac{\lambda}{n} \sqrt{np(1-p) + \frac{1}{12}}.$$

where  $\lambda > \sqrt{\frac{8}{3}}$  is the parameter of the Vysochansky–Petunin inequality [33]

$$p(|y - m(y)| \geq \lambda \sigma(y)) \leq \frac{4}{9} \frac{1}{\lambda^2}.$$

Denote

$$\tilde{\psi}(p) = \sqrt{np(1-p) + \frac{1}{12}}, \quad p \in R^1.$$

The graph of  $\tilde{\psi}(p)$ ,  $p \in R^1$  is the upper half of the ellipse  $E$  passing through the points

$$A = \left( \frac{1}{2n} \left( n + \sqrt{\frac{n}{3} + n^2} \right), 0 \right), \quad B = \left( \frac{1}{2}, \sqrt{\frac{1}{12n} + \frac{1}{4}} \right),$$

$$C = \left( \frac{1}{2n} \left( n - \sqrt{\frac{n}{3} + n^2} \right), 0 \right), \quad D = \left( \frac{1}{2}, -\sqrt{\frac{1}{12n} + \frac{1}{4}} \right)$$

with the center  $\left(\frac{1}{2}, 0\right)$ . The graph of  $\psi(p)$  is the restriction of the graph of  $\tilde{\psi}(p)$  on the segment  $[0, 1]$  stretching or shrinking the graph by  $\frac{\lambda}{n}$  and shifting it by  $\frac{1}{2n}$ .

Therefore, the graph of the function  $\psi(p)$  which does not depend on  $h$  is an arc of ellipse  $\Gamma$  passing through the points  $(0, \psi(0))$ ,  $\left(\frac{1}{2}, \psi\left(\frac{1}{2}\right)\right)$ ,  $(1, \psi(1))$ , such that the function  $\psi(p)$  reach the minimum at the point  $p = \frac{1}{2}$  and it is symmetrical with respect to this point.

The lower confidence limit  $p_1$  is a root of the quadratic equation

$$\left(1 + \frac{\lambda^2}{n}\right)p^2 - \left(\frac{\lambda^2}{n} - \frac{1}{n} + 2h\right)p + h^2 - \frac{h}{n} + \frac{1}{4n^2}\left(1 - \frac{\lambda^2}{3}\right) = 0. \quad (6)$$

If  $h > \psi(0) = \frac{1}{2n} + \frac{\lambda}{n\sqrt{12}}$ , then the lower confidence limit  $p_1$  is the least root of (6). If  $h \leq \psi(0)$ , then  $p_1 = 0$ .

Similarly, the upper confidence limit  $p_2$  is a root of the square equation

$$\left(1 + \frac{\lambda^2}{n}\right)p^2 - \left(\frac{\lambda^2}{n} + \frac{1}{n} + 2h\right)p + h^2 + \frac{h}{n} + \frac{1}{4n^2}\left(1 - \frac{\lambda^2}{3}\right) = 0. \quad (7)$$

If  $1 - h > \psi(1)$ , then the upper confidence limit  $p_2$  is the largest root of (4). If  $1 - h \leq \psi(1)$ , then  $p_2 = 1$ .

**Remark.** Since  $p_1 \leq h \leq p_2$ , the proportion of successes always is in the confidence interval  $[p_1, p_2]$ .

For the generalized Bernoulli model similar reasoning gives the following quadratic equation for lower confidence limit:

$$\left(1 + \frac{(m+n+1)\lambda^2}{(n+2)m}\right)p^2 + \left(\frac{1}{m} - \frac{(m+n+1)\lambda^2}{(n+2)m} - 2h\right)p + h^2 - \frac{h}{m} + \frac{1}{4m^2}\left(1 - \frac{\lambda^2}{3}\right) = 0 \quad (8)$$

If  $h > \frac{1}{2m} + \frac{\lambda}{m\sqrt{12}} = \gamma$ , then the lower confidence limit  $p_1$  for the generalized Bernoulli model is the least root of (8). If  $h \leq \gamma$ , then  $p_1 = 0$ .

Similar, the upper confidence limit  $p_2$  for the generalized Bernoulli model is the root of the equation

$$\left(1 + \frac{(m+n+1)\lambda^2}{(n+2)m}\right)p^2 - \left(\frac{1}{m} + \frac{(m+n+1)\lambda^2}{(n+2)m} + 2h\right)p + h^2 + \frac{h}{m} + \frac{1}{4m^2}\left(1 - \frac{\lambda^2}{3}\right) = 0 \quad (9)$$

If  $1 - h > \gamma$ , then the upper confidence limit  $p_2$  is the largest root of (9). If  $1 - h \leq \gamma$ , then  $p_2 = 1$ . By virtue of the previous results the significance level of the confidence interval does not exceed  $\frac{4}{9} \frac{1}{\lambda^2}$  (in particular, 0.05 for  $\lambda = 3$ ).

## 4. Experiments and results

To assess the true positive and true negative rates of the proposed tests, we performed numerical experiments using samples from the normal distribution  $N(\mu, \sigma)$  of various degree of overlapping. We considered 100 samples of 40 random numbers having different averages and the same variance (location shift) and as well as 100 samples of 40 random numbers having the same average value and different variance (scale shift). We calculated the average  $p$ -statistics and its lower and upper confidence limits, the average Kolmogorov–Smirnov statistics and its  $p$ -value, and the average Wilcoxon statistics and its  $p$ -value. To estimate the true positive rate of the Klyushin–Petunin test we used the relative frequency of an event when the  $p$ -statistic is less than 0.95 for different distributions. The true positive rate of the Kolmogorov–Smirnov and Wilcoxon sign rank tests is the relative frequency of an event when the corresponding  $p$ -value is less than 0.05, when the distributions are different. The true negative rate of the Klyushin–Petunin test is the relative frequency of an event when the upper confidence limit of the  $p$ -statistic is greater than 0.95 for identical distributions. The true negative rate of the Kolmogorov–Smirnov tests and Wilcoxon signed ranks is the relative frequency of an event when the value of  $p$  is less 0.05, when the distributions are identical. Thus, we tested two statistical hypotheses: location shift and scale. The null location shift hypothesis means that the mathematical expectations of both distributions are identical. The null scale hypothesis means that the variances of both distributions are identical. Alternative hypothesis, in contrast, asserts that the distribution functions are different. The results are presented in Tables 1–11.

**Table 1**

$P$ -statistics for the location shift hypothesis without ties

| Distribution | N(0,1) | N(1,1) | N(2,1) | N(3,1) | N(4,1) |
|--------------|--------|--------|--------|--------|--------|
| N(0,1)       | 1.000  | 0.752  | 0.680  | 0.457  | 0.389  |
| N(1,1)       | –      | 1.000  | 0.846  | 0.584  | 0.424  |
| N(2,1)       | –      | –      | 1.000  | 0.680  | 0.442  |
| N(3,1)       | –      | –      | –      | 1.000  | 0.570  |
| N(4,1)       | –      | –      | –      | –      | 1.000  |

**Table 2**

Exact  $p$ -statistics for the location shift hypothesis without ties

| Distribution | N(0,1) | N(1,1) | N(2,1) | N(3,1) | N(4,1) |
|--------------|--------|--------|--------|--------|--------|
| N(0,1)       | 1.000  | 0.646  | 0.459  | 0.376  | 0.374  |
| N(1,1)       | –      | 1.000  | 0.990  | 0.522  | 0.418  |
| N(2,1)       | –      | –      | 1.000  | 0.859  | 0.505  |
| N(3,1)       | –      | –      | –      | 1.000  | 0.959  |
| N(4,1)       | –      | –      | –      | –      | 1.000  |

**Table 3**

$P$ -value of the Kolmogorov–Smirnov test for the location shift hypothesis without ties

| Distribution | N(0,1) | N(1,1) | N(2,1)  | N(3,1)  | N(4,1)  |
|--------------|--------|--------|---------|---------|---------|
| N(0,1)       | 1.000  | 0.0001 | <0.0001 | <0.0001 | <0.0001 |
| N(1,1)       | –      | 1.000  | 0.0003  | <0.0001 | <0.0001 |
| N(2,1)       | –      | –      | 1.000   | <0.0001 | <0.0001 |
| N(3,1)       | –      | –      | –       | 1.000   | <0.0001 |
| N(4,1)       | –      | –      | –       | –       | 1.000   |

**Table 4***P*-value of the Wilcoxon signed rank test for the location shift hypothesis without ties

| Distribution | N(0,1) | N(1,1) | N(2,1)  | N(3,1)  | N(4,1)  |
|--------------|--------|--------|---------|---------|---------|
| N(0,1)       | 1.000  | 0.002  | <0.0001 | <0.0001 | <0.0001 |
| N(1,1)       | –      | 1.000  | 0.005   | <0.0001 | <0.0001 |
| N(2,1)       | –      | –      | 1.000   | <0.0001 | <0.0001 |
| N(3,1)       | –      | –      | –       | 1.000   | <0.0001 |
| N(4,1)       | –      | –      | –       | –       | 1.000   |

Note that the *p*-statistic is monotonically decreasing as the location shift increases. As expected, in this case the Kolmogorov-Smirnov and Wilcoxon sign rank tests work well. However, when the distribution functions are largely overlapped the discrepancy between them is not very significant. Moreover, the Wilcoxon signed-rank test poorly recognizes the inversions between largely overlapped samples. These statements are justified by the following results (Table 5–8).

**Table 5***P*-statistics for the scale shift hypothesis without ties

| Distribution | N(0,1) | N(0,2) | N(0,3) | N(0,4) | N(0,5) |
|--------------|--------|--------|--------|--------|--------|
| N(0,1)       | 1.000  | 0.726  | 0.641  | 0.581  | 0.427  |
| N(0,2)       | –      | 1.000  | 0.819  | 0.753  | 0.620  |
| N(0,3)       | –      | –      | 1.000  | 0.979  | 0.976  |
| N(0,4)       | –      | –      | –      | 1.000  | 0.998  |
| N(0,5)       | –      | –      | –      | –      | 1.000  |

**Table 6**Exact *p*-statistics for the scale shift hypothesis without ties

| Distribution | N(0,1) | N(0,2) | N(0,3) | N(0,4) | N(0,5) |
|--------------|--------|--------|--------|--------|--------|
| N(0,1)       | 1.000  | 0.725  | 0.556  | 0.585  | 0.613  |
| N(0,2)       | –      | 1.000  | 0.899  | 0.785  | 0.706  |
| N(0,3)       | –      | –      | 1.000  | 0.988  | 0.996  |
| N(0,4)       | –      | –      | –      | 1.000  | 0.998  |
| N(0,5)       | –      | –      | –      | –      | 1.000  |

**Table 7***P*-value of the Kolmogorov–Smirnov test for the scale shift hypothesis without ties

| Distribution | N(0,1) | N(0,2) | N(0,3) | N(0,4)  | N(0,5)  |
|--------------|--------|--------|--------|---------|---------|
| N(0,1)       | 1.000  | 0.011  | 0.003  | <0.0001 | <0.0001 |
| N(0,2)       | –      | 1.000  | 0.027  | 0.012   | 0.084   |
| N(0,3)       | –      | –      | 1.000  | 0.752   | 0.398   |
| N(0,4)       | –      | –      | –      | 1.000   | 0.742   |
| N(0,5)       | –      | –      | –      | –       | 1.000   |

**Table 8***P*-value of the Wilcoxon signed rank test for the scale shift hypothesis without ties

| Distribution | N(0,1) | N(0,2) | N(0,3) | N(0,4) | N(0,5) |
|--------------|--------|--------|--------|--------|--------|
| N(0,1)       | 1.000  | 0.212  | 0.341  | 0.352  | 0.920  |
| N(0,2)       | –      | 1.000  | 0.065  | 0.144  | 0.350  |
| N(0,3)       | –      | –      | 1.000  | 1.000  | 0.307  |
| N(0,4)       | –      | –      | –      | 1.000  | 0.751  |
| N(0,5)       | –      | –      | –      | –      | 1.000  |

The Kolmogorov–Smirnov test fails when samples are largely overlapped in more than almost a half of the cases, and the Wilcoxon signed-rank test has failed at all. The Klyushin–Petunin test fails in almost a third of cases of very overlapped samples following the distributions  $N(0,3)$ ,  $N(0,4)$  and  $N(0,5)$ .

To simulate the ties in samples we rounded the samples from previous experiments to two decimal digits. Due to this, every sample contained four ties. The results are provided in Tables 9–12. The construction of the Kolmogorov–Smirnov and Wilcoxon signed rank tests do not depend on ties. Thus, we provide only results for the  $p$ -statistics.

**Table 9**

$P$ -statistics for the location shift hypothesis with ties

| Distribution | $N(0,1)$ | $N(1,1)$ | $N(2,1)$ | $N(3,1)$ | $N(4,1)$ |
|--------------|----------|----------|----------|----------|----------|
| $N(0,1)$     | 1.000    | 0.672    | 0.505    | 0.355    | 0.309    |
| $N(1,1)$     | –        | 1.000    | 0.831    | 0.305    | 0.323    |
| $N(2,1)$     | –        | –        | 1.000    | 0.705    | 0.424    |
| $N(3,1)$     | –        | –        | –        | 1.000    | 0.573    |
| $N(4,1)$     | –        | –        | –        | –        | 1.000    |

**Table 10**

Exact  $P$ -statistics for the location shift hypothesis with ties

| Distribution | $N(0,1)$ | $N(1,1)$ | $N(2,1)$ | $N(3,1)$ | $N(4,1)$ |
|--------------|----------|----------|----------|----------|----------|
| $N(0,1)$     | 1.000    | 0.627    | 0.504    | 0.367    | 0.330    |
| $N(1,1)$     | –        | 1.000    | 0.804    | 0.561    | 0.348    |
| $N(2,1)$     | –        | –        | 1.000    | 0.683    | 0.480    |
| $N(3,1)$     | –        | –        | –        | 1.000    | 0.581    |
| $N(4,1)$     | –        | –        | –        | –        | 1.000    |

**Table 11**

$P$ -statistics for the scale shift hypothesis with ties

| Distribution | $N(0,1)$ | $N(0,2)$ | $N(0,3)$ | $N(0,4)$ | $N(0,5)$ |
|--------------|----------|----------|----------|----------|----------|
| $N(0,1)$     | 1.000    | 0.420    | 0.331    | 0.356    | 0.392    |
| $N(0,2)$     | –        | 1.000    | 0.799    | 0.682    | 0.588    |
| $N(0,3)$     | –        | –        | 1.000    | 0.993    | 0.965    |
| $N(0,4)$     | –        | –        | –        | 1.000    | 0.982    |
| $N(0,5)$     | –        | –        | –        | –        | 1.000    |

**Table 12**

Exact  $P$ -statistics for the scale shift hypothesis with ties

| Distribution | $N(0,1)$ | $N(0,2)$ | $N(0,3)$ | $N(0,4)$ | $N(0,5)$ |
|--------------|----------|----------|----------|----------|----------|
| $N(0,1)$     | 1.000    | 0.726    | 0.640    | 0.495    | 0.524    |
| $N(0,2)$     | –        | 1.000    | 0.987    | 0.928    | 0.774    |
| $N(0,3)$     | –        | –        | 1.000    | 0.967    | 0.953    |
| $N(0,4)$     | –        | –        | –        | 1.000    | 0.983    |
| $N(0,5)$     | –        | –        | –        | –        | 1.000    |

The  $p$ -statistics monotonically decreases as the difference between the means increases. The Klyushin–Petunin test, like the Kolmogorov–Smirnov test, does not distinguish between the distributions of  $N(0, 3)$ ,  $N(0, 4)$ , and  $N(0, 5)$ . At the same time, it turned out to be effective in cases where the Kolmogorov–Smirnov tests and the Wilcoxon sign rank test do not work. Thus, there is an advantage of the  $p$ -statistics over the Kolmogorov–Smirnov and Wilcoxon sign rank tests.

## 5. Conclusions

Correct generalization based on finite training sets depends on correctly chosen underlying hypotheses. Traditional discriminant analysis is based on the compactness hypothesis, which states that objects of one class in the feature space are located closer to each other than to objects from another class. This geometric hypothesis does not work when classifying random samples that differ from feature vectors. For samples, the concept of distance is meaningless. It should be replaced by the concept of homogeneity, meaning that features of objects have the same distribution function. The evaluation of the homogeneity of the samples is provided by the Petunin  $p$ -statistics and its variants, which demonstrate high sensitivity and specificity in experiments both when testing the hypothesis of a shift in the mean and in testing the hypothesis of a shift in the scale. The proposed method has a rigorous mathematical justification and high efficiency in practical applications.

## 6. References

- [1] R.P.W. Duin, D. de Ridder, D.N.J. Tax, Experiments with a featureless approach to pattern recognition, *Pattern Recognit Lett* 18 (1997) 1159–1166. doi: 10.1016/S0167-8655(97)00138-4.
- [2] R.P.W. Duin, E. Pekalska, D. de Ridder, Relational discriminant analysis, *Pattern Recognition Letters* 20 (1999) 1175–1181. doi: 10.1016/S0167-8655(99)00085-9.
- [3] E. Pekalska, R.P.W. Duin, On combining dissimilarity representations, in: J. Kittler, F. Roli (Eds.), *Multiple Classifier Systems, LNCS*, vol. 2096, Springer-Verlag, 2001, pp. 359–368. doi: 10.1007/3-540-48219-9\_36.
- [4] E. Pekalska, R.P.W. Duin, *The Dissimilarity Representation for Pattern Recognition, Foundations and Applications*, World Scientific, Singapore, 2005.
- [5] V. Mottl, S. Dvoenko, P. Seredin, C. Kulikowski, I. Muchnik, Featureless pattern recognition in an imaginary Hilbert space and its application to protein fold classification. *Machine Learning and Data Mining in Pattern Recognition, Lecture Notes in Computer Science*, 2123 (2001) 322–336. doi: 10.1007/3-540-44596-X\_26.
- [6] V. Mottl, O. Seredin, S. Dvoenko, C. Kulikowski, I. Muchnik, Featureless pattern recognition in an imaginary Hilbert space, in: *Object recognition supported by user interaction for service robots*, Quebec City, QC, Canada, 2002, pp. 88–91, vol.2. doi: 10.1109/ICPR.2002.1048244.
- [7] O. Seredin, V. Mottl, A. Tatarchuk, N. Razin, D. Windridge, Convex support and Relevance Vector Machines for selective multimodal pattern recognition, in: *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*, Tsukuba, Japan, 2012, pp. 1647–1650.
- [8] V. Mottl, O. Seredin, O. Krasotkina, Compactness Hypothesis, Potential Functions, and Rectifying Linear Space in Machine Learning: International Conference Commemorating the 40th Anniversary of Emmanuil Braverman's Decease, Boston, MA, USA, April 28–30, 2017, *Invited Talks*. doi: 10.1007/978-3-319-99492-5\_3.
- [9] R.I. Andrushkiw, N.V. Boroday, D.A. Klyushin, Y.I. Petunin, *Computer-aided cytogenetic method of cancer diagnosis*, New York, Nova Publishers, 2007.
- [10] B. Kulis. Metric learning: A survey. *Foundations and Trends in Machine Learning*, 5 (2013) 287–364. doi: 10.1561/22000000019
- [11] A. H. de Souza Junior, F. Corona, G. A. Barreto, Y. Miche, A. Lendasse, Minimal Learning Machine: A novel supervised distance-based approach for regression and classification. *Neurocomputing*, 164 (2015) 34–44. doi: 10.1016/j.neucom.2014.11.073.
- [12] D. P. P. Mesquita, J. P. P. Gomes, A. H. de Souza Junior, Ensemble of efficient minimal learning machines for classification and regression, *Neural Processing Letters*, 46 (2017) 751–766. doi: 10.1007/s11063-017-9587-5.
- [13] A. N. Maia, M. L. D. Dias, J. P. P. Gomes, and A. R. da Rocha Neto, Optimally selected minimal learning machine, in: H. Yin, D. Camacho, P. Novais, A. J. Tallón-Ballesteros (Eds.), *Intelligent Data Engineering and Automated Learning – IDEAL*, Springer International Publishing, Cham, 2018, pp. 670–678. doi: 10.1007/978-3-030-33617-2.

- [14] W. L. Caldas, J. P. P. Gomes, D. P. P. Mesquita, Fast Co-MLM: An efficient semi-supervised co-training method based on the minimal learning machine, *New Generation Computing*, 36 (2018) 41–58. doi: 10.1007/s00354-017-0027-x
- [15] J. A. Florêncio, M. L. D. Dias, A. R. da Rocha Neto, A. H. de Souza Júnior, A fuzzy c-means-based approach for selecting reference points in minimal learning machines, in: G. A. Barreto and R. Coelho (Eds.), *Fuzzy Information Processing*, Springer International Publishing, Cham, 2018, pp. 398–407. doi: 10.1007/978-3-319-95312-0\_34.
- [16] T. Kärkkäinen, Extreme minimal learning machine: Ridge regression with distance-based basis, *Neurocomputing*, 342 (2019) 33–48. doi:10.1016/j.neucom.2018.12.078.
- [17] H. Cao, S. Bernard, R. Sabourin, L. Heutte, Random forest dissimilarity based multi-view learning for radiomics application, *Pattern Recognition* 88 (2019) 185–197. doi: 10.1016/j.patcog.2018.11.011
- [18] J. A. Florêncio, S. A. Oliveira, J. P. Gomes, A. R. da Rocha Neto, A new perspective for minimal learning machines: A lightweight approach, *Neurocomputing* 401 (2020). doi: 10.1016/j.neucom.2020.03.088.
- [19] L. Nanni, A. Rigo, A. Lumini, S. Brahnam, Spectrogram Classification Using Dissimilarity Space, *Appl. Sci.* 10 (2020) 4176. doi: 10.3390/app10124176.
- [20] A. C. F. da Silva, F. Saïs, E. Waller, F. Andres, Dissimilarity-based approach for Identity Link Invalidation, in: *IEEE 29th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE)*, Bayonne, France, 2020, pp. 251–256, doi: 10.1109/WETICE49692.2020.00056.
- [21] M. Bicego, Dissimilarity Random Forest Clustering, in: *IEEE International Conference on Data Mining (ICDM)*, Sorrento, Italy, 2020, pp. 936–941. doi: 10.1109/ICDM50108.2020.00105.
- [22] Y. M. G. Costa, D. Bertolini, A. S. Britto, G. D. C. Cavalcanti, L. S. Oliveira, The dissimilarity approach: a review, *Artificial Intelligence Review*, 53 (2020) 2783–2808. doi: 10.1007/s10462-019-09746-z.
- [23] J. Hämäläinen, A. Alencar, T. Kärkkäinen, C. Mattos, A. Souza Júnior, J.P.P. Gomes, Minimal Learning Machine: Theoretical Results and Clustering-Based Reference Point Selection, *Journal of Machine Learning Research*, 21 (2020) 1–29.
- [24] B. Derrick, P. White, D. Toher, Parametric and non-parametric tests for the comparison of two samples which both include paired and unpaired observations. *Journal of Modern Applied Statistical Methods*, 18 (2019) eP2847. doi: 10.22237/jmasm/1556669520.
- [25] D. A. Klyushin, Yu. I. Petunin, A Nonparametric Test for the Equivalence of Populations Based on a Measure of Proximity of Samples, *Ukrainian Mathematical Journal* 55 (2003) 181–198. doi: 10.1023/A:1025495727612.
- [26] D. M. Hill, Posterior distribution of percentiles: Bayes’ theorem for sampling from a population. *Journal of American Statistician Association* 63 (1968) 677–691.
- [27] I. Madreimov, Yu. I. Petunin, Characterization of a uniform distribution using order statistics. *Theory of Probability and Mathematical Statistics*, 27 (1982) 96–102.
- [28] R.I. Andrushkiw D. A. Klyushin, Yu. I. Petunin, and V. N. Lysyuk, Construction of the bulk of general population in the case of exchangeable sample values, in: *Proceedings of the International Conference of Mathematics and Engineering Techniques in Medicine and Biological Science (METMBS'03)*, Las Vegas, Nevada, USA, 2003, pp. 486–489.
- [29] S.A. Matveichuk, Yu.I. Petunin, Generalization of Bernoulli schemes that arise in order statistics. I. *Ukrainian Mathematical Journal* 42 (1990): 459–466. doi: 10.1007/BF01058940.
- [30] S.A. Matveichuk, Yu.I. Petunin, Generalization of Bernoulli schemes that arise in order statistics. II. *Ukrainian Mathematical Journal*, 43(1991): 728–734. doi: 10.1007/BF01058940.
- [31] N. Johnson, S. Kotz, Some generalizations of Bernoulli and Polya-Eggenberger contagion models. *Statist Paper* 32 (1991) 1–17. doi: 10.1007/BF02925473.
- [32] R.I. Andrushkiw, D.A. Klyushin, Yu.I. Petunin, M. Yu. Savkina, The exact confidence limits for unknown probability Bernoulli models, in: *27th International Conference on Information Technology Interfaces*, 2005, pp. 164–168. doi:10.1109/ITI.2005.1491116.
- [33] D. Vysochanskij, Y. Petunin, Justification of the 3-sigma rule for unimodal distributions, *Theory of Probability and Mathematical Statistics* 21 (1980) 25–36.



# Implementation of probabilistic data structures in the processes of neuroevolutionary synthesis

Serhii Leoshchenko<sup>a</sup>, Andrii Oliinyk<sup>a</sup>, Sergey Subbotin<sup>a</sup>, Viktor Lytvyn<sup>a</sup> and Oleksandr Korniienko<sup>a</sup>

<sup>a</sup> National university "Zaporizhzhia polytechnic", Zhukovskogo street 64, Zaporizhzhia, 69063, Ukraine

## Abstract

Neuroevolutionary synthesis is an alternative to training a pre-designed neural network. The use of neuroevolution methods is explained by fewer free parameters, less dependence on the expert, and usually better adaptive characteristics. However, neuroevolution methods are more resource-intensive, because the synthesis uses populations consisting of a large number of networks. Various mechanisms, such as parallelization, can be used to solve problems related to synthesis time. However, they usually lead to even more memory usage in the process. Mechanisms such as selective pressure only partially solve this problem. Probabilistic data structures have proven to be a fairly compact storage mechanism. Therefore, the paper provides an example and analysis of ways to use probabilistic data structures during the neuroevolutionary synthesis of artificial neural networks.

## Keywords 1

Neuroevolutionary, synthesis, artificial neural networks, probabilistic data structures, encoding, sequencing, genetic operators

## 1. Introduction

Artificial neural networks (ANN) are one of the areas of scientific research in the field of creating artificial intelligence [1-3], which is based on the desire to imitate the human nervous system [3-5]. Including its (nervous system's) ability to correct mistakes and self-learn [5-7]. The scope of application of ANNs is constantly expanding, today they are used in such areas as:

- machine learning, which is a type of artificial intelligence [1-7]. It is based on training the ANN on the example of millions [8-11] of tasks of the same type. Nowadays, machine learning is actively implemented by Google [12-15], Bing [16], [17], and Baidu [18], [19] search engines. So based on the millions of search queries that we all enter into search systems every day, their algorithms learn to show us the most relevant results so that we can find exactly what we are looking for [14], [15];
- in robotics [20], [21], neural networks are used to develop numerous algorithms for the "brains" of robots. Here, training usually uses methods that implement a reinforcement learning strategy, such as swarm intelligence or the actor-critical method (A2C and A3C) [22-25];
- computer system architects use neural networks to solve the problem of parallel computing [4], [7];
- with the help of neural networks, scientists can solve various complex mathematical problems [6].

As already mentioned in one of the theses, before direct training of the ANN, it is necessary to train. This can be done using certain examples (supervised learning), using information about the environment and the state of the "agent" in that environment (reinforcement learning), or without any

---

CMIS-2021: The Fourth International Workshop on Computer Modeling and Intelligent Systems, April 27, 2021, Zaporizhzhia, Ukraine  
EMAIL: sergleo.zntu@gmail.com (S. Leoshchenko); olejnikaa@gmail.com (A. Oliinyk); subbotin@zntu.edu.ua (S. Subbotin); lytvynviktor.a@gmail.com (V. Lytvyn); al.korn95@gmail.com (O. Korniienko)  
ORCID: 0000-0001-5099-5518 (S. Leoshchenko); 0000-0002-6740-6078 (A. Oliinyk); 0000-0001-5814-8268 (S. Subbotin); 0000-0003-4061-4755 (V. Lytvyn); 0000-0003-4812-5382 (O. Korniienko)



© 2021 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

clear input information at all (unsupervised learning) [26-33]. Neuroevolution methods can be used for all approaches, because usually, in addition to conventional training (parametric synthesis), neuroevolution methods adjust the structure (topology) of the ANN (structural synthesis) [37], [38]. Such methods are separated into a separate group of methods for the evolution of interneuronal connections and network topologies (TWEANNs) [37], [38], which are individuals in the population.

However, one of the main disadvantages of neuroevolution methods is their resource intensity. Large amounts of synthesis time are usually explained by the search for the most optimal ANN structure, which is achieved by using evolutionary mechanisms in individual populations. On the other hand, taking into account the fact that in usual network training, using gradient methods, it is necessary to design topologies first and only then train it, it should be noted that neuroevolutionary synthesis automates this subprocess [39-42]. Moreover, the use of parallelization can significantly speed up the synthesis process [39-42]. And given the fact that almost all neuroevolution methods can be parallelized, this removes the speed problem. However, parallelization imposes additional memory requirements on the computing system. The use of selective pressure mechanisms [43] can increase the level of adaptability of final decisions and reduce memory usage, but in sequential systems it can lead to a low level of genetic diversity.

Therefore, the question arises about using certain special structures to preserve information about the population. Moreover, the use of these structures is possible at the stage of encoding genetic information about an individual, even at the beginning of synthesis.

Probabilistic data structures (PDS) are a group of data structures that are extremely useful for big data and streaming applications [44-46]. Generally speaking, these data structures use hash functions to randomize and compactly represent a set of elements. Data collisions are ignored (usually using a number of mechanisms that prevent such situations), and errors are controlled at a certain threshold. Compared to error-free approaches, these algorithms use much less memory and have a constant query time. Moreover, support for union and intersection operations provides a low level of parallelization complexity.

Therefore, let's look at what stages of synthesis the use of PDS is justified and appropriate. The main stages are shown in Figure 1.

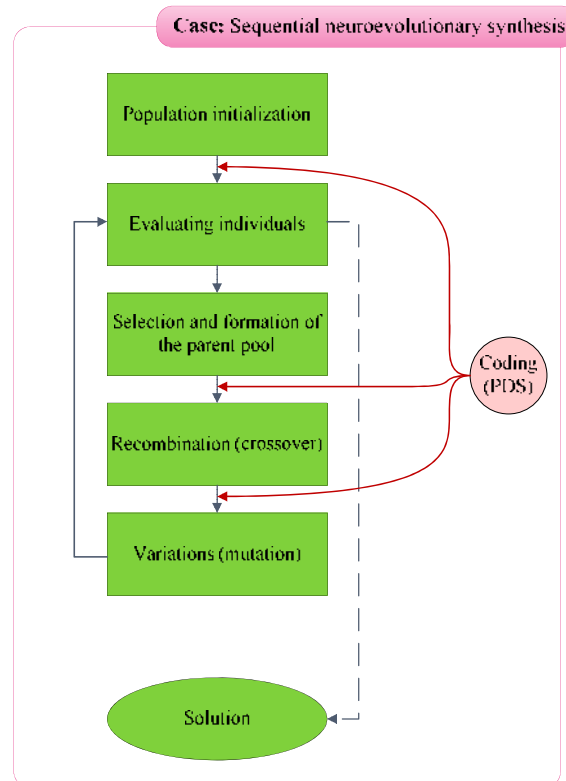


**Figure 1:** Synthesis stages where PDS can be implemented

Analyzing Figure1 note that the implementation of PDS during synthesis is advisable at the following stages:

- encoding of genetic information about ANN, as individuals of the population;
- crossover of individuals (since PDS supports unification and intersection).

Special cases can be considered situations when information about the ANN is transmitted, for example, between the cores of parallel computing systems or computing nodes, as shown in Figure 2.

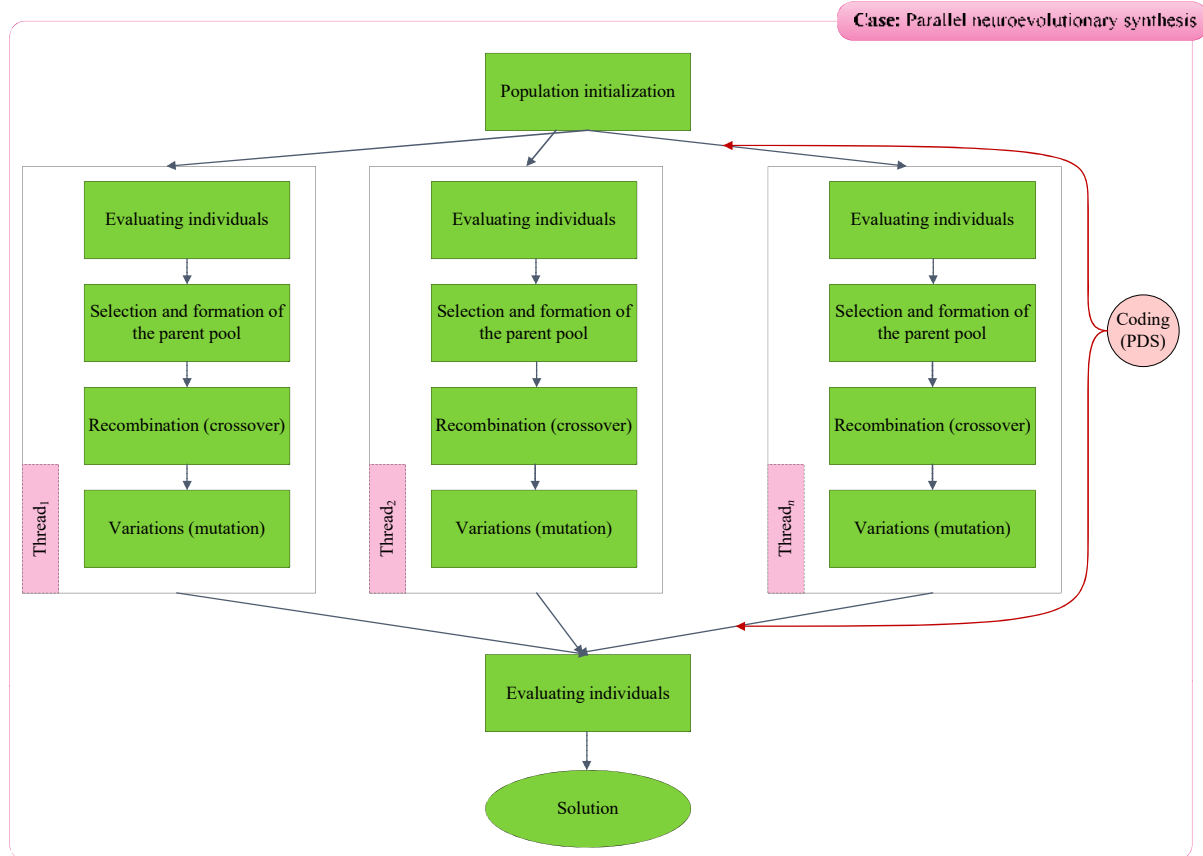


**Figure 2:** ANNs forwarding in parallel systems

## 2. Literature Review

Statistical analysis and analysis of data that related to big data is a common and quite relevant task, especially in such areas as smart systems, Internet of things (IoT), Cognitive Internet of Things (CIoT), web analytics [44-47]. One of the options for organizing data storage when analyzing big data is powerful distributed databases and data warehouses: Hydra [48], [49], Hadoop [50] and NetApp [51], [52]. For processing, the same methods can be used, as in MapReduce, to search for a HashSet [49], [52]. However, such approaches are fraught with low problems in real-time data processing, due to the excessive importance of high-latency analytical processes and poor applicability. Therefore, if simple additive metrics (such as total transactions, total page views, or average conversion price) are most valuable during analytics, it is obvious that the raw data can be effectively summed up, for example, on a daily basis or using simple counters in the flow. Calculating more complex metrics (such as average humidity, the number of unique users, or the most frequently encountered elements) is more complex and requires more resources if implemented consistently. PDS allows to evaluate these and many other metrics and manipulate [44-46] the accuracy of estimates for memory consumption. These data structures can be used as temporary data accumulators in query processing procedures, or perhaps more importantly, as a compact, sometimes surprisingly compact, replacement for raw data in streaming computing.

A number of PDS's will be analyzed in abstracts, while others are described in detail using a detailed mathematical analysis of these structures can be found in the original articles. Preliminary comments are as follows [47]:

- for some structures, such as Loglog Counter or Bloom filter, there are simple and practical formulas that allow you to determine the structure parameters based on the expected amount of data and the required error probability. Other structures, such as Count-Min Sketch or Stream-Summary, have a complex dependence on the statistical properties of data, and experiments are the only reasonable way to understand their applicability to real-world use cases;
- the applicability of the PDSs is not strictly limited to the above queries or a single set of data. Experiments and other works prove that structures filled with different data sets can often be combined to handle complex queries, and other types of queries can be supported using custom versions of the described algorithms.

## 2.1. Bloom filter

Bloom filter is probably the most well-known and widely used PDS [53-56]. Bloom filter is similar to Linear Counting, but it is designed to maintain the identity of each element, not statistics. A Bloom filter is a bit array of  $M$  bits initialized at 0. To add an element, it is processed by  $k$  hash functions to get  $k$  array positions, and set the bits in these positions to 1. To request an element, it is passed to  $k$  hash functions to get  $k$  array position. If any of the bits in these positions are 0, then the element is definitely not included in the set. If all bits are equal to 1, then the element can be in the set. That is, If the filter has a relatively large size compared to the number of individual elements, each element has a relatively unique signature and you can check a specific value – whether it is already registered in the bit set or not. If all the bits of the corresponding signature are units, then the answer is yes (with a certain probability, of course). Bloom filter with a false positive rate of 1% requires only 9.6 bits per element, regardless of the size of the elements [54-56].

Similar to Linear Counting, Bloom filter supports bit typing. However, as the from description makes clear, each value is mapped not to one, but to a certain fixed number of bits using several independent  $k$  hash functions.

Let's consider formulas that allow us to calculate the parameters of the Bloom filter as a function of error probability and capacitance. Given false positive probability  $p$  and the estimated number of insertions  $n$ , the length of the bit array can be calculated as [54-56]:

$$m = -\frac{n \ln p}{\ln^2 2}, \quad (1)$$

where  $m$  is filter (sketch) size (bits);

$p$  is an error probability (false positive);

$n$  is maximum cardinality (capacity).

The hash functions used for bloom filter should generally be faster than cryptographic hash algorithms with good distribution and collision resistance.

$$k = \frac{m}{n} \ln 2, \quad (2)$$

where  $k$  is number of hash functions.

In Figure 3 shows an example of placing elements in Bloom filter.

Consider an example: we enter  $x$ ,  $y$  and  $z$  in the filter with  $k = 3$  hash functions, as shown in the figure above. Each of these three elements has three bits, each with a value of 1 in the bit array. When we search for  $w$  in the set because one of the bits is not set to 1, Bloom filter will tell us that it is not in the set [53-56].

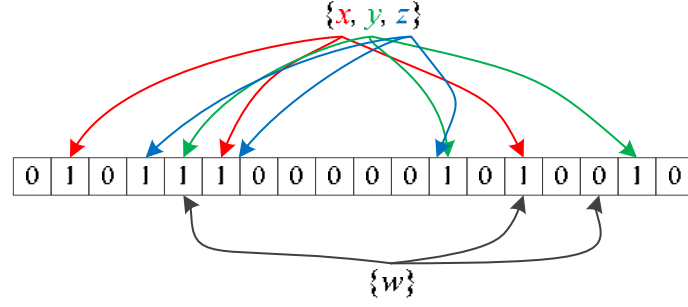


**Figure 3:** Placing elements in Bloom filter

## 2.2. MinHash

To begin with, MinHash is not just a PDS. In computer science and data mining, MinHash is a method for quickly estimating how similar two sets are. This scheme was invented by Andrei Broder (1997) [57] and was originally used in the AltaVista search engine to detect duplicate web pages and exclude them from search results [58]. It has also been used in large-scale clustering tasks, such as clustering documents based on the similarity of their word sets.

Let's introduce the concept of Jacquard similarity coefficient (Jaccard similarity coefficient) [57], [58]:

$$J[A, B] = \frac{|A \cap B|}{|A \cup B|}. \quad (3)$$

In other words, the number of elements in the cross-section is divided by the number of elements in the union. This estimate is called the Jacquard coefficient, the coefficient is zero when the sets have no common elements, and one when the sets are equal, in other cases the value is somewhere in the middle. In general, the calculation of such a coefficient boils down to the fact that at the beginning we need to divide the sets into separate parts, these will be the elements of our sets, then we need to somehow calculate the dimensions of the intersection and union. Usually, to efficiently perform the last two operations, sets are represented as hash tables without key-associated values. this structure works very quickly. Building a table:  $O(n)$ , you need two of them, calculating the cross section:  $O(n)$  and calculating the union is also  $O(n)$ , where  $n$  is the number of elements in the set [57], [58].

Here are two sets  $A$ ,  $B$  and  $h$  a hash function that can count hashes for elements of these sets. Next, we define the function  $h_{\min}(S)$ , which calculates the function  $h$  for all members of any set  $S$  and returns its lowest value. Now, let's start calculating  $h_{\min}(A)$  and  $h_{\min}(B)$  for different pairs of sets, the question is: what is the probability that  $h_{\min}(A) = h_{\min}(B)$ .

This probability must be proportional to the size of the intersection of sets: in the absence of common terms, it tends to zero, and to one, when the sets are equal, in intermediate cases it is somewhere in the middle. This is the  $J(A, B)$  so Jacquard coefficient.

However, if we just count  $h_{\min}(A)$  and  $h_{\min}(B)$  for our two sets and then compare the values, it won't give us anything, because there are only two options: equal or not equal. We need to somehow get enough statistics for our sets to calculate  $J(A, B)$  close to the truth. This is done simply, instead of one function  $h$ , we introduce several independent hash functions, or rather  $k$  functions [57]:

$$k = \left\lceil \frac{1}{\varepsilon^2} \right\rceil, \quad (3)$$

where  $\varepsilon$  is the highest error value is desired. So, to calculate  $J(A, B)$  with an error of no more than 0.1, you need 100 hash functions. On the one hand, this is a lot. So let discuss in what will be the optimization.

First, we can calculate the so-called  $H_{\min}(S)$  signature, i.e. the minima of all hash functions for the set  $S$ . The complexity of calculating it is greater than when building a hash table, but, nevertheless, it

is linear, and you only need to do this for each document once, for example, when adding it to the database, then it is enough to operate only with signatures [58].

Secondly, the signature has a fixed size for a given maximum error value. In order to compare any two sets of any size, you need to perform a fixed number of operations. In addition, in theory, signatures are much more convenient to index. In theory, because their indexing does not fit very well on relational databases, it is also not particularly suitable for full-text engines.

### 2.3. Count-Min Sketch

Count-min sketch is a PDS that is a table of event frequencies in a data stream. It uses hash functions to map events to frequencies, but unlike a hash table, it uses only sublinear space, by recalculating some events due to collisions [47], [59].

The Count-min sketch structure can be represented as a sketch with depth  $d$  and width  $w$  as at Figure 4.

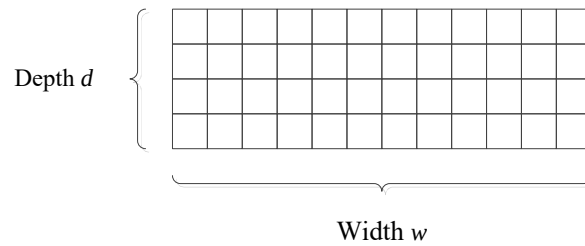


**Figure 4:** Count-min sketch with depth  $d$  and width  $w$

Then we represent the distribution of values in the Count-min sketch array, based on the principle shown in Figure 5 this distribution can be considered a sketch.

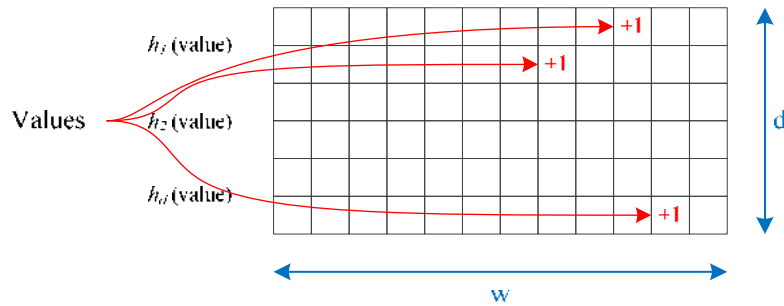


**Figure 5:** Distribution of values in the structure

In general, we will have a representation with a distribution according to the following rules (4) and (5).

$$\mathcal{E}_{estimation} \leq 2 \cdot \max cardinality_{hashfunction} / width_{sketch}, \quad (4)$$

$$\delta = 1 - 0.5^{depth_{sketch}}, \quad (5)$$

$\mathcal{E}_{estimation}$  is an evaluation error;

$\delta$  is a probability for a value with Count-min sketch;

$cardinality_{hashfunction}$  is a power of the hash function;

$depth_{sketch}$  is a sketch depth (height) ;

$width_{sketch}$  is a sketch width.

It is worth noting that the width of the sketch limits the error value (4), and the height (depth) controls the probability that the score will break this limit (5) [47], [59].

## 2.4. Comparison of PDS for further work

Let's compare the analyzed PDS. The comparison is shown in tabular form in Table. 1.

**Table 1**

PDS comparison

| Comparison criteria               | Bloom Filter           | MinHash | Count-Min Sketch |
|-----------------------------------|------------------------|---------|------------------|
| Required number of hash functions | Low                    | High    | Low              |
| Complexity of hash functions      | Low                    | Low     | Low              |
| Sketch size                       | Medium                 | Big     | Small            |
| Frequency of false positives      | Very low               | Low     | Low              |
| Ability to add / / delete items   | Requires modifications | Missing | Maybe            |

As can be seen from the tabular comparison, all the analyzed PDS have certain disadvantages, which may become certain limitations for their further practical use during neurosynthesis [47], [60]. So for Bloom Filter, there are restrictions on adding or removing elements. This problem is solved in the latest modifications, but such mechanisms significantly complicate the work. MinHash requires a large number of hash functions while running, and while they may be quite simple, calculating them still takes up some time resources. Moreover, when using a large number of such hash functions, the sketch of the resulting view also becomes more complicated. Count-Min Sketch looks like the best option in this comparison, precisely because of the low complexity of the sketch and low requirements for the number and complexity of hash functions. However, it should still be noted that there is a higher probability of false triggering (although it is also noted in other PDS cases) [47], [60].

Therefore, we will plan the further use scheme based on these shortcomings.

## 3. Materials and methods

Based on the above-mentioned features, subprocesses of neurosynthesis and analysis of the most popular PDS, we will determine the scheme of implementation of PDS in the process of neuroevolutionary synthesis of ANN:

- Bloom Filter will be used when sending information about the ANN between the cores of a parallel system;
- MinHash is used in the sequential execution of neuroevolutionary synthesis to encode genetic information about individuals, to assess acceleration at the breeding stage;
- Count-Min Sketch will also be used to encode genetic information about individuals during neuroevolution synthesis.

Previously, it has been noted that all genetic information about neural networks is encoded using direct coding sequencing [61-64]. In general, this approach is based on coding connections: the genotype of an individual presents information about the weights ( $WW$ ) of interneuronal connections of the neuromodel, but each gene will contain information about the indices of the initial ( $N_i$ ) and final ( $N_o$ ) communication neuron, as well as its weight. In the case when the method works with recurrent neural networks, an additional cell with the feedback weight is added, its index is determined by the index of the original neuron.

When using Bloom filter, four hash functions will be used, which will be used to sequentially encode values from an array of interneuronal relationships [65]:

$$H_{1(N_i)} \oplus H_{2(N_o)} \oplus H_{3(WW)} = f(WW), \quad (6)$$

So, to find the value of  $f(WW)$ , hash  $WW$  four times, perform three table searches, and exclude-or combine-four values. In practice, four  $v$  hashes can be reduced to a single 64-bit or 128-bit hash, which is divided into four 16-bit or 32-bit values, respectively.

Such an operation will be performed with the best individuals before sending them from parallel cores to the main thread of the parallel system to investigate the reduction of overhead costs.

MinHash will be used for coding and subsequent comparison in sequential neuroevolution synthesis. However, since adding or removing data from such a structure is not possible (or not rational due to computational costs), the population size will increase and the number of epochs will decrease [57].

When using Count-min Sketch, an array of data about the ANN is represented as a specific sketch and decoded afterwards as at Figure 6.

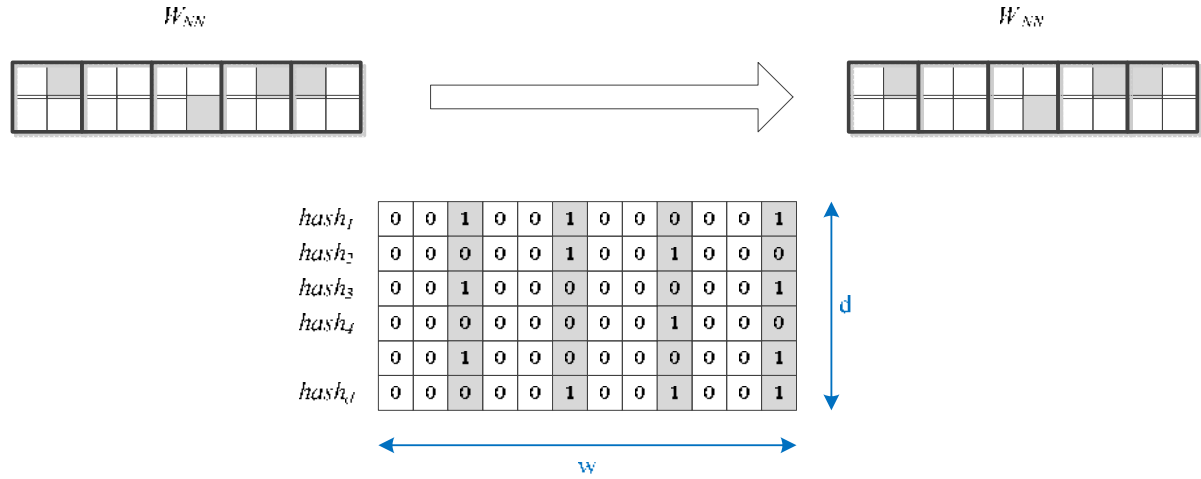


**Figure 6:** Encoding and decoding information about the ANN using Count-min sketch

Thus, the proposed encoding method begins with an initial representation of a matrix constructed from so called polymer cells containing information about the input-output connections between neurons and the weights of such connections.

Next, hash functions are defined (by default, we recommend taking 4 hash functions) and a matrix is created for their output, as shown in Figure 6.

After that, hash outputs are calculated for each element of the ANN data stream and the corresponding counter in the matrix is increased.

Thus, increasing the corresponding counts in the matrix, we get an updated matrix.

In some cases, due to a hash collision, it is possible that the received frequency of the element is slightly higher than expected. The encoding accuracy will depend on how unique the hash functions that return the element value are. Also, the larger the hash function, the more accurate the frequency will be.

In this case, the probabilistic data structure Count-min sketch allows you to calculate the frequency of big data flows in sublinear space with an estimated time complexity, that is  $O(1)$ , with a constant time component that does not depend on the parameters of the neural network model.

## 4. Experimental research

For an experimental study of the work, it was decided to choose a moderate sample of data as a sample, which would not initially require additional complications of neural networks, since this can become a problem in the case of using MinHash. test it on two tasks that would differ in scale. This is how the Tic-Tac-Toe Endgame Data Set was selected [66-69]. The main characteristics of the sample are shown in Table 2.

As a parallel computing system, the equipment of Department of the software tools the National university "Zaporizhzhia polytechnic" was used: Xeon processor E5-2660 v4 (14 cores), RAM 4x16 GB DDR4, the programming model of Java threads.



**Table 2**

Characteristics of the tic-Tac-Toe endgame data set sample

| Tic-Tac-Toe Endgame Data Set |              |                       |     |
|------------------------------|--------------|-----------------------|-----|
| Data Set Characteristics:    | Multivariate | Number of Instances:  | 958 |
| Attribute Characteristics:   | Categorical  | Number of Attributes: | 9   |

To test the MinHash case, the metaparameters specified in Table 3 will be set.

**Table 3**

Metaparameters for the MinHash use case

| Comparison criteria                     | Bloom Filter                         |
|-----------------------------------------|--------------------------------------|
| Population size                         | 100                                  |
| Number of training epochs               | 1                                    |
| Elite size                              | 5%                                   |
| Activation function (fitness functions) | hyperbolic tangent                   |
| Probability of mutation                 | 25%                                  |
| Crossover type                          | two-point                            |
| Types of mutation                       | deleting an interneuronal connection |
|                                         | removing a neuron                    |
|                                         | adding interneuronal connection      |
|                                         | adding a neuron                      |
|                                         | changing the activation function     |

For testing the Count–min sketch case, the main parameters remain unchanged, but the number of epochs increases to 5.

When testing the Bloom Filter use case, not only the meta parameters of the method were met, but also the hardware usage parameters specified in Table 4.

**Table 4**

Metaparameters for the Bloom Filter use case

| Comparison criteria                     | Bloom Filter                         |
|-----------------------------------------|--------------------------------------|
| Population size                         | 100                                  |
| Number of training epochs               | 5                                    |
| Elite size                              | 5%                                   |
| Activation function (fitness functions) | hyperbolic tangent                   |
| Probability of mutation                 | 25%                                  |
| Type of crossover on parallel threads   | two-point                            |
| Type of crossover on the main thread    | uniform (two-point)                  |
| Types of mutation                       | deleting an interneuronal connection |
|                                         | removing a neuron                    |
|                                         | adding interneuronal connection      |
|                                         | adding a neuron                      |
|                                         | changing the activation function     |
| Number of cores (threads)               | 14                                   |

The test results for the tic-tac-Toe Endgame data set sample presented using MinHash are shown in Table 5.

The test results for the tic-tac-Toe Endgame data set sample presented using Count-Min Sketch are shown in Table 6.

**Table 5**

Test results using MinHash

| Method        | Synthesis Time, s | Time at the selection stage, s | Memory usage (at the selection stage) | Error in the training sample | Error in the test sample |
|---------------|-------------------|--------------------------------|---------------------------------------|------------------------------|--------------------------|
| MGA           | 2487              | 561                            | 91%                                   | 0.11                         | 0.24                     |
| MGA + MinHash | 2189              | 287                            | 61%                                   | 0.11                         | 0.25                     |

**Table 6**

Test results using Count-Min Sketch

| Method                 | Synthesis Time, s | Time at the selection stage, s | Memory usage (at the selection stage) | Error in the training sample | Error in the test sample |
|------------------------|-------------------|--------------------------------|---------------------------------------|------------------------------|--------------------------|
| MGA                    | 5364              | 1742                           | 94%                                   | 0.024                        | 0.13                     |
| MGA + Count-Min Sketch | 4865              | 1025                           | 56%                                   | 0.029                        | 0.15                     |

The test results for the tic-tac-Toe Endgame data set sample presented using Bloom filter are shown in Table 7.

**Table 7**

Test results using Bloom filter

| Method             | Synthesis Time, s | Communication overhead | Average time of communication overhead | Error in the training sample | Error in the test sample |
|--------------------|-------------------|------------------------|----------------------------------------|------------------------------|--------------------------|
| MGA                | 2494              | 0.4                    | 199.5                                  | 0.022                        | 0.1                      |
| MGA + Bloom filter | 1867              | 0.28                   | 104.6                                  | 0.03                         | 0.174                    |

## 5. Analysis of experimental results

Analyzing the presented results, we can come to the following conclusions. When using MinHash, there were no significant improvements: it was at the selection stage that memory usage decreased from 92% to 61%, but up to this point, memory usage has not changed. Also, the acceleration was only at the selection stage, namely from 1742 s to 1025 s. However, the time for encoding and decoding networks increased significantly, precisely because of the complexity of this process, so the optimization was rather point-based and cannot be recommended for use in the future. In addition, we should mention the low accuracy due to the reduction in the number of epochs and the inability to use crossbreeding.

When using Count-min Sketch, the acceleration situation was similar, because during the execution of the method, the acceleration was noticeable, but the process of encoding and decoding individuals took too long, so the total synthesis time is almost the same. But memory usage has really become justifiably optimized from 94% to 56%. This can be explained by the fact that, unlike MinHash, almost during the entire process, in addition, such encoding made it possible to perform

mutation and crossing without significant problems. On the other hand, we should note a slight decrease in the accuracy of operation (due to false positives of the decoded network).

Using Bloom filter made it possible to reduce the share of shipments and their time, because data was sent from cores in a compact format. In addition, decoding networks on the main core did not impose additional complex calculations on parallel cores. However, even this strategy did not significantly improve the speed of work. Moreover, although uniform crossing was chosen on the main core, during the development of information modules, the problem of organizing multi-parent crossing arose. Therefore, unfortunately, it was necessary to use uniform crossover as a two-point crossover.

In general, it can be seen that the most promising is the use of Count-min Sketch, because this made it possible to perform the full synthesis process without additional complications.

## 6. Conclusions

The paper examines the variants and possibilities of using PDS in neuroevolutionary synthesis of ANNs, which can significantly reduce memory usage during synthesis and speed up this process.

The scientific novelty lies in the fact that during neuroevolutionary synthesis at different stages and with different execution schemes (sequential and parallel), it is proposed to use different PDS. This made it possible to study options for reducing resource intensity pointwise and in general. Studies have shown that despite the popularity of PDS for preserving big data in neuroevolutionary synthesis, this approach meets a number of obstacles. Moreover, during the work, recommendations were developed for the use of individual PDS for individual stages and approaches.

The practical significance lies in the fact that practical problems of coding the ANN are solved, which can later be used for diagnostics, forecasting, evaluation and modeling. The results of the experiments showed that the proposed coding methods make it possible to encode information about the ANN more compactly for its subsequent transmission to workstations for use as a model for diagnosis, prediction, evaluation and modeling.

Prospects for further research and development areas include a detailed analysis of the use of the Count-Min sketch data structure. This is due to the most acceptable results when testing various PDS and approaches to their use.

## 7. Acknowledgements

The work was carried out with the support of the state budget research projects of the state budget of the National University "Zaporozhzhia Polytechnic" "Intelligent methods and software for diagnostics and non-destructive quality control of military and civilian applications" (state registration number 0119U100360) and "Development of methods and tools for analysis and prediction of dynamic behavior of nonlinear objects" (state registration number 0121U107499).

## 8. References

- [1] M. Elgendy, Deep Learning for Vision Systems, New York, Manning Publications, 2020.
- [2] D. Hudgeon, R. Nichol, Machine Learning for Business: Using Amazon SageMaker and Jupyter, Manning Publications, 2020.
- [3] D. Barh, Artificial Intelligence in Precision Health: From Concept to Applications, Academic Press, Cambridge, 2020.
- [4] K. Warr, Strengthening Deep Neural Networks: Making AI Less Susceptible to Adversarial Trickery, O'Reilly Media, Massachusetts, 2019.
- [5] Y. Goldberg, G. Hirst, Neural Network Methods in Natural Language Processing (Synthesis Lectures on Human Language Technologies), Morgan and Claypool Publishers, California, 2017.
- [6] T. Neskorodieva, E. Fedorov, Method for Automatic Analysis of Compliance of Expenses Data and the Enterprise Income by Neural Network Model of Forecast, in: Proceedings of the

- Mathematical Modeling and Simulation of Systems (MODS'2020), Springer (2020), 156-165. doi: 10.1007/978-3-030-58124-4\_15
- [7] J.A.J. Alsayaydeh, W.A.Y. Khang, A.K.M.Z Hossain, V. Shkarupylo, J. Pusppanathan, The experimental studies of the automatic control methods of magnetic separators performance by magnetic product, ARPN Journal of Engineering and Applied Sciences, vol. 15(7) (2020) 922-927.
  - [8] J.A.J. Alsayaydeh, W.A. Indra, W.A.Y. Khang, V. Shkarupylo, D.A.P.P. Jkatisan, Development of vehicle ignition using fingerprint, ARPN Journal of Engineering and Applied Sciences, vol. 14(23) (2019) 4045-4053.
  - [9] J.A.J. Alsayaydeh, W.A.Y. Khang, W.A. Indra, J.B. Pusppanathan, V. Shkarupylo, A.K.M. Zakir Hossain, S. Saravanan, Development of vehicle door security using smart tag and fingerprint system, ARPN Journal of Engineering and Applied Sciences, vol. 9(1) (2019) 3108-3114.
  - [10] J.A.J. Alsayaydeh, W.A.Y. Khang, W.A. Indra, V. Shkarupylo, J. Jayasundar, Development of smart dustbin by using apps, ARPN Journal of Engineering and Applied Sciences, vol. 14(21) (2019) 3703-3711.
  - [11] J.A. Alsayaydeh, M. Nj, S.N. Syed, A.W. Yoon, W.A. Indra, V. Shkarupylo, C. Pellipus, Homes appliances control using bluetooth, ARPN Journal of Engineering and Applied Sciences, vol. 14(19) (2019) 3344-3357.
  - [12] Google, 2021. URL: <https://www.google.com/>
  - [13] S. Dresden, Search engine with neural network weighting based on parametric user data, 2004. Patent No. US20050004905A1, Filed March 3rd., 2003, Issued June 6th., 2005.
  - [14] Search Engines and Neural Networks, 2018. URL: <https://towardsdatascience.com/https-towardsdatascience-com-search-engines-and-neural-networks-97e0df4f088d>
  - [15] W. Serrano, Neural Networks in Big Data and Web Search. Neural networks in web search, 4(1):7 (2018) 1-41. doi: 10.3390/data4010007
  - [16] Bing, 2021. URL: <https://www.bing.com/>
  - [17] Microsoft to accelerate Bing search with neural network, 2015. URL: <https://www.extremetech.com/extreme/199814-microsoft-to-accelerate-bing-search-with-neural-network>
  - [18] Baidu, 2021. URL: <http://www.baidu.com/>
  - [19] YATI & ERNIE: Machine Learning in Yandex and Baidu, 2021. URL: <https://www.searchenginejournal.com/yati-ernie-machine-learning-yandex-baidu/392982/>
  - [20] V.Mnih, A.P. Badia, M. Mirza, A. Graves, et al., Asynchronous methods for deep reinforcement learning, in: Proceedings of the 33rd International Conference on International Conference on Machine Learning, ICML'16, 48 (2016) 1928–1937.
  - [21] L. Tai, J. Zhang, M. Liu, J. Boedecker, W. Burgard, A Survey of Deep Network Solutions for Learning Control in Robotics: From Reinforcement to Imitation, 2018. URL: <https://arxiv.org/abs/1612.07139>
  - [22] R. Gilman, K. Wang Intuitive RL: Intro to Advantage-Actor-Critic (A2C), 2018. URL: <https://hackernoon.com/intuitive-rl-intro-to-advantage-actor-critic-a2c-4ff545978752>
  - [23] J. K.H. Franke, G. Köhler, N. Awad, F. Hutter Neural Architecture Evolution in Deep Reinforcement Learning for Continuous Control NeurIPS 2019 MetaLearn Workshop, 2019. URL: <https://arxiv.org/abs/1910.12824>
  - [24] Reinforcement Learning: Deep Q Networks, 2020. URL: <https://blogs.oracle.com/datascience/reinforcement-learning-deep-q-networks>
  - [25] A. Juliani Simple Reinforcement Learning with Tensorflow Part 8: Asynchronous Actor-Critic Agents (A3C), 2016. URL: <https://medium.com/emergent-future/simple-reinforcement-learning-with-tensorflow-part-8-asynchronous-actor-critic-agents-a3c-c88f72a5e9f2>
  - [26] P. Dangeti, Statistics for Machine Learning: Techniques for exploring supervised, unsupervised, and reinforcement learning models with Python and R, Packt Publishing, Birmingham, 2017.
  - [27] B. Bateman, A. R. Jha, B. Johnston, I. Mathur, The Supervised Learning Workshop: A New, Interactive Approach to Understanding Supervised Learning Algorithms, 2nd Edition, Packt Publishing, Birmingham, 2020.
  - [28] X. Zhu, A.B. Goldberg, Introduction to Semi-Supervised Learning (Synthesis Lectures on Artificial Intelligence and Machine Learning, Morgan and Claypool Publishers, California, 2009.

- [29] G. Hinton, *Unsupervised Learning: Foundations of Neural Computation* (Computational Neuroscience), Bradford Books, Massachusetts, 1999.
- [30] A. Jones, C. Kruger, B. Johnston *The Unsupervised Learning Workshop: Get started with unsupervised learning algorithms and simplify your unorganized data to help make future predictions*, Packt Publishing, Birmingham, 2020.
- [31] R.S. Sutton, A.G. Barto, *Reinforcement Learning, second edition: An Introduction* (Adaptive Computation and Machine Learning series) second edition, Bradford Books, Massachusetts, 2018.
- [32] P. Winder, *Reinforcement Learning: Industrial Applications of Intelligent Agents*, O'Reilly Media, Massachusetts, 2020.
- [33] L. Graesser, W.L. Keng, *Foundations of Deep Reinforcement Learning: Theory and Practice in Python* (Addison-Wesley Data & Analytics Series), Addison-Wesley Professional, Boston, 2019.
- [34] K. O. Stanley, J. Clune, J. Lehman, et al., Designing neural networks through neuroevolution. *Nature Machine Intelligence* 1 (2019) 24–35. doi: 10.1038/s42256-018-0006-z
- [35] Omelianenko, *Hands-On Neuroevolution with Python: Build high-performing artificial neural network architectures using neuroevolution-based algorithms*, Packt Publishing, Birmingham, 2019.
- [36] A. Bergel, *Agile Artificial Intelligence in Pharo: Implementing Neural Networks, Genetic Algorithms, and Neuroevolution*, Apress, New York, 2020.
- [37] A. Gaier, David Ha *Exploring Weight Agnostic Neural Networks*, 2019. URL: <https://ai.googleblog.com/2019/08/exploring-weight-agnostic-neural.html>
- [38] K. O. Stanley, J. Clune *Welcoming the Era of Deep Neuroevolution*, 2017. URL: <https://eng.uber.com/deep-neuroevolution/>
- [39] A.O. Oliinyk, T.A. Zaiko, S.A. Subbotin, Factor analysis of transaction data bases. *Automatic Control and Computer Sciences* 48(2) (2014) 87-96. doi: 10.3103/S0146411614020060
- [40] A. Oliinyk, S. Skrupsky, S.A. Subbotin, Parallel computer system resource planning for synthesis of neuro-fuzzy networks. *Advances in Intelligent Systems and Computing*, 543 (2017) 88-96. doi: 10.1007/978-3-319-48923-0\_12
- [41] A. Oliinyk, S. Skrupsky, S. Subbotin, I. Korobiichuk, Parallel method of production rules extraction based on computational intelligence. *Automatic Control and Computer Sciences*, 51(4) (2017) 215-223. doi: 10.3103/S0146411617040058
- [42] A. Oliinyk, S. Skrupsky, S. Subbotin, Experimental research and analysis of complexity of parallel method for production rules extraction. *Automatic Control and Computer Sciences*, 52 (2) (2018) 89-99. doi: 10.3103/S0146411618020062
- [43] S. Leoshchenko, A. Oliinyk, S. Subbotin, T. Zaiko, N. Gorobii, Implementation of selective pressure mechanism to optimize memory consumption in the synthesis of neuromodels for medical diagnostics, in: *Proceedings of the 2nd International Workshop on Informatics & Data-Driven Medicine, IDDM 2019, CEUR-WS, 2019*, pp. 109-120. DBLP Key: conf/iddm/LeoshchenkoOSZG19
- [44] A. Gakhov, *Probabilistic Data Structures and Algorithms for Big Data Applications*, Books on Demand, 2019.
- [45] H. Knebl, *Algorithms and Data Structures: Foundations and Probabilistic Methods for Design and Analysis*, Springer, Berlin, 2020.
- [46] T.H. Cormen, C.E. Leiserson, R.L. Rivest, C. Stein, *Introduction to Algorithms*, The MIT Press, Cambridge, 2009.
- [47] *Probabilistic data structures for web analytics and data mining*, 2012. URL: <https://highlyscalable.wordpress.com/2012/05/01/probabilistic-structures-web-analytics-data-mining/>
- [48] Hydra, 2014. URL: <https://github.com/addthis/hydra>
- [49] *Big Data Hadoop Alternatives: What They Offer and Who Uses Them*, 2018. URL: <https://datafloq.com/read/Big-Data-Hadoop-Alternatives/1135>
- [50] Apache Hadoop, 2021. URL: <https://hadoop.apache.org/>
- [51] NetApp, 2021. URL: <https://www.netapp.com/>
- [52] *Big Data Analytics solutions: The bigger the better*, 2021. URL: <https://www.netapp.com/artificial-intelligence/big-data-analytics/>

- [53] B.H. Bloom, Space/time trade-offs in hash coding with allowable errors. *Communications of the ACM* T. 13(7) (1970) 422–426. <https://doi.org/10.1145/362686.362692>
- [54] Bloom Filters: Is element x in set S?, 2017. URL: <https://www.abhishek-tiwari.com/bloom-filters-is-element-x-in-set-s/>
- [55] How to count Big Data: Probabilistic data structures and algorithms, 2019. URL: <https://www.kdnuggets.com/2019/08/count-big-data-probabilistic-data-structures-algorithms.html>
- [56] Introduction to Probabilistic Data Structures, 2015. URL: <https://dzone.com/articles/introduction-probabilistic-0>
- [57] A.Z. Broder, On the resemblance and containment of documents. *Compression and Complexity of Sequences: Proceedings, IEEE* (1997) 21–29. doi:10.1109/SEQUEN.1997.666900
- [58] A.Z. Broder, M. Charikar, A.M. Frieze, M. Mitzenmacher, Min-wise independent permutations, in: *Proceedings of the 30th ACM Symposium on Theory of Computing (STOC '98)*, New York, NY, USA: Association for Computing Machinery (1998) 327–336. doi:10.1145/276698.276781
- [59] G. Cormode, Count-Min Sketch, *Encyclopedia of Database Systems*, Springer, 2009, 511–516.
- [60] I. Katsov, *Introduction to Algorithmic Marketing: Artificial Intelligence for Marketing Operations*, Books on Demand, 2017.
- [61] S. Leoshchenko, A. Oliinyk, S. Subbotin, T. Zaiko, S. Shylo, V. Lytvyn, Sequencing for encoding in neuroevolutionary synthesis of neural network models for medical diagnosis, in: *Proceedings of the 3rd International Conference on Informatics & Data-Driven Medicine, IDDM 2020, CEUR-WS*, 2020, pp. 62–71.
- [62] B.A. Pierce, *Genetics: A Conceptual Approach*, W.H. Freeman & Co. Ltd., New York, 2019.
- [63] R. Brooker, *Genetics: Analysis and Principles*, McGraw-Hill Education, New York, 2020.
- [64] S. Leoshchenko, A. Oliinyk, S. Subbotin, N. Gorobii, T. Zaiko, Synthesis of artificial neural networks using a modified genetic algorithm, in: *Proceedings of the 1st International Workshop on Informatics & Data-Driven Medicine, IDDM 2018, CEUR-WS*, 2018, pp. 1–13. DBLP Key: conf/iddm/PerovaBSKR18
- [65] B. Reagen, U. Gupta, R. Adolf, M. Mitzenmacher, A. Rush, G.-Y. Wei, D. Brooks, Weightless: Lossy Weight Encoding For Deep Neural Network Compression, *Computer Science, Mathematics*, 2017. URL: <http://proceedings.mlr.press/v80/reagan18a/reagan18a.pdf>
- [66] Tic-Tac-Toe Endgame Data Set, 1991. URL: <https://archive.ics.uci.edu/ml/datasets/Tic-Tac-Toe+Endgame>
- [67] C. J. Matheus, J.L. A. Rendell Constructive induction on decision trees. *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence*, Detroit, MI: Morgan Kaufmann, 1989, pp. 645–650.
- [68] C. J. Matheus Adding domain knowledge to SBL through feature construction. *Proceedings of the Eighth National Conference on Artificial Intelligence*, Boston, MA: AAAI Press, 1990, pp. 803–808.
- [69] D. W. Aha Incremental constructive induction: An instance-based approach. In *Proceedings of the Eighth International Workshop on Machine Learning*, Evanston, ILL: Morgan Kaufmann, 1991, pp. 117–121.

# Data Mining Methods for Evaluation and Forecasting the Mobile Internet Traffic in Roaming

Nataliia V. Kuznietsova<sup>a</sup>, Petro I. Bidyuk<sup>a</sup> and Anastasiia V. Kulinich<sup>a</sup>

<sup>a</sup> National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, 03056, Ukraine

## Abstract

This paper is dedicated to the solving such real actual practical task for telecommunication industry as forecasting the services in roaming. It is concentrated on the study of the data mining methods that allow to predict the volume of services (calls, traffic) for a particular subscriber abroad on the bases of available statistical information. The problem was shared in two tasks: forecasting the volume of the internet traffic in roaming and the clients' classification due to their behavior in roaming. The task of evaluation and forecasting was solved with the models based on the time series theory and appropriate autoregression models. The best model was selected based on the statistical criteria and used for forecasting the volume of traffic in next months. The task of classification was solved by such data mining methods as neural networks, gradient boosting, random forest and logistic regression. The model based on the gradient boosting was selected because of the highest completeness and accuracy to the input data. In accordance to received modelling results the recommendations and special strategies for the telecommunication company were developed.

## Keywords 1

Data Mining, Gradient Boosting, Random Forest, Neural Networks, Logistic Regression, Bootstrap analysis, Time Series, Mobile Internet Traffic, Forecasting, Roaming

## 1. Introduction

Today mobile operators are actually interested in resuming their services including mobile roaming, which was observed before the Covid-19 epidemic. In order to predict the mobile Internet volume, including abroad [1], it is necessary to determine optimal packages and behavior of subscribers [2, 3], i.e. how they use certain services within their own country. In the future, it will be possible to predict how such subscriber will behave abroad using modern methods of data mining and forecasting [4 – 12].

Special mention should be paid to virtual mobile operators (for example, LycaMobile), which use roaming technology and rent telecommunication towers and equipment of other mobile operators to provide services. Virtually every subscriber of a virtual operator is in roaming within his own country, and therefore information on the Ukrainian services usage can be used to predict his behavior abroad.

Roaming is a situation when a mobile operator subscriber (of the home network) uses the network of another mobile operator outside the geographical coverage area of the home network. The peculiarity of doing business in the field of international roaming is the need to make a huge number of deals (agreements) with major international telecommunication operators providing services in more than a hundred countries [1]. The national operator should take into account that certain services in some telecommunication companies abroad are extremely expensive for both the operator and the subscriber. Thus, roaming tariffs in such countries should be formed and allocated

CMIS-2021: The Fourth International Workshop on Computer Modeling and Intelligent Systems, April 27, 2021, Zaporizhzhia, Ukraine  
EMAIL: natalia-kpi@ukr.net (N. Kuznietsova); pbidyuk\_00@ukr.net (P. Bidyuk); a.kulinich2509@gmail.com (A. Kulinich)  
ORCID: 0000-0002-1662-1974 (N. Kuznietsova); 0000-0002-7421-3565 (P. Bidyuk); 0000-0001-7156-109X (A. Kulinich)



© 2020 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

in a separate group, because for the mobile operator one minute can cost much more than the company can receive from the subscriber.

The national mobile operator must also provide appropriate conditions for the usage of mobile services for its subscribers around the world. Services that subscribers use abroad can be divided into several types including voice communication, messages and mobile Internet.

Usually Ukrainian subscribers abroad prefer voice calls and mobile data usage. It was expected that with providing new 4G standard the traditional telecommunication services such as voice calls and short messages (SMS) remained in the past. Now there is an actual decrease in the number of subscribers who use standard short messages but with voice calls it is not so clear. The people needs in live communication which is impossible in today's lockdown and quarantine situations have led to a huge increase in demand for voice calls. Therefore, now this service remains one of the most developed and necessary for users and therefore relevant in roaming. Let's take a closer look at the tasks of forecasting services and tariff packages faced by telecommunication companies to provide users in roaming. To do this, first the problem of the traffic in roaming amount forecasting will be solved. Next the subscribers' classification according to mobile Internet tariff packages in roaming will be performed. As a result, it will be possible to determine which tariff packages are relevant now for subscribers, whether it is needed to develop one or more different proposals.

## 2. Problem statement

This work is concentrated on the study and research of the information tools that allow one to predict the volume of services (calls, traffic) for a particular subscriber abroad on the bases of available statistical information. Mathematical models need to be developed to analyze consumption and forecast the services that will be used abroad. It is necessary to predict the mobile Internet traffic volume in roaming using the statistics on subscriber behavior and perform the subscribers' classification in order to develop relevant packages and offers. Finally it is needed to develop special techniques and approaches to increase the number of subscribers who will use these services in roaming.

## 3. The main methods used in the research

According to the seventh international workshop held by SAS in September 2020 "Real Time Analytics & Cyber Security 2020", the "big four" of the most promising and relevant methods for solving analytical problems in financial sector were as follows: neural networks [11– 13], gradient boosting, random forest and support vectors machine [14 – 16]. It was also noted that the high results for classification shows the logistic regression method. So, based on the international experience we chose these methods to solve the problem of classification for our tasks. For the internet traffic forecasting the theory of time series analysis [8, 9, 17, 18] was applied and some autoregressive models were explored and built after preliminary data processing.

### 3.1. Logistic regression

Logistic regression is a statistical method used to analyze a data set consisting of one or many independent characteristics that affect the outcome. The result is evaluated using a dichotomous variable, which indicates with what probability the result belongs to a particular class [10]. The logistic regression algorithm uses a linear equation (boundary function) with independent variables to determine which class the data belongs to [4]. This equation describes the linear boundary that separates the input data space.

The limit function can be generally written as follows:

$$\text{Score}(x_i) = w_0 + w_1x_1 + \dots + w_nx_n = w^T \cdot x_i, \quad (1)$$

where  $x_i$  is input variable feature,  $w_0$  is decision making threshold,  $w_1, \dots, w_n$  is vector of weights.



In order to obtain the probability value  $[0; 1]$  of the class membership from the boundary function, the following logistic function is used:

$$\text{logit}(\text{Score}) = \frac{1}{1 - e^{-\text{Score}}}. \quad (2)$$

First of all, to construct the limit function it is necessary to find the coefficients,  $w_1, \dots, w_n$ . To do this, it is necessary to determine the training sample, which consists of independent variables (characteristics) and the corresponding values of the dependent variable  $y$  (initial result). Formally, it is a set of pairs,  $(x^{(1)}; y^{(1)}) \dots (x^{(m)}; y^{(m)})$ , where  $x^{(i)} \in R^n$  is the vector of independent variables values and  $y^{(i)} \in \{0, 1\}$  is the corresponding value of  $y$ . Each such pair is called the learning example. Usually the method of maximum likelihood is used and the parameters are selected so that to maximize the value of the likelihood function in the training sample [4]:

$$W = \arg \max_w L(W) = \arg \max_w \prod_{i=1}^m P\{y = y^{(i)} \mid x = x^{(i)}\}. \quad (3)$$

Maximizing the likelihood function is equivalent to maximizing its logarithm:

$$\begin{aligned} \log L(W) &= \sum_{i=1}^m \log P\{y = y^{(i)} \mid x = x^{(i)}\} = \\ &= \sum_{i=1}^m y^{(i)} \log f(w^T x^{(i)}) + (1 - y^{(i)}) \log(1 - f(w^T x^{(i)})) \end{aligned} \quad (4)$$

where  $x^{(i)} = w_0 + w_1 x_1 + \dots + w_n x_n$ .

To maximize this function the gradient descent method could be used. By setting some initial value  $w_0$  a maximum can be found iteratively [4]:

$$w = w_0 + \alpha \nabla \log L(w) = w_0 + \alpha \sum_{i=1}^m (y^{(i)} - f(w^T x^{(i)})) x^{(i)}, \alpha > 0. \quad (5)$$

### 3.2. Neural networks

The most modern neural networks are constructed of formal neurons that resemble their biological prototype. The structure of the neuron consists of  $x_1, \dots, x_n$  are the values that are fed to the inputs (synapses) of the neuron;  $w_1, \dots, w_n$  are weighting coefficients of synapses, which can have both slowing down and strengthening effect;  $S$  is the weighted sum of the input characteristics:

$$S = \sum_{i=1}^n w_i \cdot x_i - T, \quad (6)$$

$T$  is neuron threshold (omitted in many models),  $F$  is the neuron activation function that converts the weighted sum into an output signal:  $y = F(S)$  [20].

The neurons are regularly organized into layers, and the elements of a layer are associated only with the neurons of the previous layer and the information spreads from the previous layers to the next. The input layer, which consists of sensitive (sensory) S-elements, which receives the input signals,  $X_i$ , does not perform any information processing and performs only distribution functions. Each S-element is associated with a set of associative elements (A-elements) of the first intermediate layer, and the A-elements of the last layer are connected to the reacting elements (R-elements) [20].

Weighted combinations of R-element outputs determine the system response which indicates that the evaluated object belongs to a certain image. If only two images are recognized, then one R-element is installed in the perceptron, which has two reactions – positive and negative. If there are more than two images then for each image its R-element is set and each such element output is a linear combination of output A-elements.

Neural networks are now one of the most common methods that is being developed, and adapted, using recurrent algorithms (RNN), long short-term dependency learning (LSTM), to solve regression problems (GRNN). Therefore, the use of neural networks is appropriate for our problems.

### 3.3. Random Forest

The random forest method is based on the large number (ensemble) of decision trees (this number is a parameter of the method) construction, each of which is built on a sample obtained from the original training sample using bootstrap (i.e. sample with return), in contrast to classical construction algorithms decision trees [15].

The *bootstrap procedure* is randomly retrieving repeated samples from the empirical distribution multiple times. Specifically, if we have an initial sample of  $n$  terms,  $x_1, x_2, \dots, x_{n-1}, x_n$ , then by using the random numbers generator evenly distributed on the interval  $[1, n]$ , we can extract from it an arbitrary element  $x_k$ , which will be returned to the original sample for possible re-extraction. This procedure is repeated  $n$  times. A bootstrap sample is formed where some elements can be repeated two or more times, while other elements are absent. For example, for  $n = 6$  one of such bootstrap combinations has the following form:  $x_1, x_2, x_2, x_1, x_4, x_5$  [21].

Bootstrap samples are performed evenly and with a return, so some initial samples will be missing while others will be duplicated: on average, one such sample contains about  $2/3$  of unique initial observations. Bootstrap was particularly useful in models ensemble formation especially in combination with tree-like structures which are very sensitive to small changes in training data [6].

As the averaging of several observations reduces the data variance estimation the same reasonable way to reduce the variance of the forecast is to obtain a large number of data from the general population, building a predictable model for each training sample and averaging of the obtained forecasts. If instead of separate training samples to perform bootstrap and based of the generated pseudo-samples to build  $B$  regression trees, the average collective forecast will have a lower variance:

$$\hat{f}_{bag} = (f^1(x) + f^2(x) + \dots + f^B(x)) / B. \quad (7)$$

This procedure is called *bagging* (short abbreviation for bootstrap aggregating). Bagging can be performed not only for regression trees but also to other models [10].

The random Forest is an improvement of decision tree bagging which aims to eliminate the correlation between trees. As with bagging we build several hundred decision trees based on training bootstrap samples. However, at each iteration of constructing the trees are randomly selected  $m$  from  $p$  as to be considered as predictors and it is allowed to perform partitioning only by one of these  $m$  variables [15]. The meaning of this procedure is quite effective for improving the quality of the obtained solutions and it is that with the probability  $\frac{p-m}{p}$  any potentially dominant predictor

that seeks to enter each tree is blocked. If the dominance of such predictors is allowed then all the trees as a result will be very similar to each other. Also the obtained on their basis forecasts will be strongly correlated and the decrease in variance will not be so obvious. By blocking the dominants other predictors will get their chance and the tree variation increases. Choosing a small value of  $m$  when constructing a random forest will be useful in case of a large number of correlating predictors. Naturally, if a random forest is built using  $m = p$  then the whole procedure is reduced to a simple bagging [20].

Random forests provide a significant increase in accuracy while the trees in the ensemble are weakly correlated due to the double injection of randomness into the inductive algorithm – by bagging and random subspaces methods for splitting each vertex; they don't exhibit the overfitting problem. They are easy for usage: the only algorithm parameters are the trees number in the ensemble and the number of traits randomly selected for splitting at each top of the tree.

### 3.4. Gradient Boosting

Now let's consider the problem of recognizing objects from the multidimensional space  $X$  with the label space. Let a training sample,  $\{x_i\}_{i=1}^N$ , where  $x_i \in X$  is given. And let are known the true

values of the labels for each object,  $\{y_i\}_{i=1}^N$ , where  $y_i \in Y$ . It is necessary to build a recognition operator that can predict the labels for each new object  $x \in X$  as accurately as possible. Let the family of the basic algorithms  $H$  is given, each element of  $h(x; a) \in H : X \rightarrow R$  is determined by some vector of parameters  $a \in A$  [7].

We will search for the final classification algorithm in the form of the composition

$$F_M(x) = \sum_{m=1}^M b_m h(x; a_m), b_m \in R, a_m \in A. \quad (8)$$

However, the selection of the optimal set of parameters  $\{a_m, b_m\}_{m=1}^M$  is a very time-consuming task. Therefore, we will try to build such a composition by greedy manner building, each time adding to the sum a term which is the most optimal parameter from all possible. We assume that we have already constructed a classifier  $F_{m-1}$  of length,  $m-1$ . Thus, the problem is to find a pair of the most optimal parameters  $\{a_m, b_m\}$  for the classifier of length,  $m$ :

$$F_m(x) = F_{m-1}(x) + b_m h(x; a_m), b_m \in R, a_m \in A. \quad (9)$$

The idea of boosting can be also applied to classification. In case of binary classification, this means  $Y = \{-1; +1\}$ . Then it is often assumed that each algorithm  $h \in H$  returns the actual «degree» of object belonging to a certain class, and the resulting answer  $\tilde{F}$  is obtained by applying a boundary rule to the composition [16].

### 3.4.1. Multiclass classification

The idea of boosting for binary classification could be easily generalized to the case of  $K$  classes [22]. Now the following loss function is introduced:

$$L(y, F) = -\sum_{i=1}^k y_i \log p_i(x). \quad (10)$$

Here  $y_i \in \{0,1\}$  shows the affiliation of the object of class  $i$ , and  $p_i$  shows the probability of belonging the object to the class  $i$ , obtained during application of the logistic regression. Write down the formulas for the class  $K$  classifier of multiclass logistic regression:

$$f_k(x) = \log p_k(x) - \frac{1}{K} \sum_{i=1}^K \log p_i(x). \quad (11)$$

It is possible to receive after transformations that:

$$\nabla Q_i = y_{ik} - p_{k,m-1}(x_i). \quad (12)$$

If the problem is too complicated for calculations then in case of computational trees it is possible to use the first step of the Newton-Raffson algorithm as an approximation:

$$c_{jm} = \frac{K-1}{K} \frac{\sum_{x_j \in R_{jkm}} \nabla Q_{ik}}{\sum_{x_j \in R_{jkm}} |\nabla Q_{ik}| (1 - |\nabla Q_{ik}|)}. \quad (13)$$

The  $i$ -th classifier search for such object class means that the probability of belonging of other classes was minimum:

$$c_{jm} = \arg \min_{k \in [1, K]} \sum_{k=1}^K c(k, \tilde{k}) p_{\tilde{k}m}(x). \quad (14)$$

In this formula,  $p_{\tilde{k}m}(x)$  denotes the probability of belonging to class  $k$  as a result of the  $m$ -th iteration of the boosting algorithm. The value of  $c(k, \tilde{k})$  denotes the error cost function if it is assumed that the object belongs to the class,  $k$ , although in fact it belongs to the class  $\tilde{k}$  [16].

## 4. Input data features and characteristics description

To analyze and predict subscribers who are using roaming services, the mobile operator provided a sample of 120,000 records – data for subscribers traveling abroad. For 10,000 random records from

each month for the period from August 2017 to July 2018, i.e. for twelve consecutive months. It is necessary to predict, first of all, whether the subscriber traveling abroad will use communication services as well as to determine which services (calls or mobile internet) the subscriber will use.

The input data contains the following characteristics: the previous month before going abroad (used to display the history of using services in Ukraine); and a set of characteristics that reflect the features and types of services used by the subscriber in Ukraine (subscriber's internal tariff plan, the number of calls minutes, the number of short messages, the amount of GPRS traffic, the cost spent by the subscriber to use services in Ukraine, the amount and quantity replenishment of the balance in the specified month, administrative region of Ukraine, in which the subscriber was more than 90 days before the moment of departure abroad). We also add characteristics for our task: the month in which the subscriber was left abroad, the code of the country tariff group and the name of the country to which the subscriber left.

The following characteristics describe the amount of services that the customer used in roaming during his previous trip abroad. Such information is stored by the mobile operator for the entire period of customer service, but for our task it is advisable to take into account the previous visit not earlier than the year of 2016. This is due to significant changes in tariffs and the high cost of roaming services, and therefore there was very infrequent use of this service before 2016. This is significantly different from today's situation and the mobile operator faces the task of determining the subscriber behavior and relevant services in modern conditions.

- HISTORY\_ROAM\_GPRS is the amount of GPRS traffic in roaming.
- HISTORY\_ROAM\_MINS is the amount of call minutes in roaming.
- HISTORY\_ROAM\_SMS is the number of short messages in roaming.

If the subscriber has not traveled abroad in the last three years, the characteristic values will be empty. Empty values of these characteristics will be further processed and filled by one of the methods for incomplete data recovering (average or median value for the tariff group of countries) [23].

In total the sample contains 15 characteristics that describe the subscriber behavior in Ukraine and abroad. An experimental study was conducted and such characteristics from behavioral characteristics in Ukraine were generalized and selected for further modeling:

- MO\_UKR is the sum of outgoing calls made inside the network to other operators and also outgoing calls abroad.
- SMS\_UKR is the sum of outgoing short messages in the network sent on the phone numbers of the national operators and international short messages.
- AMOUNT is the total sum of all services interaction into calls, GPRS and short messages.

## **5. Data mining methods application to solve the classification and forecasting problems**

### **5.1. The task of predicting the services amount that will be used by the subscriber in roaming**

Based on subscribers' statistics it is necessary to forecast and offer the outgoing subscriber an appropriate services package for roaming. It is also necessary to predict the amount of mobile GPRS traffic in roaming, which will be used by the subscriber – the target variable  $Y_{mi}$ . It is advisable to try to solve this problem by using different types of regression models [5, 8, 18].

To forecast the mobile data in roaming usage was used the dataset for the period from January 2016 to July 2018. Altogether, the data was provided for 31 consecutive months. A graphical representation of the time series for mobile internet usage in this period is shown in Figure 1.

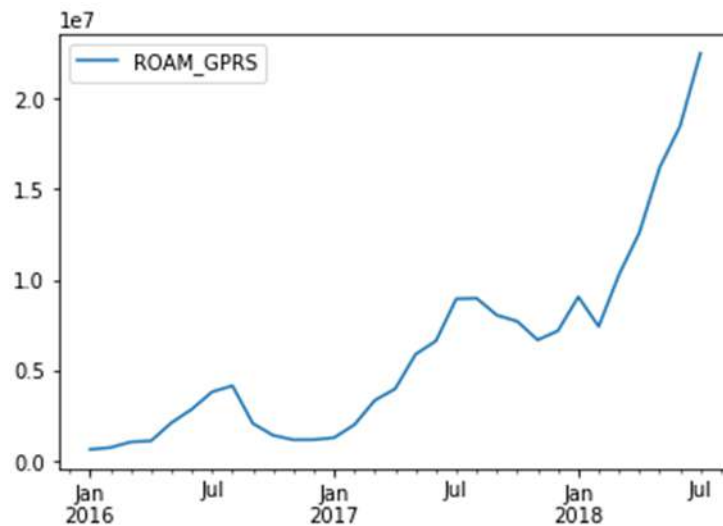


**Figure 1:** The volume of GPRS mobile data traffic, MB

The usage of the GPRS service abroad is growing all this period due to a significant reduction in the prices for the service of subscribers and international partners. Visually, the time series shown in Figure 1, has an exponential trend, but we still firstly conducted a series of studies on stationarity, constructed a moving average and standard deviation, performed Dickey-Fuller and KPSS tests, which indicated that the time series is not stationary (Dickey-Fuller: p-value = 1.00, KPSS criterion: p-value = 0.05) [24, 25].

In order to transform the original series of mobile data in roaming to stationary, it is necessary to remove the exponent by usage the second differences. After this transform the time series became stationary so autoregression models and autoregression with moving average (MA) models could be used. For building this models we need to find the order of the model and the order for moving average.

The process of parameter selection for the models  $AR(p)$ ,  $MA(q)$  and  $ARMA(p,q)$  for the series is based on partial autocorrelation functions to determine the lag and search for the parameters  $p$  and  $q$ . The results of comparative analysis of estimating statistical characteristics for the selected models are presented in table 1.

**Table 1**

Comparison of the models which describe the process of using mobile data traffic in roaming

| Models      | Criteria |     |     |      |
|-------------|----------|-----|-----|------|
|             | $R^2$    | AIC | BSC | DW   |
| AR(2)       | 0.95     | 843 | 847 | 1.97 |
| MA (6)      | 0.95     | 699 | 709 | 1.76 |
| ARMA (2, 6) | 0.98     | 676 | 689 | 2.12 |

The best model was defined the model ARMA (2, 6). In Figure 2 are shown the results of simulation based on ARMA (2, 6) with the input data values for a given series. Next based on this model the forecast for the next 7 months ahead was built and provided to the mobile operator for estimation of further tariff strategy.

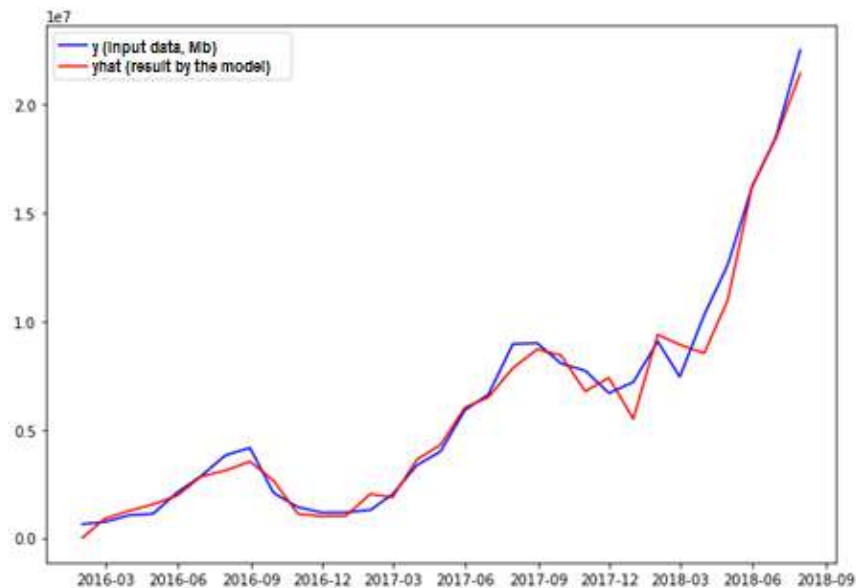


**Figure 2:** The results of the ARMA model (2, 6) in comparison with the original series

## 5.2. Classification models construction to determine the best package of services for roaming subscribers

To develop new tariffs and service packages, a sample of 120,000 records which contained subscriber's data who were staying abroad in the period from August 2017 to July 2018 was provided. The task was to classify the target variable – a package of services for mobile data transmission in roaming by classification models.

Let's preliminary define the main service packages for GPRS mobile internet in roaming. The target variable  $Y_{p\_mi}$  for predicting GPRS traffic in roaming can take the following values:

- $gr\_0$  – the subscriber did not use the service at all;
- $gr\_100$  – the subscriber used up to 100 MB;
- $gr\_500$  – the subscriber used from 100 to 500 MB;
- $gr\_over\_500$  – the subscriber used more than 500 MB.

It is known that subscribers when traveling abroad often choose not to use mobile services at all. For calls in roaming such group of subscribers is near 66%, for GPRS services – 42% (in 50,623 trips subscribers will not use any of the internet service packages). Figure 3 shows the distribution of classes for different mobile Internet services in roaming packages.

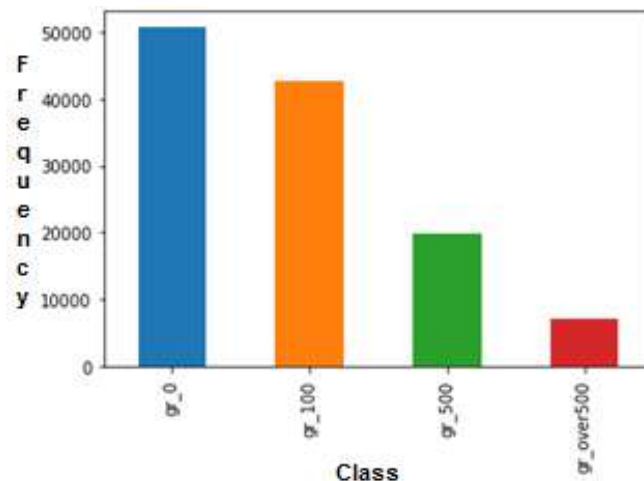


**Figure 3:** The frequency of the mobile data transmission in roaming usage classes

The data on the above mentioned 15 characteristics were formed by using the appropriate technical systems of the mobile operator, the numerical values are accurate, and the term data has a limited finite number of values and can be immediately taken as the categorical features.

### 5.2.1. Data processing and preparation

For continuous variables in the input dataset it was used discretization – splitting a continuous variable into some categories [25]. To convert the categorical features represented by string literals the direct coding or as it is also called one-hot-encoding (coding with one active state) was used [4]. The basic idea is to replace a category variable with a new variable by the formula for a linear binary classification. In such a way  $N$  new features are created for a categorical feature, where  $N$  is the number of categories. Each new feature is a binary characteristic of a certain category.

After preliminary data processing initial sample was divided into training (75%) and test (25%) data set. While the sample is unbalanced the class stratification method was used to ensure that sufficient values were obtained from each group.

On the next step the sample was balanced by increasing the number of records of smaller classes to the number of the majority class, increasing their weight for the training sample. A modified method of over-sampling the SMOTE (Synthetic Minority Over-sampling Technique) [19] was used, in which records are not simply duplicated but artificially generated based on examples of real representatives of the class with minor deviations.

### 5.2.2. The statistical characteristics for checking the classification quality used in our task

The most well-known statistical indicators for assessing the quality of classification are general accuracy, error matrix (Confusion Matrix), first and second kind of errors, index GINI [25]. Usually the error matrix is determined for binary classification, size  $2 \times 2$ . In general, the error matrix  $N \times N$ , where  $N$  is the number of classes, which shows the correct model predictions, as well as forecast errors. For our classification problem into four classes the matrix is presented in the form of the following table 2. Here the classification errors are divided into the following groups:

FN (False Negative) – first kind of errors, false negative value. For our task such error means that the forecast value implies a package of services smaller than actually needed.

FP (False Positive) – second kind of errors, false positive value. Here, this error means that the forecast value implies a package of services greater than actually needed [22].

**Table 2**  
Confusion Matrix for mobile internet in roaming

| Real values | Forecasted values |               |               |               |
|-------------|-------------------|---------------|---------------|---------------|
|             | $\hat{y} = 1$     | $\hat{y} = 2$ | $\hat{y} = 3$ | $\hat{y} = 4$ |
|             | $y = 1$           | TP            | FP            | FP            |
|             | $y = 2$           | FN            | TP            | FP            |
|             | $y = 3$           | FN            | FN            | TP            |
|             | $y = 4$           | FN            | FN            | TP            |

The peculiarity of the classification task of the package services is that the fact that the second kind errors are more acceptable for marketing decisions than errors of the first kind. Because the first kind of errors will mean potentially unearned income, and mistakes of the second kind – only a failed attempt to sell a package of services.

The values of the confusion matrix are used to calculate the main metrics for estimating the classification model.

The overall *accuracy* of the model is calculated as [26]:

$$accuracy = \frac{TP}{TP + FP + FN}. \quad (15)$$

To evaluate a model with unequal classes this metric will not fully characterize the model correctness. On example of the subscribers' behavior in roaming classification it means that if 70% of subscribers do not use roaming services, then even if all instances are classified as class 0, so the accuracy of the model equal to 0.7 will be obtained. Obviously, the purpose of the forecast does not coincide with this result and the model doesn't have significant practical value.

The quality assessment of the multiclass classification model is performed using the following metrics, which are calculated for each class separately.

*Precision* shows that some of the objects which were called positive by the classifier and in fact they were indeed positive.

$$precision = \frac{TP}{TP + FP}. \quad (16)$$

*Recall* shows which part of the positive class objects out of all the positive class objects were found by the algorithm.

$$recall = \frac{TP}{TP + FN}. \quad (17)$$

The recall demonstrates the ability of the algorithm to find this class in general, and precision – the ability to distinguish this class from other classes [26].

*F-measure (f1-score)* is the average harmonic between precision and recall:

$$f_{\beta} = (1 + \beta^2) \cdot \frac{precision \cdot recall}{(\beta^2 \cdot precision) + recall}, \quad (18)$$

where  $\beta$  determines the accuracy measure in this metric and for  $\beta = 1$  it is the harmonic mean (with the factor  $(1 + \beta^2)$  equal to 2, so that in the case when  $precision = 1$  and  $recall = 1$ ) the f1-score reaches its maximum at completeness and accuracy equal to one. The measure f1-score approaches zero if one of the indicators approaches zero [26].

The *Matthews correlation coefficient* (MCC) in contrast to the previously considered estimates takes into account all the values of the confusion matrix and is calculated by the following equation:

$$MCC = \frac{c \cdot s - \sum_{k=1}^K p_k t_k}{\sqrt{(s^2 - \sum_{k=1}^K p_k^2)(s^2 - \sum_{k=1}^K t_k^2)}}, \quad (19)$$

where  $t_k$  is the number of instances in class  $k$ ;  $p_k$  – how many times the model predicted the class  $k$ ;  $c$  – is the number of correctly classified instances;  $s$  – is the total number of instances.

The MCC can take the values from -1 to +1. A model rated to +1 is considered as an ideal. The model which received a score of -1 is considered very weak.

## 6. Simulation results

The simulation was performed on the basis of such methods as: logistic regression, neural networks, random forest, and gradient boosting. Let's consider more in detail the results of application of the methods described above to classify the proposals of the mobile data service in roaming. Visually the results of each model in the test sample are shown by using the confusion matrix in Figure 4.



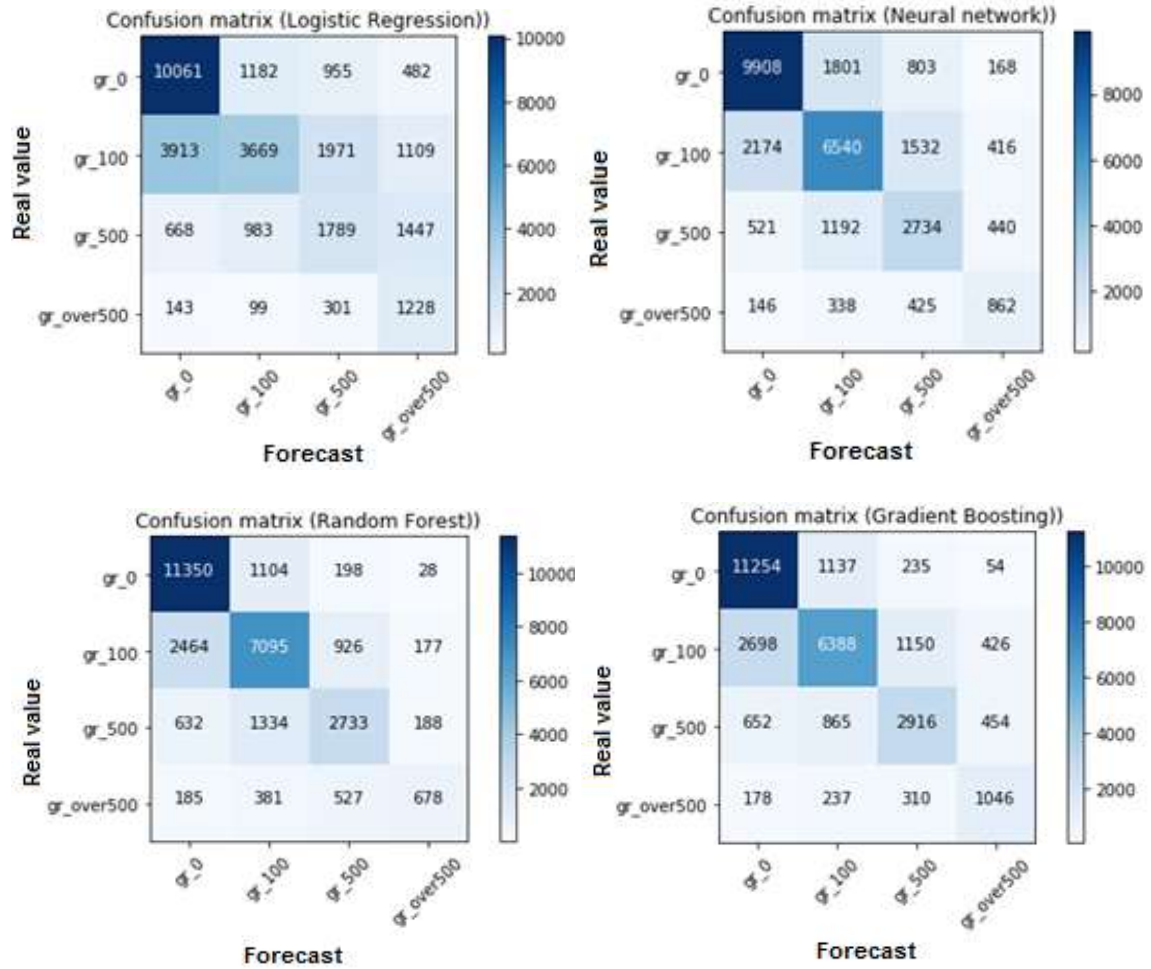


**Figure 4:** Confusion matrixes for the mentioned above models for mobile data services in roaming classification

GPRS service is gaining more popularity among the national mobile operator subscribers not only in Ukraine but also in roaming. In the case of offering such services one should expect a better response on this offers from subscribers even if they are from the erroneously predicted model classes (FP - errors). To select the best model we will use similar indicators of the classification models quality. The results for all models are gathered together in Table 3.

Final comparisons of the classifiers were made using the Matthews coefficient: logistic regression shows MCC=0.36; Neural Network MCC=0.51; Random forest – MCC=0.58; and Gradient Boosting – MCC = 0.58. The results of the Gradient Boosting and Random forest models are very similar in all respects. Only f1-score for Gradient Boosting has a few percent better performance, so this model can be considered as the best one.

## 7. Discussion

According to the simulation results received, several possible improvements in the formation of proposals for subscribers traveling abroad were proposed.

The first scenario takes into account the current state of the roaming communications market and aims to analyze customer preferences and offer appropriate service packages. Such proposals may provide some extra benefits for clients. For example, such as:

**Table 3**

Comparison of classification models for the use of GPRS in roaming

| Class          | Method              | precision | recall | f1-score |
|----------------|---------------------|-----------|--------|----------|
| gr_0           | Logistic regression | 0.68      | 0.79   | 0.73     |
|                | Neural Network      | 0.78      | 0.78   | 0.78     |
|                | Random forest       | 0.78      | 0.90   | 0.83     |
|                | Gradient Boosting   | 0.76      | 0.89   | 0.82     |
| gr_100         | Logistic regression | 0.62      | 0.34   | 0.44     |
|                | Neural Network      | 0.66      | 0.61   | 0.64     |
|                | Random forest       | 0.72      | 0.67   | 0.69     |
|                | Gradient Boosting   | 0.74      | 0.60   | 0.66     |
| gr_500         | Logistic regression | 0.36      | 0.37   | 0.36     |
|                | Neural Network      | 0.50      | 0.56   | 0.53     |
|                | Random forest       | 0.62      | 0.56   | 0.59     |
|                | Gradient Boosting   | 0.63      | 0.60   | 0.61     |
| gr_over<br>500 | Logistic regression | 0.29      | 0.69   | 0.41     |
|                | Neural Network      | 0.46      | 0.49   | 0.47     |
|                | Random forest       | 0.63      | 0.38   | 0.48     |
|                | Gradient Boosting   | 0.53      | 0.59   | 0.56     |

- for subscribers whose behavior is classified in the group gr\_100 to offer 150 megabytes in the second purchased package. This encourages the subscriber to use more than 100 megabytes and to order one more package;
- for subscribers who are classified by the model in the group gr\_500 (the subscriber used from 100 to 500 megabytes) to offer to purchase till promotional date a package of 500 megabytes with a 5% discount;
- for subscribers who have been classified by the model in the gr\_over500 group to offer a 1 GB service package instead.

This will encourage subscribers who use roaming services to use a little more services, and thus improve the experience and encourage the quantity of people who use the roaming services and thus to increase the total amount of mobile traffic in roaming.

The first scenario can be implemented using classification models. The best model based on the Gradient Boosting provides satisfactory accuracy and completeness of classification, which gives confidence that the target audience for the marketing campaign will be defined qualitatively.

The second scenario involves a significant marketing campaign aimed at reducing the percentage of subscribers who do not use mobile services when traveling abroad. The first assumption: subscribers cannot assess their own need to use services, estimate costs and therefore prefer not to use services at all, sometimes even turn off their phones.

A successful offer for a fixed package of services, for example, 100 megabytes with the specified cost and a guarantee that exceeding this limit is impossible and will not incur excessive costs can positively affect the decisions of subscribers. The main task is to determine among subscribers who do not use roaming services, their behavior characteristics in Ukraine which will give us assumption which package should be offered and to whom. For this purpose were developed and trained the models mainly on the subscribers' behavior who use roaming services.

For the second scenario, one can also use the proposals similar to the first scenario. The equality of the second scenario is that you need to understand that the cost of a marketing campaign can be much higher and the conversion much lower. The second scenario is designed for the long term campaign.

All developed models help to qualitatively form and determine the target audience for the marketing proposals. The decision to provide special offers or inform subscribers is made by marketing department depending on the available budget.

## 8. Conclusions

Mobile operators perform behavioral users' analysis in order to develop new tariff packages and retain existing subscribers among their active users. The Covid-19 crisis has shown that the modern world is adapting to new conditions at an extremely rapid pace and there is a need to support and develop telecommunications and Internet services. After the borders are opening between countries and intensification of the international travel which may take place in early 2022 the tourist and business flow will resume to the pre-crisis level. It is expected that during the year it will reach the indicators of 2018 year, so the simulation and forecasting of roaming traffics will allow mobile operators to develop appropriate tariff policies for roaming subscribers.

From the all four built classification models for predicting the subscribers behavior in roaming a model based on the gradient boosting was selected because of the highest completeness and accuracy to our input data. This model should be used for prudent marketing strategies. For bolder marketing campaigns such classification model which will contain more second kind errors should be used. In such case it is predicted that the subscriber should be offered a tariff but in fact he will not use the service. Such mistakes will allow to conduct a marketing campaign to activate roaming subscribers who do not use the services. In conditions of increasing numbers of subscribers that get used Internet the chances of success for such marketing campaign will increase day by day.

In further research it is planned to analyze and forecast the usage by clients other services in roaming (calls and SMS), as well as in cooperation with other departments of the mobile operator to give recommendations on the company's development and distribution of various services in national and international vectors.

## 9. References

- [1] European Commission Press Release. Brussels, 17 February 2014. URL: [http://europa.eu/rapid/press-release\\_IP-14-152\\_en.htm](http://europa.eu/rapid/press-release_IP-14-152_en.htm).
- [2] N. V. Kuznietsova, Information Technologies for Clients' Database Analysis and Behaviour Forecasting, in: CEUR Workshop Proceeding, 2017, pp. 56-62. URL: <http://ceur-ws.org/Vol-2067/>.
- [3] M. Havrylovych, N. Kuznietsova, Survival analysis methods for churn prevention in telecommunications industry, in: CEUR Workshop Proceeding, 2020, pp. 47-58. URL: <http://ceur-ws.org/Vol-2577/paper5.pdf>.
- [4] G. James, D. Witten, T. Hastie, R. Tibshinari, An Introduction to Statistical Learning with Applications in R, Springer-Verlag, New York (2013). doi: 10.1007/978-1-4614-7138-7.
- [5] N. S. Papageorgiou, V. D. Radulescu, D. D. Repovs, Nonlinear Analysis – Theory and Methods, Springer, Cham, Switzerland, 2019.
- [6] L. Breiman, Random forests, Machine Learning 45 (2001), 5–32. URL: <https://doi.org/10.1023/A:1010933404324>.
- [7] J. Friedman, Greedy Function Approximation: A Gradient Boosting Machine, The Annals of Statistics, vol.29, No.5 (2001), pp. 1189-1232. URL: <https://www.jstor.org/stable/2699986?origin=JSTOR-pdf&seq=1>.
- [8] T. M. Rassias, Applications of Nonlinear Analysis, Springer, Cham Switzerland, 2018.
- [9] J. Beran, Mathematical Foundations of Time Series Analysis, Springer, Cham Switzerland, 2017.
- [10] V. K. Shitikov, S. E. Mastitsky, Classification, regression and other data mining algorithms using R(in Russian), 2017. URL: <https://ranalytics.github.io/data-mining/index.html>.
- [11] N. V. Kuznietsova, M. Seebauer, S. Zabelin, Some methods for estimating financial risks in banking, IEEE 1st Conf. on System Analysis and Intelligent Computing, SAIC 2018 (2018), pp. 271–274. doi: <https://ieeexplore.ieee.org/document/8516873>.
- [12] S. Osovsky, Neural networks for information processing (translation in Russian by I.D. Rudinsky), Finance and statistics, Moscow, 2002.

- [13] O.V. Gorokhovatsky, O. O. Peredriy, Multilayer perceptron as the primary instrument for image clustering (in Ukrainian) / Registration, storage and data processing 18 (2016) 33–43.
- [14] C. Cortes, V. Vapnik, Support-vector networks, Machine learning (1995), vol. 20, no. 3, 273–297.
- [15] S.P. Chistyakov, Random forests: an overview (in Russian), Works of Karelian scientist center of RAS, 2013, Issue. 1, pp. 117–136. URL: [http://resources.krc.karelia.ru/transactions/doc/trudy2013/trudy\\_2013\\_1\\_117-136.pdf](http://resources.krc.karelia.ru/transactions/doc/trudy2013/trudy_2013_1_117-136.pdf).
- [16] P. Tseng, On accelerated proximal gradient methods for convex-concave optimization, 2008. URL: <http://www.math.washington.edu/~tseng/papers/apgm.pdf>.
- [17] R. S. Tsay, Analysis of financial time series, John Wiley & Sons, Inc., New York, NY, 2010.
- [18] P. Bidyuk, A. Gozhyj, Y. Matsuki, N. Kuznetsova, I. Kalinina, Modeling and Forecasting Economic and Financial Processes Using Combined Adaptive Models, in: Babichev S., Lytvynenko V., Wójcik W., Vyshemyrskaya S. (Eds.), Lecture Notes in Computational Intelligence and Decision Making, ISDMCI 2020, Advances in Intelligent Systems and Computing, vol 1246, Springer, Cham, 2021. URL: [https://doi.org/10.1007/978-3-030-54215-3\\_25](https://doi.org/10.1007/978-3-030-54215-3_25).
- [19] V. N. Nikulin, I. S. Kanishchev, I.V. Bagaev, Methods of balancing and normalization of data to improve the quality of classification, Computer tools in education 3 (2016) 16–24.
- [20] V. K. Shitikov, G. S. Rosenberg, T.D. Zinchenko, Quantitative hydroecology: methods of systemic identification, IEVB RAS, Togliatti, 2003.
- [21] J.H. Friedman, On Bias, Variance, 0/1—Loss, and the Curse-of-Dimensionality, Data Mining and Knowledge Discovery 1, 55–77 (1997). <https://doi.org/10.1023/A:1009778005914>.
- [22] F. Herrera, F. Charte, A.J. Rivera, M.J. del Jesus, Multilabel Classification Problem Analysis, Metrics and Techniques, Springer International Publishing, Switzerland, 2016. doi: 10.1007/978-3-319-41111-8.
- [23] Kuznetsova N. V. Analytical Technologies for Clients' Preferences Analyzing with Incomplete Data Recovering, in: CEUR Workshop Proceeding, 2018, pp.118-128. URL: <http://ceur-ws.org/Vol-2318/>.
- [24] D. Kwiatkowski, P. Phillips, P. Schmidt, Y. Shin, Testing the null hypothesis of stationarity against the alternative of a unit root, Journal of Economics, n. 54 (1992) 159–178.
- [25] N. V. Kuznetsova, P. I. Bidyuk, Theory and practice of financial risk analysis: systemic approach, Lira-K, Kyiv, 2020.
- [26] M. Hossin, M.N, Sulaiman, A Review on Evaluation Metrics for Data Classification Evaluations, International Journal of Data Mining & Knowledge Management Process 5 (2015) 1-11. doi: 10.5121/ijdkp.2015.5201.

# Method for Visual Video Defects Detection using Machine Learning

Dmytro Fedasyuk<sup>a</sup>, Roman Lukomskyi<sup>a</sup>, Tetyana Marusenkova<sup>a</sup>

<sup>a</sup>Lviv Polytechnic National University, St. Bandery str, 28 a, Lviv, 79013, Ukraine

## Abstract

The main problem to be solved by this research is the imperfection of the video testing process. Nowadays this process involves mainly manual testing, which is inefficient due to the high probability of human errors, significant time, and material costs. In order to improve this process, we have created a convolutional neural network that can detect defects in video frames with high probability. We have also built a prototype of a software system that can automatically detect defects in video using the created and trained convolutional neural network.

## Keywords

Defects detection, image recognition, machine learning, deep learning, convolutional neural networks, automation, video testing.

## 1. Introduction

The evolution of digital communication systems has led to the active development of multimedia systems and applications such as IPTV (Internet protocol television), mobile multimedia, social networks, virtual reality games, video conferencing, and educational multimedia presentations. These multimedia applications have now become an integral part of everyday life, and they are expected to grow rapidly in the future. Video is being widely used in the above-mentioned application areas, which set strict requirements to its quality, i.e., video content should be free of defects.

A visual defect or a perceptual artifact in a video is a noticeable frame distortion. Such distortion can appear as a result of errors in compression, transmission, or encoding. An example of a visual defect is shown in Figure 1.



Figure 1: An example of a perceptual artifact

CMIS-2021: The Fourth International Workshop on Computer Modeling and Intelligent Systems, April, 27, 2021, Zaporizhzhia, Ukraine  
EMAIL: dmytro.v.fedasyuk@lpnu.ua (D. Fedasyuk); roman.lukomskyi.mnpz.2019@lpnu.ua (R. Lukomskyi); tetiana.a.marusenkova@lpnu.ua (T. Marusenkova)  
ORCID: 0000-0003-3552-7454 (D. Fedasyuk); 0000-0002-0345-8290 (R. Lukomskyi); 0000-0003-4508-5725 (T. Marusenkova)



© 2021 Copyright for this paper by its authors.  
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).  
CEUR Workshop Proceedings (CEUR-WS.org)

Even minor perceptual artifacts can have a significant impact on the satisfaction of users from watching videos and using multimedia services. With that in mind, multimedia providers are paying more and more attention to ensure the high quality of their video content. Nowadays, the major methods for video quality assessment involve manual testing which is inefficient due to the human factor.

The purpose of this research is to improve the process of video testing. The use of the developed method should reduce the focus on manual testing of video content quality because manual testing has many disadvantages. For example, such disadvantages include the relatively high cost of this process due to the significant cost of viewing equipment, premises, etc. Another important drawback is the high probability that a certain defect will be missed by testers because in order to save time, they usually review only small test fragments, which may not have a particular visual defect in them.

It is expected that the use of the method will increase the percentage of detected defects during testing and, accordingly, will help to improve the overall quality of video and service. An important advantage is that the input of the method is only the video that is being tested. This method, unlike some existing methods of objective quality determination, will allow usage of the software system in cases where the original video file is not available. This is an important factor because in real conditions there is often no access to the original video.

Given the rapid development of the video services market, we can conclude that the development of a method for detecting visual defects is quite promising. The presence of even minor and short defects can significantly affect the satisfaction of users, so companies that provide services related to video distribution and viewing will be interested in improving their video testing process.

The object of research is the process of testing video with the detection of visual defects. The subject of research is methods of detecting defects in the video.

## **2. Related work**

Visual quality assessment refers to gauging a probability of a visual perceptual artifact. Since humans are the ultimate consumers of video content, the subjective methods of testing video quality require manual testing, i.e. involving people who view fragments of video and give a probably biased assessment of its quality on a particular rating scale [1, 2]. Because video content is very large, testers usually are not able to view it entirely.

There are various methods of subjective video quality assessment which determine the rules for selecting an optimal test video fragment duration, number of people that will conduct the assessment as well as metrics that should be used to get the final result. For example, there is a recommendation that suggests that an optimal test fragment length is 10 seconds [3]. However, studies have been conducted to optimize this time indicator [4]. These studies claim that in some cases, reducing the duration of test fragments may not have a significant impact on the resulting testing, but at the same time significantly reduce the time spent on testing. This is evidenced by previous studies that show that, firstly, testers become less attentive if the fragment length exceeds 10 seconds [5], secondly, shorter test sequences contribute to more consistent results [6] and, thirdly, the average shooting time in most modern films is less than 10 seconds [7]. However, when optimizing test suites, there are risks associated with the possibility that defects will be contained in fragments that have been removed from the test suite. Besides, there are various laboratory factors, including, for example, screen size, lighting, viewer-to-screen distance, and so on [8].

Subjective video testing provides reliable results in some cases since it is conducted by humans. On the other hand, this approach has a number of drawbacks for the same reason. To name a few, they include but not limited to:

- the need for significant human resources;
- significant costs for the viewing equipment;
- it takes a lot of time;
- there is a possibility that a certain fragment that actually contains a defect will not be selected to the test fragments set or a short defect will go unnoticed by the tester.

Moreover, this approach cannot be used in continuous quality assessment systems. That is, if a new file is added to the video file database, it requires a separate testing session, which often cannot

take place immediately after receiving the file, but must be scheduled for the future due to the human factor. This also contributes to the risk that a video file will be skipped.

Another significant disadvantage is that very often the same video file is encoded with different quality parameters for adaptive transmission, which greatly increases the number of test videos and, accordingly, the resources spent on testing. Currently, there are methods of objective video quality testing which are being developed in order to solve these problems. They are often referred to as Objective Quality of Experience (QoE) methods.

The structural information of the video is especially important for the comprehension of a person's visual perception. There is an approach to image quality assessment based on structural information, which determines the degradation of structural similarity derived from the statistical properties of local information between the reference and the distorted images [9]. This approach has also been proposed for video. In addition, a criterion of information reliability was proposed to assess the quality of images [10], which is based on statistical characteristics of the structural information of natural images.

Image motion information is also important for optimal, realistic quality assessment. Various researchers have proposed methods for determining quality based on motion information [11, 12]. In particular, in the Motion-based Video Integrity Evaluation (MOVIE) [13], proposed for natural video, distorted and reference content is decomposed using Gabor spatio-temporal filters, therefore the quality index consists of two components – spatial and temporal.

The spatio-temporal quality method provides a significant improvement in performance in terms of statistical correlation between the results obtained with this objective method and the subjective data obtained from humans. The authors note that this method can be used mainly for videos with natural content. The authors also acknowledge that the method is not computationally efficient. This is due to the fact that, since this method involves the decomposition of a video using a large number of Gabor filters, which, in turn, require a large number of frames, the use of the method requires the availability of significant computing resources. Taking this into account, we can conclude that the possibilities for using this method are limited.

Despite some progress in the objective detection of defects in the video, there are a number of issues that need to be addressed [14]. In particular, it is difficult to use the existing objective models in general cases, as each has its own characteristics, which are associated with certain types of content, context, or defects. Because of this, a particular model may do well with some types of video, but have very poor results, in cases of use for another video because it was not configured for it, or this video does not have the statistical characteristics on which the model is built.

Given the different nature of multimedia data, the creation of a generalized model, the logic of which will not depend on the above factors will significantly improve the state of modern video quality assessment.

In addition, traditional methods of quality assessment are often based on explicit modeling of human perception. As a result, systems based on these methods are prone to so-called overfitting and therefore have questionable results on real data. Instead, machine learning-based methods could mimic human perception of quality, rather than developing a precise model of the human visual perception system. Moreover, such methods will not need the original media file in order to assess quality.

### **3. Using the convolutional neural networks**

#### **3.1. The neural network architecture**

In order to check if visual defects are present in the video, we can divide it into frames and analyze them separately. The use of such an approach gives us an ability to get detailed results that not only show if the defect is present or not but also, in case if it is present, it can indicate the time when the defect occurred. This might be especially handy if there is a need for a human to double-check if there is a real defect before marking the video as defective.

We can use artificial neural networks (ANN) in order to detect visual defects in specific frames. Artificial neural networks are software implementation of the logic of brain structures. We can train them by specifying input information and examples of relevant output information [15].



As a matter of fact, the biggest limitation of artificial neural networks is that they inevitably require a big amount of computational resources when working with a large amount of input data. This might not be an issue for small black and white images, but the image data in real-world videos always contains a lot of input information (taking into account its size and color) which renders traditional neural networks useless when applied in image recognition tasks. In order to resolve this problem, there needs to be an optimization of the input data that is passed to the fully connected layers. We need to separate image features that are truly important in the given recognition task.

Convolutional Neural Networks (CNN) are a class of artificial neural networks for deep learning that is often used in visual image analysis [16]. The architectural structure of a conventional neural network is fairly simple – each connection has its weight which is used during the process of propagation. The convolutional neural networks try to optimize their input before passing it to a fully connected layer. When it receives large input data (such as a colored image), it uses a number of convolution kernel matrices to retrieve important information from the given data. The important data in the case of a graphical input can be relative locations of certain shapes such as circles, arcs, and lines, distribution of color across the image, etc. When a fully connected layer receives information about the presence or absence of these features as well as their relative location information, it can effectively conduct the image recognition process since the volume of this input is tremendously smaller compared to the full input. A basic CNN architecture is shown in Figure 2. Usually, a CNN architecture would contain several convolutional and pooling layers.

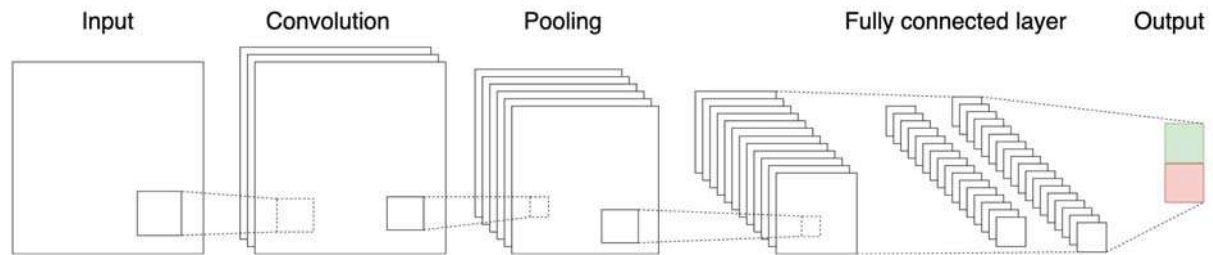


**Figure 2:** Basic CNN architecture

Naturally, in a convolutional neural network, a set of weights encodes important visual features of the image (color distribution, angles, lines as well as their relative location information). The network's convolution kernels are formed during the training process [17]. They cannot be set in advance, since each image recognition task has its own important features (recognition of certain objects would require a different set of graphical features than recognition of defects).

The pooling operation reduces the size of the formed feature maps. In this network architecture, it is considered that information about the presence of the specific feature is more important than accurate knowledge of its coordinates, so the maximum of several neighboring neurons of the feature map is selected and taken as one neuron of a compact feature map of smaller size [16]. Because of this operation, in addition to accelerating further computations, the neural network becomes more adaptable to the scale of the input image.

A convolutional neural network consists of input and output layers, as well as several hidden layers. Hidden CNN layers typically consist of convolutional, pooling, fully connected, and normalization layers. Convolutional layers apply a convolution operation to the input data, passing the result to the next layer. The convolution operation simulates the response of a single neuron to a specific visual stimulus. Each convolutional neuron processes data for its receptive field. The convolution operation solves the problem of the increasing number of connections, as it reduces the number of parameters, allowing the network to be deeper [16]. The convolutional layer is the main building block of CNN. The parameters of this layer consist of a set of filters for learning, which have a small receptive field. During the forward propagation, each filter performs a convolution on the width and height of the input layer, calculating the scalar product of the filter and input data, and forming a 2-dimensional activation map of this filter. As a result, the network learns which filters are activated when it detects a particular type of feature in a particular spatial position in the input data.

When processing multidimensional inputs (such as images), it is impractical to connect all neurons with all neurons of the previous layer, because such a network architecture does not take into



account the spatial structure of the data. Convolutional networks use spatial-local correlation by providing a scheme of local connection between neurons of adjacent layers: each neuron is connected to only a small area of the input layer. The receptive field of a neuron is a hyperparameter that is the degree to which neurons connect. Connections are local in space (along width and height) but always propagate along the entire depth of the input layer.

The spatial size of the output volume is calculated using a formula (1).

$$N = (W - F + 2P) / 2 + 1, \quad (1)$$

where  $N$  – the output volume size,  
 $W$  – the input volume size,  
 $F$  – the kernel size,  
 $P$  – the amount of zero padding,  
 $S$  – the stride.

In order to ensure that the dimensions of the input and output matrices are equal (provided that the step of the filter area is equal to one), the size of the zero padding is determined by the formula (2).

$$P = (F - 1) / 2, \quad (2)$$

where  $P$  – the amount of zero padding,  
 $F$  – the kernel size.

Convolutional and pooling layers are followed by fully connected layers (Figure 3). Neurons in the fully connected layer are connected to all neurons in the previous layer, just as they do in conventional artificial neural networks [18].

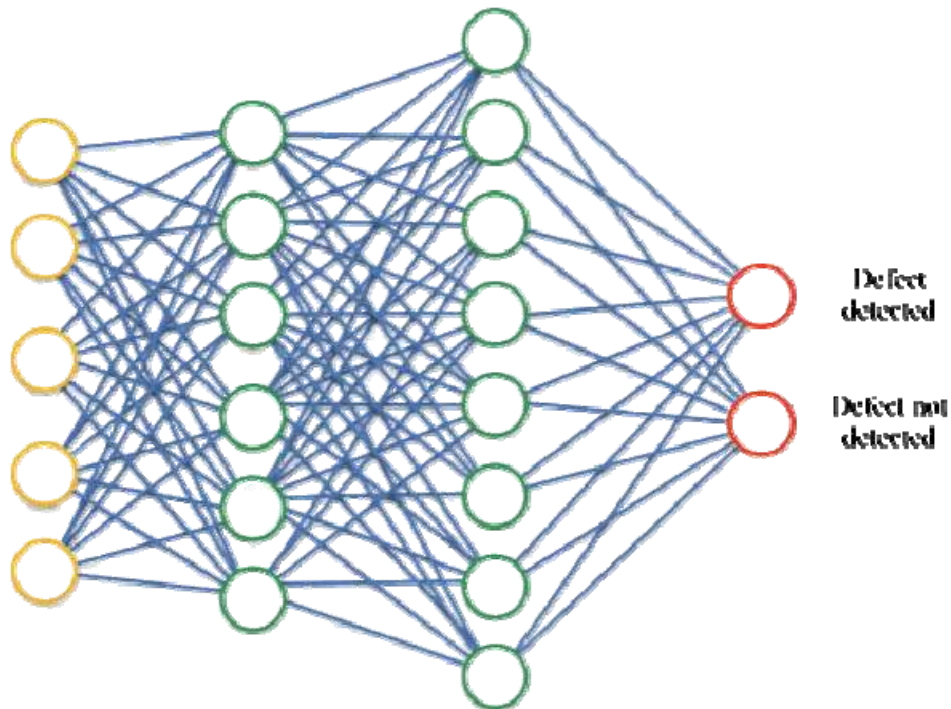


**Figure 3:** Fully connected layers

The loss layer determines how the training process penalizes the deviation between the predicted and expected results and is usually the final layer. It can use different loss functions for different tasks. For example, normalized exponential (softmax).

Because the fully connected layer has the most parameters, it is prone to overfitting. One of the most common methods of reducing overfitting is a dropout. At each stage of training, individual nodes with a certain probability are either "excluded" from the network or remain in it, thus, as a result, the network is reduced. Inbound and outbound links of excluded nodes are also deleted. In the

next step, only a reduced version of the neural network is trained on the data. The removed nodes are then re-inserted into the network with their previous weights.

Taking into account the above facts it was decided to choose the neural network architecture which is shown in Figure 4.

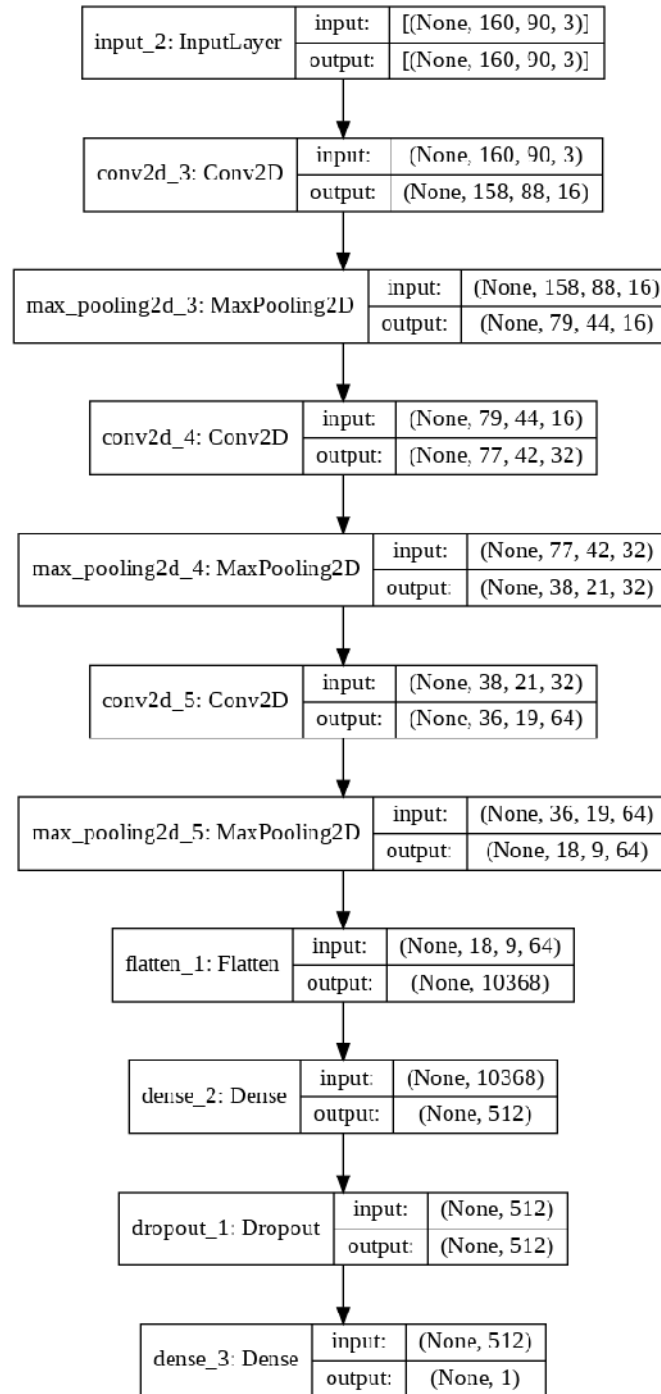


**Figure 4:** Architecture of the created convolutional neural network

### 3.2. Dataset construction

In order to construct the dataset, we selected several creative commons licensed videos from the Internet. To provide the best learning results during the neural network training process, the dataset needs to have a wide variety of training data. In order to achieve this, we made sure that the selected video content includes a diverse set of content genres. In addition to that, the scenes in the videos include different types of camera motion such as static, moving, and zoom. Additionally, we

have ensured that the videos are of pristine quality by carefully inspecting each of them. Making sure that the training data is labeled correctly will help to avoid confusing the convolutional neural network during the training process.

In order to have distorted versions of the same videos, we have simulated errors in the original videos. After that, we have once again ensured that the distorted videos indeed have visual defects in each frame. Later, we broke down the videos into separate frames so that they could be given to the CNN as its input.

The final dataset consists of 129,092 video frames. Each frame from the dataset is labeled as either “damaged” or “original” as shown in Figure 5. For each damaged frame there is an original frame in the dataset and vice versa. This way the convolutional neural network can learn most effectively since it has similar reference images for both classes.

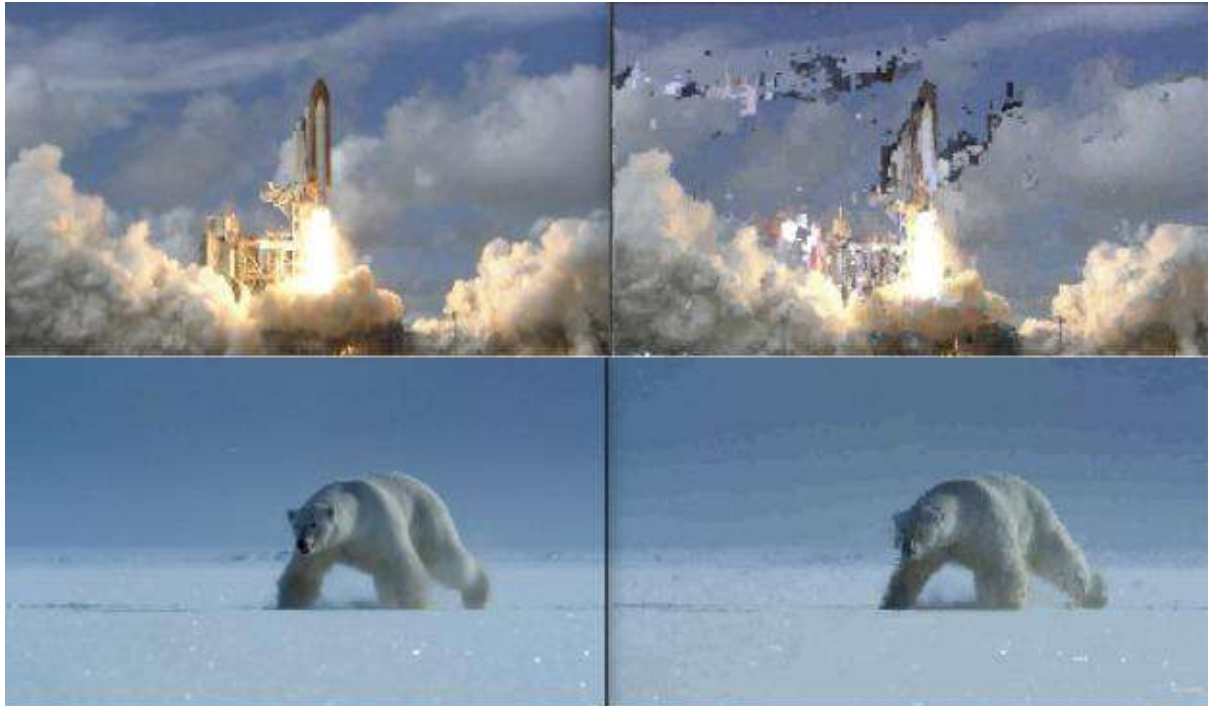


**Figure 5:** Original (on the left) and distorted (on the right) frames from the dataset

### 3.3. Training

Since convolutional neural networks require a lot of training data to perform well, it is common to use the existing entries in the dataset in order to generate new training examples. This technique is called data augmentation [18, 19]. Advanced data augmentation might include the usage of generative adversarial network (GAN) [20] or balancing GAN (BAGAN) [21]. However, taking into account the nature of our data, especially the importance to keep all of our training examples as close to natural samples as possible, these techniques are not needed in our case. Traditional transformations consist of using a combination of transformations on the original data. Since we are working with images, such transformations might include:

- horizontal/vertical image flip;
- rotation;
- cropping;
- translation;
- shift;
- zoom;
- shading.

Experimentally we have discovered that in the case of visual defects detection, image flip performs the best out of all of the above approaches. This might be explained by the fact that other of the above techniques might introduce distortions in the dataset images marked as “original” (for

example, excessive pixelation after zooming in or out). This would have a significant negative effect on the training process and the final CNN accuracy. Taking this into account, we have decided to apply the data augmentation using the image flip technique.

In order to conduct the training, we used Tensorflow and Keras with Python. As a result of training on the previously created dataset, the neural network managed to achieve an accuracy of 98.5%. The process of training with the respective changes in both training and validation loss and accuracy is shown in Table 1.

**Table 1**

The training process

| Epoch | Training loss | Training accuracy | Validation loss | Validation accuracy |
|-------|---------------|-------------------|-----------------|---------------------|
| 1     | 0.2121        | 0.9146            | 0.3368          | 0.9148              |
| 2     | 0.0973        | 0.9719            | 0.0710          | 0.9785              |
| 3     | 0.0845        | 0.9786            | 0.0663          | 0.9835              |
| 4     | 0.0826        | 0.9806            | 0.0443          | 0.9850              |

As can be seen from Table 1, the created convolutional neural network can detect visual artifacts in frames with decent probability.

We can now use the created convolutional neural network to detect visual defects in the video. Using a trained model, we have built a prototype of a software system that checks for visual defects in the video by cyclically performing the following steps:

- retrieving a video frame – in this step of the algorithm, the prototype of the software system forms a video frame in the final image form – the same form in which it would be shown to the viewer;
- analyzing the video frame – the convolutional neural network analyzes the video frame which was obtained in the previous step. In this step the CNN will output the probability of visual distortion presence;
- saving the result of the analysis – the software system stores the result of the analysis in a particular video frame in order to be able to later show it to the user on the graphical interface;

The described algorithm is shown in Figure 6.

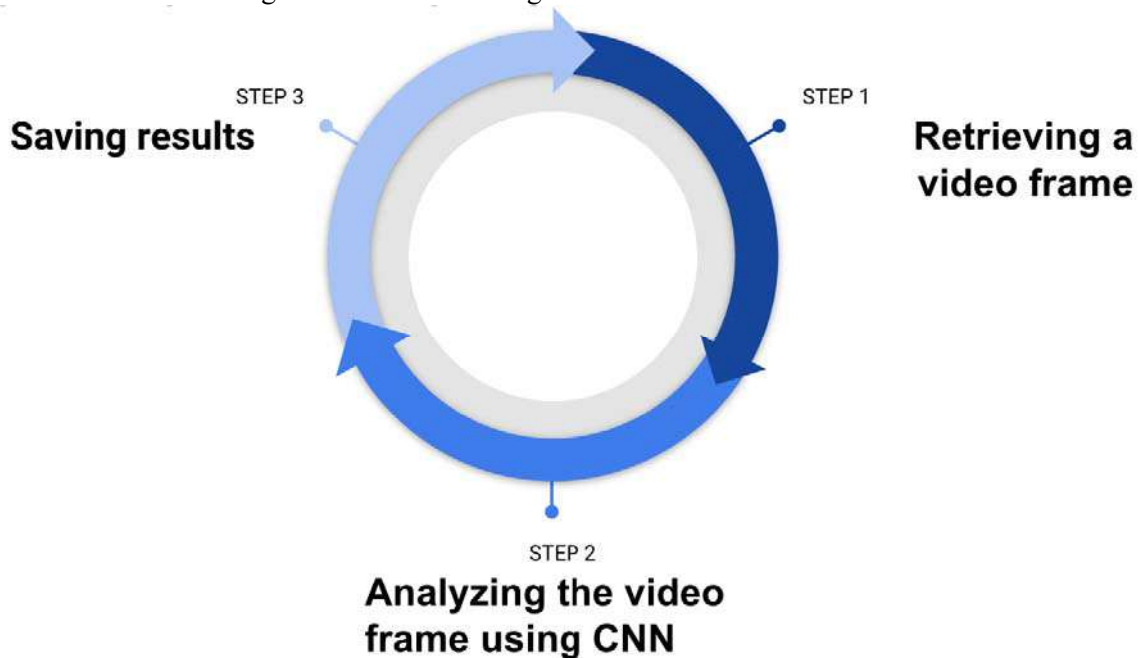


**Figure 6:** Algorithm of the software system operation

After performing the steps described above, the prototype of a software system builds a diagram of the defects as shown in Figure 7. Normally, at this point, a person who is responsible for the testing process would need to take a look at the diagram and investigate the defects that have been found.

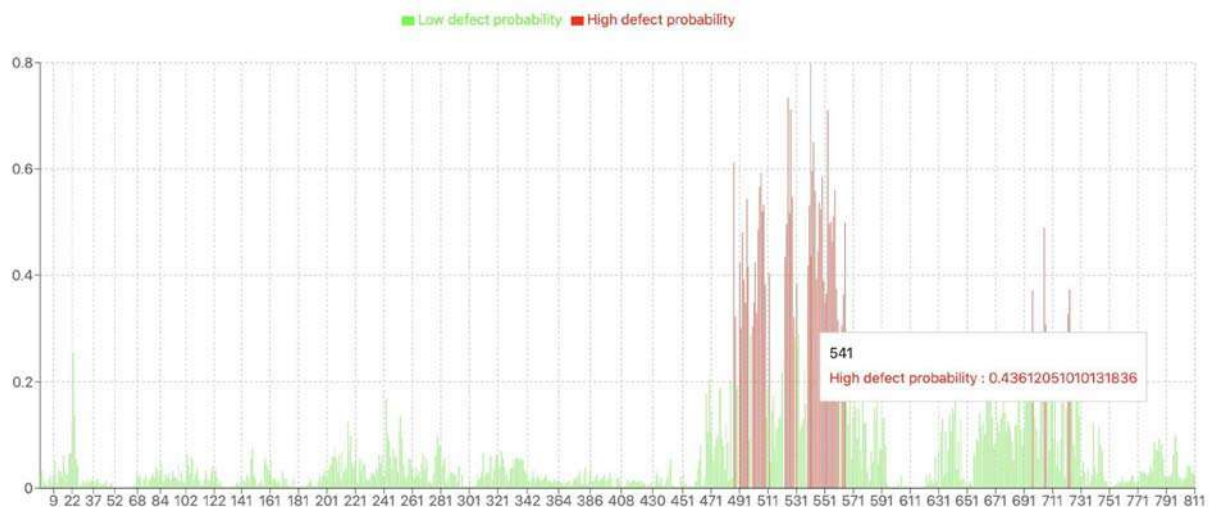


**Figure 7:** The defects probability diagram

#### 4. Conclusion

In this paper, we analyzed the problem of video testing. The video content in modern video content delivery systems undergoes multiple stages that can introduce visual distortions. Such distortions can have tremendous effects on the end-user experience.

Manual testing takes an unreasonable amount of time and leads to a high probability of human error. In spite of different methods and ideas for improving defect detection, there are a number of issues such as the need for significant human resources, significant costs for the viewing equipment, limited usage for continuous quality assessment systems. To add, traditional methods of quality assessment are often based on explicit modeling of human perception. Therefore, systems based on such an approach are prone to overfitting and have questionable results on real data. Machine learning-based methods could solve this issue. Moreover, such methods will not need the original media file in order to assess quality.

In order to solve this issue, we proposed using a convolutional neural network. Machine learning-based methods can mimic human perception of quality and it can increase the percentage of detected defects during testing. In order to conduct the training of the CNN, we have created a dataset from pristine and distorted videos which consists of 129,092 frames. In order to provide even more training data to the CNN, we used the data augmentation technique. As a result of training, the convolutional neural network achieved an accuracy of 98.5%. Taking into account the high accuracy of the trained model we can see that it is possible to use it in order to detect visual defects in video frames.

The created model was used to build a prototype of a software system that can detect defects in video with quite a high accuracy. The use of the developed software system can reduce the human factor in the process of video testing, as well as significantly speed up and reduce the cost of this process. Saving the result of the analysis and diagram of the defects helps the user to easily analyze the detected defects.

#### 5. References

- [1] A. K. Moorthy, K. Seshadrinathan, A. C. Bovik, Image and Video Quality Assessment: Perception, Psychophysical Models, and Algorithms, *Perceptual Digital Imaging: Methods and Applications* (2017) 55-81.
- [2] M. H. Pinson, S. Wolf, Comparing subjective video quality testing methodologies, *Communications and Image Processing* (2003) 573-582. doi:10.1117/12.509908.

- [3] ITU-R, Methodology for the Subjective Assessment of the Quality of Television Pictures, 2002. URL: <http://www.gpds.ene.unb.br/databases/2012-UNB-Varium-Exp/Exp3-Delft/00-report-alexandre/Papers---Judith/Subjective%20Studies/ITU-Recommendation---BT500-11.pdf>
- [4] F. M. Moss, K. Wang, F. Zhang, R. Baddeley, D. R. Bull, Moss, Felix Mercer, Ke Wang, Fan Zhang, Roland Baddeley, and David R. Bull. "On the optimal presentation duration for subjective video quality assessment, IEEE Transactions on Circuits and Systems for Video Technology (2015) 1977-1987.
- [5] P. Fröhlich, S. Egger, R. Schatz, M. Mühlegger, K. Masuch, and D. Gardlo, QoE in 10 seconds: Are short video clip lengths sufficient for quality of experience assessment?, 2012 fourth international workshop on quality of multimedia experience. IEEE (2012) 242-247.
- [6] B. W. Tatler, R. J. Baddeley, and I. D. Gilchrist, Visual correlates of fixation selection: Effects of scale and time, Vision research 45, no. 5 (2005) 643-659.
- [7] J. E. Cutting, K. L. Brunick, J. E. DeLong, C. Iricinski, A. Candan, Quicker, faster, darker: Changes in Hollywood film over 75 years, i-Perception 2 (2011) 596-576. doi:10.1068/i0441aap.
- [8] Q. Huynh-Thu, M. Ghanbari, D. Hands, M. Brotherton, Subjective video quality evaluation for multimedia applications, Human Vision and Electronic Imaging XI, vol. 6057, p. 60571D. International Society for Optics and Photonics (2006). doi: 10.1117/12.641703.
- [9] Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE transactions on image processing 13(4) (2004) 600-612. doi:10.1109/TIP.2003.819861.
- [10] H. R. Sheikh, A. C. Bovik, G. De Veciana, An information fidelity criterion for image quality assessment using natural scene statistics, IEEE Transactions on image processing 14, no. 12 (2005) 2117-2128. doi:10.1109/TIP.2005.859389.
- [11] K. Seshadrinathan, A. C. Bovik, Motion tuned spatio-temporal quality assessment of natural videos, IEEE transactions on image processing 19.2 (2009) 335-350. doi:10.1109/TIP.2009.2034992.
- [12] R. Soundararajan, A. C. Bovik, Video quality assessment by reduced reference spatio-temporal entropic differencing, IEEE Transactions on Circuits and Systems for Video Technology (2013) 684-694. doi:10.1109/TCSVT.2012.2214933.
- [13] K. Seshadrinathan, A. C. Bovik, Motion-based perceptual quality assessment of video, Human Vision and Electronic Imaging XIV (2009). doi:10.1117/12.811817.
- [14] Z. Akhtar, T. H. Falk, Audio-Visual Multimedia Quality Assessment: A Comprehensive Survey, IEEE Access (2017). doi:10.1109/ACCESS.2017.2750918.
- [15] C. Aggarwal, Neural Networks and Deep Learning: A Textbook, Springer, New York, NY, 2018.
- [16] K. O'Shea, R. Nash, An Introduction to Convolutional Neural Networks, arXiv preprint arXiv:1511.08458 (2015).
- [17] F. Millstein, Convolutional Neural Networks In Python: Beginner's Guide To Convolutional Neural Networks In Python, Scotts Valley New York, NY, 2018.
- [18] L. Perez, J. Wang, The effectiveness of data augmentation in image classification using deep learning, arXiv preprint arXiv:1712.04621 (2017).
- [19] C. Shorten, T. M. Khoshgoftaar, A survey on image data augmentation for deep learning, Journal of Big Data 6.1 (2019) 1-48. doi:10.1186/s40537-019-0197-0.
- [20] I. Goodfellow, Nips 2016 tutorial: Generative adversarial networks, arXiv preprint arXiv:1701.00160 (2016).
- [21] G. Mariani, F. Scheidegger, R. Istrate, C. Bekas, C. Malossi, Bagan: Data augmentation with balancing gan, arXiv preprint arXiv:1803.09655 (2018).



# Control and Diagnostics of TV3-117 Aircraft Engine Technical State in Flight Modes Using the Matrix Method for Calculating Dynamic Recurrent Neural Networks

Serhii Vladov<sup>a</sup>, Yurii Shmelov<sup>a</sup> and Ruslan Yakovliev<sup>a</sup>

<sup>a</sup> *Kremenchuk Flight College of Kharkiv National University of Internal Affairs, vul. Peremohy, 17/6, Kremenchuk, Ukraine, 39605*

## Abstract

In this paper, one of the methods of using dynamic recurrent neural networks for solving applied problems of control and diagnosing of aircraft engines technical state, including TV3-117, is considered. A method for coding neural networks into signal graphs is proposed, and it is shown that their adjacency matrices can be used as associative memory in step matrix algorithms for solving dynamic recurrent neural networks. It is shown that in fully connected recurrent neural networks, any neuron can be input or output, and one neuron can simultaneously be input and output. Examples of teaching by the evolutionary optimization algorithm for multiextremal problems of recurrent dynamic neural networks intended for control and diagnosing of TV3-117 aircraft engine technical state are given. The functions of activation blocks of neurons in dynamic recurrent neural networks in this work are used difference expressions of simulation models of linear dynamic links. It is shown that for identification in the time domain of transient processes in dynamic systems of the third order, satisfactory accuracy is achieved at the output of any neuron of a recurrent dynamic neural network with four neurons, while it was found that useful information about the dynamic properties of the dynamic system under study can be simultaneously obtained from the output of any neuron network.

## Keywords 1

Aircraft engine, neural network, signal graph adjacency matrices, synaptic weights, associative memory.

## 1. Introduction

Flight safety of aircraft, including helicopters, is one of the key problems of aviation, largely depends on the reliability of the operation of their power plants. In turn, the reliable operation of power plants and, in particular, a gas turbine engine (GTE) (for example, TV3-117, which is part of the power plant of the Mi-8MTV helicopter and its modifications) is ensured by a whole range of measures, among which an important place is given to its diagnostics technical state. At present, the technical diagnostics of a GTE, in the broad sense of this concept, is carried out only on the ground by an engineering and technical staff, who have at their disposal the appropriate tools and methodological apparatus. But most of the failures that affect the safety of a helicopter operation arise and manifest themselves in flight, which requires an effective on-board diagnostic system that determines the technical condition of the gas turbine engine in real time directly on board the helicopter. The difficulty of carrying out diagnostics of the power plant in automatic mode in flight is associated with the high complexity of aircraft GTE, due to the multiparametry, multi-connectivity, nonlinearity of the processes occurring in them, the multi-mode application, which requires significant machine and time resources [1].

CMIS-2021: The Fourth International Workshop on Computer Modeling and Intelligent Systems, April 27, 2021, Zaporizhzhia, Ukraine  
EMAIL: ser26101968@gmail.com (S. Vladov); nviddil.klk@gmail.com (Yu. Shmelov); klk.univd@gmail.com (R. Yakovliev)  
ORCID: 0000-0001-8009-5254 (S. Vladov); 0000-0002-3942-2003 (Yu. Shmelov); 0000-0002-3788-2583 (R. Yakovliev)



© 2021 Copyright for this paper by its authors.  
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).  
CEUR Workshop Proceedings (CEUR-WS.org)

Today engines of the 4<sup>th</sup> generation have on-board parameter control systems, which solve a wide range of tasks related to the assessment of the residual life of aircraft engines and control of their parameters. In the process of performing the task of engine life development, the problem of forecasting (short-term) the current state of the engine and trend analysis of its parameters is solved. In modern on-board system for control and diagnosing the following tasks of control of a large number of parameters of the engine [1] are solved: temperature and pressure of gases behind and in front of the turbine; rotational speed of the compressor rotor of low and high pressure; the efficiency of the oil system and engine prompting on the limit values of the parameters of the engine oil system; oil consumption to change the level in the engine oil tank; efficiency of the engine fuel system; vibration level; temperature in the engine cavities; bearing temperature; mass fuel consumption; temperature and air pressure behind the compressor; technical state of the flowing part of the engine according to its thermogas-dynamic parameters [2, 3].

However, with a wide variety of tasks, there are some difficulties: limited computing resources (amount of RAM, speed, accuracy of the results); difficulties in formalizing classical control and diagnostic algorithms and, as a consequence, their practical implementation; the need to implement the above algorithms in low-level language (assembler); difficulties in recovering information when sensors fail.

Taking into account the specific features of the diagnostic object (aircraft GTE), as a complex technical system with essential nonlinear characteristics, the above difficulties can be effectively eliminated by using the mathematical apparatus of artificial neural networks, which have a number of advantages given in [4].

## 2. Mathematical model of the change of aircraft engine TV3-117 technical state

To study the possible of TV3-117 aircraft engine technical states according to [5, 6], it is assumed that the engine in the helicopter flight mode can be in four states:  $S_0$  – engine idling (here we mean idling),  $S_1$  – engine works in the nominal mode (here we mean takeoff, nominal, cruising I, cruising II),  $S_2$  – the engine is working properly, readjustment is in progress,  $S_3$  – the engine is faulty (emergency mode of the engine). According to these data, it is possible to obtain a discrete Markov network based on the following possible transition states:  $S_0 - S_1$ ,  $S_1 - S_0$ ,  $S_0 - S_3$ ,  $S_1 - S_3$ ,  $S_2 - S_0$ ,  $S_2 - S_1$ ,  $S_2 - S_3$ ,  $S_3 - S_2$ , the probability of the engine being in each of the above states must be taken into account (fig. 1) [5, 6].

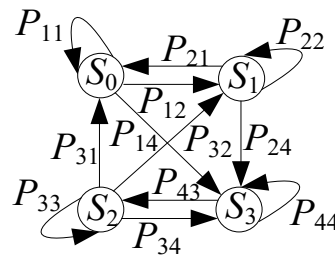


**Figure 1:** Markov discrete network showing all kinds of changes of TV3-117 aircraft engine technical state as a part of a helicopter power plant during flight tests [5, 6]

According to fig. 1, transition probabilities matrix has the form:

$$P = \begin{pmatrix} P_{11} & P_{12} & 0 & P_{14} \\ P_{21} & P_{22} & 0 & P_{24} \\ P_{31} & P_{32} & P_{33} & P_{34} \\ 0 & 0 & P_{43} & P_{44} \end{pmatrix}. \quad (1)$$

To find the probabilities of the stay of Markov chain in certain states as  $n \rightarrow \infty$  (final probabilities) in [5, 6] an equations system of the form is solved:



$$\begin{cases} P_{11}\pi_1 + P_{21}\pi_2 + P_{31}\pi_3 = \pi_1; \\ P_{12}\pi_1 + P_{22}\pi_2 + P_{32}\pi_3 = \pi_2; \\ P_{33}\pi_3 + P_{43}\pi_4 = \pi_3; \\ P_{14}\pi_1 + P_{24}\pi_2 + P_{34}\pi_3 + P_{44}\pi_4 = \pi_4. \end{cases} \quad (2)$$

According to [5, 6], in matrix form, system (2) has the form  $\pi = P^T\pi$  and  $(P^T - E)\pi = 0$ , where

$$P^T = \begin{pmatrix} P_{11} & P_{21} & P_{31} & 0 \\ P_{12} & P_{22} & P_{32} & 0 \\ 0 & 0 & P_{33} & P_{43} \\ P_{14} & P_{24} & P_{34} & P_{44} \end{pmatrix} - \text{transposed matrix}; \quad E = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} - \text{identity matrix}.$$

According to [5, 6], adding to this equation the normalization condition of the form:

$$\pi_1 + \pi_2 + \pi_3 + \pi_4 = 1; \quad (3)$$

the final matrix is obtained,  $P^T - E = \begin{pmatrix} P_{11}-1 & P_{21} & P_{31} & 0 \\ P_{12} & P_{22}-1 & P_{32} & 0 \\ 0 & 0 & P_{33}-1 & P_{43} \\ P_{14} & P_{24} & P_{34} & P_{44}-1 \end{pmatrix}.$

Replacing the fourth row of the matrix  $(P^T - E)$  with a single row, that is,  $P^T - E = \begin{pmatrix} P_{11}-1 & P_{21} & P_{31} & 0 \\ P_{12} & P_{22}-1 & P_{32} & 0 \\ 0 & 0 & P_{33}-1 & P_{43} \\ 1 & 1 & 1 & 1 \end{pmatrix}$ , a final system of linear algebraic equations is obtained:

$$\begin{cases} (P_{11}-1)\pi_1 + P_{21}\pi_2 + P_{31}\pi_3 = 0; \\ P_{12}\pi_1 + (P_{22}-1)\pi_2 + P_{32}\pi_3 = 0; \\ (P_{33}-1)\pi_3 + P_{43}\pi_4 = 0; \\ \pi_1 + \pi_2 + \pi_3 + \pi_4 = 1. \end{cases} \quad (4)$$

As a result of solving (4), expressions were obtained for determining the elements of the vector of state probabilities  $\pi(\pi_1, \pi_2, \pi_3, \pi_4)$  (shows the probability that the engine will be in the  $i$ -th state), which will allow in the presence of values of the transition probabilities  $P_{11} \dots P_{33}$  to obtain prognostic indicators of changes of TV3-117 aircraft engine technical state. As can be seen from system (4), to determine the probability of engine failure  $\pi_4$ , it is enough to know the transition probabilities for the other three states, as well as the transition probabilities  $P_{34}, P_{43}$ , which can be obtained as a result of technical diagnostics of the helicopter.

The adequacy of the developed method is confirmed in [5] by definition. The paper proposes an implementation of this method using dynamic recurrent artificial neural networks.

### 3. Matrix method for determining dynamic recurrent artificial neural networks

In this paper, the topology (internal connections) of the neural network is proposed to be encoded by signal graphs [7, 8]. A model of a neuron of a neural network is proposed to be represented by a node in the signal graph with the assignment of a serial number 1, 2, ...,  $N-1, N$ . The nodes are connected by directed arcs. The arcs are identified by synaptic weights  $w_{ij}$ , where  $i$  – number of the donor neuron,  $j$  – number of the acceptor neuron.

A convenient descriptor of a neural network is the adjacency matrix of the signal graph  $W$ . Elements of the adjacency matrices are the storage of the associative memory of neural network for various purposes. The fig. 2 shows the signal graphs and the corresponding adjacency matrices: for the forward-directed and recurrent neural network. From the theory of graphs it follows that if the column of the adjacency matrix is zero (fig. 2,  $a$ ), then the vertex of the signal graph is a sink, and in a

neural network this neuron is an input one. The zero-line points to the top of the source of the signal graph, and in the neural network this neuron is the output. In a recurrent neural network (fig. 2, *b*) there is no formal concept of layers. Any neuron can be both input and output.

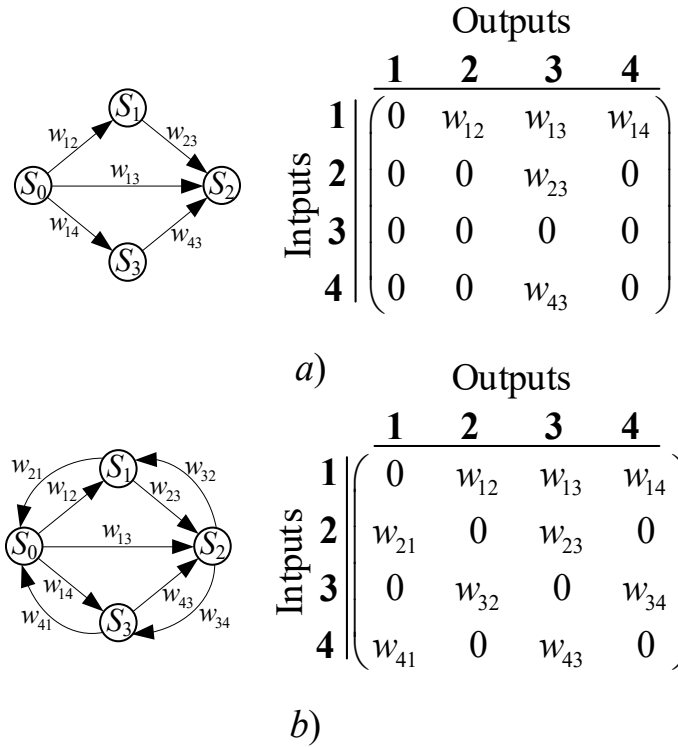


**Figure 2:** Signal graphs and adjacency matrices: *a* – direct neural network; *b* – recurrent neural network

The most widespread at present are direct-directional neural networks – perceptron's [9–12], which are currently widely used in technology to solve problems of classifying the state of equipment and diagnosing events. For training forward-directed neural networks with a large number of neurons, there is a layer-by-layer backpropagation method [9–12]. For modeling and synthesis of control systems, dynamic recurrent neural networks are more attractive (fig. 2, *b*).

Due to intranet feedbacks, to calculate each neuron in a recurrent neural network, signal values are required not only from the associated outputs of the neurons of previous layers, as in forward-directed neural networks, but also the values of signals received on this neuron itself, and others associated with him neurons. Layer-by-layer calculation of recurrent neural networks becomes impossible, in which each neuron can be both input and output at the same time. Currently, there are no formalized algorithms for calculating recurrent neural networks, including those designed to simulate dynamic systems.

The known algorithms aim at adapting the backpropagation algorithm for calculating forward-directed neural networks with feedback [13, 14]. Dynamic properties (associative memory) in forward-directed neural networks are achieved by applying input signals to various layers of the neural network and increasing the total number of neurons.

An analogy is seen with neural networks of existing structures of aggregate and controller control algorithms, and in the control algorithms there are real nonlinear elements (analogs of activation blocks). But they have a very small number of neurons (functional blocks). This necessitates, in contrast to multi-neural networks of scanners-classifiers, the study of dynamic neural networks from below, from a small number of neurons. And their complexity, topology and the number of neurons, must meet the requirements of the assigned tasks – high-quality implementation of various types of classical, adaptive, single-channel and hybrid multichannel, linear with switching and non-linear algorithms for regulation and logic control.

Studies have shown that the listed tasks can be implemented by dynamic recurrent neural networks, for which there are currently no formalized calculation algorithms. In this article, we

propose a matrix algorithm for the numerical step-by-step calculation of recurrent neural networks. In the proposed algorithm, the models of neural network neurons are represented by static equations of linear adders [15]:

$$s_{jk} = \sum_{i=1}^N x_{ik} \cdot w_{ij}; j = 1, 2, \dots, N \quad (5)$$

and activation block

$$\gamma_n = \varphi(s_{nk}); \quad (6)$$

where  $x_{1n}, x_{2n}, \dots, x_{Nn}$  – signal values at the input of the  $n$ -th neuron;  $w_{1n}, w_{2n}, \dots, w_{Nn}$  – synaptic weights of channels at the input of the  $n$ -th neuron;  $(s_{nk})$  – value of the signal at the output of the adder at the  $k$ -th step of the neural network calculation.

In the processes of training or operation of a neural network, the proposed matrix algorithm requires [15]:

1) adjacency matrix of the signal graph of the neural network  $W$  filled with the values of synaptic weights;

2) vector of functions of activation blocks  $\Phi(S)$ :

$$\Phi(S) = \{\varphi_1(s_1), \varphi_2(s_2), \dots, \varphi_{N-1}(s_{N-1}), \varphi_N(s_N)\}; \quad (7)$$

3) vector of values supplied to the inputs of all neurons of the neural network of external signals  $X_0^{in}$ :

$$X_0^{in} = \{x_{10}^{in}, x_{20}^{in}, \dots, x_{N-10}^{in}, x_{N0}^{in}\}; \quad (8)$$

4) vector of signal values at the outputs of the activation blocks:

$$\Gamma_0 = \{\gamma_{10}, \gamma_{20}, \dots, \gamma_{N-10}, \gamma_{N0}\}. \quad (9)$$

In the process of calculating neural networks, the elements of vectors  $X^{in}$  and  $\Gamma$  change from step to step. Suppose at the  $k$ -th step they have values  $X_{k-1}^{in}$  and  $\Gamma_{k-1}$ . Then the signals that at the  $k$ th step must be fed to the inputs of all neurons are summed up from vectors  $X_{k-1}^{in}$  and  $\Gamma_{k-1}$  [15]:

$$X_k = X_k^{in} + \Gamma_{k-1}. \quad (10)$$

The procedure for calculating the signal at the output of the adder of the  $j$ -th neuron of the neural network in accordance with (5) consists in multiplying the elements of the columns of the adjacency matrix  $W$  by the elements of the vector  $X_k$ , followed by summing the multiplication results with the elements of the  $j$ -th column of this matrix.

It follows from the above that at the  $k$ th step, the neural network calculation procedure is reduced to three computation stages. At the first stage, using the values of the input signals  $X_{k-1}^{in}$  and the results of calculating the neural network at the  $(k-1)$ -th step  $\Gamma_{k-1}$ , from expression (10), the elements of the vector of actual inputs of the neurons of the neural network for the  $k$ -th step  $X_k$  are calculated:

$$X_k = X_k^{in} + \Gamma_{k-1}.$$

At the second stage, the transposed adjacency matrix  $W^T$  is multiplied by the vector  $X_k$ :

$$S_k = W^T \cdot X_k. \quad (11)$$

At the third stage, after the transformation of the vector elements by the functions of activation blocks:

$$S_k = \{s_{1k}, s_{2k}, \dots, s_{N-1k}, s_{Nk}\}; \quad (12)$$

where for the  $k$ -th step the elements of the signal vector at the outputs of the neurons of the neural network are calculated:

$$\Gamma_k = \{\gamma_{1k}, \gamma_{2k}, \dots, \gamma_{N-1k}, \gamma_{Nk}\}. \quad (13)$$

Then the algorithm returns to the first stage to calculate the values of signals at the outputs of neurons at the  $(k+1)$ -th step of the neural network.

The fig. 3 shows graphs of four fully connected recurrent neural networks as test results of the proposed matrix algorithm for calculating a fully connected dynamic recurrent neural network. The arrows in fig. 3 indicate the contacts between the variables (in this case, the possible TV3-117 aircraft engine technical states), while the signal passes only in the direction of the graph arrows.

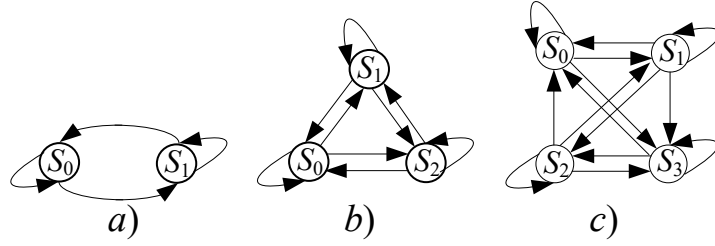


**Figure 3:** Signal graphs of three fully connected recurrent neural networks

#### 4. Training a neural network with an evolutionary algorithm based on the matrix method

The topology of fully connected neural networks is interesting for maximum intelligence when learning. It is well formalized and convenient for the algorithmization of matrix transformation operations of expressions (10) – (13).

In matrix form [15], these calculations are convenient because in a fully connected recurrent neural network, the number of synaptic weights is equal to the square of the number of neurons, and in the network any neuron, including the input one, can be used as an output one.

For dynamic neural networks, oriented to work in regulation and control systems, an important property is their inertia. In the proposed stepwise algorithm, to calculate the elements of the vector  $\Gamma$  at the current  $k$ -th step, the values at the outputs of all neurons at the previous  $(k - 1)$ -th step are required. Then, if the increment of the signal value at the output of the  $n$ -th neuron from step to step [15]:

$$\Delta\gamma_{kn} = \gamma_{kn} - \gamma_{(k-1)n} \quad (14)$$

multiply by the accepted time interval between calculation steps

$$\Delta t = t_k - t_{k-1} \quad (15)$$

then the operation of integrating the primitive of the simulation dynamic model of the neuron with  $N$  inputs will be reproduced in the neuron:

$$\gamma_{kn} = \gamma_{(k-1)n} + \Delta t \cdot \sum_{z=1}^N \gamma_{(k-1)z}; \quad n = 1, 2, \dots, N-1, N. \quad (16)$$

In this case, the step size (the time interval between calculations  $\Delta t$ ) will determine the accuracy of the transient simulation. Then the elements of the vector  $\Gamma$  will be the stack storage of the associative memory of the dynamic neural network.

As activation blocks in the proposed algorithm, we tested simplified primitives of recurrent expressions with unit parameters of one-dimensional simulation models of a linear integral link:

$$\gamma_k = \gamma_{k-1} - \Delta t \cdot s_{k-1}; \quad (17)$$

and linear inertial (aperiodic) link

$$\gamma_k = (1 - \Delta t) \cdot \gamma_{k-1} - \Delta t \cdot s_{k-1}. \quad (18)$$

For comparison, the algorithm was tested with the sigmoid function as an activation block:

$$\gamma_k = \frac{e^{s(k-1)} - e^{-s(k-1)}}{e^{s(k-1)} + e^{-s(k-1)}}. \quad (19)$$

It was found that expressions (17) and (18) have almost the same positive effect on the dynamic properties of the trained neural network, and with expression (19) it was not possible to train the neural network for a dynamic mode of operation. Obviously, the sigmoid is applicable only for neural network classifiers.

The neural network was trained in the identifier mode using test graphs of the object's response to a step effect and to harmonic oscillations of the same amplitude at three frequencies. Test signals were generated by a circuit from recurrent expressions of sequential connection of simulation models of aperiodic links ( $A$ -links) with parameters:  $T_a = 10$  s and  $k_a = 3$  units out / units in:

$$\begin{cases} y_{k1} = \left(1 - \frac{\Delta t}{T_a}\right) \cdot y_{(k-1)1} + k_a \cdot \frac{\Delta t}{T_a} \cdot x_{k-1} \\ y_{k2} = \left(1 - \frac{\Delta t}{T_a}\right) \cdot y_{(k-1)2} + k_a \cdot \frac{\Delta t}{T_a} \cdot y_{k1} \\ y_{k3} = \left(1 - \frac{\Delta t}{T_a}\right) \cdot y_{(k-1)3} + k_a \cdot \frac{\Delta t}{T_a} \cdot y_{k2} \end{cases} \quad (20)$$

and a circuit of recurrent expressions of the serial connection of the  $A$ -link and the integrating ( $I$ -link) with negative feedback

$$\begin{cases} y_{k1} = y_{(k-1)2} + k_a \cdot \frac{\Delta t}{T_u} \cdot (x_{k-1} - y_{(k-1)2}) \\ y_{k2} = \left(1 - \frac{\Delta t}{T_a}\right) \cdot y_{(k-1)2} + k_a \cdot \frac{\Delta t}{T_a} \cdot y_{k1} \end{cases} \quad (21)$$

In system (21), the parameters of the  $A$ -link are  $T_a = 10$  s and  $k_a = 3$  units out / units in, and for the  $I$ -link the neural network is trained for two options:  $T_i = 50$  s, when the transient process is monotonic, and  $T_i = 3$  s, when the transient is oscillatory.

A probabilistic evolutionary algorithm for optimizing multiextremal problems is used as a tool for tuning a neural network. The algorithm consists of a probabilistic Monte Carlo procedure for creating a set (population) of starting points for finding an optimal solution  $R\{r_1, r_2, \dots, r_{z-1}, r_z\}$  [16], where  $Z$  – given number, a genetic procedure for eliminating (removing) 10 % of unpromising points from a set  $R$ , an operation of replenishing a set  $R$  with new, randomly selected points and their distribution by a regular optimization algorithm over the coordinates of local extrema [17–19].

## 5. Description of input data

The input data for control and diagnostics of TV3-117 aircraft engine technical state are the results of modeling the transient process in it. The papers [20, 21] describe in detail the general approaches to modeling aviation gas turbine engines together with their automation in the simulator DVIG\_OTLADKA. The TV3-117 aircraft engine automatic control system operates as follows: depending on the engine control joystick (ECJ) installation angle, the rotor speed is maintained with a temperature correction at the engine inlet (i.e., the following control laws are implemented ( $n = f(\alpha_{ECJ}, T_H)$ ),  $G_T = f(n)$ ), the limiting values of the rotor speed  $n_{max}$  and the gas temperature in front of the compressor turbine  $T_G^*$  are limited.

According to [22], the dynamic characteristics of an aircraft engine are determined by the moment of inertia of the compressor rotor. With the help of the structural element «Pump-regulator», the set rotor speed is maintained by changing the fuel consumption in the combustion chamber. For the simulated TV3-117 aircraft engine, the rotor moment of inertia is taken equal to  $0.00045 \text{ kg} \cdot \text{m}^2$ . As described in [23], the dynamic characteristics of various automation elements are unified in the form of coefficients of a second-order nonlinear differential equation:

$$kx(t) = T_K^2 y''(t) + T_D y'(t) + y(t); \quad (22)$$

where  $k$  – gain;  $T_K$  – oscillatory time constant, s;  $T_D$  – differentiating time constant, s;  $x(t)$  – value of the function (for the structural element «Pump-regulator» – the value of the fuel consumption);  $y(t)$  – value of the controlled parameter (for the structural element «Pump-regulator» – the reduced or physical rotor speed);  $y''(t)$  and  $y'(t)$  – respectively the second and first time derivatives of the value of the controlled parameter.

According to [20], the schedule of the transient process in the TV3-117 aircraft engine has the form shown in fig. 4.

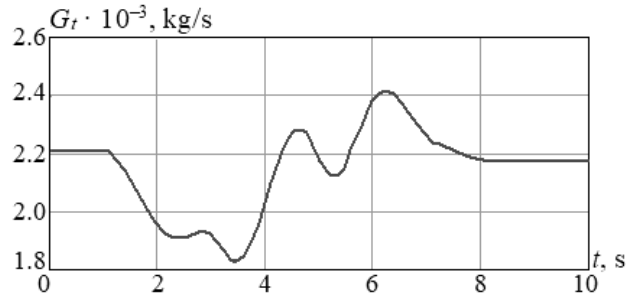


**Figure 4:** Fuel consumption graph in the transient process with the settings of the structural element “Pump-regulator”  $T_D = 0.15 \text{ s}$ ,  $T_K = 0.25 \text{ s}$ .

## 6. Results and discussion

The above-considered evolutionary algorithm trained fully connected recurrent neural networks with two, three, and four neurons, which set the possible operating modes of the TV3-117 aircraft engine. For all options, the same cardinality of the set of initial points was set  $\mu = 800$ , and the criterion for exiting the solution was the admissible difference between the minimum and maximum values of the optimized function  $\varepsilon = 10^{-3}$ .

The fig. 5 shows the results of the optimization program; adjacency matrices with optimal values of synaptic weights for a neural network: with two (fig. 5, *a*), with three (fig. 5, *b*) and with four neurons (fig. 5, *c*).

$$\begin{array}{ccc}
 \begin{pmatrix} 0.356 & 0.158 \\ -0.832 & 1.093 \end{pmatrix} & \begin{pmatrix} 1.089 & -0.235 & 0.031 \\ 0.426 & 0.633 & -1.225 \\ -0.144 & 0.282 & -0.399 \end{pmatrix} & \begin{pmatrix} -0.171 & -0.188 & 0.375 & 0.031 \\ -0.127 & 0.554 & 0.193 & 0.623 \\ 0.833 & 1.119 & 0.745 & 1.251 \\ 1.613 & 0.674 & -0.841 & -0.113 \end{pmatrix} \\
 a) & b) & c)
 \end{array}$$

**Figure 5:** Adjacency matrices of synaptic weights for recurrent neural networks: *a* – with two neurons; *b* – with three neurons; *c* – with four neurons

To train a neural network with four neurons (search for 16 values of synaptic weights for the matrix according to fig. 5, *c*), the program takes from 10 to 20 seconds.

The fig. 6, 7 and 8 show graphs of the results of studying the learning processes of a neural network: with two neurons (*a*), with three neurons (*b*) and with four neurons (*c*). It can be seen that the identification accuracy depends on the number of neurons, and the dynamic error of the neural network depends on the rate of change of the modeled process. The dynamic error of modeling transient processes is no more than  $10^{-4}$  attained by a neural network with four neurons.

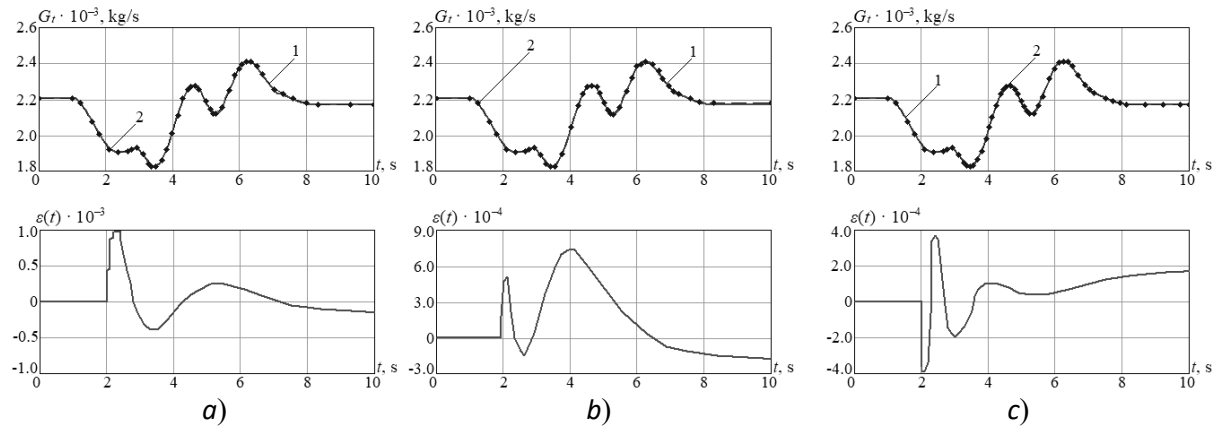


**Figure 6:** Graphs of learning outcomes of dynamic recurrent neural networks for system identification (20): 1 – neural network; 2 – analytical calculation

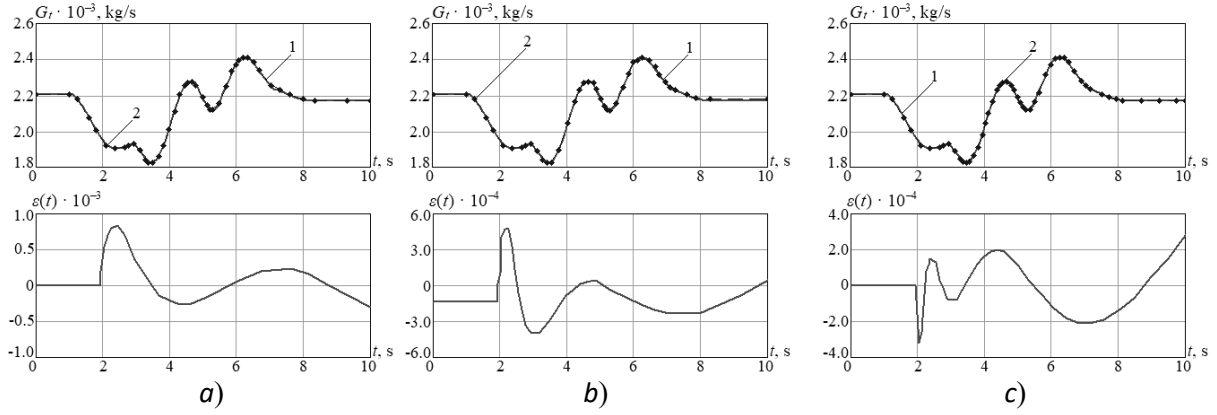


**Figure 7:** Graphs of learning outcomes of dynamic recurrent neural networks for the identification of system (21) with A-link parameters  $T_a = 10$  s and  $k_a = 3$  units out / units in and I-link:  $T_i = 50$  s: 1 – neural network; 2 – analytical calculation

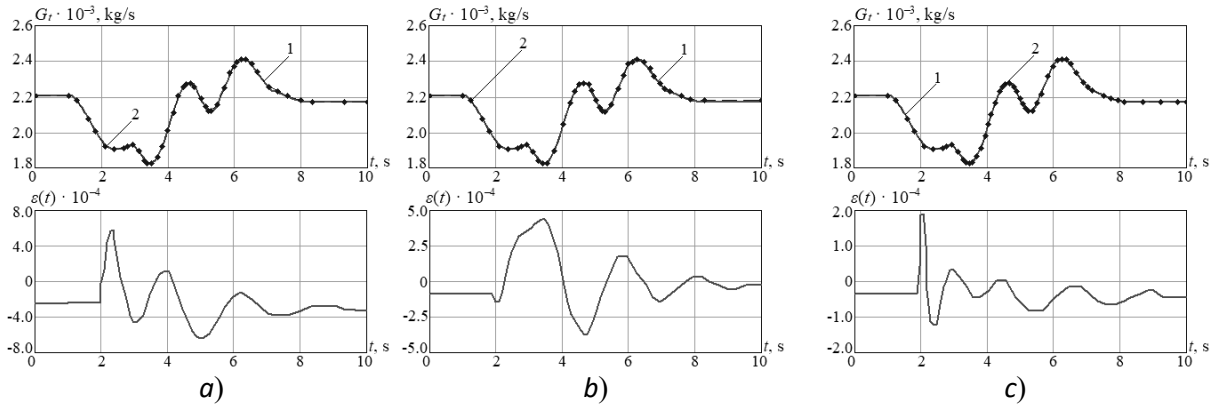


**Figure 8:** Graphs of learning outcomes of dynamic recurrent neural networks for system identification (21) with A-link parameters  $T_a = 10$  s and  $k_a = 3$  units out / units in and I-link:  $T_i = 3$  s: 1 – neural network; 2 – analytical calculation

The fig. 9 and 10 show graphs of the results of studying the learning processes of dynamic recurrent neural networks with four neurons to identify the transient process (fig. 4) at the output:  $a$  – first neuron,  $b$  – second neuron,  $c$  – third neuron.

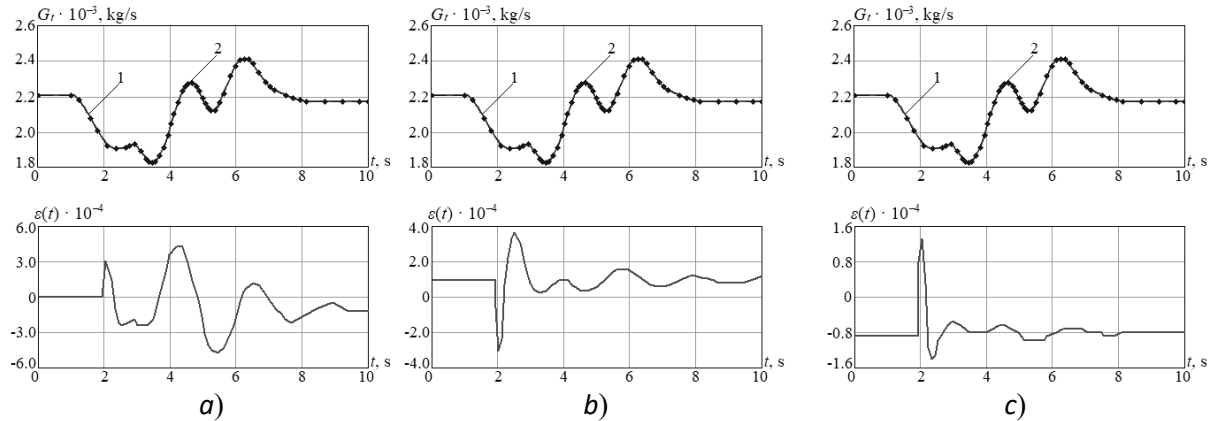


**Figure 9:** Graphs of the results of studying the learning processes of a dynamic recurrent network with four neurons to identify the transient process at the output:  $a$  – first neuron,  $b$  – second neuron,  $c$  – third neuron for transient processes in system (21) with parameters  $T_a = 10$  s and  $k_a = 3$  units out / units in and I-link:  $T_i = 50$  s: 1 – neural network; 2 – analytical calculation

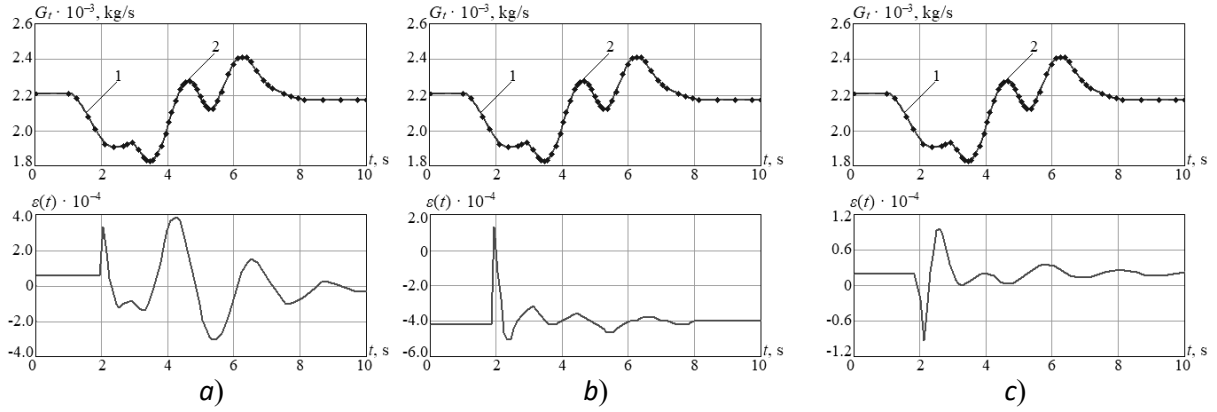


**Figure 10:** Graphs of the results of studying the learning processes of a dynamic recurrent network with four neurons to identify the transient process at the output: *a* – first neuron, *b* – second neuron, *c* – third neuron for transient processes in system (21) with parameters  $T_a = 10$  s and  $k_a = 3$  units out / units in and *l*-link:  $T_l = 3$  s: 1 – neural network; 2 – analytical calculation

In the work, a neural network with four neurons is tested for the identification of graphs under various stepwise disturbances. The fig. 11 shows the test results under disturbances: *a* – 0.5; *b* – 1.0; *c* – 2.0 units. Good agreement of the neural network with analytical calculations with a satisfactory error is seen.

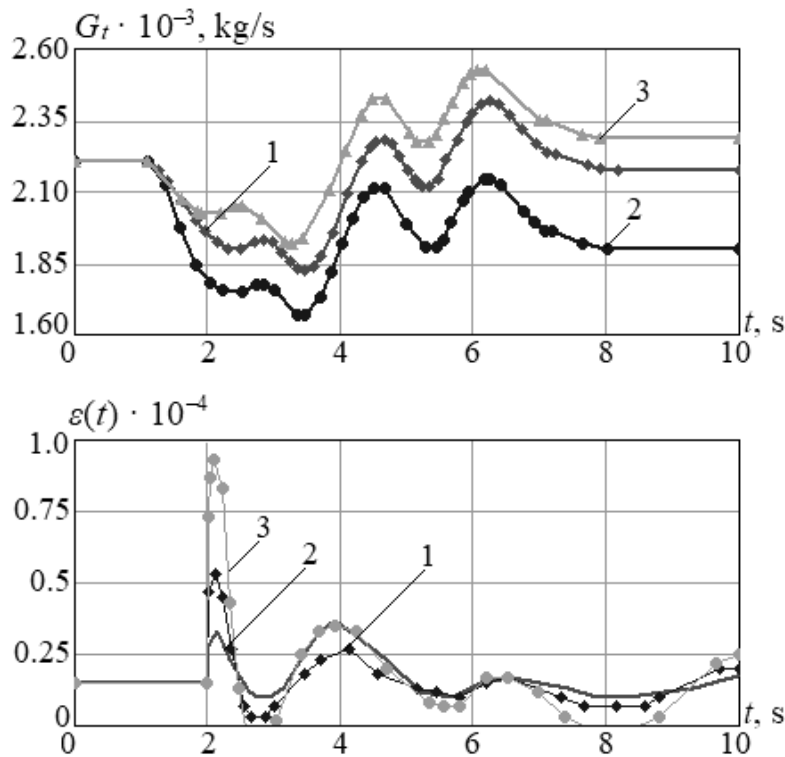


**Figure 11:** Graphs of transient processes and dynamic errors of identification of a dynamic system (21) of a recurrent neural network under stepwise disturbances, where “—” – neural network; “•” – analytical calculation: 1 – original signal; 2, 3 – modified signal (0.5 and 2.0 units, respectively)

The fig. 12 shows neural networks and graphs of transient processes obtained at the outputs of all neurons, obtained after training the neural network to identify transient processes at the output of the second neuron: 1 – output from a neuron characterizing the operating mode of the engine  $S_i$ ; 2 – output from neuron 2, characterizing the normal operation of the engine  $S_1$  (nominal, I cruising, II cruising)



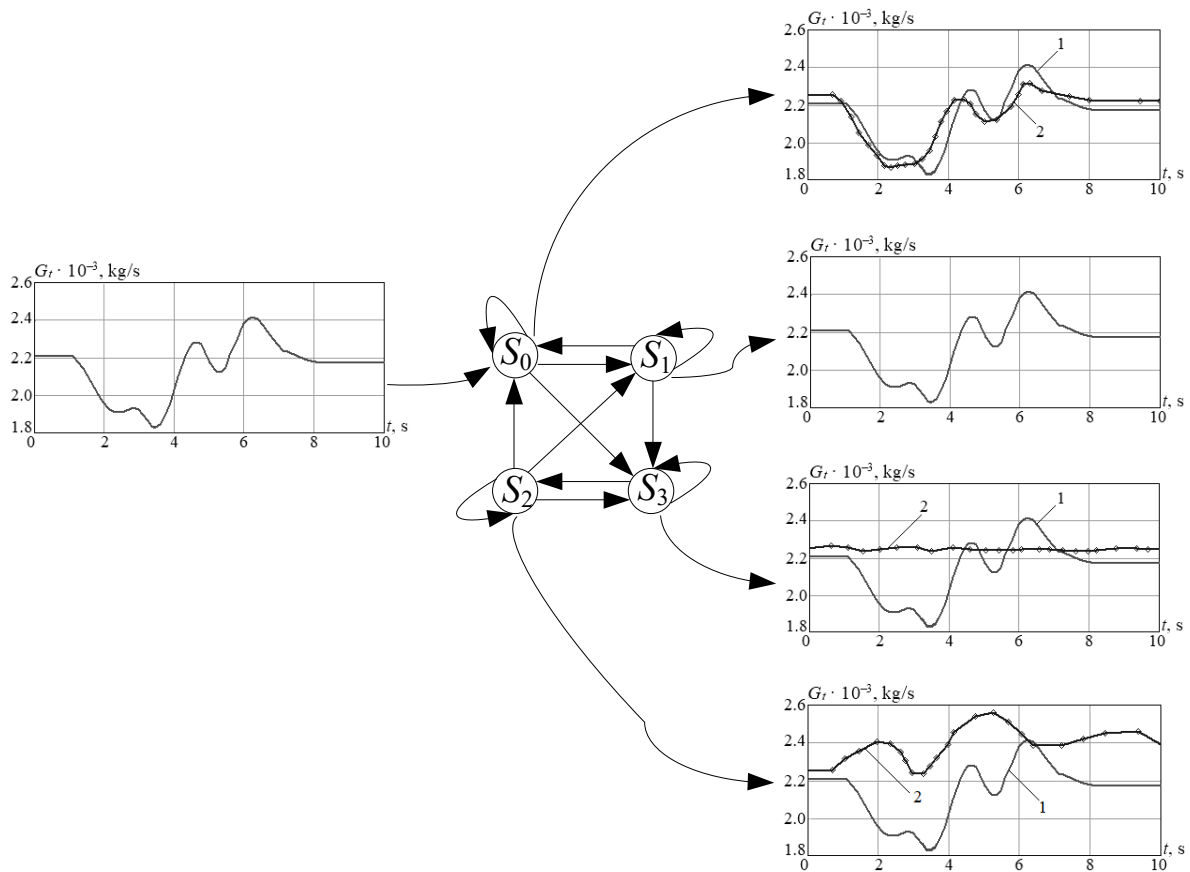


**Figure 12:** Signal graph of a recurrent neural network trained to identify the transient process at the output of the second neuron, signaling the normal operation of aircraft engine TV3-117, and graphs of the corresponding transient processes at the outputs of the remaining neurons

## 7. Conclusions

This work is devoted to the implementation of Markov discrete network showing all kinds of changes of TV3-117 aircraft engine technical state as a part of a helicopter power plant during flight tests using neural network technologies, which allows using, along with quantitative mathematical models of the TV3-117 aircraft engine, qualitative and experimental information obtained during flight tests.

For modeling and implementation Markov discrete network showing all kinds of changes of TV3-117 aircraft engine technical state as a part of a helicopter power plant during flight tests as dynamic systems in control algorithms, a dynamic recurrent neural network, encoded by signal graphs, is proposed.

To calculate dynamic recurrent neural networks, a step-by-step matrix method is proposed that uses the signal graph adjacency matrix as an associative memory storage, which made it possible to obtain sufficient accuracy in approximating the thermogas-dynamic parameter signal, for example, fuel consumption in the transient process with the settings of the structural element «Pump-regulator».

It was found that to identify a third-order dynamic system, it is sufficient to train a fully connected recurrent neural network with four neurons, symbolizing the vector of probabilities of TV3-117 aircraft engine technical states, in which the number of synaptic weights is equal to the square of the number of neurons, and any neuron in the network, including the input one, can be used as an output.

Studies have shown that in a neural network trained for the selected (second) output neuron, active signals dynamically connected with the signal on the second neuron are observed at the outputs of its other neurons.

Research prospects are the use of the obtained properties of recurrent neural networks as virtual intermediate signals in the control and diagnostics system of aircraft GTEs technical state, including TV3-117, as dynamic systems.

The results of this work can be introduced into an intelligent on-board system for control and diagnosing of aircraft GTEs technical state, including TV3-117 [24].

## 8. References

- [1] D. S. Legkonogikh, Using of neural networks technologies in diagnostics Systems of aviation powerplants, *Izvestia of Samara Scientific Center of the Russian Academy of Sciences*, 2012, vol. 14, no. 4 (2), pp. 639–643.
- [2] E. L. Ntantis, Diagnostic methods for an aircraft engine performance, *Journal of Engineering Science and Technology*, 2015, vol. 8, no. 4, pp. 64–72.
- [3] A. M. Pashayev, D. D. Askerov, C. Ardil, R. A. Sadiqov, P. S. Abdullayev, Condition monitoring system of aircraft gas turbine engine complex, *International Journal of Aerospace and Mechanical Engineering*, 2007, vol. 1, no. 11, pp. 689–695.
- [4] S. Vladov, K. Kotliarov, S. Hrybanova, O. Husarova, I. Derevyanko, S. Gvozdk, Neuro-mechanical methods of control and diagnostics of the technical state of aircraft engine TV3-117 in film regions, *Visnyk of Kherson National Technical University*, 2020, no. 1 (72), part 1, pp. 141–154.
- [5] S. Vladov, Yu. Shmelov, T. Shmelova, Modeling of the TV3-117 aircraft engine technical state as part of the helicopter power plant in the form of the Markov process of death and reproduction, in: *CEUR Workshop Proceedings*, 2020, vol. 2740, pp. 400–407.
- [6] S. I. Vladov, Ya. R. Klimova, O. I. Korniienko, Method for determining the probabilities of Mi-8MTV helicopter engine state under flight test conditions, in: *Aviation and cosmonautics: state, achievements and prospects*, 2018, pp. 26–27.
- [7] J. Bang-Jensen, G. Gutin, *Digraphs Theory, Algorithms and Applications*, Springer-Verlag GmbH, 2007.
- [8] M. Needham, A. E. Hodler, *Graph Algorithms: Practical Examples in Apache Spark and Neo4j*, O'Reilly Media, 2019.
- [9] M. Zgurovsky, V. Sineglazov, E. Chumachenko, *Artificial Intelligence Systems Based on Hybrid Neural Networks: Theory and Applications*, Springer, 2021.
- [10] E. V. Bodyansky, O. G. Rudenko, *Artificial neural networks: architectures, training, applications*, Teletex, Kharkiv, 2004.
- [11] S. V. Zhernakov, A. T. Gilmanshin, New algorithms for on-board diagnostics of aircraft gas turbine engine based on fuzzy networks, *Bulletin of USATU*, 2015, vol. 19, no. 2 (68), pp. 63–68.
- [12] S. V. Zhernakov, Control and diagnostics of complex failures of ACS GTE using recurrent neural networks, *Aerospace Instrument-Making*, 2009, no. 2, pp. 33–49.
- [13] V. B. Lila, E. V. Puchkov, Methodology of training recurrent artificial neural network with dynamic stack memory, *Software & Systems*, 2014, no. 4 (108), pp. 132–135.
- [14] N. I. Selvesyuk, A. S. Ostrovskiy, R. S. Aristov, A. A. Gladkikh, Object-oriented design of a neural network to automate the process of computer architecture determination in the information security problems, *Science Bulletin of the NSTU*, 2016, vol. 62, no. 1, pp. 133–145.
- [15] V. R. Sabanin, D. A. Dementyev, I. Yu. Kazmiruk, A. I. Repin, Matrix method for calculating dynamic recurrent artificial neural networks, *Young scientist*, 2018, no. 13 (199), pp. 1–9.
- [16] S. P. Li, A guided Monte Carlo method for optimization problems, *International Journal of Modern Physics C*, 2002, vol. 13, no. 10, pp. 1365–1374.
- [17] V. R. Sabanin, N. I. Smirnov, A. I. Repin, Modified genetic algorithm for optimization and control problems, *Exponenta Pro. Mathematics in applications*, 2004, no. 3–4, pp. 78–85.
- [18] A. A. Andreyev, Use of the genetic algorithms in optimization of non-linear functions, *Tomsk State University Journal*, 2009, vol. 14, no. 5, pp. 1036–1040.

- [19] A. N. Koshev, V. V. Salmin, A. A. Generalova, D. S. Bichkov, The development of genetic algorithm with adaptive mutations to determine the global extremum function of n-variables, The Eurasian Scientific Journal, 2016, vol. 8, no. 6, URL: <https://naukovedenie.ru/PDF/32TVN616.pdf>
- [20] D. A. Akhmedzyanov, I. A. Krivosheev, A. E. Kishalov, Simulation of the operation of aircraft GTE with elements of control systems, Bulletin of USATU, 2008, vol. 11, no. 2 (29), pp. 3–11.
- [21] D. A. Akhmedzyanov, A. E. Kishalov, Simulation of an aircraft GTE together with elements of its automation, Youth Bulletin of USATU, 2012, no. 2 (3), pp. 54–60.
- [22] D. A. Akhmedzyanov, A. E. Kishalov, K. V. Markina, Z. N. Bakirova, Experimental study of static and dynamic characteristics of gas turbine engines on automated laboratory units SR-30 and TJ-100 with various engine control algorithms, Young scientist, 2013, no. 3 (50), pp. 26–34.
- [23] D. A. Akhmedzyanov, A. E. Kishalov, An integrated approach to joint modeling and research of working processes of aircraft GTE and its automation, Young scientist, 2011, vol. 1, no. 10 (33), pp. 16–21.
- [24] Y. Shmelov, S. Vladov, Y. Klimova, M. Kirukhina, Expert system for identification of the technical state of the aircraft engine TV3-117 in flight modes, in: System Analysis & Intelligent Computing : IEEE First International Conference on System Analysis & Intelligent Computing (SAIC), 2018, pp. 77–82.

# On Classification Hidden Concepts Language in Specialized Texts Based on Methods of the Intellectual Data Processing

Iurii Krak<sup>a,b</sup>, Valentina Petrovych<sup>a</sup>, Vladislav Kuznetsov<sup>a</sup>, Eduard Manziuk<sup>c</sup>, Olexander Barmak<sup>c</sup>, and Anatoliy Kulias<sup>a</sup>

<sup>a</sup> Glushkov Cybernetics Institute, Kyiv, 40, Glushkov ave., 03187, Ukraine

<sup>b</sup> Taras Shevchenko National University of Kyiv, Kyiv, 64/13, Volodymyrska str., 01601, Ukraine

<sup>c</sup> National University of Khmelnytskyi, Khmelnytskyi, 11, Instytska str., 29000, Ukraine

## Abstract

In this article discussed and solved the problems of comparing language concepts in specialized texts, in particular, scientific texts in the Ukrainian language. A corpus of scientific texts and dictionaries as well as stop words and affixes has been formed for processing specialized texts. The resulting texts were analyzed and converted into text frequency-inverse document frequency (TF-IDF) feature representation. To transform the original vector of features, it is proposed to use an algorithm for the synthesis of linear systems, in combination with the T-stochastic neighbor embedding (T-SNE). A series of experiments were performed on test examples for the determination of informational density in the text and classification by keywords in specialized texts using the method of random samples consensus (RANSAC). A method of classification of hidden language concepts was proposed, use of clustering methods (K-means). As a result of the experiment, the structure of the classifier of hidden language concepts was obtained in structured texts. The stability of the proposed method is investigated by using the perturbation of the original data by a variational autoencoder. The obtained structure allowed to achieve a relatively high recognition accuracy (97%-99%) using decision trees and machines of extreme gradient amplification.

## Keywords 1

Texts analysis, language concepts, pseudo inversion, clustering, feature extraction

## 1. Introduction and problems statement

One of the important problems in scientific research in the specialized texts are to compare texts according to a certain criterion, namely the inclusion of a certain phrase or set of phrases in a given scientific text to obtain results that contain the desired set of phrases [1-3]. This method is currently used successfully in searching for information by keywords, but it has disadvantages - first, the need to search all the text and search for each keyword by a given criterion, which causes search results that are not relevant to the searched text [2,4]. Note that among the many known methods of textual information research are the most popular: the method of frequency analysis of the term, taking into account the inversion of frequency to other documents - TF- IDF [4], the linear method of reference vectors (LSVM) [5] and the method of the Bag of Words [6]. The advantages of TF-IDF and Bag of Words include good speed and great applications; the main advantage of the LSVM method is accuracy. The disadvantages are the slowness of execution, "rejection" of stop words, which

---

CMIS-2021: The Fourth International Workshop on Computer Modeling and Intelligent Systems, April 27, 2021, Zaporizhzhia, Ukraine  
EMAIL: yuri.krak@gmail.com (I. Krak); filonval63@gmail.com (V. Petrovych); kuznetsowlad@gmail.com (V. Kuznetsov);  
eduard.em.km@gmail.com (E. Manziuk); alexander.barmak@gmail.com (O. Barmak); anatoly016@gmail.com (A. Kulias)  
ORCID: 0000-0002-8043-0785 (I. Krak); 0000-0002-5982-8983 (V. Petrovych); 0000-0002-1068-769X (V. Kuznetsov); 0000-0002-7310-2126 (E. Manziuk); 0000-0003-0739-9678 (O. Barmak); 0000-0003-3715-1454 (A. Kulias)



© 2021 Copyright for this paper by its authors.  
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).  
CEUR Workshop Proceedings (CEUR-WS.org)

inevitably leads to loss of meaning; lack of consideration of the position of each word in the text, which can provoke difficulties in finding content in a particular text.

Based on the analysis of the subject area and the selection of research issues, the following problems are formulated:

- to form a sample of scientific texts on various topics;
- get an idea of the vectors of the features of individual sentences in the text;
- analyze the affinity of texts from the annotations of texts with the source texts;
- to investigate the representation of feature vectors with the involvement of different methods of data dimension reduction, clustering and grouping of features;
- assess the information concentration of the content of different samples of text on the representation of the vectors of the characteristics of individual sentences of the text;
- to classify individual sentences of texts by subject;
- to investigate the selection of hidden concepts by clustering methods;
- evaluate the stability of classification algorithms to transformations of dimensionality of features.

## **2. Related works**

A large number of publications are devoted to the study of specialized and scientific texts, among which the following should be noted: in [7] the amplitude and phase characteristics of terms in the text were investigated to assess the concentration of categories of terms in the text, in [8], experimental information technology was developed analysis of frequency characteristics of semantic terms and word combinations in the text and built a model Sub-Verb-Sub [8,9]. Despite the availability of these studies on this topic, a number of problems have not been resolved, in particular, the analysis of the frequencies of individual sentences in the text, the relationship of the meaning of the abstracted text (annotations, abstracts) with the text, statistical proximity [10,11] different authors on one topic and the study of ways to improve the results of text classification in the context of the problem of analysis of hidden concepts of language, in particular using the methods of grouping and reducing the dimensionality of the vector of characteristics [12,13,14].

## **3. Implementation**

### **3.1. Getting data**

The study formed three samples of scientific texts for modeling and recognition of communicative information on three topics: facial expressions, Ukrainian sign language and texts in Ukrainian (about 1500 sentences) [15]. After detailed processing these sets will be presented as open dataset.

Thus, using the proximity of themes, it is possible to determine how the representation of these texts in the space of characteristic features changes, which will allow to determine the common and different concepts of language in these texts. To solve this problem, scientific texts were presented in the form of a matrix in which each line corresponds to a separate sentence, including the title, annotation, captions, conclusions, and other textual elements that contain the text.

### **3.2. Data analysis environment**

To analyze this data, a module for intelligent analysis of scientific texts in Python was created with the involvement of the scikit-learn data processing library in the Jupiter development environment [16, 17]. This module implements the following operations:

- parsing of text data;
- deletion of uninformative data and word endings;
- obtaining statistical characteristics of the text;
- transformation of text elements into a vector of characteristic features;
- transformation of the vector of features into a vector of reduced dimension;
- training and testing of methods of classification of text data by a set of features;

- visualization of the vector of characteristic features.

The resulting module was tested on a hardware platform with Windows and the following characteristics: Intel Core i5-6600k processor, 8 GB of DDR4 RAM.

### 3.3. Data organisation

To solve this problem, we present scientific texts in the form of a matrix, in which each line corresponds to a separate sentence, including title, abstract, captions, conclusions, and other elements that characterize this text, presented as a vector of features, where each of the features corresponds to a particular term and the frequency of its appearance in the text. To compare several texts with different topics, we will form a body of scientific texts that limits the area of interest (for example, scientific articles that contain the required information).

Texts (body of texts and compared text) are analyzed by a parser, which eliminates all words that do not carry significant meaning (stop words) and cuts off affixes (suffixes and endings) of words. On the basis of the received set of terms frequency characteristics are formed. Then both texts (or corpora of texts) are compared and all terms and, accordingly, features that are not included in at least one of the texts are cut off.

### 3.4. Features extraction

Each row of the matrix was presented in the representation TF-IDF (text frequency inverse document frequency) [8,9,14], where each of the elements corresponded to a separate term and frequency of its occurrence. The obtained frequency characteristics were compared and all terms and, accordingly, features that are not included in at least one of the texts were cut off from the vector of features. This allowed to preserve the dimensionality of the data and to take into account only the ratio of the number of scientific terms common to texts with different topics.

In addition, when using some methods to reduce the dimensionality of the data (in particular, Karunen-Loeve [9]), this allowed to reduce the number of zero elements and the degree of sparseness of the sample matrix.

### 3.5. Statistic relationship of abstracts in specialized texts

To estimate the statistical affinity in the first group of experiments, it was proposed to use the following indicators: Cartesian distance, Pearson's test and standard deviation. These values indirectly indicate the heterogeneity of sentences and therefore at the preliminary stage of the process allow you to indirectly check the validity of the sample [18,19]. As a result of tests when comparing individual sentences from a single scientific text and annotation from this text, it was noted that sentences that had different meanings differed in the proposed parameters.

The example of the 1st sentence from the annotation shows the similarity and difference of sentences on three parameters are given in Table. 1.

**Table 1**  
Statistic features of the studied sentences

| №<br>n/n | Abstract |          |          | Article |          |          |
|----------|----------|----------|----------|---------|----------|----------|
|          | Dist     | Pearson  | St. Dev  | Dist    | Pearson  | St. Dev  |
| 1        | 0        | 1        | 0.490698 | 11      | 0.425414 | 0.347540 |
| 2        | 10       | 0.618363 | 0.529891 | 4       | 0.854699 | 0.325396 |
| 3        | 11       | 0.409801 | 0.300327 | 12      | 0.501629 | 0.490698 |
| 4        | 17       | 0.002661 | 0.300327 | 14      | 0.334416 | 0.418213 |
| 5        | 12       | 0.541444 | 0.529891 | 11      | 0.425414 | 0.34754  |

The calculated values in Table 1. show that for sentences with common concepts, there is a pattern of similarity of statistical indicators for texts with similar content, for example, in sentences that had the same Cartesian distance, but for the 1st sentence of the annotation and the 2nd of the main content, which contained close in content sentences, these values did not allow to argue about the unambiguous similarity of these sentences in content.

### 3.6. Feature vector dimensionality reduction

To reduce the dimensionality of the original vector of characteristic features of individual sentences, the Karunen-Loeve transformation was applied. Before performing the decomposition of the original data into eigenvectors, insignificant features were discarded, in particular, features with a low frequency of occurrence in the text to reduce the sparseness of the matrix and to exclude zero rows or columns in the covariance matrix.

As a result of this experiment, feature sets were obtained for sampling, which were analyzed further. This schedule of text sampling for the first body of texts showed that 95% of the energy of eigenvectors was within first three vectors, which allowed to visualize the spatial arrangement of feature vectors of data elements in the form of a three-dimensional diagram, shown below in Fig. 1.

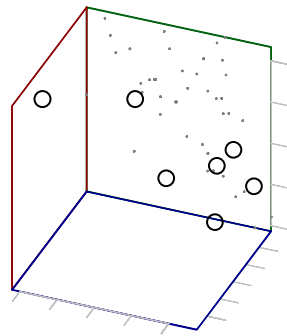


**Figure 1.** First three eigenvalues of the studied sentences in Karhunen-Loeve representation

As can be seen from the diagram, first, the first three eigenvalues form separate clusters of points, which allows us to argue about the difference of concepts in these sentences. Second, the distance between the elements of the test sample (circled) and the training sample (other points) allows you to visually assess the proximity of data samples to each other, and, therefore, used to identify hidden text parameters.

### 3.7. Dimensionality reduction using grouping of features

Since the three-dimensional representation is not convenient for visualization, it was proposed to perform feature grouping for visualization in the form of a flat image, which additionally used T-stochastic neighbor embedding [20] to reduce the dimensional vector of features to two. For clarity, Fig. 2 below shows the representation of the feature vector for the second body of texts by T-stochastic neighbor embedding and clustering of points by the method of K-means.

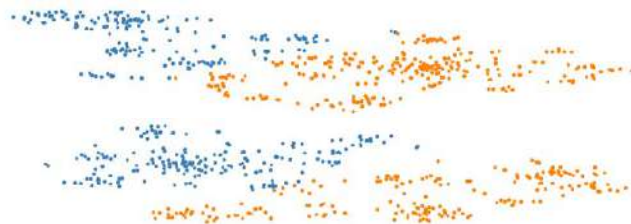


**Figure 2.** Representation of a feature vectors for the second body of texts by T-stochastic neighbor embedding, grouping of features and clustering of points by the method of K-means

At the Fig. 3 shows histograms of the distribution of feature values for the first and second corpora of texts, respectively.

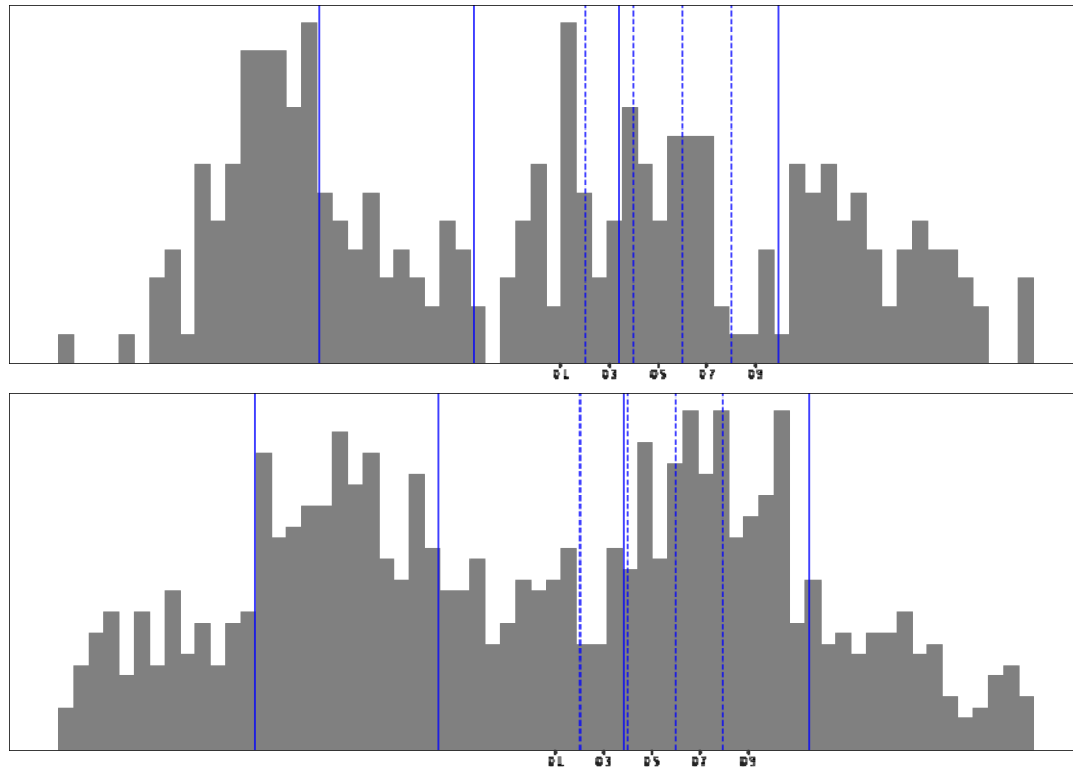


**Figure 3.** Histograms of the distribution of feature values for the body of texts 1 and 2

### 3.8. Informational density of sentences in text

With a relatively large number of samples in the body of texts and, accordingly, the educational sample, the sentences and their location in the feature space should come to the fore [21]. Accordingly, texts that differ significantly in content will have a large number of non-occurrences of sentence elements in clusters, which indicates the difference between the subject matter of the text and hidden concepts.

To test this hypothesis, it was proposed to use a regression function - the method of consensus of random samples (RANSAC) [22,23]. In Fig. 4 shows the regressions on three samples of scientific texts in the representation of T-SNE features.

Dark gray in Fig. 3 shows which points are accepted by the method of the most informative and used to build a regression. Accordingly, comparing the graphs, we can indicate that the texts have a common theme, because the location of the points intersect. In addition, the dark gray dots in Fig. 3 also indicate the areas with the greatest information density of the content of each of the corpora.

### 3.9. Sentences classification using different text corpora

A series of tests was performed on T-SNE representations of texts involving methods of intelligent data processing [24,25]. Thus, the following methods were studied: random forest, decision tree, nearest neighbors method, reference vectors method, naive Bayesian classifier, single-layer neural network. As part of the experiment, the Bayesian classifier performed best on the obtained data set. The learning results of the algorithms are illustrated in Table 2.



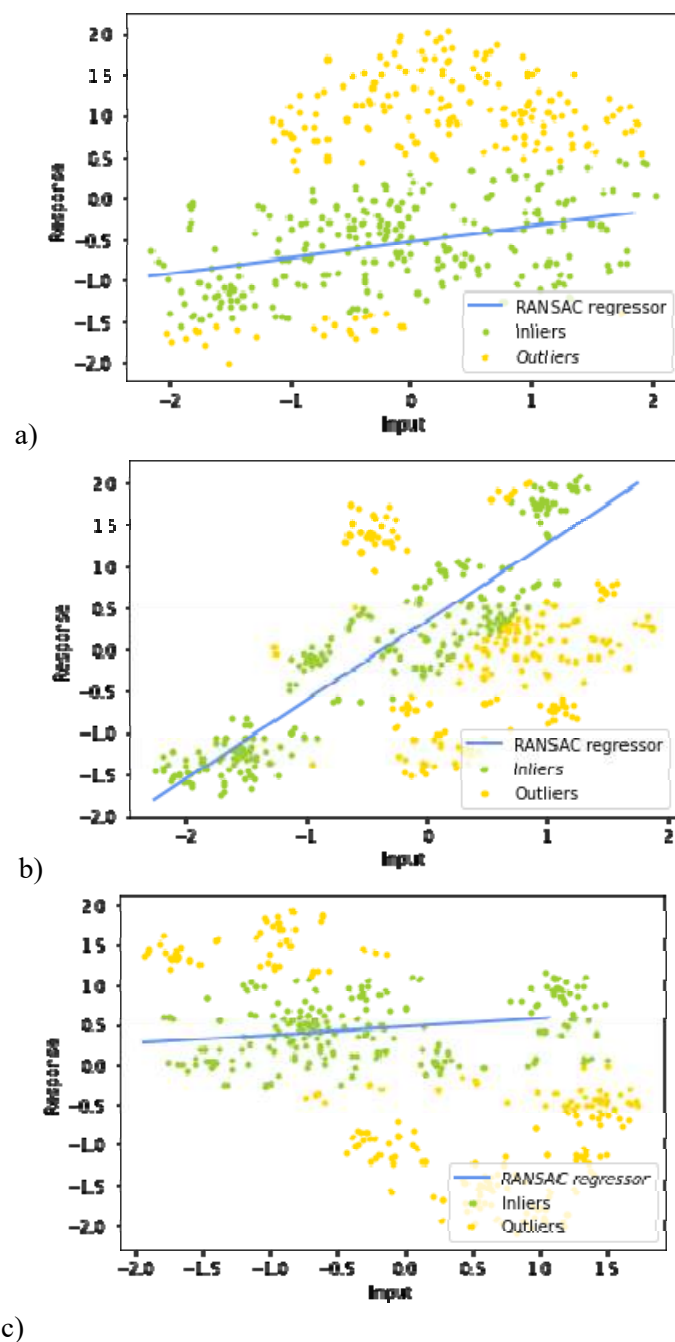


**Figure 4.** Sentences from 1 to 3 (a to c) and their informational density

**Table 2**

Precision, recall, score and support for Bayesian classifier on given datasets

| Dataset      | Precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| Emotion      | 0.94      | 0.87   | 0.90     | 138     |
| Gesture      | 0.82      | 0.94   | 0.87     | 163     |
| NLP          | 0.95      | 0.87   | 0.91     | 159     |
|              |           |        |          |         |
| accuracy     | 0.90      | 0.89   | 0.89     | 460     |
| macro avg    | 0.90      | 0.89   | 0.89     | 460     |
| weighted avg | 0.90      | 0.89   | 0.89     | 460     |

Based on the experiment, it was shown that the obtained set of TF-IDF features allows to classify texts with high reliability (87%). The presence of recognition errors of the 1st and 2nd kind (see Fig.

4) is explained by the great affinity of the texts and the closeness of the author's styles in scientific texts, which involves the use of common terms, and this can be seen from the similarity of representations in Fig. 3.

### **3.10. Classification of text using help of data clustering**

In order to study the possibility of improving the results of classification, the following approach is proposed: since the studied texts include hidden concepts (subsets) from the category of sentences that are close in content, the selection of such hidden concepts allows to correctly set the task of classifying texts, namely to classify sentences according to their similarity to a particular concept of the text [26].

To do this, it was proposed to use one of the implementations of the method of K-means (mini-batch K-means), which is best suited for experimental data. The obtained data clusters were selected as data labels, which thus allowed to move from the classification by subject of texts to the classification by the method of clustering [27,28], taking into account the internal structure of the data, namely the hidden concepts of language.

The following methods were chosen as classification methods: the method of reference vectors with linear and nonlinear hypothesis, single-layer neural network, Bayesian classifier, decision trees and related methods, including adaptive amplification machine and extreme gradient amplification. As a result of the test it was shown that the obtained set of data from the studied methods of the highest accuracy of recognition of hidden concepts is achieved by the method of reference vectors with a linear hypothesis (97.4%), single-layer neural network (97.4%), random forest (99.1%) , decision tree with extreme gradient boost (99.1%).

### **3.11. Algorithm stability testing using perturbations of feature matrix**

To verify the algorithm, an additional experiment was performed to study the stability of determining data classes when introducing perturbations into the vector of characteristic features. Perturbation was performed by converting the original representation of T-SNE (two-dimensional) into a latent feature space (two-dimensional), in which the data points from the source space corresponded to the location in the latent feature space. A variational autoencoder (VAE) was used for this purpose [28]. This method minimizes the rms error between the input (in the T-SNE representation) and the output data set (in the latent feature space), and generates additional data elements that have the same distribution as the sample of training samples in Fig. 5.

In addition, in order to achieve high accuracy of data representation, the hidden layer of the autocoder has a dimension much higher than in conventional applications.

Using the dimensional transformation obtained by the autocoder, it was noted that the representation of features in the latent space (see Fig. 1) allows us to build a linear hypothesis when classifying each other (one class versus another class). Below in Fig. 6 shows hypotheses and data labels for representation of features in the latent space.

Considering Fig. 5 and Fig. 6 it can be noticed that the separation band has decreased. Accordingly, this led to a decrease in the accuracy of recognition by classification methods by 12% - up to 85% for the method of reference vectors and single-layer neural network and by 2% - up to 97.1% for random forest and decision tree methods with extreme gradient enhancement. Despite the decrease in the accuracy of recognition, this experiment is interesting to study the effect of perturbation on the accuracy of recognition by different methods of classification.

### **3.12. Determining the nature of perturbations in a sample matrix using pseudo inversion**

Perturbation of input data affects the convergence of the classifier learning algorithm. The measure of perturbation in the original and transformed matrix in the latent space of features is the

value of entropy. Entropy can be determined indirectly by the decrease in energy of its eigen mean square error vectors of the sample matrix.

A variational auto encoder minimizes the average quadratic error between data elements (MSE). MSE is also an expression of the magnitude of data. Thus, it is possible to associate the value of variance change with the nature of the decline in energy of its own numbers of the source and transformed matrix.

Knowing the magnitude of the variance change, we can estimate the number of vectors responsible for the informative part of the data under study. The advantage of latent representation is the reduction of the distance between data classes (and separation band), which ultimately affects the number of iterations of the optimization algorithm and its convergence of the classification algorithm.

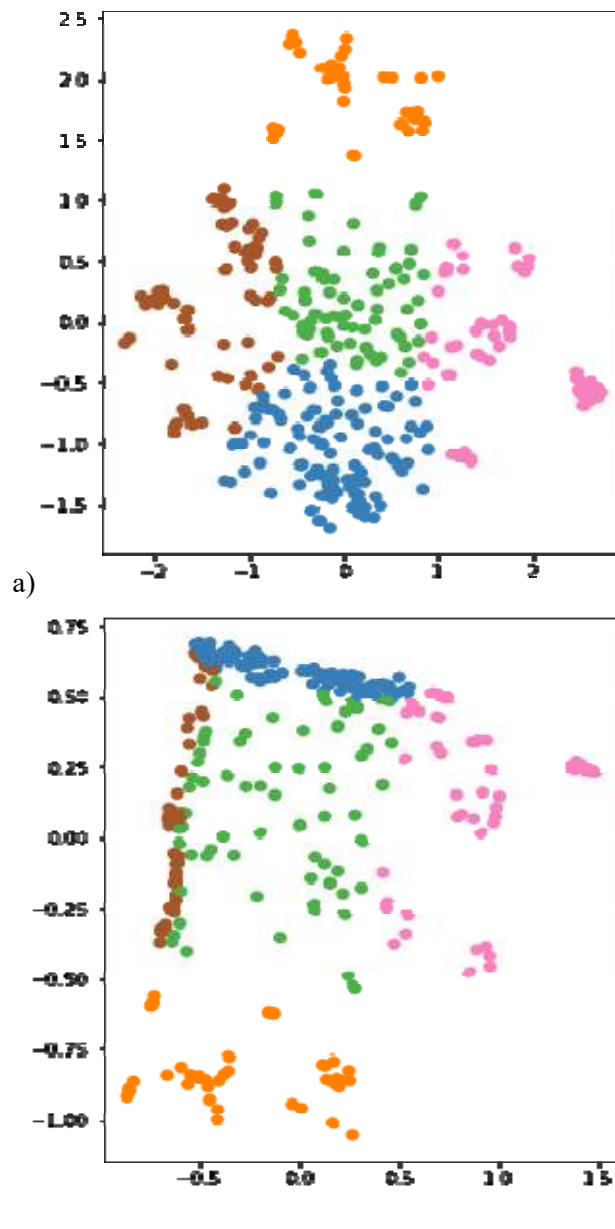


**Figure 5.** T-SNE a) and b) representation of feature space of higher dimension

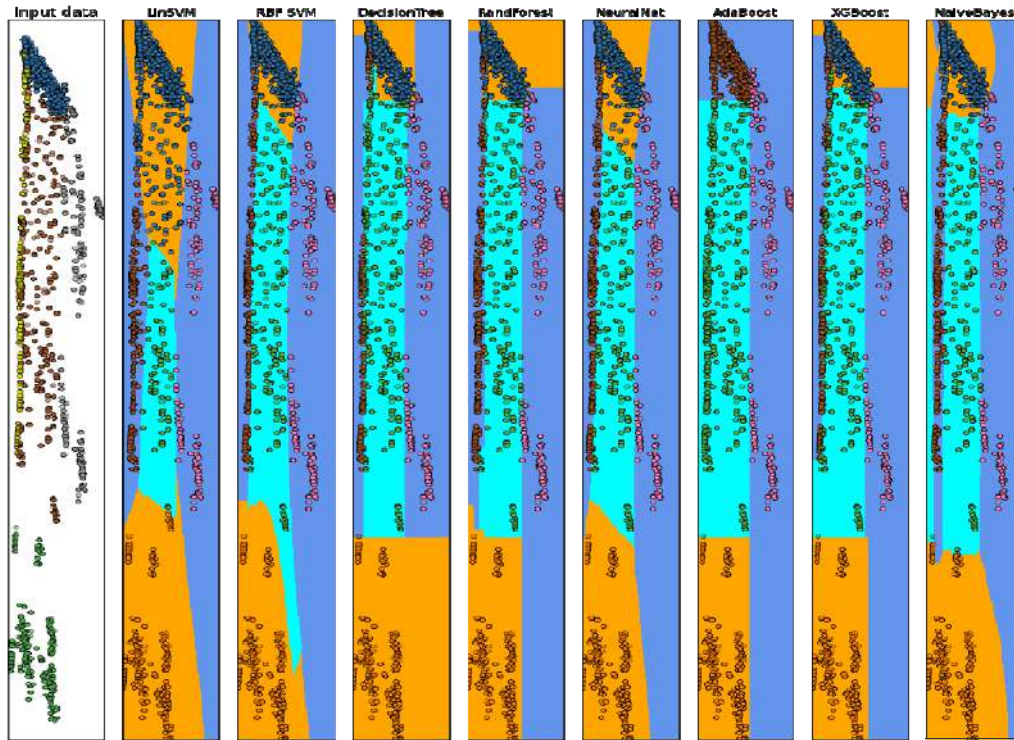


**Figure 6.** Hypotheses visualizing for different classifiers in latent feature space

## 4. Conclusion

The presented work proposes an approach to text analysis using text mining methods, including TF-IDF feature extraction method, feature dimensionality reduction based on Karhunen-Loeve transform and T-SNE data representation in 2-dimensional space and classification of acquired data using decision tree models. Also the paper discusses how the stability of the methods is affected by perturbation of data using variational autoencoder.

In order to study proposed method an experimental dataset was obtained; this dataset consists of three text corpora on three different topics – facial expression recognition, gesture recognition and text mining; these datasets related to similar topics of research and thus giving a possibility to study proposed methods in condition where the datasets have some features in common.

The experimental results of algorithms is shown in tables and figures; most important results are depicted on the Fig. 4, Fig.5 and Table 2. Given data is separable and has alignment axes in feature vector in reduced dimensionality representation; moreover, application of variational autoencoder copes well with data noise and reduces unnecessary variance but decreases separability of the data.

Feature vectors of individual sentences formed from the appearance of individual words are very effective for the classification of texts by subject, but the presence of common themes of scientific texts and typical sentences reduces the efficiency of recognition.

The use of dimensional reduction and grouping methods indicated the presence of data point concentrations that could be used to assess the author's style using typical sentence constructs. Two results should be pointed out separately: in the classification of hidden concepts (clusters of data points) with the use of clustering methods and the introduction of perturbations into the original data.

Thus, given results make it possible to determine the further direction of research, namely the study of a sample of scientific texts with other topics and authorial styles, the use of other methods of classification of structured texts, what will be the goal of our further investigation.

## 5. References

- [1] N. Firoozeh, A. Nazarenko, F. Alizon, B. Daille. Keyword extraction: Issues and methods. *Natural Language Engineering*, 26(3) (2020): 259-291. doi:10.1017/S1351324919000457
- [2] J. Ventura, J. Silva. (2007). New techniques for relevant word ranking and extraction. In: Neves J., Santos M.F., Machado J.M. (eds) *Progress in Artificial Intelligence. EPIA 2007. Lecture Notes in Computer Science*, 4874. Springer, (2007), pp.691–702. <https://doi.org/10.1007/978-3-540-77002-2>
- [3] M. Ortuno, P. Carpena, P. Bernaola, E. Munoz, A.M. Somoza. Keyword detection in natural languages and DNA. *Europhys. Lett*, 57 (5) (2002): 759-764.
- [4] B. Das, S. Chakraborty. An Improved Text Sentiment Classification Model Using TF-IDF and Next Word Negation. 2018. arXiv preprint arXiv:1806.06
- [5] M. Labbé, L.I. Martínez-Merino, A.M. Rodríguez-Chía. Mixed Integer Linear Programming for Feature Selection in Support Vector Machine. *Discrete Applied Mathematics*, 261. Elsevier, (2019), pp.276-304. [ff10.1016/j.dam.2018.10.025f](https://doi.org/10.1016/j.dam.2018.10.025).
- [6] B. Heap, M. Bain, W. Wobcke, A. Krzywicki, S. Schmeidl. Word Vector Enrichment of Low Frequency Words in the Bag-of-Words Model for Short Text Multi-class Classification Problems, 2017. arXiv:1709.05778
- [7] Y. Krak, O.Barmak, O. Mazurets. The practical implementation of the information technology for automated definition of semantic terms sets in the content of educational material. *CEUR WS*, Vol. 2139, (2018):245-254. DOI:10.15407/pp2018.02.245
- [8] S. Robertson. Understanding inverse document frequency: On theoretical arguments for IDF, *Journal of Documentation*. 60 (5) (2004): 503–520.
- [9] A. Aizawa. An information-theoretic perspective of tf-idf measures, *Information Processing and Management*. 39 (1) (2003): 45–65.
- [10] M. Farouk. Measuring Sentences Similarity: A Survey. *Indian Journal of Science and Technology*, 12(25) (2019): 1-11. DOI: 10.17485/ijst/2019/v12i25/143977
- [11] W.H. Gomaa, A. A. Fahmy. A survey of text similarity approaches. *International Journal of Computer Applications*, 68(13) (2013): 13-18.
- [12] A.V. Barmak, Y.V. Krak, E.A. Manziuk, V.S. Kasianiuk. Information technology of separating hyperplanes synthesis for linear classifiers. *Journal of Automation and Information Sciences*, 51(5) (2019): 54-64. doi: 10.1615/JAutomatInfScien.v51.i5.50
- [13] Iu.V. Krak, G.I. Kudin, A.I. Kulyas. Multidimensional scaling by means of pseudoinverse operations. *Cybernetics and Systems Analysis*, 55(1) (2019): 22-29. doi: 10.1007/s10559-019-00108-9
- [14] E.L. Shimomoto, L.S. Souza, B.B. Gatto, K. Fukui. Text classification based on word subspace with term frequency. 2018. arXiv:1806.03125v1
- [15] Iu.V. Krak, O.V. Barmak, S.O. Romanyshyn. The method of generalized grammar structure for text to gesture computer-aided translation, *Cybernetics and Systems Analysis*, 50(1) (2014): 116-123. doi: 10.1007/s10559-014-9598-4
- [16] S. Bird, E. Klein, E. Loper. *Natural Language Processing with Python*. O'Reilly Media, 2009
- [17] J. Perkins. *Python Text Processing with NLTK 2.0 Cookbook*. Packt Publishing, 2010.
- [18] T. Mikolov, K. Chen, G. Corrado, J. Dean. Efficient Estimation of Word Representations in Vector Space. 2013. arXiv:1301.3781.
- [19] A. Globerson, G. Chechik, F. Pereira, N. Tishby. Euclidean Embedding of Co-occurrence Data, *Journal of Machine Learning Research*, 8 (2007): 2265-2295.
- [20] L. Van der Maaten, G. Hinton. Visualizing Data using t-SNE, *Journal of Machine Learning Research*, 9 (2008): 2579-2605.
- [21] T. Mikolov. Distributed representations of words and phrases and their compositionality, *Advances in Neural Information Processing Systems*. 2013. arXiv:1310.4546.
- [22] M.A. Fischler, R.C. Bolles. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography, *Comm. of the ACM*, 24(6) (1981): 381-395. <https://doi.org/10.1145/358669.358692>.

- [23] A. Hast, A. Nysjö, A. Marchetti. Optimal RANSAC – Towards a Repeatable Algorithm for Finding the Optimal Set, *Journal of WSCG*, 21(1)(2013): 21-30.
- [24] *Survey of Text Mining I: Clustering, Classification, and Retrieval*. Ed. by M. W. Berry. 2004, . <https://www.springer.com/gp/book/9780387955636>.
- [25] *Emerging Technologies of Text Mining: Techniques and Applications*. Ed. by H. A. Do Prado, E. Ferneda. IGI Global, 2007.
- [26] G.E. Hinton, R.R. Salakhutdinov. Reducing the Dimensionality of Data with Neural Networks, *Science*, 313(5786) (2006):504-507. doi: 10.1126/science.1127647.
- [27] E.A. Manziuk, A.V. Barmak, Y.V. Krak, V.S. Kasianiuk. Definition of information core for documents classification, *J. Autom. Inf. Sci.* 50(4) (2018): 25-34.
- [28] S. Ioffe, C. Szegedy. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. 2015. arXiv:1502.03167[cs].



# The Neural Network for Emotions Recognition under Special Conditions

Mariia Tiahunova<sup>a</sup>, Olesia Tronkina<sup>a</sup>, Galina Kirichek<sup>a</sup> and Stepan Skrupsky<sup>a</sup>

<sup>a</sup> National University "Zaporizhzhia Polytechnic", Zhukovsky str., 64, Zaporizhzhia, 69063, Ukraine

## Abstract

This paper describes a convolutional neural network for emotions recognition considering such special cases like low light conditions, objects on a human face (glasses) and various foreshortening. For correct recognition a dataset that suits the task was found and used, the data from the dataset was normalized, a model was built and trained, metrics with optimizers and a loss function were chosen. The model was tested on test data, its parameters were chosen in accordance with the reference values. All emotion (happy, surprise, angry, fear, neutral, sad and disgust) was recognized with a minimum accuracy of 76% and maximum accuracy of 100%.

## Keywords 1

Artificial Intelligence, Deep Learning, Facial Emotion Recognition, Convolutional Neural Network

## 1. Introduction

Emotion is a process that reflects a person's subjective attitude to a stimulus. It is emotions that induce a person to take action, determine his health, and sometimes block thinking. Thanks to emotions, we understand what mental state a person is, on the basis of these conclusions, we choose a further model of behavior [1]. Emotions are also a signal of a person's desire and need for something. Emotions can save lives during an emergency [2].

In 1992, Paul Ekman, a psychologist at the University of California, San Francisco, described 6 basic emotions that people of all cultures experience and can recognize. This is happy, anger, fear, sadness, surprise, disgust [3]. Any emotion cannot be hidden because for every basic emotion there is an unmistakable facial expression. Even if someone deliberately tries to cheat, or if social norms prohibit the display of basic emotions, micro-expressions always betray them.

The current level of development allows you to create new algorithms and computer vision systems. With the increase in the power of graphics processors, the increase in Random Access Memory (RAM), it became possible to create emotion recognition systems on an ordinary computer [4, 5].

Thanks to the development of computer vision technologies, emotion recognition has become more accurate and accessible for a wide range of people. The software can be integrated into cameras and read people's emotions. Also, emotion recognition systems can be used not only in the field of security, but also in robotics, marketing, targeting, biometrics, bioengineering, as well as contactless payments. Using this system, you can check people for honesty in interviews or interrogations. These developments have long been used on street cameras in developed countries, which helps to reduce the crime rate. Also, many applications, thanks to access to the camera, read emotions and, draw conclusions about the interests of a person, and then, based on these interests, offer content.

---

CMIS-2021: The Fourth International Workshop on Computer Modeling and Intelligent Systems, April 27, 2021, Zaporizhzhia, Ukraine  
EMAIL: mary.tyagunova@gmail.com (M. Tiahunova); olesyatronkina17@gmail.com (O. Tronkina); kirgal08@gmail.com (G. Kirichek); sskrupsky@gmail.com (S. Skrupsky)  
ORCID: 0000-0002-9166-5897 (M. Tiahunova); 0000-0001-5239-2358 (O. Tronkina); 0000-0002-0405-7122 (G. Kirichek); 0000-0002-9437-9095 (S. Skrupsky)



© 2021 Copyright for this paper by its authors.  
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).  
CEUR Workshop Proceedings (CEUR-WS.org)

## 2. Literature review

The idea of using machine learning is quite old and widespread. Many scientists tried to achieve better results by increasing the complexity of the algorithms, creating their own databases, using powerful hardware, training models for months. In general, the positions of researchers can be divided into two conditional groups. The first group includes studies in which the authors' attention is focused on the study of algorithms [6, 7]. The second group includes studies that analyze databases for better algorithm performance. Two of the most representative examples from each group are described below.

In the article by Shan Li and Weihong Deng "Deep Facial Expression Recognition: A Survey" [8] datasets for building deep learning algorithms. In this article, Shan Li and Weihong Deng discuss the available datasets, the work of modern algorithms on this data. The paper also describes the problems of using each algorithm, and assesses their quality. They made a diagram of the recognition accuracy on each of the known datasets. Each stage of the implementation of the system is described on the basis of basic knowledge and based on suggestions for frequent uses of the implementation for each stage. The uniqueness lies precisely in the overview of databases, their assessment and what data is better to choose for modern deep neural networks. Learning strategies are also described in the article based on static and dynamic images, the advantages and disadvantages are discussed. Based on this scientific work, a database was selected for development in this article, as well as the deep neural network algorithm that is used in this article.

For the second example, the article by Byoung Chul Ko "A Brief Review of Facial Emotion Recognition Based on Visual Information" [9]. It describes the classic and the simplest way to determine emotion from a face image is based on the classification of facial landmarks, the coordinates of which can be obtained using various algorithms. A method is described in which usually from 5 to 68 points are marked, tying them to the position of the eyebrows, eyes, lips, nose, jaw, which allows you to partially capture facial expressions. Normalized coordinates of points can be directly fed into a classifier (for example, SVM or Random Forest) and get a basic solution. Naturally, the position of the faces should be aligned. The simple use of coordinates without a visual component leads to a significant loss of useful information, therefore, to improve the system, various descriptors are calculated at these points: LBP, HOG, SIFT, LATCH, etc. After concatenating the descriptors and reducing the dimension using PCA, the resulting feature vector can be used for classification emotions.

Thus, the task of determining emotions is not fully understood and realized. This is due to insufficient data, high quality images are difficult to find and require a lot of storage space, difficult lighting conditions, and every emotion is subjective. It is especially difficult to pinpoint an emotion if the face is tilted. It is because of the angle of inclination that the key point recognition algorithm does not correctly identify the emotion. It also takes time and resources to find features, and neural networks differ from other machine learning models in that they skip the feature engineering stage. The subject of research is information models artificial neural networks, as well as the implementation and description of the algorithm convolutional neural networks. The complexity of the implementation lies in the training of the neural network.

## 3. Facial landmarks method

The classic way to identify emotions is a method based on facial landmarks, key points. In total, the algorithm has 2 stages. At the first stage, you need to identify a feature vector, this is done by a machine learning model based on input data. At the second stage, the already trained model assigns new images of the object to a certain class.

The vector of facial features can be anything, nose, eyes, lips, facial contour, eyebrows, etc. (Fig.1). It is these characteristics that serve as input to the model. Next, two statistical models are built: a shape model and a texture model.

Shape model – a parametric linear model describing the possible options for the position of key points.

Shape – a set of points on the image that make up the contours of the object under study. For



example, the shape of a human hand can be defined through points on the contours of the palm and fingers, and the shape of the face through points on the eyebrows, nose, lips, and outer contour [11].

A texture model is a similar model, but it describes the possible variations in pixel intensity. The texture is the intensity of pixels in the image inside the outer contour of the shape. The number of pixels inside the outer contour of the shape may differ for different images [12].

Therefore, all face images are reduced to a single averaged shape using piecewise affine transformation (Fig. 2).

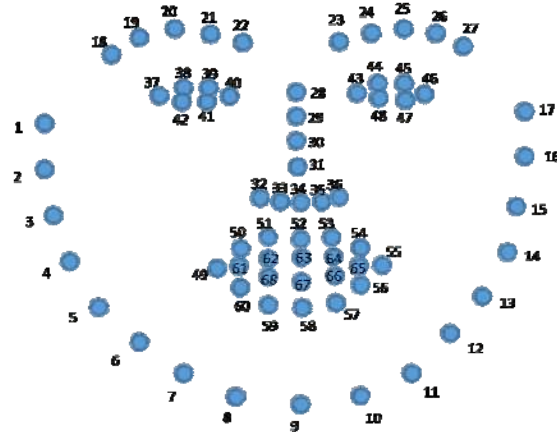


**Figure 1: Facial landmarks**

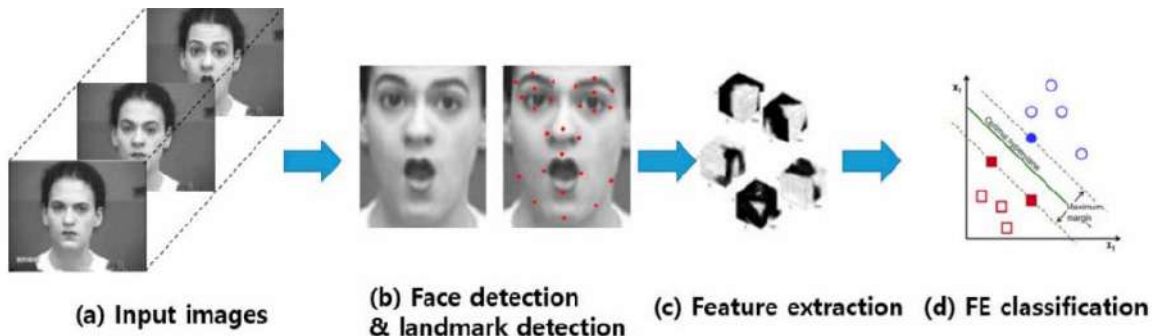


**Figure 2: Facial landmarks method**

Based on the review of this algorithm, a model is required to find the coordinates of key points. This complicates the algorithm by searching for a large number of features. The position of the images and faces should be aligned with ideal lighting and no noise. For these points need to calculate descriptors, otherwise this will lead to the loss of useful information. To date, the effectiveness of this method is an average of 72.65%. There are a large number of algorithms that make it possible to obtain key points with an accuracy sufficient for further classification of human emotions by these points [13]. But, to use this approach, it is necessary that the position of the face in the image be aligned, and in real conditions, this is almost unrealistic. This method is deprecated. Based on this, convolutional neural networks are the best algorithm for analyzing visual data.

## 4. Convolutional Neural Networks

Since the method of key points has a number of disadvantages, a replacement was proposed. This is using convolutional neural networks. This is a deep learning algorithm that is a good approach for image classification, based on the input image, importance is assigned to each feature using weights and biases. Image preprocessing requires fewer weights, which means it requires less processing power compared to other machine learning algorithms. For example, in fully connected neural

networks, filters are created manually for sufficient training, and a convolutional neural network is able to identify and learn these filters and features on its own.

This architecture was invented as a result of studying the work of the brain of cats. It turned out that neurons respond to stimuli / lines only in a limited field of view and from these lines they collect a complete picture. These lines are the signs, each sign has a different importance. That is, the convolutional network generalizes all features. Each filter has different features. The more filters there are, the deeper the signs.

The main advantage of such networks over fully connected neural networks is the decrease in the number of trained neurons, and as a result, the acceleration of network learning and a decrease in the required amount of training data [17]. Due to a large number of abstract layers, they provide partial resistance to changes in scale, displacement, rotation, change of perspective, and other distortions [18]. Convenient parallelization of computations, and, consequently, the possibility of implementing algorithms for working and training the network on GPUs.

## 4.1 Facial Expressions Databases

To start developing a neural network, we will need a sufficient amount of labeled data for training, including 6 classes of basic emotions. Large companies have the resources and the ability to independently collect and tag data. In this paper, only publicly available databases are considered, with already marked up images and representative data.

The Extended CohnKanade (CK+) [19] is the most widely used database for evaluating emotions. Contains about 600 videos and 5876 tagged images of 123 people. The sequences have durations from 10 to 60 frames, these frames show the transition from neutral emotions to the most pronounced ones. The videos are tagged with Ekman's seven main emotions: anger, contempt, disgust, fear, happiness, surprise, sadness. Images are mostly single-channel with the same background and size 640 \* 490 pixels. But CK + does not provide a dataset for training, validating, and testing algorithms. Because of this, it is not possible to use this dataset in work.

EMOTIC or EMOTIon [20] this unique dataset contains images of people in their natural environment and is tagged with emotions that are intuitively identified. Most of the images are collected by hand from the internet and search queries. The database contains about 30,000 images of different people, emotions are divided into 2 types, 26 discrete categories. In the three-dimensional coordinate system of emotions, valence, arousal, and dominance. This project was created in order for machines to understand what a person is experiencing based on a coordinate system. Machines with these capabilities interact better with humans.

The FER-2013 [21] This is a classic database that was created for Kaggle competitions, and after the competition, the data was shared. There are 35,000 images in the dataset. The data is divided into training, private test and public test data. Images are labeled with a tag of 7 seven basic emotions: happiness, fear, anger, surprise, disgust, neutral, sadness. The images have already been converted to an array of numbers and placed in a .csv file. Based on this, the obvious advantage is the use of a small amount of memory. Also, this dataset can be freely downloaded to google drive. This will save space. Unfortunately, the size of the images is only 48 \* 48 pixels, but all images were automatically registered, the face is in the center and occupies the same space in each image.

## 4.2 Preparing data

First, we need to look at the data structure using the Pandas library. There are 3 columns in fer2013.csv (Fig. 3). In the first, the number of the emotion, in the second, a set of an array of image pixels, and the third column indicates which data type the image belongs to. Images are NumPy 28x28 arrays with pixel values ranging from 0 to 255.

|   | emotion | pixels                                            | Usage    |
|---|---------|---------------------------------------------------|----------|
| 0 | 0       | 70 80 82 72 58 58 60 63 54 58 60 48 89 115 121... | Training |
| 1 | 0       | 151 150 147 155 148 133 111 140 170 174 182 15... | Training |
| 2 | 2       | 231 212 156 164 174 138 161 173 182 200 106 38... | Training |
| 3 | 4       | 24 32 36 30 32 23 19 20 30 41 21 22 32 34 21 1... | Training |
| 4 | 6       | 4 0 0 0 0 0 0 0 0 0 0 3 15 23 28 48 50 58 84...   | Training |

**Figure 3:** Structure of FER2013

Labels are an array of integers ranging from 0 to 6. They correspond to the class of emotion that the image represents. There are three types of sampling in total: training, test, and validation. The training subsampling contains the standard, the input data along with the correct / expect result. The model does not see the test dataset during training, and therefore, on this dataset, we can evaluate the quality of the model. Validation subsampling is needed to select the hyperparameters and select the best model.

Divide the data into 3 subsamples. There are 3 classes of the “Usage” column: “Training” contains 28709 images, “PublicTest” contains 3589 images, “PrivateTest” images. Data from the “Training” column can be used to train the model, data from “PrivateTest” for model validation, and “PublicTest” for test images. The data needs to be transformed so that it has the following structure: a feature vector  $X_i = [x_1, \dots, x_n]$  and a label (target variable)  $Y_i$  [Fig. 4]. The graph shows the number of pictures belonging to each class, on the train and validation subsample.

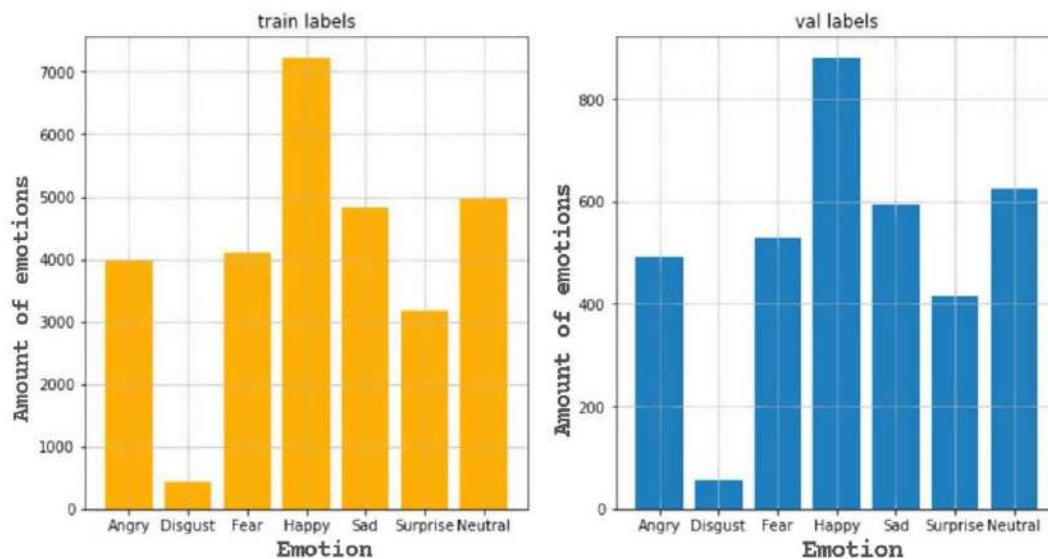


**Figure 4:** Distribution of data between classes

Since the pixel values fall in the range from 0 to 255, these values need to be scaled. To do this, divide the image array by 255. It is important that the training set and the test set are pre-processed in the same way. Also, the data needs to be converted to the float32 data type. Next, the data must be transformed into a size of 48 \* 48 and the number of channels must be 1. In color images the number of channels is 3, and in black and white 1. The array of labels is converted into a one-dimensional vector. The method `to_categorical()`, converts an array or vector with integers with different categories to a new array with binary values and columns equal to the number of categories in the data. Below you can see the data converted from array to image and also normalized.

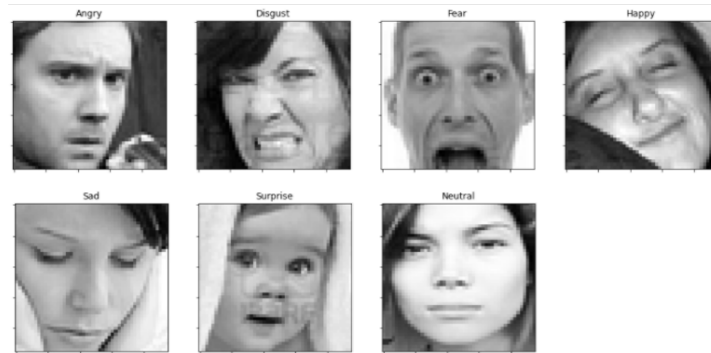


**Figure 5:** Normalized data

### 4.3 Neural Network Architecture

In total, the neural network will have 3 convolutional layers, 3 max-pooling layers, a Flatten layer, and 2 fully connected layers (Fig. 7) [22]. The first layer will be convolutional. It is he who is responsible for the convolution operation and the selection of convolution kernels. There are one-dimensional and two-dimensional analogs. On the first convolutional layer, the number of different convolution kernels to be applied will be 32. At the same time, this is the number of feature maps that we get on the next layer after convolution. Next comes the size of the convolution kernel. The most commonly used sizes are  $3 \times 3$ ,  $5 \times 5$ , and  $7 \times 7$ , but there may be others, including non-square ones. This architecture uses  $3 \times 3$ . The size of the input data is determined from the image size, that is,  $48 \times 48 \times 1$ . Used only in the first convolutional layer. Further, the input dimension is determined automatically from the output dimension of the previous layer. In the case of a convolutional layer, the dimension necessarily contains three parameters: the number of rows of pixels in the image matrix, the number of columns, the number of color channels (in the case of a black-and-white image, this is 1, and not the absence of channels). Next comes padding. Padding can have two values: "valid" and "same". The default is "valid", which minimizes the image. If you want the image not to shrink after the convolution operation, then set the "same" parameter, which, by adding zero rows and columns around the image, allows you to keep the image size the same. The indent can be filled with units and mirrored [23].

Next comes the layer that normalizes the input data – batchnormalization. When the weights of a layer change, the distribution of its outputs changes, so the next layer will have to start over. This interferes with learning. This problem is called intrinsic variable shifting. The solution is that we normalize each input separately, but not for all images in the dataset, but for mini batches. Next, you need to enter new parameters: shift and stretch. These parameters are also trained by gradient descent. [24]. Relu [25] is used as the activation function. An activation function is a function that accepts the result of an adder as input and performs some kind of transformation to turn the sum of the weighted inputs into an adequate output that can be interpreted in terms of solving the problem. Since in the case of a small value of the summation function, the activation function can return 0, that is, "nothing", then it is a kind of analog of the mechanism in a real neuron that is responsible for the excitation of the neuron (its "activation"). Relu returns  $x$  if  $x$  is positive, 0 otherwise. It is a good approximator - any function can be approximated by a combination of ReLU.

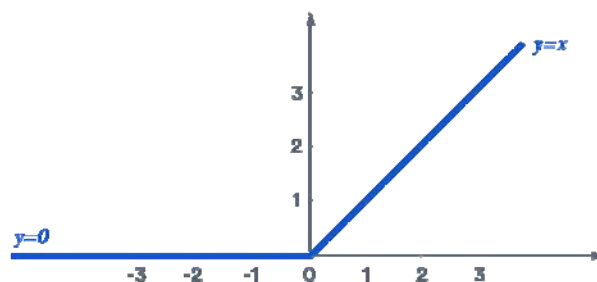


**Figure 6:** Relu

It also has the advantage that this function creates sparsity when neurons are activated. That is, unlike the sigmoid and the hyperbolic tangent, not all neurons will be activated, which will reduce the computational complexity. This property stems from the fact that all negative values vanish. Using ReLU significantly improves the convergence rate of gradient descent compared to sigmoid and hyperbolic tangent. The next is the pooling layer. The maxpooling layer is  $2 \times 2$ , other values are possible, most often this value is used. Typically, a convolutional layer followed by a pooling layer is called a convolutional block. The neural network architecture is visualized using the Tensorboard tool. Figure 7 shows the location of each layer based on the above architecture.

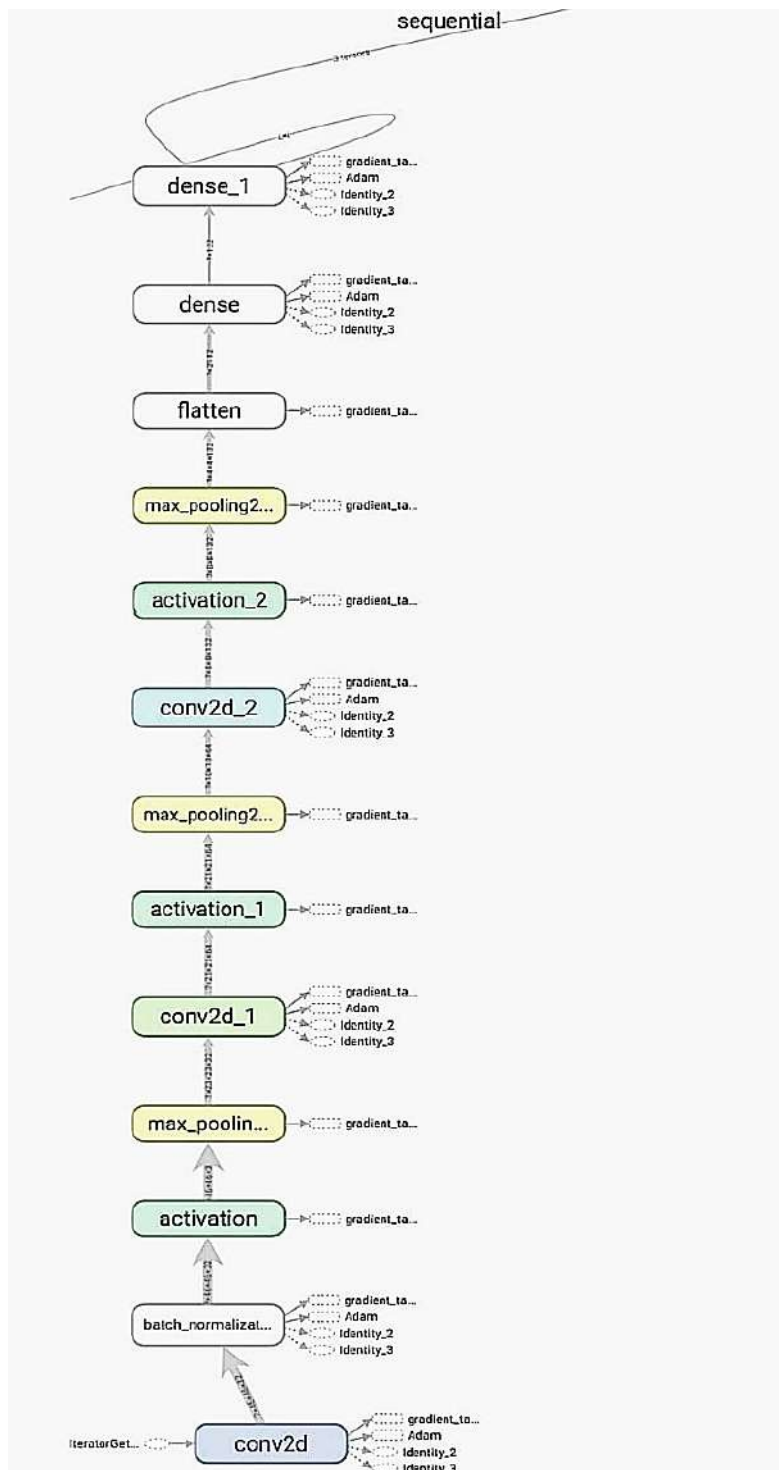


**Figure 7:** Neural network architecture

After the convolutional blocks come to the Flatten layer. It does not have any parameters, but simply converts multidimensional data into a flat vector, which can already be given as input to a fully connected layer. Convolution blocks will not give us the probability of assigning an image to any class. The fully connected layer is responsible for this. In total, this architecture has 2 fully connected layers. In the first fully connected layer 132 inputs to the neuron. In the second, there are 7 inputs to the neuron, according to the number of classes that need to be predicted. On the last layer there is a softmax function [25]. Since there are 7 inputs to the neuron on the last layer, this is the number of classes. Each neuron should give the probability of how much the object belongs to the desired class, a value from 0 to 1. All neurons should add up to one. Softmax function (1):

$$\sigma(z)_i = \frac{e^{z_i}}{\sum_{k=1}^K e^{z_k}} \quad (1)$$

## 5. Model configuration

After creating the model, it needs to be configured with losses and metrics. The categorical cross-entropy [26] was chosen as the loss function. Categorical crossentropy (2) this loss function is commonly used to classify 3 or more classes. The function simply calculates the difference between the two probability distributions and quantifies it. The categorical crossentropy loss function calculates the loss of an example:

$$Loss = \sum_{i=1}^{output\ size} y_i \times \log \hat{y}_i \quad (2)$$

where  $\hat{y}_i$  is the i-th scalar value in the model output,  $y_i$  is the corresponding target value, and output size is the number of scalar values in the model output. This function is good at showing how two vectors differ from each other.

Adam [25] with learning rate  $lr = 1e-3$  is used as an optimizer. This is a good default algorithm, good convergence. Adam combines the benefits of RMSProp, AdaGrad and SGD + Momentum. We first estimate the first pulse and the weighted sum of the gradients, and then we estimate the second pulse and the square of the gradients. The first impulse plays the role of speed, and the second serves to optimize the parameters. Initially, both parameters are equal to zero, so a correction offset is added to them to avoid too large a step at the very beginning.

Accuracy (3) [25] is used as a quality metric. Accuracy is one metric for evaluating classification models. Informally, accuracy is the fraction of predictions our model got right. Formally, accuracy has the following definition:

$$accuracy = \frac{number\ of\ correct\ predictions}{total\ number\ of\ predictions} \quad (3)$$

## 6. Training and testing

A training dataset was used to train the neural network, and a validation dataset was used for validation. Batch size = 64, epoch = 20. The model works ineffectively if it is given as input and trained from images at once. Therefore, the data was divided into 64. In order for the network to learn, many iterations of 64 must go through. The larger the batch sample size, the more memory is required from the computer. Below are the training schedules. They show that 20 epochs, even 12 epochs can be taken, is quite enough. A little more and there would be overfitting (Fig. 8-9). Figure 8 shows the accuracy at each epoch of the training and validation data. Figure 9 shows the change in the growth of the loss function on the data.

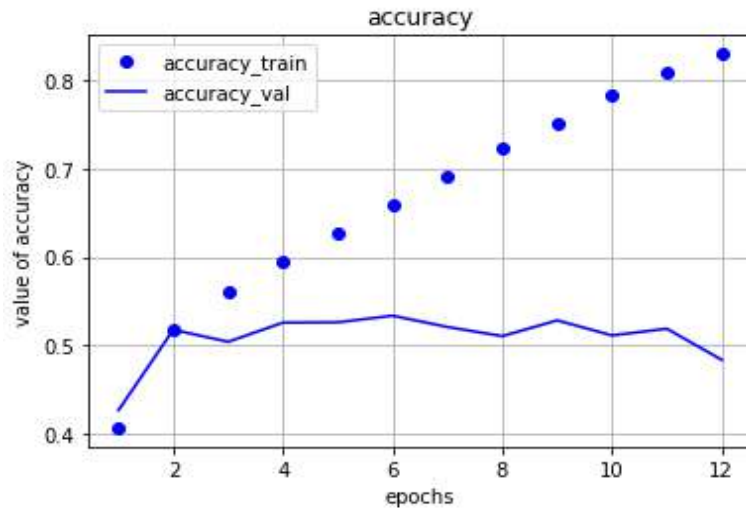


**Figure 8: Accuracy**

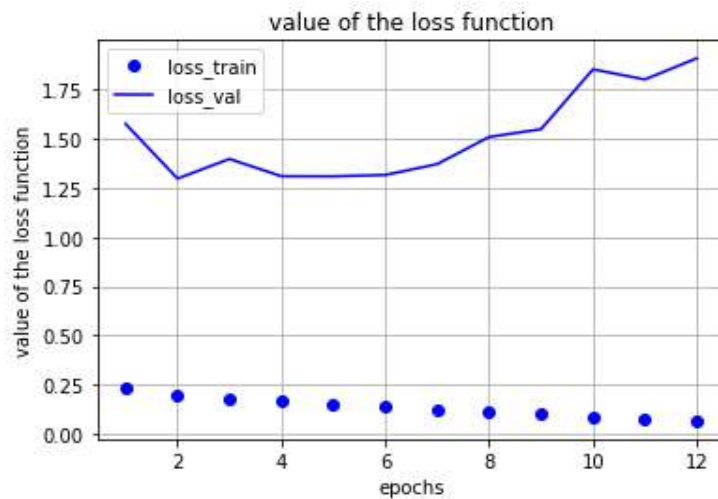


**Figure 9: Loss**

Figures 10 - 16 show the testing stage. On the 10th figure, the neural network recognized the emotion "happiness" with a probability of 1.0 on the test sample. Figure 11 shows that the neural network recognized the "surprise" emotion with a 1.0 probability. On the 12th figure, the neural network recognized the "anger" emotion with a probability of 0.78 and the "fear" emotion with a probability of 0.22. This is because emotions are similar and subjective and can be similar to each other. Figure 13 shows a similar situation, the neural network recognized 2 emotions, "fear" with a probability of 0.78 and "sad" with a probability of 0.21. In figure 14, the "neutral" emotion was recognized with a probability of 0.76, the "sad" emotion with a probability of 0.22, and the "angry" emotion with a low probability of 0.05. In Figure 15, the convolutional neural network recognized the "sad" emotion with a probability of 0.78 and the "angry" emotion with a probability of 0.18, and a low probability of other emotions. In Figure 16, the neural network recognized the "angry" emotion with a probability of 0.82. Only in figure 16 did the predictions turn out to be wrong. In the remaining figures, the neural network correctly predicted all emotions. Above is a label for each emotion. The model does not always make accurate predictions, it is especially difficult to recognize the disgust emotion, and there were few examples of this emotion in the dataset.

Based on other articles, such as [11], the accuracy on the JAFFE dataset ranged from 0.249 to 0.502. In this article, the accuracy is higher even with the minimal resources of the graphics processing unit (GPU Tesla k80). In the article [27] the accuracy on the training data reached 0.96, which is 0.03 less than what was obtained in this article.



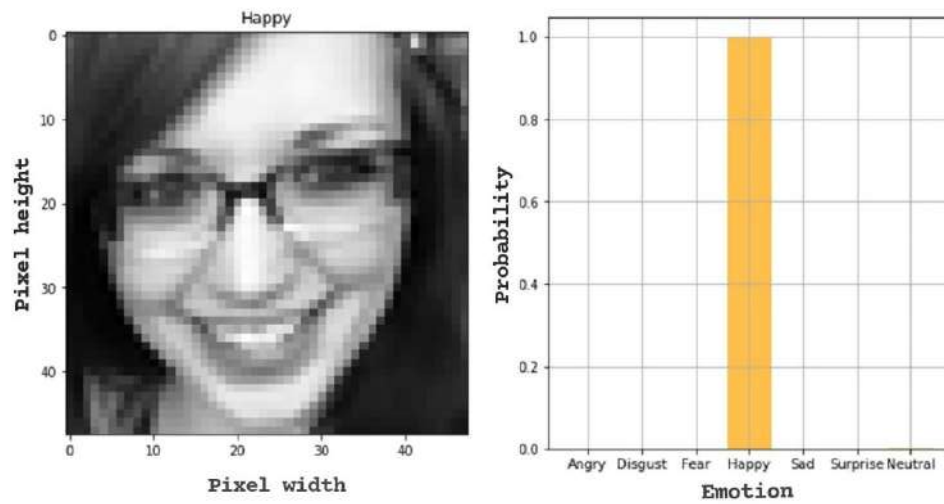


Figure 10: Testing emotion "happy"

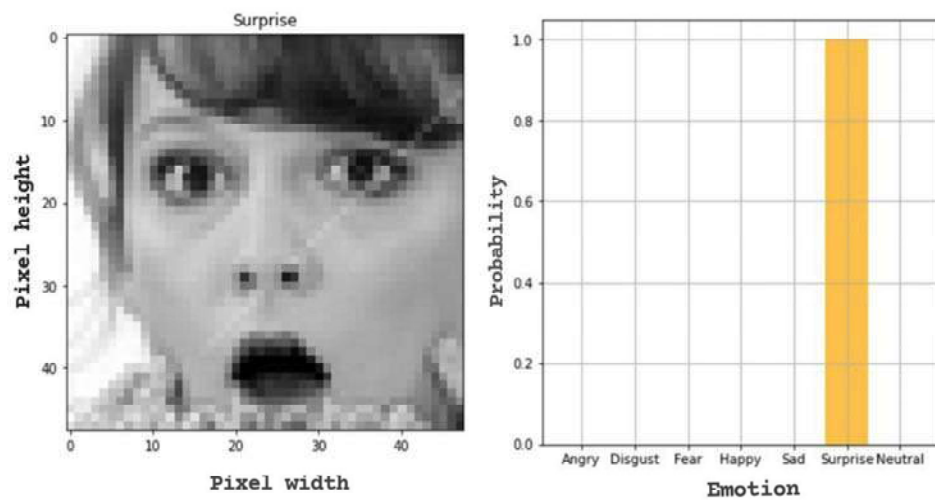


Figure 11: Testing emotion "surprise"

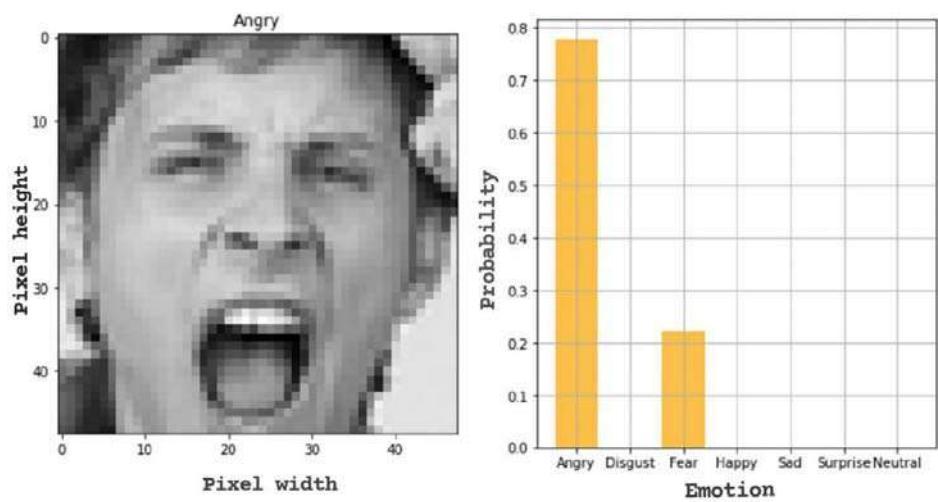


Figure 12: Testing emotion "angry"



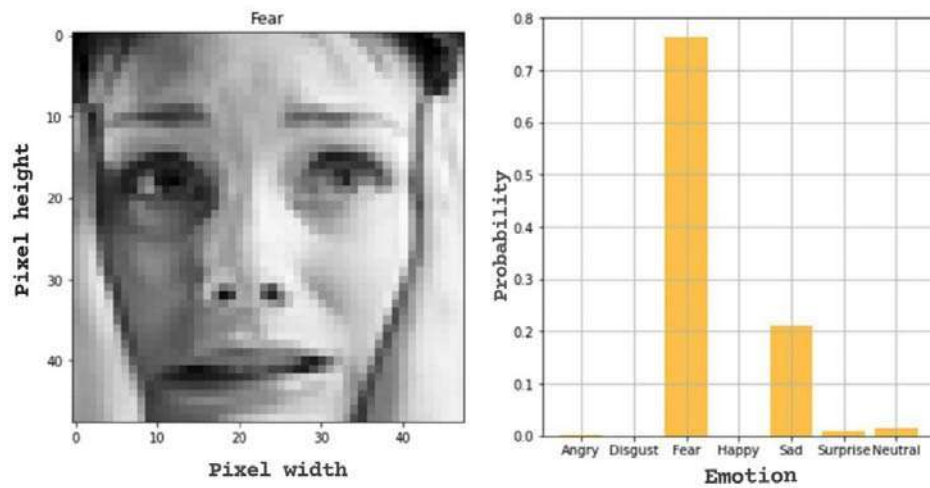


Figure 13: Testing emotion "fear"

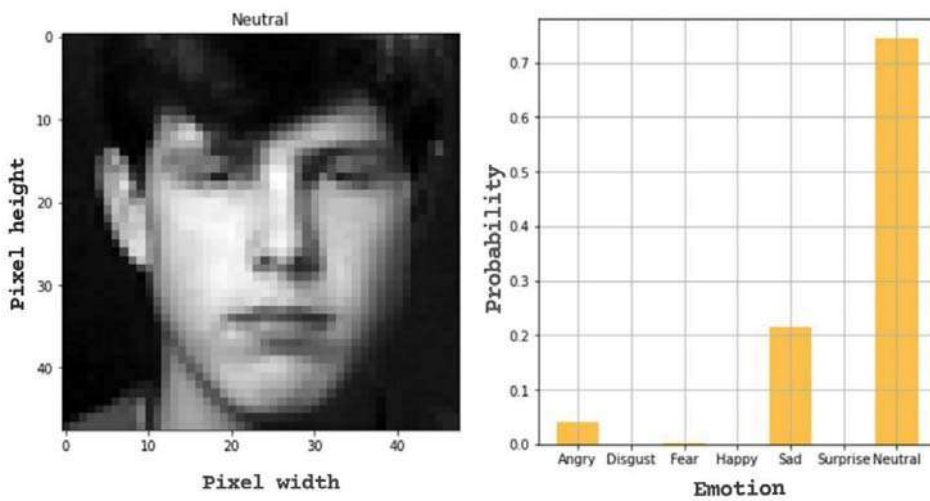


Figure 14: Testing emotion "neutral"

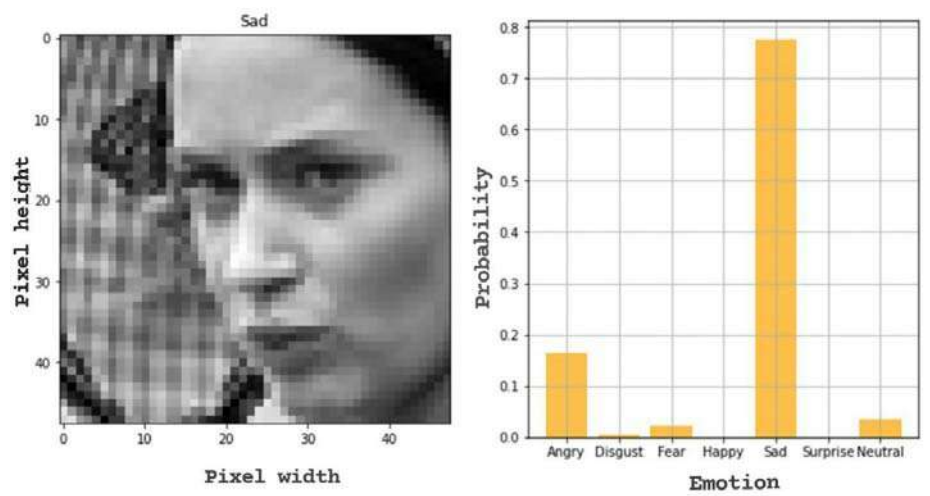


Figure 15: Testing emotion "sad"

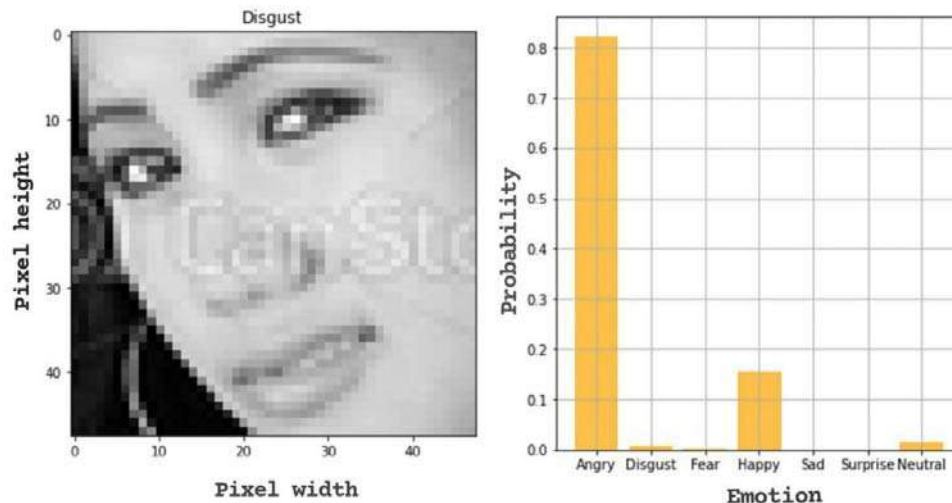


**Figure 16:** Testing emotion “disgust”

## 7. Conclusion

The article implements a convolutional neural network capable of recognizing human emotions. This is a solution to a technically and mathematically complex problem, requiring in-depth study of the subject area, own analysis, skills and abilities in the processing and selection of experimental data, deep knowledge in the field of discrete mathematics, geometry, programming, as well as psychology and many other areas of science. In this work, a convolutional neural network is implemented that is capable of detecting emotion in low light conditions, with objects on the face (glasses) and at an angle of inclination. This is what distinguishes this algorithm from the previous ones. For this, was found a dataset suitable for this task, all data from the dataset was normalized, a model was built and trained, metrics with optimizers and a loss function were selected. The model is also tested on test data, and the model parameters are selected on the validation data.

The scientific novelty consists in the fact that this algorithm has a high accuracy, learns quickly and does not require large computing resources. Also, the model can be saved and integrated into any system, such as a video camera. This can be used for commercial purposes. Training this network proved to be optimal for the available machine resources, and at the same time it trained 10 times faster than on a central processor unit. The convolutional neural network model is also capable of identifying features and reducing unnecessary features. Also, a convolutional neural network is able to adapt to image resizing, scaling, rotations, unlike other algorithms.

The practical value lies in the fact that the algorithm is implemented in the form of software that allows you to detect emotions in people in low light conditions, with a small number of computing resources. This algorithm detects a person's emotion every second with a very low probability of error.

## References

- [1] C. Darwin, P. Prodger, The expression of the emotions in man and animals, 3rd ed, Oxford University Press, 1998.
- [2] Y. I. Tian, T. Kanade, J. F. Cohn, Recognizing action units for facial expression analysis, IEEE Transactions on pattern analysis and machine intelligence 23.2 (2001) 97-115. doi:10.1109/34.908962.
- [3] P. Ekman, W. V. Friesen, Constants across cultures in the face and emotion, Journal of personality and social psychology 17 (1971) 124–129. doi:10.1037/h0030377.
- [4] A. Oliynyk, S. Skrupsky, S. A. Subbotin, Parallel computer system resource planning for synthesis of neuro-fuzzy networks, in: R. Szewczyk, M. Kaliczyńska (Eds.), Recent Advances in Systems, Control and Information Technology, volume 543 of Advances in Intelligent Systems

- and Computing, Springer, Cham, 2016, pp. 88-96. doi:10.1007/978-3-319-48923-0\_12.
- [5] G. Kirichek, S. Skrupsky, M. Tiahunova, A. Timenko, Implementation of web system optimization method, in: Proceedings of the Third International Workshop on Computer Modeling and Intelligent Systems, CMIS '2020, CEUR Workshop Proceedings 2608, Zaporizhzhia, Ukraine, 2020, pp. 199–210.
  - [6] S. Subbotin, A. Oliinyk, V. Levashenko, E. Zaitseva, Diagnostic rule mining based on artificial immune system for a case of uneven distribution of classes in sample, Communications - Scientific Letters of the University of Zilina 18.3 (2016) 3–11. URL: <http://komunikacie.uniza.sk/index.php/communications/article/view/301>.
  - [7] A.O. Oliinyk, T.A. Zaiko, S.A. Subbotin, Factor analysis of transaction data bases, Automatic Control and Computer Sciences 48.2 (2014) 87-96. doi:10.3103/S0146411614020060.
  - [8] S. Li, W. Deng, Deep Facial Expression Recognition: A Survey, IEEE Transactions on Affective Computing (2020) 1-1. doi:10.1109/TAFFC.2020.2981446.
  - [9] B. C. Ko, A brief review of facial emotion recognition based on visual information, Sensors 18.2 (2018) 401. doi:10.3390/s18020401.
  - [10] W. J. Baddar, J. Son, D. H. Kim, S. T. Kim, Y. M. Ro, A deep facial landmarks detection with facial contour and facial components constraint, in: 2016 IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, USA, 2016, pp. 1–5. doi:10.1109/ICIP.2016.7532952.
  - [11] I. Tautkute, T. Trzcinski, A. Bielski, I know how you feel: Emotion recognition with facial landmarks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 2018, pp. 1878-1880. doi:10.1109/cvprw.2018.00246.
  - [12] Kostinger, M., Wohlhart, P., Roth, P.M., Bischof, H.: Annotated facial landmarks in the wild: A large-scale, real-world database for facial landmark localization, in: 2011 IEEE International Conference on Computer Vision Workshops (ICCVW), Barcelona, Spain, 2011, pp. 2144–2151, doi:10.1109/ICCVW.2011.6130513.
  - [13] Y. Sun, X. Wang, X. Tang, Deep learning face representation by joint identification-verification, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), arXiv preprint arXiv:1406.4773, 2014, pp. 1891-1898.
  - [14] A. Krizhevsky, I. Sutskever, G. E. Hinton, Imagenet classification with deep convolutional neural networks, Advances in neural information processing systems 25 (2012) 1097–1105. URL: <https://kr.nvidia.com/content/tesla/pdf/machine-learning/imagenet-classification-with-deep-convolutional-nn.pdf>.
  - [15] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, arXiv preprint arXiv:1409.1556, 2014.
  - [16] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, in: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 2015, pp. 1–9. doi:10.1109/cvpr.2015.7298594.
  - [17] B. Hasani, M. H. Mahoor, Spatio-temporal facial expression recognition using convolutional neural networks and conditional random fields, in: 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition, FG '2017, IEEE, Washington, DC, 2017, pp. 790–795. doi:10.1109/fg.2017.99.
  - [18] G. Kirichek, V. Harkusha, A. Timenko, N. Kulykovska, System for detecting network anomalies using a hybrid of an uncontrolled and controlled neural network, in: Proceedings of the Computer Science Software Engineering: Proceedings of the 2nd Student Workshop, CSSE@SW '2019, CEUR Workshop Proceedings 2546, Kryvyi Rih, Ukraine, 2019, pp. 138–148.
  - [19] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, I. Matthews, The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression, in: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition – Workshops, IEEE, San Francisco, CA, USA, 2010, pp. 94–101. doi:10.1109/cvprw.2010.5543262.
  - [20] C. F. Benitez-Quiroz, R. Srinivasan, A. M. Martinez, Emotionet: An accurate, real-time algorithm for the automatic annotation of a million facial expressions in the wild, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016.

doi:10.1109/cvpr.2016.600.

- [21] I. J. Goodfellow, D. Erhan, P. L. Carrier, A. Courville, M. Mirza, B. Hamner, W. Cukierski, Y. Tang, D. Thaler, D.-H. Lee et al., Challenges in representation learning: A report on three machine learning contests, in: International Conference on Neural Information Processing. Springer, Berlin, Heidelberg, 2013, pp. 117–124. doi:10.1007/978-3-642-42051-1\_16.
- [22] Z. Meng, P. Liu, J. Cai, S. Han, Y. Tong, Identity-aware convolutional neural network for facial expression recognition, in: 12th IEEE International Conference on Automatic Face & Gesture Recognition, FG '2017, IEEE, Washington, DC, USA, 2017, pp. 558–565. doi:10.1109/fg.2017.140.
- [23] G. Levi, T. Hassner, Emotion recognition in the wild via convolutional neural networks and mapped binary patterns, in: Proceedings of the 2015 ACM on international conference on multimodal interaction, 2015, pp. 503–510. doi:10.1145/2818346.2830587.
- [24] A. T. Lopes, E. de Aguiar, A. F. De Souza, T. Oliveira-Santos, Facial expression recognition with convolutional neural networks: coping with few data and the training sample order, Pattern Recognition 61 (2017) 610–628. doi:10.1016/j.patcog.2016.07.026.
- [25] S.O. Subbotin Neironni merezhi teoriia ta praktyka (Neural networks: theory and practice), 2020.
- [26] Peltarion.com, Categorical crossentropy, 2021. URL: <https://peltarion.com/knowledge-center/documentation/modeling-view/build-an-ai-model/loss-functions/categorical-crossentropy>.
- [27] M. A. Ozdemir, B. Elagoz, A. Alaybeyoglu, R. Sadighzadeh, A. Akan, Real time emotion recognition from facial expressions using CNN architecture, in: 2019 Medical Technologies Congress (TIPTEKNO), IEEE, Izmir, Turkey, 2019, pp. 529–532. doi:10.1109/TIPTEKNO.2019.8895215.

# Object detection algorithm for high resolution images based on convolutional neural network and multiscale processing

Rykhard Bohush<sup>a</sup>, Sergey Ablameyko<sup>b,c</sup>, Sviatlana Ihnatsyeva<sup>a</sup> and Yahor Adamovskiy<sup>a</sup>

<sup>a</sup> Polotsk State University, 29 Blokhin st., Novopolotsk, 211440, Belarus

<sup>b</sup> Belarusian State University, 4 Nezavisimosti Avenue, Minsk, 220030, Belarus

<sup>c</sup> United Institute for Informatics Problems, National Academy of Sciences of Belarus, 6 Surganova st., Minsk, 220012, Belarus

## Abstract

In this article we propose an effective algorithm for small object detection in high resolution images. We look at the image at different scales and use block processing by convolutional neural network. Pyramid layers number is defined by input image resolution and convolutional layer size. On each layer of pyramid except the highest we perform splitting overlapping blocks to improve small object detection accuracy. Detected areas are merged into one if they belong to the same class and have high overlapping value. In the paper experimental results using YOLOv4 for 4K and 8K images are presented. Our algorithm shows better detecting small objects results in high-definition video than YOLOv4.

## Keywords 1

Small objects detection, Image pyramid, Convolutional neural networks, YOLO.

## 1. Introduction

Video surveillance systems are used to solve various problems such as identification of pedestrians, vehicles, car numbers detection, in security system and other. Now day 4K and 8K camcorders are able of recording high-definition video, which can identify object details to improve and increase the possibility of detection tiny objects [1]. These notions refer to any digital images containing resolution of approximately 4000 or 8000 pixels [2]. The advantage of using high resolution systems is the ability not only to accurately small objects detect but also to display the correct large objects shape. It allows to improve detection accuracy. However, there is also a negative sides, the main one of which is a large information capacity in each frame or image. This leads to the need to use fast and efficient image processing algorithms to process high resolution images and video in real time.

Convolutional neural networks (CNNs) are very effective for object detection and became widespread due to the rapid growth in computing power. Convolutional layers with different filter types allow to create effective feature sets to describe images that have unclear structure and many variations within a class. However, GPU computing power and memory size is still insufficient for processing high-resolution images. Therefore, when using a CNN, input image is scaled to the input layer size at the first stage. After that, input image is less informative for further processing. It is important to take into account that image resolution increasing reduces percentage of object size if it is constant in pixels. At present, most known systems do not consider input image resolution. Such systems scale to the input layer size and if image or video frame with high-resolution, tiny objects can

---

CMIS-2021: The Fourth International Workshop on Computer Modeling and Intelligent Systems, April 27, 2021, Zaporizhzhia, Ukraine  
EMAIL: r.bogush@psu.by (R. Bohush); ablameyko@bsu.by (S. Ablameyko); s.ignatieva@psu.by (S. Ihnatsyeva); e.adamovsky@psu.by (Y. Adamovskiy)

ORCID: 0000-0002-6609-5810 (R. Bohush); 0000-0001-9404-1206 (S. Ablameyko); 0000-0002-9780-5731 (S. Ihnatsyeva); 0000-0003-1044-8741 (Y. Adamovskiy)



© 2020 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

be missed. Consequently, the algorithms development for detecting small objects in high-resolution images is an urgent task.

This article is structured as follows. In Section 2, related algorithms based on CNNs for object detection in images are illustrated. In Section 3 our algorithm by multiscale and block processing is presented. In Section 4 experimental results using database of 4K images and discussion are shown. Finally, the conclusion and future work details are provided in Section 5.

## 2. Related Works

Generally, most of the image object detection algorithms for high definition images are carried out using two principals: image dividing into blocks and CNNs application for each block. In [3], the algorithm for the accurate of small objects detection in high-definition video are presented. Each image is separated into overlapping blocks and object detection in each block is performed with CNN YOLO. After that, the detected objects post-processing is performed in each subframe and the same class probabilities adjacent areas are merge. Compared to the classic YOLO, the algorithm shows the better results when detecting tiny objects in high-definition video. But the main disadvantage of this algorithm is sufficiently high probability of combining several different objects that are located at the processed blocks boundaries into one region and skipping the object due to its fragmentation into blocks.

Yongxi and Javidi [4] use neighbor visual features and high-level image features to predict bounding boxes to guide a two-stage search process that adaptively focuses on areas that may contain tiny objects. Objects are detected on a reduced image copy using the AlexNet model. The original image was divided into non-overlapping blocks to detect small objects. Sizes of such blocks varied from 600 to 75 pixels and blocks were processed using the SCN (Spatial Correlation Network). The overlap varied from 25 to 50% relative to the block sizes. For this algorithm, authors obtained a recall value of 80% based on SUN2012 dataset [5]. It is known that the recall metric doesn't allow evaluating false positives, and the dataset used by the authors contained almost no images with 4K resolution.

In [6] the algorithm for detecting human uses a two-stage approach. First, the original image is scaled down to low-resolution space. Low resolution object detection guides the attention of the model to the important areas of the input image. In the second step, the algorithm is returned to high resolution only, only reviewing the selected areas. Large blocks are split into smaller blocks to refine the bounding boxes of the detected regions of interest. Found regions merging is based on the vertically elongated (rectangular) person shape. This means that the algorithm is not adapted for other object classes. For both stages model YOLO v2 is applied. Algorithm accuracy was estimated using PEVID-UHD dataset [7] and a closed dataset containing a slightly larger number of objects. Average precision is 90.7% for the first dataset and 75.4% for the second. At the same time, in some cases, the use of block overlap in the range of 20 to 50 pixels is not sufficient; in addition, it is not adaptive to input image size.

Unel et al. [8] use CNN PeleeNet for detecting small objects of the "vehicle" and "pedestrian" classes in images with 2K resolution and dimensions [1920×1080] onboard a micro aerial vehicle.

The algorithm uses block image processing of the PeleeNet model with 25% overlap between the tiles. For this algorithm mAP metric on VisDrone2018 dataset is 36.67%. The proposed technique is implemented on Nvidia Jetson TX1 and TX2. However, this paper does not provide an estimate of the block overlap effect on the algorithm accuracy. The number of missed objects of the "pedestrian" class is significantly higher than the "vehicle" class. Consequently, this algorithm detects small objects much worse even for images with a resolution below 4K.

As we can see, it is important to use CNN with high accuracy but with low computational costs. CNN YOLO favorably differentiate by other models because it processes entire image without previous RoI extraction. It means that CNN considers contextual information about whole image and it's important. YOLO is a one-stage object detector. In addition, as stated in [9] YOLO is the fast object detection system that performs better than famous Faster R-CNN. This is of great importance because computational costs is a significant parameter when processing high-resolution image and video frames. YOLO is more effective by computational cost. On the downside, there are many

localization errors particularly for tiny objects, due to the input image scaling. At paper [10], considers YOLO modifications: YOLO v2, YOLO9000. Improvement in segmentation and classification has been achieved through the use of new Darknet-19 classification model with anchor boxes; batch normalization from [10]; using k-means; direct location prediction; RPN usage; fine-grained features; multi-scale training. Top-5 accuracy was 91.2% but proposed algorithm took a big computational cost. Also multi-scale training gave better opportunity to detect small objects, but it is not enough for 4K resolution systems. YOLOv2 [10] uses five anchor-boxes, CNN Darknet-19, and Batch Normalization for predictions on each convolutional layer, which increased the detection accuracy. The next version as YOLOv3 [11] has further improved its detection accuracy indicators. This was achieved using CNN darknet-53, residual connection, multi-scale object detection through the FPN (Feature Pyramid Network) use. YOLOv3 uses  $1 \times 1$  detection kernels on feature maps for object detection in three different network layers for three different sizes. This improves detection accuracy and detects smaller objects. However, YOLOv3 loses out in terms of speed compared to YOLOv2. A solution to this problem is proposed in YOLOv4.

YOLOv4 represents state-of-the-art architecture, with one of the best ratio accuracies and detection object speed [12]. The approach is based on the Darknet53 convolutional neural network that uses Cross-Stage-Partial-connections (CSP), pyramids for feature extraction, anchor-box for detection. To achieve high detection speed rates and accuracy, as well as the ability to use available training equipment, YOLOv4 uses several methods to increase efficiency without computational cost or with negligible computational complexity. These include: using genetic algorithms to optimize the learning rate, Mosaic and CutMix dataset augmentation methods, SAT (Self-Adversarial Training), modified SAM (Spatial Attention Module) and PAN (Path aggregation Network) modules, SPP (Spatial Pyramid Pooling), class label smoothing, CmBN - using Cross mini-Batch Normalization, Mish activation, Clou-loss, etc. Various optimization methods combinations allow to achieve an average precision AP of about 43%. Thus, as a model of CNN for object detection in high-resolution images, we will use YOLO v4.

### 3. Object detection algorithm based on multiscale representation and convolutional neural network

Our algorithm is a group of the following steps as showed in Figure 1. The algorithm requires a pyramidal image representation as a decreasing scale copies set. Scale decreases as we go higher on the pyramid. Images are broken into blocks at each pyramid level. There is object detection in each block with an application CNN. It is necessary to divide the image into minimal number of parts, so that the object on the CNN input block is not fragmented. Therefore, overlapping blocks are used. Moreover, the higher overlap amount, the larger object can be detected without the probability of its division into parts. However, it will require additional computational cost. After image processing using CNN, a procedure for combining the found regions at all levels is required. Thus, our algorithm includes four basic steps.

1. Determination of the number of levels in the image pyramid. It should be borne in mind that the top-level size should be close to the input layer size for used CNN. Number of levels is determined considering possible image rotation as:

$$P = \lceil \max(W, H) / l \rceil - 1, \quad (1)$$

where  $W, H$  the input image width and height;  $l \times l$  - the used CNN input layer size;  $\lceil * \rceil$  - nearest integer.

This approach eliminates objects fragmentation at a minimum scale and allows them to increase their correct classification as a whole in image.

2. The image is separated into blocks with overlapping on the  $p$  layer. Their number is determined by the expression:

$$B_p = \left( \frac{W / p - l}{k} + 1 \right) \cdot \left( \frac{H / p - l}{k} + 1 \right) = \lceil B_{W_p} \rceil \cdot \lceil B_{H_p} \rceil, \quad (2)$$

where  $p$  is the pyramid level number,  $p=1, \dots, P$ ;  $k$ -block shift amount, in pixels.



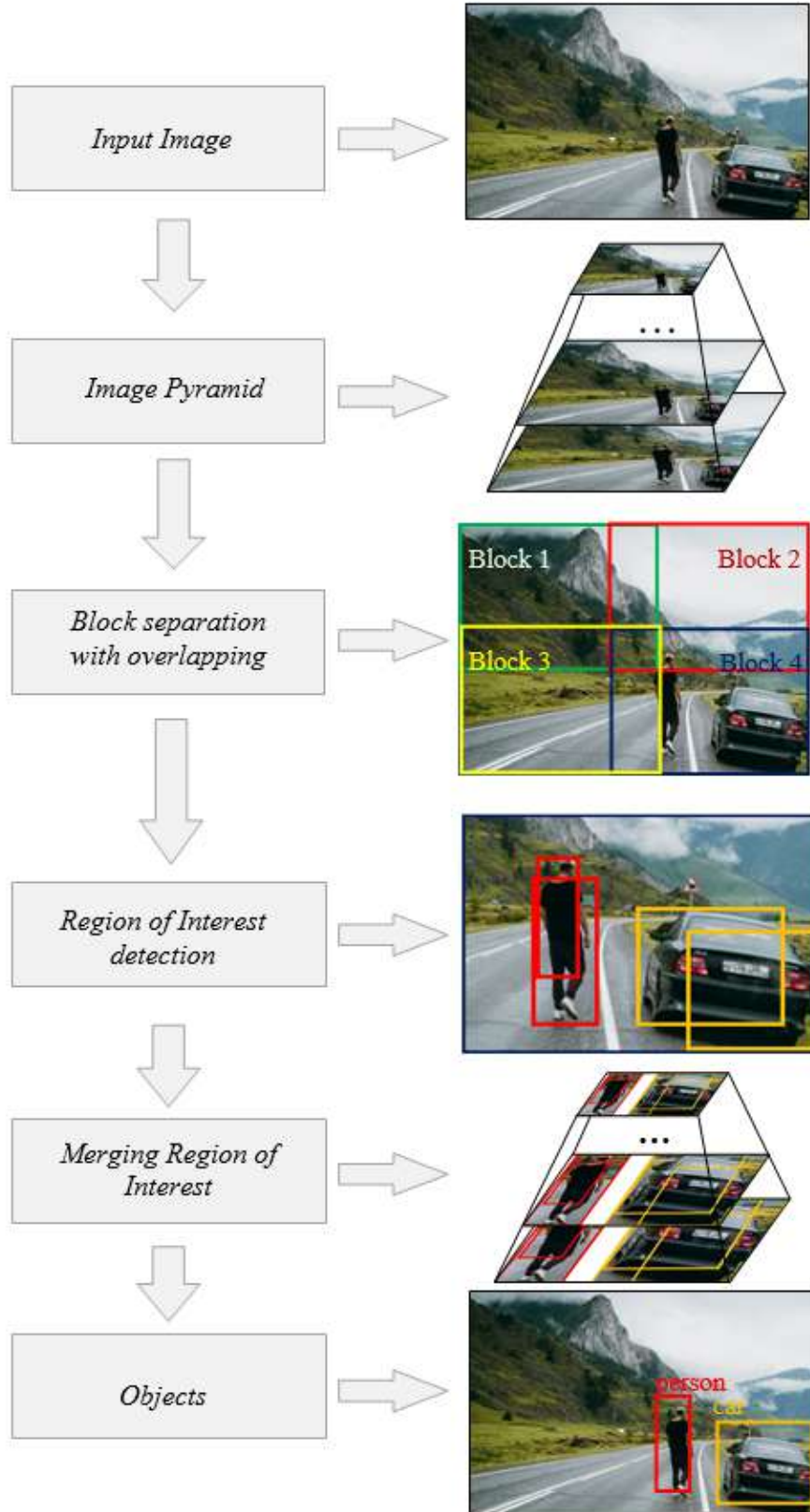


**Figure 1:** Flow chart of our proposed algorithm

The last incomplete block is merged with the previous block and scaled to the input CNN layer size when the  $B_{w_p}$  or  $B_{h_p}$  is rounded down. Otherwise, if the block size is rounded up, then the block values are padded with zeros.



The block overlap number can be calculated using the equation:

$$\alpha = \frac{k}{l} \cdot 100\%, \quad (3)$$

3. Region of interest (RoI) detection that may contain an object or the object fragment. For this, each block is processed using CNN, and the detected RoI is described by a feature set:

$$F = (x_1, y_1, x_2, y_2, E, Cl), \quad (4)$$

where  $x_1, y_1$  - upper left corner coordinates of the found region in the image;  $x_2, y_2$  - lower right corner coordinates of the found region in the image;  $Cl$  - the selected object class;  $E$  is CNN confidence in the correct classification.

4. Merging ROI.

4.1. Localization of coordinates of the object at different levels of pyramid can be determined with slight deviations. In addition, if the object is large, then its fragment can be identified as an independent ROI at some middle or lower levels. Therefore, we use an analysis of the overlaps number at all levels and blocks, as well as belonging to the same class to combine ROI  $O_i$  and  $O_j$ :

$$IoU(O_i, O_j) > 0.5, Cl_i = Cl_j, \quad (5)$$

where

$$IoU(O_i, O_j) = \frac{O_i \cap O_j}{O_i \cup O_j}, \quad i, j \in 1, N \mid i \neq j, \quad (6)$$

The merged region  $F(O')$  descriptor is formed as follows:

$$F(O') = \left( \begin{array}{l} \min(x_1^{O_i}, x_1^{O_j}), \min(y_1^{O_i}, y_1^{O_j}), \max(x_2^{O_i}, x_2^{O_j}), \\ \max(y_2^{O_i}, y_2^{O_j}), \max(E^{O_i}, E^{O_j}), Cl(\max(E^{O_i}, E^{O_j})) \end{array} \right). \quad (7)$$

4.2. The same object or its fragments can be detected with a coordinate shift at different scales or in neighboring cells during block detection results processing. In addition, the smaller size object fragment features may be different from the original object features and even be closer to another class descriptors. Therefore, post-processing is additionally applied when combining process. To do this, it is proposed to use the rule according to which all areas obtained in the previous step are analyzed and combined:

$$IOU(O'_i, O'_j) > 0.2, Cl_i = Cl_j, d_x > t_x, d_y > t_y, \quad (8)$$

where  $d_x$  and  $d_y$  are the difference values between the candidate regions sizes along the  $x$  and  $y$  axes, respectively;  $t_x$  and  $t_y$  are threshold levels.

The false objects merging probability of the same class that are located at an insignificant distance from each other will be increased if the threshold value is decreased at step 4.1.

## 4. Experimental results and discussion

The object detection algorithm was tested on a sample of 820 4K images. These images were taken under different weather conditions, with different mounting heights and camera tilt angles. The annotated image dataset was prepared using labellmg tool [13]. We used YOLOv4 implementation from [14]. The annotated images dataset includes 6125 objects of two classes ("person" and "vehicle"). The objects sizes vary from  $[24 \times 14]$  to  $[322 \times 780]$ . The "vehicle" class is a composite class because it includes the "bus", "car", and "truck" classes. In Figure 2 total marked 11 objects, of which 3 belong to the class "person", and 8 belong to the class "vehicle".

The proposed algorithm is implemented in the Python programming language using the PyTorch machine learning framework for experiments. Basic operations on images are implemented using the OpenCV computer vision library. To increase the detection speed, batch image blocks processing for each pyramid level is applied using CUDA technology.

We calculate mAP (mean Average Precision) metric to assess the performance of the algorithm. Metric mAP is calculated by taking the mean AP over all classes. At the same time, at the first stage, the number of true positive  $T_p$  and false positive  $F_p$  detections were calculated. For this, intersection over union (IoU) metric was applied for the detected and annotated objects on the images of the dataset with a threshold level  $T = 50\%$  for making a decision. Then the precision  $p$  and recall  $r$  were calculated:

$$p = \frac{T_p}{T_p + F_p}, \quad (9)$$

$$r = \frac{T_p}{T_p + F_n}, \quad (10)$$

where  $F_n$  - false negative detections

Based on the obtained values, the AP metric was calculated according to the formula (9) for each class for 11 threshold levels in the range from 0 to 1 with a step (0.1) [15]:

$$AP = \frac{1}{11} \cdot \sum_{i=0}^{10} p_{\text{int}}(r_i) \quad (11)$$

where  $p_{\text{int}}(r_i) = \max p(\tilde{r}), \tilde{r} \geq r_i$ .

At the last stage, we calculate the mAP metric obtained in the previous step for two object class.



**Figure 2:** Examples of annotated images for test

To use the CNN capabilities more optimally when detecting and identifying small size objects, it is necessary to determine the CNN threshold confidence and the input layer size, when the greatest work efficiency. Experiments have shown that the smaller the scaling, the greater the accuracy in detecting small objects in high-resolution images. So, with the input layer size  $[608 \times 608]$ , mAP is 22%, and with the size  $[1024 \times 1024]$  the highest value for mAP is 48% is achieved on a test images set with 4K resolution, while the confidence threshold is  $T = 50\%$ . We used computer with CPU Intel Core i9-10900 2.8 GHz, 64 GB RAM, NVIDIA GeForce RTX 2080 Ti. Based on step one of our algorithm and more effective layer size 4K resolution image must be presented in the pyramid form of three levels for processing.

At the experiments next stage, the algorithm accuracy was evaluated for various block overlap and CNN threshold value with a step of 5%. It should be noted that for the minimum original image scale,

threshold value of  $T = 50\%$ , which corresponds to the most effective for the used CNN. The experimental results are presented in Table 1. The analysis shows that for any  $\alpha$  and  $T$  values, the considered algorithm's mAP is higher than the mAP of the CNN used. The best result is provided with  $\alpha = 50\%$  and  $T = 75\%$ , and  $mAP = 68.3\%$ .

Figure 3 shows processed 4K resolution parts of images from the prepared dataset. We can see in Figure 3a two detected objects using YOLO v4 and four detected objects using our algorithm in Figure 3b.

**Table 1**

Influence of overlap and confidence threshold on object detection accuracy, mAP, %

| T  | $\alpha$ |              |       |       |
|----|----------|--------------|-------|-------|
|    | 40       | 50           | 60    | 70    |
| 65 | 64.6%    | 68.0%        | 62.7% | 61.6% |
| 70 | 66.6%    | 67.6%        | 63.5% | 62.8% |
| 75 | 66.3%    | <b>68.3%</b> | 65.8% | 64.4% |
| 80 | 62.0%    | 62.4%        | 65.2% | 63.7% |



a)



b)

**Figure 3:** There are 4K-image parts with examples of object detection: a) using YOLOv4; b) using the proposed algorithm

Also, the proposed algorithm was tested on 4K [16] and 8K [17] images. Figure 4 shows the object detection result by the proposed algorithm with parameters which were determined as a result of experiments: image 4K with resolution  $[3840 \times 2160]$ , input layer size  $[1024 \times 1024]$ , number of pyramid levels  $P=3$ ,  $\alpha = 50\%$  and  $T = 75\%$ . In this case, 90 objects were detected, their minimum size is  $[33 \times 11]$  pixels. Small objects give the greatest gain in object detection based on presented algorithm for 4K images (Figure 4).



**Figure 4:** Examples of object detection on a 4K image based on proposed algorithm

Figure 5a shows a reduced copy for 8K image with resolution  $[7360 \times 4912]$ . YOLOv4 based image processing result is proposed in Figure 5b. On this image 27 objects were detected. The 8K image has a huge size, so object detection result is presented only for bottom right part of the image. Our algorithm detected 114 objects and their minimum size is  $[26 \times 31]$  pixels on this image. Figure 5c shows the object detection result by the proposed algorithm with parameters which were determined as a result of experiments: input layer size  $[1024 \times 1024]$ , number of pyramid levels  $P=4$ ,  $\alpha = 50\%$  and  $T = 75\%$ . Figure 5c shows that small objects give the greatest gain in object detection on image 8K resolution. We can see (Figure 5c) that the presented algorithm even detected the drivers of vehicles, who stopped in front of the pedestrian crossing.

## 5. Conclusion

This article proposes the algorithm for detecting objects in 4K and 8K images, which includes the following basic steps: forming an image pyramid until the top level dimensions are close to the input layer dimensions of the used CNN; overlapping block splitting for all resulting layers; using CNN for each block object detection; post-processing for the results obtained in the previous step. We used a YOLOv4 CNN for object detection with an input layer size of  $[1024 \times 1024]$ . It gives high efficiency in detecting small objects in 4K and 8K resolution images. The proposed algorithm was implemented using Python with PyTorch machine learning framework, computer vision library OpenCV and CUDA technology. To conduct experiments to assess the proposed algorithm effectiveness, a high resolution images dataset with annotated small and medium size of objects of two classes ("person" and "vehicle") was prepared. Experiments have shown that mAP metric can reach 68.3% for the data used. The obtained results indicate that the proposed approach is promising for solving applied problems of small object detecting in high resolution images. In some cases, if objects of one class are located nearby, then possibly our algorithm will mistakenly merge them. Therefore, our further work will be done to solve this problem.





**Figure 5:** Examples of object detection on 8K image: a) reduced copy of 8K image; b) using YOLOv4 algorithm; c) using the proposed algorithm

## 6. References

- [1] S.A. Rajjak, A. K. Kureshi, Recent Advances in Object Detection and Tracking for High Resolution Video: Overview and State-of-the-Art, in: Proceedings of the 5th International Conference On Computing, Communication, Control And Automation (ICCUBEA), Pune, India, 2019, pp.1-9. doi:10.1109/ICCUBEA47591.2019.9128812.
- [2] K. Goulekas: Visual Effects in a Digital World: A Comprehensive Glossary of over 7000 Visual Effects Terms, Morgan Kaufmann, San Francisco, CA, 2001.
- [3] D. Vorobjov, I. Zakharova, R. Bohush, S. Ablameyko, An effective object detection algorithm for high resolution video by using convolutional neural network, volume 10878 of Lecture Notes in Computer Science, Springer-Verlag, Cham, 2018, pp. 503–510. doi:10.1007/978-3-319-92537-0\_58.

- [4] L. Yongxi, T. Javidi, Efficient object detection for high resolution images, in: Proceedings of 53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton), Monticello, IL, USA, 2015, pp. 1091–1098. doi:10.1109/ALLERTON.2015.7447130.
- [5] J. Xiao, J. Hays, K. Ehinger, A. Oliva, A. Torralba, Sun database: large-scale scene recognition from abbey to zoo, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, CVPR'10, San Francisco, CA, USA, 2010, pp. 3485–3492. doi:10.1109/CVPR.2010.5539970.
- [6] V. Ruzicka, F. Franchetti, Fast and accurate object detection in high resolution 4K and 8K video using GPUs, in: Proceedings of IEEE High Performance Extreme Computing Conference (HPEC), Waltham, MA, USA, 2018, pp. 1–7. doi:10.1109/HPEC.2018.8547574.
- [7] P. Korshunov, T. Ebrahimi, UHD video dataset for evaluation of privacy, in: Proceedings of 6th International Workshop on Quality of Multimedia Experience, QoMEX'14, Singapore, 2014, pp. 232–237. doi:10.1109/QoMEX.2014.6982324.
- [8] F.O. Unel, The power of tiling for small object detection, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition Workshop (CVPRW 2019), Long Beach, CA, 2019. Doi: 10.1109/CVPRW.2019.00084.
- [9] J. Redmon, S. K. Divvala, R. B. Girshick, A. Farhadi, You only look once: unified, real-time object detection, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, CVPR'16, Las Vegas, NV, USA, 2016, pp.779–788. doi:10.1109/CVPR.2016.91.
- [10] J. Redmon, A. Farhadi, YOLO9000: Better, Faster, Stronger, in: IEEE Conference on Computer Vision and Pattern Recognition, IEEE Press, Washington DC, 2017, pp. 6517–6525. doi:10.1109/CVPR.2017.690.
- [11] J. Redmon, A. Farhadi, YOLOv3: An Incremental, 2018. URL: <https://arxiv.org/abs/1804.02767>.
- [12] A. Bochkovskiy, Ch.-Y. Wang, H.-Y. M. Liao, YOLOv4: Optimal Speed and Accuracy of Object Detection, 2020. URL: <https://arxiv.org/abs/2004.10934>.
- [13] LabelImg is a graphical image annotation tool and label object bounding boxes in images, 2018. URL: <https://github.com/tzutalin/labelImg>.
- [14] Pytorch-YOLOv4. URL: <https://github.com/Tianxiaomo/pytorch-YOLOv4>.
- [15] Everingham M., Van Gool L., Williams C., Winn J., Zisserman A. The pascal Visual Object Classes (VOC) challenge. International Journal of Computer Vision 88, 2010, pp. 303–338. doi:10.1007/s11263-009-0275-4.
- [16] HIKVISION DS-2CD2T85FWD-I8 8MP camera H265+ 4K, 2020. URL: [https://www.youtube.com/watch?v=nG\\_wGUi-Ozc](https://www.youtube.com/watch?v=nG_wGUi-Ozc).
- [17] Tomo.yun, 2021. URL: <http://www.yunphoto.net/he/photobase/hr/hr22577.html>.

# Using of Intelligence Analysis of Technological Parameters Database for Implementation of Control Subsystem of Hot Blast Stoves Block ACS

Oleksiy Koyfman<sup>a</sup>, Oleksandr Simkin<sup>a</sup>, Yevhenii Klimov<sup>a</sup>, Sergey Scherbakov<sup>a</sup>

<sup>a</sup> Pryazovskyi State Technical University, str. Universytets'ka 7, Mariupol, 87555, Ukraine

## Abstract

One of the main trends in the development of automation control systems of blast-furnace production units is the use of the intelligence analysis of the data in order to identify new dependencies between technological parameters. In automated control of the heating of the hot blast stove checkerwork, its thermal state depending on the different duration of mode switching and heat loss to the environment is not taken into account.

The goal of the work is the application of the intelligence analysis of technological information from a database about the state of the hot blast stove during the periods of its switching from mode to mode for timely adjustment of the set values of parameters in controlling of the hot blast heating.

An algorithm and program have been developed for extracting from the archive database the values of the parameters of technological process of heating the hot blast to analyze the states of the hot blast stove when it switches from mode to mode. The structure of an automatic control system for hot blast stoves block using the subsystem of the intelligence analysis of the data has been proposed, it is based on the algorithm with the capability to adjust the operating modes of the block.

The application of the proposed intelligence analysis algorithm in the automatic control system will allow to increase the hot blast temperature without significant capital investments in the reconstruction of the hot blast stoves block.

## Keywords 1

Intelligence analysis, database, algorithm, software, hot blast stove, mode

## 1. Introduction

Most of the hot blast stove blocks are equipped with automatic control systems, in which the values of technological parameters are recorded in a database which allows to study the state of the hot blast stove over a long period of operation, including the state of the checkerwork when changing its operating modes. At the same time, the actual problem is the stabilization of the hot blast temperature at the set value, taking into account the aging of the hot blast stoves.

A review of scientific publications over the past few decades related to the automation of blast-furnace production units shows that one of the main trends in the development of control systems is the use of the intelligence analysis in order to identify new dependencies between technological parameters.

The goal of the work is the application of the intelligence analysis of technological information from a database about the state of the hot blast stove during the periods of its switching from mode to mode for timely adjustment of the set values of technological parameters in controlling of the hot blast heating.

---

CMIS-2021: The Fourth International Workshop on Computer Modeling and Intelligent Systems, April 27, 2021, Zaporizhzhia, Ukraine  
EMAIL: koifman\_o\_o@pstu.edu (O. Koyfman); simkin@ukr.net (O. Simkin); 24evgeniy03@gmail.com (Ye. Klimov); sherbakov@ukr.net (S. Scherbakov)

ORCID: 0000-0003-2075-7417 (O. Koyfman); 0000-0002-9939-7866 (O. Simkin); 0000-0002-5345-0510 (Ye. Klimov) 0000-0002-3360-3137 (S. Scherbakov)



© 2021 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)



## 2. Formal problem statement

To intensify blast furnace smelting, hot blasting is supplied to the furnace, which is heated in the block of hot blast stoves to a temperature of 1200-1300 °C. The required value of the blast temperature is determined by the blast furnace technology. The hot blast stove (hereinafter referred to as HBS) is a regenerative unit of periodic action up to 40 m high. During the “on-gas” mode, the heat from the combustion of the fuel gas is transferred to the refractory checkerwork. During the “on-blast” mode, the heat from the checkerwork is transferred to the blast. A standard block of hot blast stoves (hereinafter referred to as HBSB) consists of three to four units. The main task of the HBSB control system is to maintain the hot blast temperature at a given value.

The operating mode of the hot blast stoves is determined by the following set of parameters: fuel consumption, air consumption, dome temperature, maximum smoke temperature in the under-checkerwork unit, blast consumption, time mode, i.e. the duration of the blasting period (hereinafter referred to as “on-blast”) and the heating period (hereinafter referred to as “on-gas”), the duration of the separation period, when gases do not move through the hot blast stove, as well as the “on-switch” mode [1]. The main operating modes of the block: sequential, pair-parallel, mixed, parallel.

The operating modes are determined by the parameter chart, in which the time characteristics of the “on-gas” and “on-blast” modes, of the transfer of the hot blast stove from mode to mode are determined. In this case, the deviation of one of the parameters of the parameter chart from the established ones leads to the change in the operating modes of all hot blast stoves and fluctuations in the value of the blast temperature [2, 3].

Consequently, the task of controlling the HBSB is to implement the adopted parameter chart with a minimum consumption of blast furnace gas, a high-calorific additive, and combustion air in such a way that the maximum possible hot blast temperature is achieved.

The main disturbing influences on the hot blast stove operation:

- the blast furnace gas calorific value – it changes during the day by 20-30% which directly affects the heating time of the checkerwork;
- the incorrect fuel-air ratio due to inaccurate air flow control or its complete absence;
- the time of switching the hot blast stove from mode to mode - very often the switch of the hot blast stove from mode to mode is carried out in the remote control of the gasman / operator, which affects the duration of the switching.

Most of the hot blast stoves blocks are equipped with automatic control systems (ACS) of the dome temperature, the fuel-to-air ratio, stabilization of the hot blast temperature, at the same time, switching of modes is performed by operators remotely using the appropriate switches on the control panel. The values of technological parameters are recorded in the current database, and then in the archive, which allows to study the state of the hot blast stove when changing its operating modes.

To build graphs of changes in the main technological parameters (trends), the values are taken mainly from the archive database.

In the automation system of the studied block of hot blast stoves, the values of technological parameters are recorded in two types of database files in \*.dbf format. In a file of the first type (Fig. 1) 175 values of technological parameters of four hot blast stoves of the block are recorded every 10 seconds for the whole day. One file contains over 1.5 million records. The data is stored in a rather inconvenient way - in a column. The file of the second type, the so-called "wide" base (Fig. 2), already has several dozen of columns of data, which greatly simplifies the work, but the data recording period is now 20 seconds.

|   | A          | B        | C      | D         | E      | F      | G        | H | I |
|---|------------|----------|--------|-----------|--------|--------|----------|---|---|
| 1 | Date       | Time     | TagInc | Value     | Status | Marker | Internal |   |   |
| 2 | 12.10.2019 | 00:00:20 | 0      | 5,326     | B      |        |          |   |   |
| 3 | 12.10.2019 | 00:00:20 | 1      | 1,000     | B      |        |          |   |   |
| 4 | 12.10.2019 | 00:00:20 | 2      | 10,327    | B      |        |          |   |   |
| 5 | 12.10.2019 | 00:00:20 | 3      | 72429,563 | B      |        |          |   |   |
| 6 | 12.10.2019 | 00:00:20 | 4      | 0,141     | B      |        |          |   |   |
| 7 | 12.10.2019 | 00:00:20 | 5      | 6.563     | B      |        |          |   |   |

Figure 1: Database of the common type



|                        | A          | B        | C    | D  | E       | F         | G         | H         | I   | J   | K   | L   | M   | N   | O   | P   | Q   | R   | S          | T     | U       | V | W | X   | Y |  |
|------------------------|------------|----------|------|----|---------|-----------|-----------|-----------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------------|-------|---------|---|---|-----|---|--|
| 1                      | Date       | Time     | Mill | IV | GSS\DDG | S\GSS\RDG | S\GSS\TDG | S\BVN\DSG | S\B | S\B | S\B | S\B | S\B | S\B | S\B | S\B | S\B | S\B | S\BVN\DSGZ | S\B   | S\BVN\P |   |   |     |   |  |
| 2                      | 01.09.2020 | 00:00:23 | 75   | B  | 8,914   | 68636,805 | 0,000     | 3,414     | 1   | 0   | 0   |     |     |     |     |     |     |     | 56,506     | 3,300 | 0       |   |   | 10, |   |  |
| 3                      | 01.09.2020 | 00:00:35 | 434  |    | 9,001   | 69071,422 | 0,000     | 3,431     | 1   | 0   | 0   |     |     |     |     |     |     |     | 56,519     | 3,300 | 0       |   |   | 7,  |   |  |
| 4                      | 01.09.2020 | 00:00:55 | 434  |    | 8,846   | 68445,164 | 0,000     | 3,380     | 1   | 0   | 0   |     |     |     |     |     |     |     | 56,600     | 3,300 | 0       |   |   | 3,  |   |  |
| 5                      | 01.09.2020 | 00:01:15 | 465  |    | 8,752   | 67937,992 | 0,000     | 3,358     | 1   | 0   | 0   |     |     |     |     |     |     |     | 56,400     | 3,300 | 0       |   |   | 1,  |   |  |
| 6                      | 01.09.2020 | 00:01:35 | 621  |    | 8,753   | 68309,508 | 0,000     | 3,367     | 1   | 0   | 0   |     |     |     |     |     |     |     | 56,587     | 3,300 | 0       |   |   | 3,  |   |  |
| 7                      | 01.09.2020 | 00:01:55 | 653  |    | 9,082   | 69976,766 | 0,000     | 3,462     | 1   | 0   | 0   |     |     |     |     |     |     |     | 56,444     | 3,300 | 0       |   |   | 13, |   |  |
| 2010 09 01 0000 (Wide) |            |          |      |    |         |           |           |           |     |     |     |     |     |     |     |     |     |     |            |       |         |   |   |     |   |  |

**Figure 2:** Database of the "wide" type

In this case, the temperature of the hot blast is recorded in a separate file of the "wide" database with an interval of 10 seconds.

In the base of the first type, there are certain variables corresponding to the "on-gas" mode and the "on-blast" mode, therefore it is easy to determine in which of the three modes the HBS is located.

However, in the "wide" database there is only one variable characterizing the "on-blast" mode, which complicates the determination of the "on-gas" mode and the "on-switch" mode. In doing so, nothing is written to this variable for one of the hot blast stoves of the block.

Throughout the entire campaign of the hot blast stove, its technical state deteriorates, the thermal characteristics of the checkerwork change, it requires adjustments to the control algorithms of the automation system.

The authors propose a new approach that allows to analyze in real time the state of the checkerwork of the hot blast stoves when they are switched from mode to mode using the current technological information recorded in the database.

To achieve this goal, it is necessary to solve the following tasks: to develop an algorithm for the intelligent analysis of the technological base and its software implementation, to conduct a study of changes in the temperatures of the dome and exhaust gases during periods of change in operating modes, to develop the structure of the hot blast stove checkerwork heating ACS based on the subsystem of the intelligence analysis of the data.

The authors have developed an algorithm that allows to accurately determine in which mode the hot blast stoves are according to the values of other technological parameters. The following parameters were used: gas consumption for combustion; combustion air pressure; parameter of the "on-blast" mode.

### 3. Literature review

In paperwork [4], an intelligent expert system is considered, which automatically sets the fuel gas and air flow rates for heating the checkerwork of a hot blast stove and when the set temperature of the flue gases is reached, automatically calculates the duration of the "on-gas" period and automatically sets the gas flow rate in accordance with the estimated duration.

Work [5] proposes a predictive control scheme for hot blast stoves using a linear model based on the results of experiments with a step-by-step response on a detailed dynamic model of the process. The model is also supplemented with an integrated disturbance model. The developed optimizing control scheme allows to minimize energy consumption in hot blast stoves.

The authors of [6] have developed two models that allow predicting the consumption of blast furnace gas by a block of hot blast stoves. The first model is the network Echo-state model, which is more complex and sensitive to variations in operating methods. The second model is a simple switch model that requires no training and is very easy to use. The implementation of the proposed models contributes to both the reduction of the exploitation of natural resources, for example, natural gas, and the reduction of the facility's impact on the environment.

In work [7], it is shown a modern strategy for heating a hot blast stove based on fuzzy clustering of C-means. The strategy is a dynamic management of the gas-air ratio based on cluster analysis of a large amount of production data. Compared with the traditional combustion control method, the improved method has better optimization effect and better stability.

The authors of [8] presented a self-correcting computer system for controlling the temperature of a blast-furnace blast, consisting of three separate subsystems: calculation of the gas consumption ratio; dome temperature control; regulation of combustion gas temperature.

The paper [9] proposes the use of a neural network in the optimization of fuel consumption control and stabilization control of the dome temperature and the temperature of the combustion gases. Using the neural network decoupling control technique, the fuel consumption control in the hot blast stove was disabled. To obtain the ability to control the neural network decoupling, the authors used a direct link compensator in the structure with the HBS neural network.

In the course of an intelligent study of the control philosophy used for the heating cycle of a hot blast stove, it became obvious that the combustion process is not efficient enough and there are ways to improve it [10]. The actual aggregation and intelligent evaluation of production data allowed the identification of measures to increase the temperature of the hot blast. A new control system was developed and implemented that allows combustion to be regulated depending on the current chemical compositions of the blast furnace and coke gases, the required combustion temperature and the correct air-to-fuel ratio, maintaining a minimum oxygen level in the flue gases to achieve the maximum dome temperature.

In the study of difficult operating conditions of a hot blast stove [11], a hybrid intelligent algorithm is used for preliminary clustering processing of technological data sources and obtaining reliable forecasting results. Experimental results show that the algorithm has good convergence and stability. The improved predicted Markov value is close to the actual value and in line with the target value.

Article [12] proposes a new deep memory Echo-state network (DMESN) for predicting the flue gases temperature. Data preprocessing, including deviation of blowouts, missing data processing and lag time computation, is performed to obtain the best dynamic characteristics of the dataset. To increase the accuracy of prediction, an improved structure of the hidden layer is proposed, which consists of two parts: the formation of an Echo-state and the formation of a hidden state.

#### 4. Development of the Algorithm and Program Description

The complexity of information processing is that in databases of various types, due to the imperfection of instrumentation installed on hot blast stoves, not all parameter values are presented for the same period of time, and part of the parameters are recorded with a period that does not coincide with the record period in another database. To process such huge and complex amount of information, it is necessary to create a specialized software tool for opening a database of two types, extracting the necessary data from them, extrapolating parameter values from one database to synchronize time intervals with the records of another database. The result should be an array of data for data mining of the parameters and operating modes of the hot blast stove. For each HBS of the block, a separate MS Excel file should be generated, containing three tabs with the corresponding information about the “on-gas” mode, the “on-blast” mode and the “on-switch” mode of the HBS from mode to mode.

In the base of the first type, there are corresponding variables  $VN\backslash N$  and  $VN\backslash D$  to determine the “on-gas” mode and the “on-blast” mode, so there was no difficulty in dividing the database into three operating modes of hot blast stoves.

Preliminary statistical processing of the database showed that the values of technological parameters were correctly recorded in the database. Fig. 3 shows a block diagram of the analysis of the base of the first type.

However, in the database of the second type, there is one variable describing the “on-gas” mode  $VN\backslash D$ , while the open hot blast valve corresponds to the value  $VN\backslash D=1$ , and the closed valve corresponds to  $VN\backslash D=0$ .

Based on the valve switching cyclogram, if the parameter values are correctly recorded in the database, the periods can be determined as follows:

1. The “on-blast” mode:  $VN\backslash D=1$  - the hot blast valve is open, the blast is supplied for heating to the hot blast stove checkerwork;  $VN\backslash RSG=0$ , fuel gas is not supplied for combustion in the hot blast stove burner, gas consumption is 0;  $VN\backslash PVOZ=0$ , combustion air pressure is 0, air flow is 0.
2. The “on-gas” mode:  $VN\backslash D=0$  the hot blast valve is closed, the blast is not supplied for heating to the hot blast stove checkerwork;  $VN\backslash RSG>0$ , fuel gas is supplied for combustion in the hot blast stove burner, gas consumption is more than 0;  $VN\backslash PVOZ>0$ , air is supplied for combustion, air consumption is more than 0.

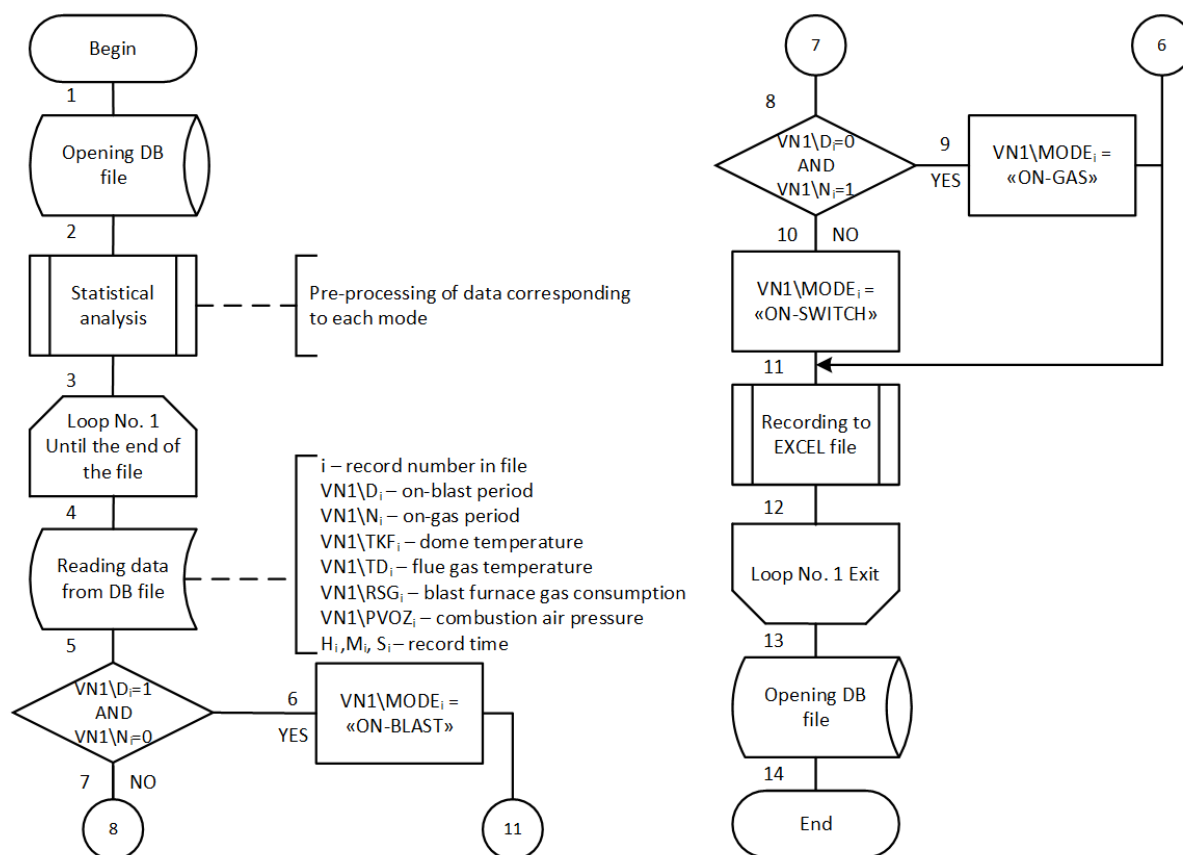


**Figure 3:** Block diagram of the analysis of the base of the first type

3. The “on-switch” mode of the hot blast stove from mode to mode:  $VN\backslash D=0$  when switching the hot blast valve must be closed;  $VN\backslash RSG=0$ , fuel gas is not supplied for combustion in the hot blast stove burner, gas consumption is 0;  $VN\backslash PVOZ \geq 0$ , at the beginning of switching from the “on-gas” mode to the “on-blast” mode, air is supplied to the ventilation of the hot blast stove, the air flow is more than or equal to 0.

When compiling the algorithm for dividing the database into modes, preliminary analysis revealed incorrect records of parameter values, which are summarized in Table 1.

**Table 1**  
Analysis of the "wide" database

| Parameter     | Mode                    | VN1                     | VN2                    | VN3                     | VN4                     |
|---------------|-------------------------|-------------------------|------------------------|-------------------------|-------------------------|
| Current value |                         |                         |                        |                         |                         |
| VN\D          | «On-blast», «On-switch» | <u>0</u>                | 1                      | 1                       | 1                       |
|               | «On-gas»                | 0                       | 0                      | 0                       | 0                       |
| VN\RSG        | «On-blast», «On-switch» | <u>0&lt;CV&lt;8000</u>  | <u>0&lt;CV&lt;8800</u> | <u>0&lt;CV&lt; 6000</u> | 0                       |
|               |                         | >0                      | >0                     | >0                      | >0                      |
|               | «On-gas»                | MIN=13030               | MIN=8968               | MIN=11791               | MIN=13919               |
|               |                         | AVG=30666               | AVG=18732              | AVG=34905               | AVG=43752               |
|               |                         | MAX=41535               | MAX=28059              | MAX=37951               | MAX=52552               |
|               |                         |                         |                        |                         |                         |
| VN\PVOZ       | «On-blast», «On-switch» | <u>0&lt;CV&lt;0,106</u> | 0                      | 0                       | <u>0&lt;CV&lt;0,049</u> |
|               |                         | >0                      | >0                     | >0                      | >0                      |
|               | «On-gas»                | MIN=1,211               | MIN=2,422              | MIN=0,572               | MIN=0,985               |
|               |                         | AVG=3,841               | AVG=3,432              | AVG=1,333               | AVG=2,205               |
|               |                         | MAX=6,078               | MAX=4,672              | MAX=2,138               | MAX=2,506               |
|               |                         |                         |                        |                         |                         |

Fig. 4 shows a block diagram of the analysis of the second type base.

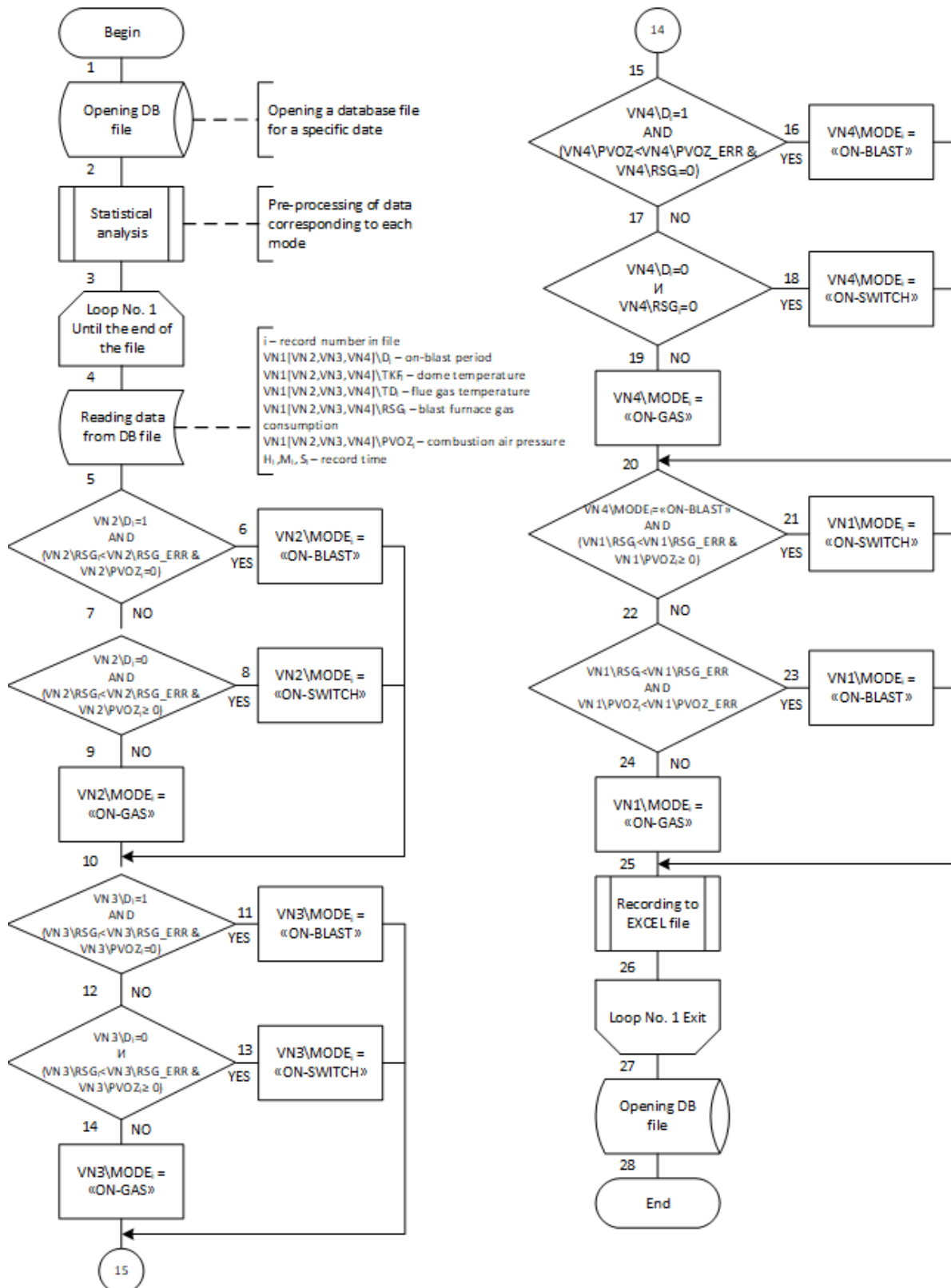


Figure 4: Block diagram of the analysis of the second type base

It was decided to implement the program in the modern programming environment IDE MS Visual Studio 2019 in the object-oriented programming language C #, which is quite simple, but with a lot of convenient features and extensions for various tasks. The program connects to the directory in

which the database files are stored, extracts the necessary information from it and submits it in an easy-to-read form to a new structured database or to an MS Excel file for further intellectual processing.

When developing the program as a provider for connecting to the database and extracting the necessary records from it, the database management system MS Visual FoxPro was used - an environment for developing database systems, including an object-oriented relational DBMS and a programming language for developing database applications and a system for reporting [13].

ClosedXML, a library for reading, processing and writing MS Excel files, was used to create a MS Excel file and to output the results of database processing into it [14]. It has a huge set of powerful functions for creating and processing MS Excel files, therefore, with its help, working with files through the program code is almost the same as working in MS Excel.

According to the algorithms included in the program (Fig. 3, 4), the information from the database is divided into files of \* .xlsx format, each of which contains information on the operating modes of the hot blast stoves block: the “on-gas” mode, the “on-blast” mode or the “on-switch” mode. The values of the dome and gas temperature in the checkerwork, as well as the hot blast temperature were chosen as the studied parameters characterizing the thermal state of the hot blast stove checkerwork.

After opening the program window (Fig. 5), the user needs to use the switch to select the type of databases for processing, and, if necessary, enter the tag numbers (except for the "wide" database) and select the step (interval) through which the data will be displayed.

**Figure 5:** The main window of the program

After operation, the program for verification displays data for each hot blast stove in the appropriate tabs in the window (Fig. 6) and reports on the successful creation of Excel files with the results of calculations for each hot blast stove.

| Date               | Time     | VN/D_TAG | VN/D_VALUE | VN/N_TAG | VN/N_VALUE | Mode | TKF_TAG | T  |
|--------------------|----------|----------|------------|----------|------------|------|---------|----|
| 26.06.2020 0:00... | 00:00:21 | 92       | 0          | 122      | 1          | 0    | 6       | 13 |
| 26.06.2020 0:00... | 00:00:27 | 92       | 0          | 122      | 1          | 0    | 6       | 13 |
| 26.06.2020 0:00... | 00:00:37 | 92       | 0          | 122      | 1          | 0    | 6       | 13 |
| 26.06.2020 0:00... | 00:00:47 | 92       | 0          | 122      | 1          | 0    | 6       | 13 |
| 26.06.2020 0:00... | 00:00:57 | 92       | 0          | 122      | 1          | 0    | 6       | 13 |

**Figure 6:** Data output for verification by tabs in the program subwindows

## 5. Research of the obtained results

### 5.1. Research of the change in the dome temperature during the “on-switch” mode of hot blast stoves

In the course of analyzing the database for a long period of operation of the hot blast stove block using the developed software, in addition to typical switchings (Fig. 7 a, b, 8 a, b), which correspond to the correct sequence of actions of the gas operator when switching operating modes, atypical switching were also found (Fig. 7 c, d, 8 c, d).

The graph of the dome temperature change for each switching (Fig. 7, 8) consists of three parts: in addition to the switching period, 50 values of the previous and subsequent modes are added.

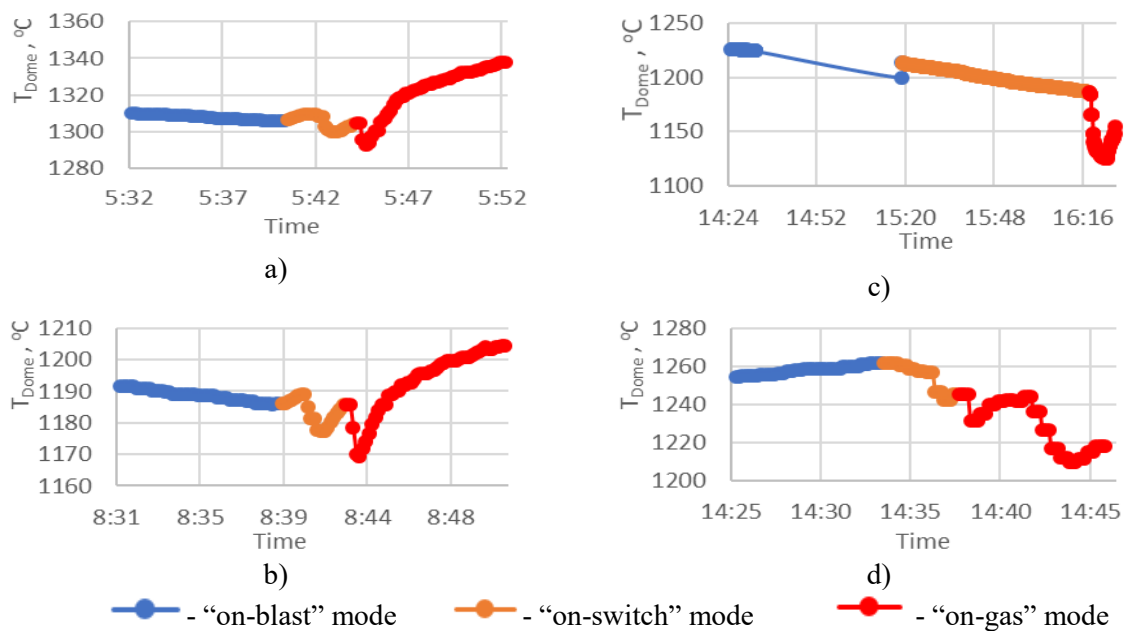


**Figure 7:** Switching of the hot blast stove from blast mode to heating mode

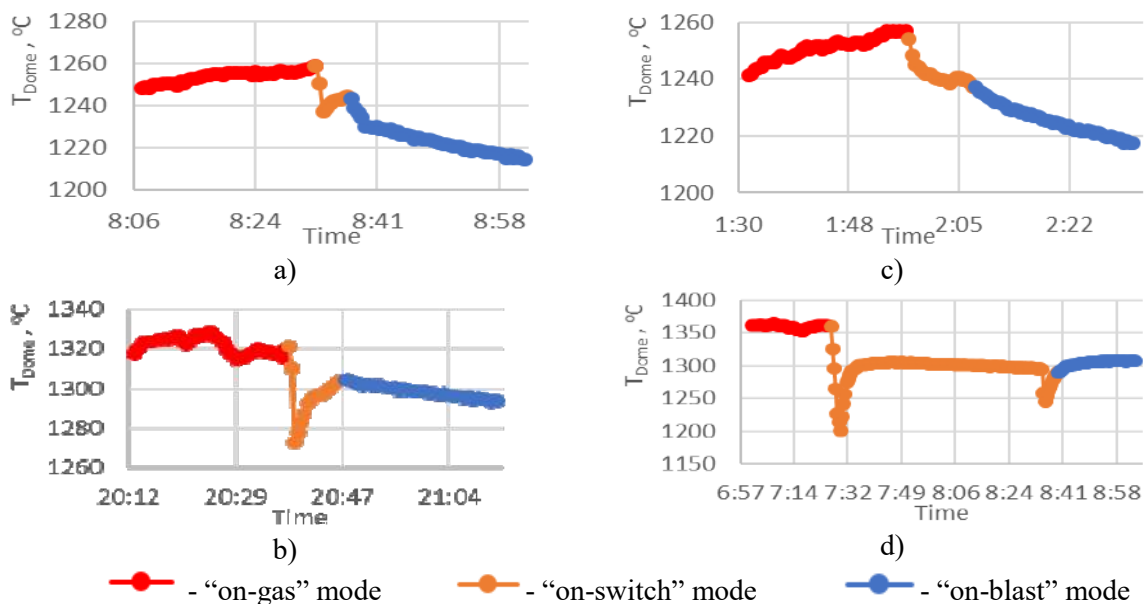


**Figure 8:** Switching of the hot blast stove from heating mode to blast mode

Typical switchings from the “on-gas” mode to the “on-blast” mode (Fig. 7 a, b) have a similar character, while lasting about 8-10 minutes, the dome temperature decreases by 15-25 °C;

Typical switching from the “on-blast” mode to the “on-gas” mode (Fig. 8 a, b) have the same character for all hot blast stoves, switching takes about 5 minutes, while the dome temperature drops by 15-20 °C.

During the “on-switch” mode, a drop in temperature can be observed within the range of 30-40 °C, and sometimes up to 50 °C, which must be taken into account when automatically controlling the subsequent operating mode of the hot blast stove.

## 5.2. Research of the “on-switch” mode of hot blast stoves of several different blocks

Using the developed program, the database has been studied over a long period to obtain and further analyze the averaged graphs of the temperature variation of the dome of hot blast stove of two blocks of different blast furnaces during the “on-switch” modes (Fig. 9).

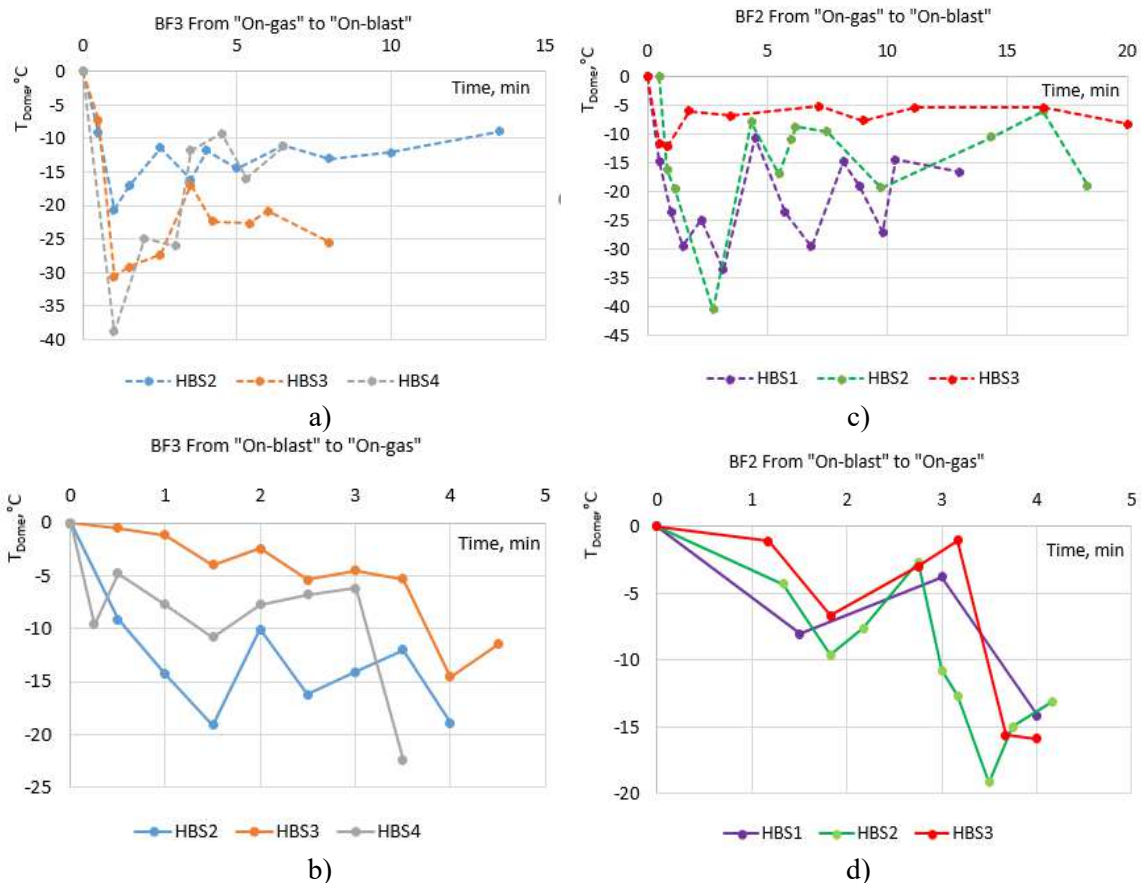


**Figure 9:** Average graphs of the dome temperature change during the “on-switch” mode

The change in the dome temperature during the “on-switch” mode from the “on-blast” to the “on-gas” modes has the same character and the duration of the “on-switch” mode corresponds to the parameter chart of operation of blocks adopted for the period under study (Fig. 9 b, d).

The duration of the “on-switch” mode from the “on-gas” to “on-blast” modes for HBS4 and HBS3 of the blast furnace No. 3 corresponds to the parameter chart, and for HBS it lasts twice as long as in the parameter chart.

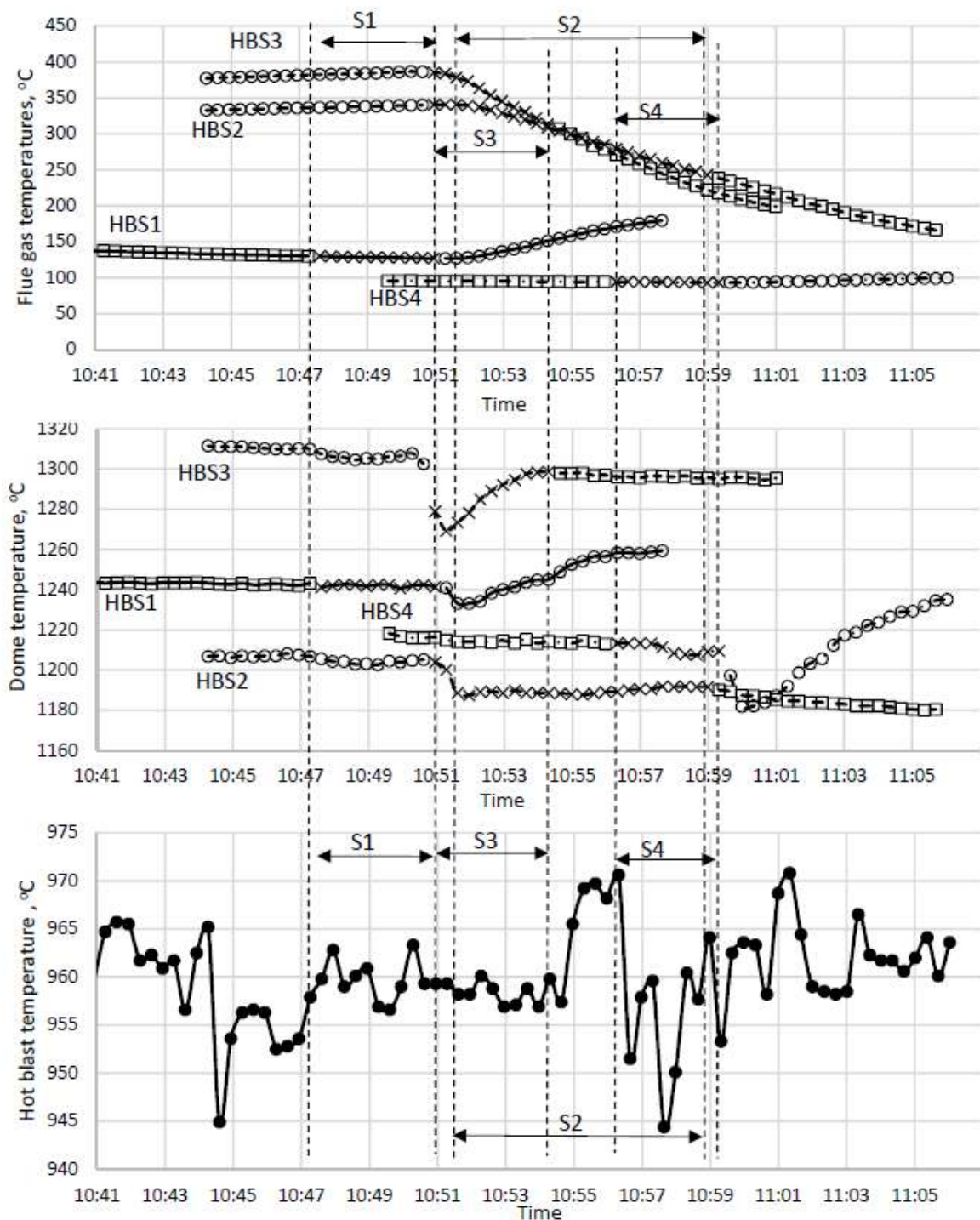
The duration of the “on-switch” mode from the “on-gas” mode to the “on-blast” mode for all hot blast stoves of the blast furnace No. 2 is much longer, which is indicated in the parameter chart.

The obtained results must be taken into account in drawing up a new corrected parameter chart to optimize the operation of the hot blast stove block.



### 5.3. Impact of the “on-switch” on hot blast temperature

Of particular interest is a complex analysis of the “on-switch” modes of the block operating in parallel mode and their impact on the hot blast temperature (Fig. 10).



«On-gas»: —○— - HBS1; - -○- - HBS2; .....○..... - HBS3; - ○ - - HBS4

«On-blast»: —×— - HBS1; - -□- - HBS2; .....□..... - HBS3; - □ - - HBS4

«On-switch»: —□— - HBS1; - -×- - HBS2; .....×..... - HBS3; - × - - HBS4

S1 - switching HBS1 from “on-gas” to “on-blast”, S2 - switching HBS2 from “on-gas” to “on-blast”,

S3 - switching HBS3 from “on-gas” to “on-blast”, S4 - switching HBS4 from “on-blast” to “on-gas”

**Figure 10:** The impact of the “on-switch” of hot blast stoves on the temperature of the hot blast



Prior the “on-switch mode, the block of hot blast stoves was in parallel operation mode: HBS1 and HBS4 operated in the “on-gas” mode, HBS2 and HBS3 - in the “on-blast” mode.

Let us consider the “on-switch” mode sequence:

- - HBS1 was the first to switch to the “on-gas” mode - no blast temperature fluctuations were observed;
- then HBS2 and HBS3 began to switch to the “on-blast” mode - there is a sharp jump in the blast temperature by about 15 °C for 2 minutes;
- HBS4 was the last to switch to the “on-gas” mode - in this case, a sharp drop in the blast temperature by 25 °C is observed;
- after the end of the “on-switch” mode, the temperature of the hot blast stabilizes.

Stabilization of the blast temperature during the “on-switch” is one of the most important tasks of today, since fluctuations in the blast temperature negatively affect the course of blast furnace smelting [15], therefore, information on the nature of fluctuations in the hot blast temperature should be transmitted to the automatic control system.

#### 5.4. Structure of Automated Process Control System on the Basis of the Proposed Algorithm

The authors propose a new structure of the automatic control system for the hot blast stoves block based on the data mining subsystem (Fig. 11), developed on the basis of information (dome and flue gas temperatures, hot blast temperatures, valve positions) obtained from the archive database, and the results of the performed research of the state of the hot blast stove during the “on-switch” mode.

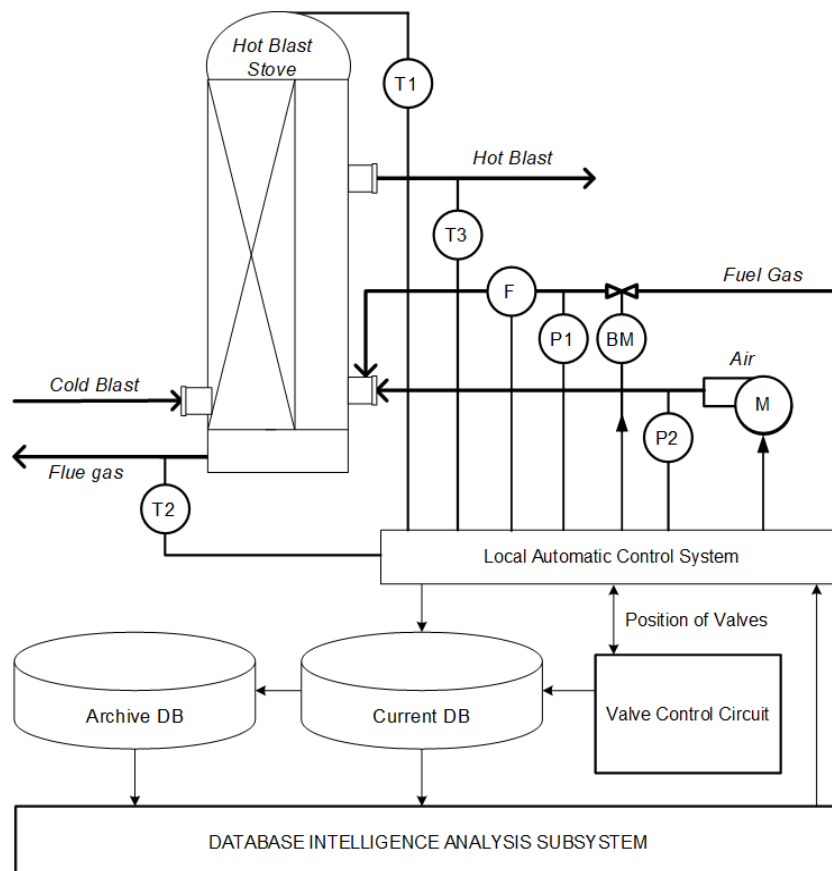


**Figure 11:** The structure of the hot blast stove checkerwork heating ACS based on the data mining subsystem

Current values of temperatures of the dome (T1), flue gases (T2) and hot blast (T3), flow rates (F) and mixed gas pressure (P1), combustion air pressure (P2), information on the position of valves from the valve control circuit are controlled in the local automatic control system (LACS) and are recorded in the current and archive database.

The data mining subsystem, based on the current technological information, monitors the change in the operating modes of the hot blast stoves, analyzes the state of the hot blast stoves during “on-switch”, compares with the archived information and makes adjustments to the operating modes of the hot blast stoves block.

If an increase in the duration of the hot blast stove switching to the “on-gas” mode is detected (a decrease in the dome temperature and an increase in heat losses to the environment, which leads to an increase in the duration of the “on-gas” mode), the data mining subsystem will send corrections to the set mixed gas flow rate to the LACS system, which will ensure the achievement of the set heating duration of the checkerwork.

Subject to availability of a mathematical model of the operation of the hot blast stove [16] in the automatic control system of the block, based on the information on switching and data mining of the durations of the “on-gas” and the “on-blast” modes [17], it becomes possible to correct the parameter chart of the hot blast stoves block for the set period of time of operation of the hot blast stoves block.

The application of the proposed method of data mining of the current technological information about the heating of the hot blast is a new approach in the design of automatic control systems for the hot blast stoves block, which allows to maintain the set temperature of the hot blast without impressive capital costs for the reconstruction of the hot blast stoves block.

## 6. Conclusion

Data mining of the database containing current and archived information on the technological process of blast-furnace heating must be used in automatic control systems. The usage of the obtained results of the analysis of the “on-switch” mode of the hot blast stove will allow to make timely adjustments to the parameter chart of the hot blast stoves block, to clarify the initial values of modeling the operation of the hot blast stove.

The authors have developed an algorithm for data mining of technological information, which allows to assess the state of the hot blast stove during the periods of the “on-switch” mode. A software tool for automated processing of databases has been developed, which significantly accelerates the process of study of hot blast stoves and minimizes the risk of errors during processing.

The analysis of the change in the dome temperature and the smoke temperature during the “on-switch” mode of the hot blast stove and the effect of switching on the blast temperature has been performed.

The structure of an automatic control system for blast furnace heating using a data mining subsystem has been proposed.

Application of the proposed data mining algorithm in the automatic control system will allow to increase the hot blast temperature without significant capital investments in the reconstruction of the hot blast stoves block.

## 7. References

- [1] S. Biswas, D. Sarkar, Introduction to Refractories for Iron- and Steelmaking, Springer, Jamshedpur, Jharkhand, India, 2020. doi:10.1007/978-3-030-43807-4
- [2] M. Helle, H. Saxén, Simulation and Optimization of Hot Stoves in a Steel Plant, in: SCANMET V - 5th Int. Conf. Process Dev. Iron Steelmak., Luleå, Sweden, 2016, paper 15.
- [3] R. Motta, W. Costa, Th. Soares, L. Souza, Blast Furnace Hot Stove Changeover by Thermal Control, In: 42º Seminário de Redução de Minério de Ferro e Matérias-primas / 13º Seminário Brasileiro de Minério de Ferro / 6th International Congress on the Science and Technology of Ironmaking, Rio de Janeiro (2012): 80-90. doi:10.5151/2594-357X-21537

- [4] M. A. Zhu-wu, Full Automatic Control Expert System of Hot Blast Stoves for Blast Furnace, *Basic Automation* 4 (2002) 57-62 URL: [https://en.cnki.com.cn/Article\\_en/CJFDTotol-JZDF200204017.htm](https://en.cnki.com.cn/Article_en/CJFDTotol-JZDF200204017.htm)
- [5] A. Şahin, M. Morari, Optimizing control of hot blast stoves in staggered parallel operation, *IFAC Proceedings Volumes*, 41 (2008) 1982-1987. doi:10.3182/20080706-5-KR-1001.00337
- [6] I. Matino, Two innovative modelling approaches in order to forecast consumption of blast furnace gas by hot blast stoves, *Energy Procedia*, Vol. 158 (February 2019) 4043–4048. doi:10.1016/j.egypro.2019.01.834
- [7] C.N. Zhang, Y.R. Li, A combustion control strategy of hot blast stove based on kernel fuzzy c-means (FCM), *Metalurgija* Vol. 58. 3/4 (2019) 179-182. URL: <https://hrcak.srce.hr/218345>
- [8] S. Chengcheng, Development of new technology of high blast temperature in hot blast stove of Baosteel No. 3 blast furnace, *Ironmaking* Vol. 24 (2005) 63-66. URL: [http://en.cnki.com.cn/Article\\_en/CJFDTotol-TIEN2005S1014.htm](http://en.cnki.com.cn/Article_en/CJFDTotol-TIEN2005S1014.htm)
- [9] X. Yuan, X. Wang, L. Bai Ling, Neural network decoupling control for fuel flow of hot blast stove. *Metallurgical Industry Automation* Vol. 3 (2009) 63-70. URL: [https://en.cnki.com.cn/Article\\_en/CJFDTotol-YJZH200903022.htm](https://en.cnki.com.cn/Article_en/CJFDTotol-YJZH200903022.htm)
- [10] J. Curilla, M. Dulovic, G. Guidugli, M. Manolov, Optimization of stoves gassing system in Slovakia U. S. Steel Kosice Plant in: *SCANMET V - 5th Int. Conf. Process Dev. Iron Steelmak.*, Lulea, Sweden, 2016
- [11] C. N. Zhang, Y. R. Li, Markov hot blast stove prediction based on hybrid intelligent algorithm *Metalurgija* Vol. 59. 4 (2019) 451-454. URL: <https://hrcak.srce.hr/241157>
- [12] Y. Yang, X. Zhao, X. Liu, A Novel Echo State Network and Its Application in Temperature Prediction of Exhaust Gas from Hot Blast Stove. *IEEE Transactions on Instrumentation and Measurement* 69.12 (2020): 9465 – 9476. doi:10.1109/TIM.2020.3003975
- [13] Microsoft Visual FoxPro (VFP) URL: <https://docs.microsoft.com/en-us/sql/odbc/microsoft/visual-foxpro-terminology?view=sql-server-ver15>
- [14] ClosedXML is a .NET library for reading, manipulating and writing Excel 2007+ (.xlsx, .xlsm) files URL: <https://github.com/ClosedXML/ClosedXML>
- [15] Y. Liu, *The Operation of Contemporary Blast Furnaces*, Beijing, China, 2021. doi:10.1007/978-981-15-7074-2
- [16] A. Koifman, A. Simkin, Development and Software Implementation of the Hot Blast Stove Computer Model, in: D. Luengo, S. Subborin (Eds.), *Proceedings of The Second International Workshop on Computer Modeling and Intelligent Systems (CMIS-2019)*, Zaporizhzhia, Ukraine, April 15-19, 2019. *CEUR Workshop Proceedings*, Vol. 2353: pp. 440-454 (2019). URL: <http://ceur-ws.org/Vol-2353/paper35.pdf>.
- [17] O. Koyfman, O. Simkin, Intelligence analysis method of automation control system archive database for controlling hot blast stove, in . Subborin (Eds.), *Proceedings of The Third International Workshop on Computer Modeling and Intelligent Systems (CMIS-2020)*, Zaporizhzhia, Ukraine, April 27-May 1, 2020. *CEUR Workshop Proceedings*, Vol. 2608: pp. 102-117 (2020). URL: <http://ceur-ws.org/Vol-2608/paper9.pdf>.

# Local Quality Improvement of Multispectral Imagery Classification with Radiometric-spatial Feedback

Iryna Piestova<sup>a</sup>, Anna Kozlova<sup>a</sup>, Artem Andreiev<sup>a</sup> and Jan Rabcan<sup>b</sup>

<sup>a</sup> Scientific Centre for Aerospace Research of the Earth, Institute of Geological Sciences, National Academy of Sciences of Ukraine, O. Gonchar str., 55-b, Kyiv, 01054, Ukraine

<sup>b</sup> University of Zilina, Univerzitna, 8215/1, Zilina, 01026, Slovakia

## Abstract

The essential requirement for accurate classification is the high resolution of input images. Among known classification problems, which caused by low-resolution images, are the mixing of training samples and the absence of boundaries between objects of different classes. The mentioned above problems were reduced by imagery spatial resolution enhancement and a hybrid approach to classification, which allows unmixing training samples and improving the quality of images and their classifications.

## Keywords 1

Remote sensing, spatial resolution enhancement, supervised classification, training samples clustering, multi-valued logic, subpixel reallocating

## 1. Introduction

Classification is the process of dividing a mass of data into classes (groups) according to some criterion. In the process of computer classification, each pixel of the image is assigned to one of the selected classes. Classification of satellite images is a widely-used remote sensing tool for solving such tasks as land-cover mapping and change detection [1], forecasting gas and oil potential of subsoil plots [2], etc. For example, when compiling a map of mountain vegetation using the satellite image, you can divide the entire territory depicted on it into areas covered with forests, meadows, glaciers, etc. The resulting image is called a "classification map", or simply "classification".

Computer classification is used to automatically separate objects displayed on images and obtain a map of the area. For computer processing, each image is presented as a table, each cell of which - an element of the image resolution (pixel) - contains a number indicating the brightness of this element. In multispectral imagery, the values of the brightness of objects are recorded in separate rather narrow areas - spectral bands - of the visible and infrared parts of the spectrum. Then, during computer processing, for each pixel of the image, several digital values of brightness are used at once in different spectral bands.

The main difference between visual decryption and automated processing is that a person (decoder) sees the entire or almost the entire image, whereas a computer in most cases analyzes digital values for only one pixel or a small group of pixels, comparing them with the rest. A person can use different decryption signs to recognize terrain objects: size, shape, length, relative position of objects, etc., but a computer can simultaneously analyze several images in different spectral bands, and, as a rule, much faster than a person can.

When carrying out any classification, some obstacles affect the quality and accuracy of the classification:

---

CMIS-2021: The Fourth International Workshop on Computer Modeling and Intelligent Systems, April 27, 2021, Zaporizhzhia, Ukraine  
EMAIL: [ipestova@casre.kiev.ua](mailto:ipestova@casre.kiev.ua) (I. Piestova); [ak@casre.kiev.ua](mailto:ak@casre.kiev.ua) (A. Kozlova); [a.a.andreev@casre.kiev.ua](mailto:a.a.andreev@casre.kiev.ua) (A. Andreiev); [jan.rabcan@fri.uniza.sk](mailto:jan.rabcan@fri.uniza.sk) (J. Rabcan)  
ORCID: 0000-0003-2981-7826 (I. Piestova); 0000-0001-5336-237X (A. Kozlova); 0000-0002-6485-449X (A. Andreiev) ; 0000-0003-2835-9114 (J. Rabcan)



© 2021 Copyright for this paper by its authors.  
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).  
CEUR Workshop Proceedings (CEUR-WS.org)

- the use of supervised classification causes errors associated with the human factor, uneven samples, a discrepancy between the selected thematic classes and their spectral characteristics, etc. - this can be called a radiometric obstacle [3];
- dependence of the quality of classification (detail, separability of classes) on the spatial resolution of the input images, as well as the frequent use of images of different resolutions or non-uniform spatial resolution of images (such as Sentinel-2 data, where there is a spatial resolution of 10, 20 and 60 m) - what can be called a spatial obstacle [4].

## 2. Methods

This study presents tools for solving the problems described above: applying a hybrid approach to classification to correct radiometry and the spatial resolution equalization of multispectral satellite images.

### 2.1. Hybrid approach to classification

The reasonable interpretation of land cover classes could only be achieved by supervised classification since it applies expert knowledge, which describes each type of land cover. Supervised classification requires training samples of classes selected by an expert. However, those classes, as a rule, are subjective as well as expert-selected training samples are not accurate. Unlike supervised classification, unsupervised one provides objective classes, obtained by clustering. The hybrid approach to classification [5] is applied to form the interpretable and objective classes. The scheme of the described above approach to classification is shown in Figure 1.

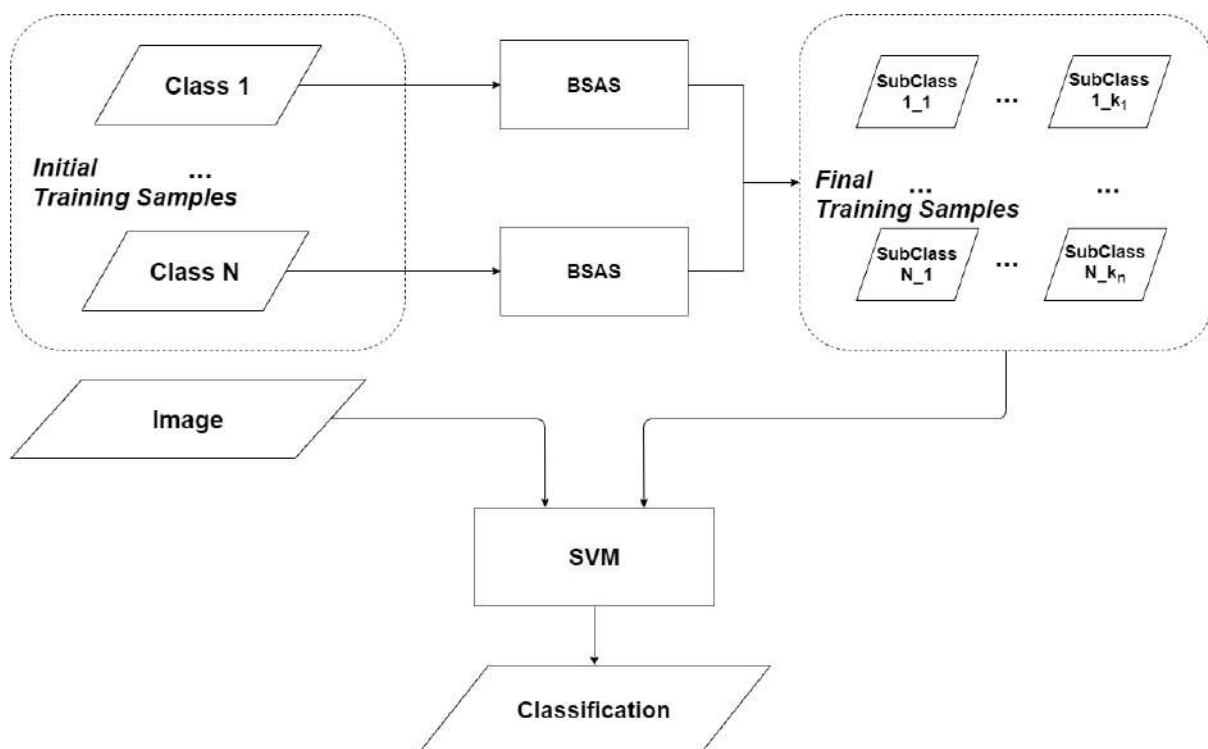


**Figure 1:** The scheme of the hybrid approach to classification

This approach to classification implies subdividing training samples of expert-selected classes into objective clusters by unsupervised classification. After that, formed clusters are used as training samples for supervised classification. Those steps are aimed to reduce the inaccuracy and

subjectiveness of the selected samples. In turn, it increases the classification accuracy in comparison with both supervised and unsupervised types of classification.

### **2.1.1. Training samples clustering**

The first step is the training samples clustering by unsupervised classification. Clustering is necessary to avoid the subjectiveness of expert-selected classes. Also, due to the high heterogeneity of some land cover classes, they overlap each other and it means that their training samples, as a rule, are mixed. This problem could be also solved by the initial training samples clustering. Basic Sequential Algorithmic Scheme (BSAS) [6] is applied as a method of unsupervised classification. The main benefit of this method among others is that the number of clusters may not be known in advance. However, instead of the number of clusters, this scheme requires two other input parameters, which are the threshold of the dissimilarity and the maximum allowed number of clusters. The first one is defined as a distance between each cluster and feature vector, which corresponds to the set of training samples. The second parameter is required not to divide the input data into a greater number of clusters than it's defined by an expert. This parameter should be defined taking into account the computational costs of those operations and/or limit of clusters, overcoming of which will provide inadequate and not interpretable training samples subdividing. This method applies separately to the training samples of each class. As a result of this step, initial expert-selected training samples would be subdivided into objective subclasses and their heterogeneity would be reduced by transforming them into dense clusters.

### **2.1.2. Supervised classification**

The next step is the supervised classification of the study area. Support vector machine [7] is applied as a method of supervised classification for this task. This method is proved its reliability in conditions of high heterogeneity of land cover classes [8]. The required input data for this procedure are an image of the study area and training samples of each class. Subclasses obtained at the previous step are used as training samples. Therefore, each subclass is interpreted as a particular class. After this step, the input image will be classified into that number of classes, which corresponds to the number of subclasses obtained after clustering of initial training samples.

In case if the task is to obtain classification, which consists only of initial classes, then merging of subclasses into initial classes is required. The subclasses' merging defines the initial class of pixels subclass, and then the value of a pixel is converted to the value of the initial class. This procedure is performed separately for each pixel of classification obtained at the second step.

## **2.2. Spatial resolution equalization**

It is very important to carry out the spatial resolution equalization procedure when satellite images or separate image bands of different spatial resolutions are used together.

A widely used method is to simply divide a pixel into subpixels according to the nearest neighbour rule [9], while the resulting subpixels retain the value of the output pixel. Although the use of the nearest neighbour rule preserves the average radiometric value of subpixels in a pixel, it does not increase the information content of the resampled bands.

Another widely used method for calculating subpixel values is the application of the selected type of interpolation based on a certain number of adjacent pixels in the original image. Consider, for example, a bicubic interpolation technique [10] that slightly smoothes the image, giving a sense of more detail. However, this procedure does not keep the radiometry within the original pixel. The average value of the subpixels in pixels is different from the initial value of the input image.

It is proposed to use a method [11] based on image segmentation by spectral signatures, as well as optimization of decision making when redistributing values in subpixels, taking into account both the similarity of the original spectral signatures and the spatial relationships of the topology for each type of land cover.

### 2.2.1. Scanning pattern

The Nearest Neighbour Oversampling procedure quadruples a pixel and ultimately generates 4 identical subpixels. To improve the overall physical resolution of the image bands, but not degrade their quality, it is necessary to correctly redistribute the signals in these subpixels, while maintaining their average radiometric value.

This should be done using the scanning window and taking into account the adjacent territory, that is, the nearest eight subpixels (numbered 1 .. 4 and 6 .. 9) around the current one (numbered as 5) as shown in Figure 2.

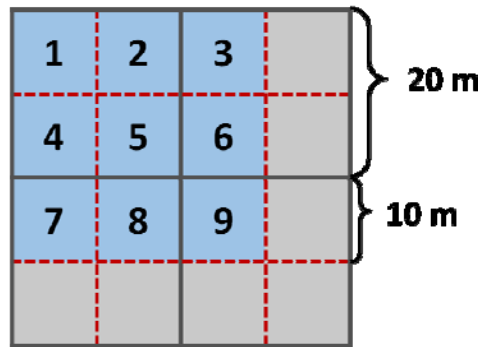


**Figure 2:** Neighbourhood of processed subpixel (5) within scanning window (1 .. 4 and 6 .. 9) including 4 subpixels (1, 2, 4, 5) of one low-resolution pixel

When processing images, the process of determining different types of the earth's surface is necessary. This is done by spectral segmentation of the image by spectral characteristics. Fuzzy logic methods are widely used to solve such problems [12].

### 2.2.2. Classes reallocating

For the correct spatial redistribution of subpixels, it is necessary to analyze the topological properties of the main classes and obtain appropriate topological descriptions. Multiple valued logic (MVL) methods are used to solve the problem of changing the subpixel class according to the relationships and classes of the nearest surrounding subpixels.

MVL is a type of logic in which the level of truth can be m-valued or infinite, not just binary, as in Boolean logic. MVL requires a mathematical approach to express the relationships between input logic values and the result of certain phenomena. A logical function with a value having the following form was used:

$$f_m : M^n \rightarrow M, \quad (1)$$

where  $n$  is the number of multivalued variables, and the set  $M = \{0, 1, \dots, m - 1\}$  is the set of certain truth levels.

MVL is used to reclassify surface types depending on their spatial distribution, which is used in our analysis of remote sensing data because this approach will allow a fairly efficient reclassification.

The pixel redistribution procedure takes into account the following topological properties of the analyzed segment:

- compactness - when subpixels of one class are localized by compact inseparable groups;
- orientation (linearity) - subpixels of one class are arranged linearly: horizontally, vertically, diagonally if you can determine the orientation of these lines;
- texture (homogeneity) - subpixels of one or more classes are arranged in a checkerboard pattern or close to it, as well as when several different classes fall within the window and compact or linear structures are not defined.

Five significant types of the relative location of subpixels of one class are defined in Figure 3. These five types represent specific shapes of subpixels, such as all columns and rows - T5, diagonals (T2), and right triangles with a certain location (T1 - the right angle at the edges, T3 - the right angle in the middle of the outer columns and rows in the matrix, T4 - the right angle in the middle of the matrix). The number of occurrences of such types can be defined as an input variable for a 3-digit logical function, which will give a decisive result for the central pixel of the class change matrix.

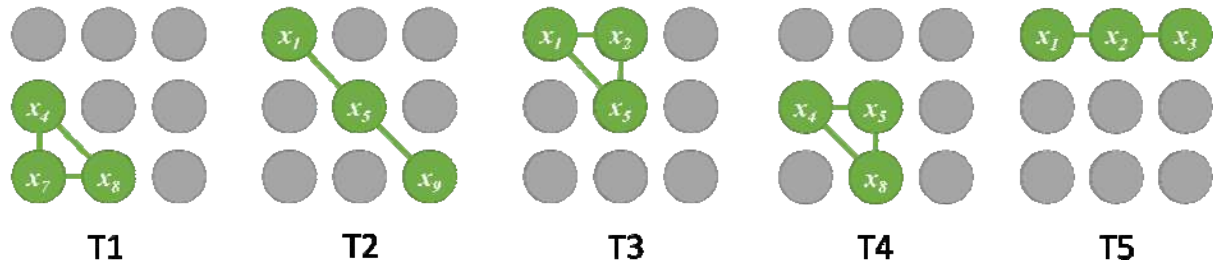


**Figure 3:** The main types of mutual arrangement of subpixels.

The values of the types of mutual arrangement of subpixels are consistent with the possible number of this type of arrangement in the considered block.

The next step is performed in a matrix consisting of subpixels of a  $3 \times 3$  sliding window, where the analyzed pixel is central. The classification in the scan window is checked sequentially. A logical function is used to decide on the replacement of the central subpixel class (Figure 4). According to this function, the central pixel changes if the value of the function is 2, and does not change if the value of the function is 0. Changing the class of the central pixel requires additional analysis if the value of this function is 1.

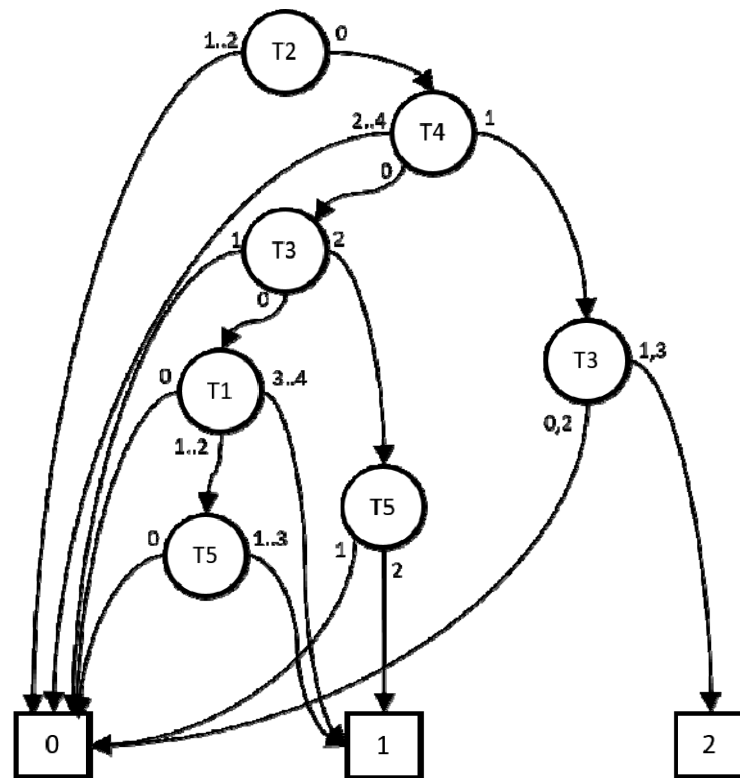


**Figure 4:** A logical function to change the class of the central pixel depending on its location.



### 2.3. Radiometric-spatial Feedback

The hybrid classification method described above requires high-resolution source images. In turn, to increase the resolution of the bands of a multispectral image, stable clustering is required, which is based on the spectral signatures of the earth's covers. This leads us to the solution of using radiometric spatial feedback. This method is an iterative process described in Figure 5.

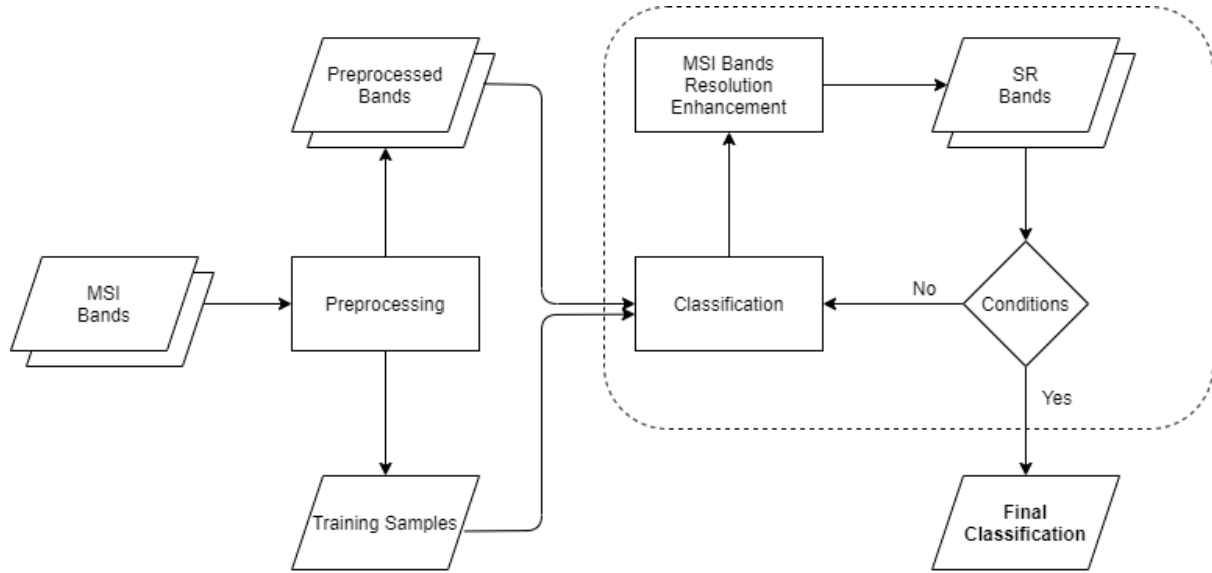


**Figure 5:** General scheme of multispectral imagery classification local quality improvement with radiometric-spatial feedback.

The iterative process continues until the spatial resolution of the bands is increased. As soon as the increase becomes insignificant compared to the previous iteration, the process stops and the final classification is carried out.

## 3. Test

The method of multispectral imagery classification local quality improvement with radiometric-spatial feedback in the selected study area was tested.

### 3.1. Study area

The study area was located around the city of Novomyrhorod in Kirovohrad Oblast (administrative province) of central Ukraine (Figure 6). It encompassed agricultural, wetland, and urban landscapes which elements substantially vary in spatial characteristics. For instance, extensive and spatially homogeneous croplands have sharp delineations in a form of roads and tree lines, while wetlands and wet grassland have meandering and vague boundaries. Relatively small built-up areas extended along streets and surrounded by highly heterogeneous household plots. However, most of these elements have similar spectral characteristics especially, during the mid-summer season. For their better recognition from satellite images, the application of spectral indices strongly recommended [13-15].



**Figure 6:** The study area involved various urban, rural, and complex natural landscapes around Novomyrhorod, Ukraine. It is shown on the fragment of the Sentinel-2 Multispectral Instrument (MSI) image acquired on 6 July 2020. The image represents a true-colour composite of Red, Green, and Blue bands of 10 m spatial resolution.

In this study, we focused on seven land cover types. They are artificial surfaces, croplands, tree-covered areas, grasslands, wet grasslands, wetlands, and water bodies (Table 1).

**Table 1**

The classification scheme used in the study

| Land Cover Class    | Description                                                                                                                       |
|---------------------|-----------------------------------------------------------------------------------------------------------------------------------|
| Artificial surfaces | Urban public and industrial built-up areas, transport units, and construction sites                                               |
| Croplands           | Arable land, permanent crops, fallow lands, heterogeneous agricultural areas, household plots                                     |
| Tree-covered areas  | Broadleaved and coniferous forest stands, ravine and floodplain forests, roadside tree lines, areas with tree cover more than 30% |
| Grasslands          | Natural herbaceous vegetation, permanent grasslands of natural origin, pastures                                                   |
| Wet grasslands      | Grassland that is periodically flooded or waterlogged by freshwater with typical plant communities of grass, sedge, and rush      |
| Wetlands            | Inland marshes, reed beds, riparian cane formations                                                                               |
| Water               | Rivers, reservoir, streams                                                                                                        |

### 3.2. Sentinel-2A data and pre-processing

Cloud-free Sentinel-2A multispectral instrument (MSI) image acquired on 06 July 2020 was downloaded from the U.S. Geological Survey (USGS) archive through the EarthExplorer interface (<https://earthexplorer.usgs.gov/>). The images were obtained at Level 1C: top of the atmosphere (TOA) reflectance (<https://sentinel.esa.int/web/sentinel/user-guides/sentinel-2-msi>).

The MSI measures the Earth's reflected radiance in 13 spectral bands from the visible near-infrared (VNIR) to shortwave infrared (SWIR) wavelength regions, with spatial resolutions from 10 to 60 m, as shown in Table 2.

**Table 2**  
Spatial and spectral resolutions of Sentinel-2 MSI

| Sentinel-2A/MSI ( $\mu\text{m}$ ) | Band            | Resolution (m) |
|-----------------------------------|-----------------|----------------|
| Band 1 (0.43–0.45)                | Coastal aerosol | 60             |
| Band 2 (0.46–0.52)                | Blue            | 10             |
| Band 3 (0.54–0.58)                | Green           | 10             |
| Band 4 (0.65–0.68)                | Red             | 10             |
| Band 5 (0.7–0.71)                 | Red-edge-1      | 20             |
| Band 6 (0.73–0.75)                | Red-edge-2      | 20             |
| Band 7 (0.76–0.78)                | Red-edge-3      | 20             |
| Band 8 (0.78–0.90)                | NIR             | 10             |
| Band 8A (0.85–0.87)               | Narrow NIR      | 20             |
| Band 9 (0.93–0.95)                | Water vapor     | 60             |
| Band 10 (1.36–1.39)               | SWIR/Cirrus     | 60             |
| Band 11 (1.56–1.65)               | SWIR-1          | 20             |
| Band 12 (2.10–2.28)               | SWIR-2          | 20             |

To provide atmospherically corrected images essential to calculations of spectral indices, Level 1C data was processed to Level 2A: bottom of the atmosphere reflectance (BOA).

The Sen2Cor tool (<https://step.esa.int/main/snap-supported-plugins/sen2cor/>) from the European Space Agency (ESA) Sentinel Application Platform (SNAP) was used to perform the corrections for the Sentinel-2 image. During the processing, Sen2Cor discarded the three bands (B1, B9, and B10) that consider the effects of aerosols and water vapour on reflectance. Then, the Sentinel-2 bands acquired at 20 m data were previously resampled using the nearest neighbour method to obtain a layer stack of 10 spectral bands at 10 m. At the final stage of data pre-processing, the obtained image was resized by an area of  $1500 \times 1000$  pixels (Figure 6) for testing the method proposed in the study.

### 3.3. Spectral indices use

Spectral indices, being nonlinear transformations of original spectral bands, substantially enhance the classification quality of complex classes [1, 4, 16]. The red-edge (RE) is the prominent spectral feature of vegetation, including wetlands [13, 14]. The SWIR range is extensively used in many applications related to water bodies and urban surfaces [15]. Sentinel-2 MSI image of Level 2A provides the RE (B5, B6, and B7) and the SWIR (B11 and B12) ranges at 20m spatial resolution.

The SI used in this study are:

- (a) the normalized difference vegetation index - NDVI [17];
- (b) the red-edge NDVI -RENDVI [18];
- (c) the red edge ratio vegetation index – RERVI [19, 20];
- (d) the normalized difference water index -NDWI [21];
- (e) the modified normalized difference water index - MNDWI [22];
- (f) the normalized difference moisture index - NDMI [23];
- (g) the normalized difference built-up index – NDBI [24].

The formulations and the bands used to calculate the spectral indices from Sentinel-2A MSI are shown in Table 3.

**Table 3**

The spectral indices calculated from Sentinel-2A MSI image data in the study

| Vegetation Index | Formulation                                               | Sentinel-2 Bands Used               |
|------------------|-----------------------------------------------------------|-------------------------------------|
| NDVI             | $\frac{R_{NIR} - R_{red}}{R_{NIR} + R_{red}}$             | $\frac{B_8 - B_4}{B_8 + B_4}$       |
| RENDVI           | $\frac{R_{NIR} - R_{red\_edge}}{R_{NIR} + R_{red\_edge}}$ | $\frac{B_8 - B_6}{B_8 + B_6}$       |
| RERVI            | $\frac{R_{NIR}}{R_{red\_edge}}$                           | $\frac{B_8}{B_6}$                   |
| NDWI             | $\frac{R_{green} - R_{NIR}}{R_{green} + R_{NIR}}$         | $\frac{B_3 - B_8}{B_3 + B_8}$       |
| MNDWI            | $\frac{R_{green} - R_{SWIR}}{R_{green} + R_{SWIR}}$       | $\frac{B_3 - B_{11}}{B_3 + B_{11}}$ |
| NDMI             | $\frac{R_{NIR} - R_{SWIR}}{R_{NIR} + R_{SWIR}}$           | $\frac{B_8 - B_{12}}{B_8 + B_{12}}$ |
| NDBI             | $\frac{R_{SWIR} - R_{NIR}}{R_{SWIR} + R_{NIR}}$           | $\frac{B_{11} - B_8}{B_{11} + B_8}$ |

#### 4. Local quality improvement

Guided by the data flow diagram described in Figure 5, a two-iteration process was carried out, which included 2 hybrid classification blocks, 2 blocks of spectral bands spatial resolution enhancement, and the final classification.

The results of evaluating the spatial resolution of spectral bands after the second iteration using the edge spread function are presented in Table 4 and show a steady increase of spatial resolution in bands 5-7 and 8a. It is assumed that a more stable increase of the spatial resolution in bands 11 and 12 can be achieved after attracting additional spectral signatures for reference clustering.

**Table 4**

Digital image bidirectional Gaussian edge spread function values after the second iteration

| Band           | Basic            | Cubic        | Enhanced     |
|----------------|------------------|--------------|--------------|
| Band 5         | $1.778 \times 2$ | 2.806        | 2.658        |
| Band 6         | $2.313 \times 2$ | 3.092        | 1.417        |
| Band 7         | $2.157 \times 2$ | 3.047        | 1.759        |
| Band 8a        | $2.239 \times 2$ | 3.128        | 1.893        |
| Band 11        | $2.561 \times 2$ | 3.379        | 4.113        |
| Band 12        | $2.043 \times 2$ | 3.009        | 4.125        |
| <b>Average</b> | <b>4.364</b>     | <b>3.077</b> | <b>2.661</b> |

An illustration of the multispectral imagery classification local quality improvement with radiometric-spatial feedback is shown in Figure 7.

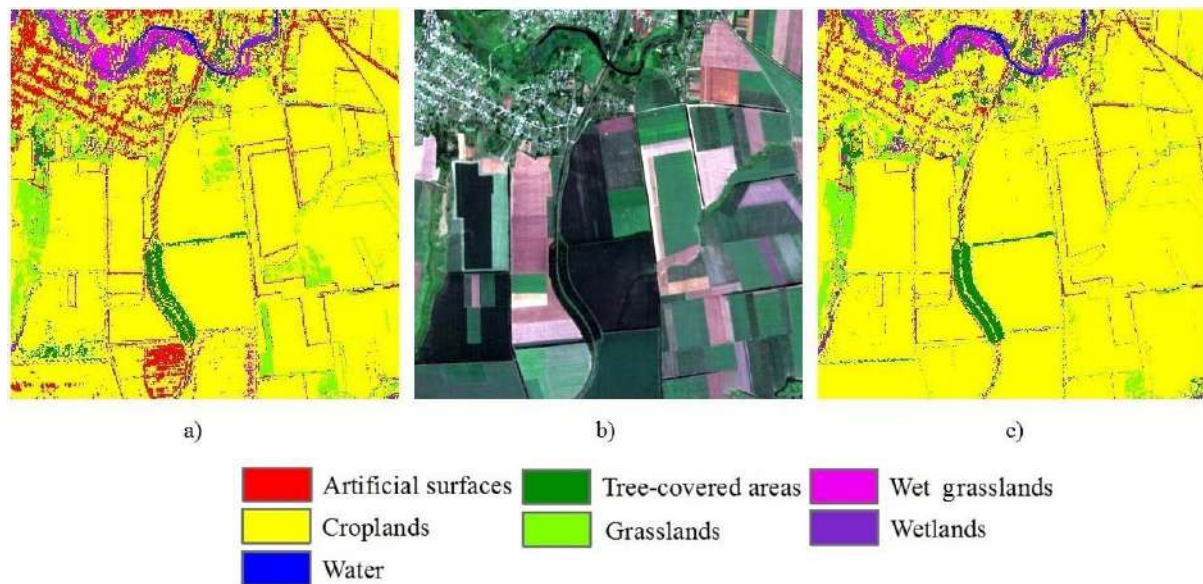


**Figure 7:** Test example: a - reference classification before the spatial resolution enhancement of spectral bands, b - a subset of the area of interest, c - final classification after the spatial resolution enhancement of spectral bands

In the presented test section, you can see that the class divergence has increased, as a result of which there are no obvious errors in the final classification when the pixels of the artificial object class appear in the middle of the field class. All linear extended features, such as forest belts and roads, retained their shape and got rid of distortions caused by larger pixels from bands with a lower spatial resolution.

## 5. Conclusion

The article presents a solution to a radiometric obstacle by using a hybrid approach to classification, as well as a solution to a spatial obstacle by the spatial resolution enhancement of lower-resolution spectral bands. Radiometric-spatial feedback has also been established.

Further research should be aimed at developing an algorithm for determining the most appropriate clustering method for each particular class. And to increase the quality of the spatial resolution enhancement in all spectral bands, it is planned to use spectral signatures databases.

## 6. References

- [1] M. Popov, S. Michaelides, S. Stankevich, A. Kozlova, I. Piestova, M. Lubskiy, O. Titarenko, M. Svideniuk, A. Andreiev, S. Ivanov, Assessing long-term land cover changes in watershed by spatiotemporal fusion of classifications based on probability propagation: The case of Dniester river basin, *Remote Sensing Applications: Society and Environment*, vol. 22, 2021, 100477. doi:10.1016/j.rsase.2021.100477.
- [2] M. O. Popov, M. V. Topolnytskyi, O. V. Titarenko, S. A. Stankevich, A. A. Andreiev, Forecasting Gas and Oil Potential of Subsoil Plots via Co-analysis of Satellite, Geological, Geophysical and Geochemical Information by Means of Subjective Logic, *WSEAS Transactions on Computer Research*, vol. 8, 2020, pp. 90–101.
- [3] G.M. Foody, Impacts of ignorance on the accuracy of image classification and thematic mapping, *Remote Sensing of Environment*, vol. 259, 2021, 112367. doi:10.1016/j.rse.2021.112367.
- [4] S.I. Toure, D.A. Stow, H. Shih, J. Weeks, D. Lopez-Carr, Land cover and land use change analysis using multi-spatial resolution data and object-based image analysis, *Remote Sensing of Environment*, vol. 210, 2018, pp. 259–268. doi:10.1016/j.rse.2018.03.023.

- [5] A.A. Andreiev, Hybrid approach to classification of remote sensing data, *CERes Journal*, vol. 6 (2), 2020, pp. 32–37.
- [6] S. Theodoridis, K. Koutroumbas, *Pattern Recognition*, fourth edition, Elsevier, 2009, pp. 633–635.
- [7] B. Tso, P. M. Mather, *Classification Methods for Remotely Sensed Data*, 2nd ed., CRC Press, Boca Raton, USA, 2009.
- [8] L. Bruzzone, B. A. Demir, A review of modern approaches to classification of remote sensing data, *Land use and land cover mapping in Europe*, I. Manakos, M. Braun, Eds. Springer: Dordrecht, Netherlands, 2014, pp. 127–143.
- [9] O. Rukundo, H. Cao, Nearest neighbor value interpolation, *International Journal of Advanced Computer Science and Applications*, 2012, pp. 25–30. doi:10.14569/IJACSA.2012.030405.
- [10] R. Keys, Cubic convolution interpolation for digital image processing, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 29, 1981, pp.1153–1160.
- [11] I. Piestova, S. Stankevich, J. Kostolny, Multispectral imagery superresolution with logical reallocation of spectra, *Proceedings of the International Conference on Information and Digital Technologies (IDT 2017)*, IEEE, Žilina, 2017, pp. 322–326. doi:10.1109/DT.2017.8024316.
- [12] E. Zaitseva, I. Piestova, J. Rabcan, P. Rusnak, Multiple-Valued and Fuzzy Logics Application to Remote Sensing Data Analysis, *26th Telecommunications Forum (TELFOR)*, Belgrade, Serbia, 2018, pp. 1–4, doi:10.1109/telfor.2018.8612109.
- [13] S. Taddeo, I. Dronova, N. Depsky, Spectral vegetation indices of wetland greenness: Responses to vegetation structure, composition, and spatial distribution, *Remote Sensing of Environment*, 2019, 111467. doi:10.1016/j.rse.2019.111467.
- [14] M. Amani, B. Salehi, S. Mahdavi, B. Brisco, Spectral analysis of wetlands using multi-source optical satellite imagery, *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 144, 2018, pp.119–136. doi:10.1016/j.isprsjprs.2018.07.005.
- [15] X. Yang, S. Zhao, X. Qin, N. Zhao, L. Lian, Mapping of Urban Surface Water Bodies from Sentinel-2 MSI Imagery at 10 m Resolution via NDWI-Based Image Sharpening, *Remote Sensing*, vol. 9, 2017, 596. doi:10.3390/rs9060596.
- [16] S. Kokhan, A. Vostokov, Using Vegetative Indices to Quantify Agricultural Crop Characteristics, *Journal of Ecological Engineering*, vol. 21, 2020, pp.120–127. doi:10.12911/22998993/119808.
- [17] T. N. Carlson, D. A. Ripley, On the relation between NDVI, fractional vegetation cover, and leaf area index, *Remote Sensing of Environment*, vol. 62, 1997, pp. 241–252.
- [18] J. Chen, C. Yang, S. Wu, Y. Chubg, A. Linton, C. Chen, Leaf chlorophyll content and surface spectral reflectance of tree species along a terrain gradient in Taiwan's Kenting National Park, *Botanical Studies*, vol. 48, 2007, pp.71–77.
- [19] Y. Kanke, B. Tubana, M. Dalen, D. Harrell, Evaluation of red and red-edge reflectance-based vegetation indices for rice biomass and grain yield prediction models in paddy fields, *Precision Agriculture*, vol. 17, 2016, pp. 507–530. doi:10.1007/s11119-016-9433-1.
- [20] Q. Cao, Y. Miao, H. Wang, S. Huang, S. Cheng, R. Khosla, R. Jiang, Non-destructive estimation of rice plant nitrogen status with Crop Circle multispectral active canopy sensor, *Field Crops Research*, vol. 154, 2013, pp. 133–144.
- [21] S.K. McFeeters, The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features, *International Journal of Remote Sensing*, vol. 17, 1996, pp. 1425–1432.
- [22] H. Xu, Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery, *International Journal of Remote Sensing*, vol. 27, 2006, pp. 3025–3033.
- [23] B. C. Gao, NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space, *Remote Sensing of Environment*, vol. 58, 1996, pp. 257–266.
- [24] Y. Zha, Y. Gao, S. Ni, Use of normalized difference built-up index in automatically mapping urban areas from TM imagery, *International Journal of Remote Sensing*, vol. 24, 2003, pp. 583–594.



# Face Recognition in the Presence of Non-Gaussian Noise

Oleg Rudenko<sup>a</sup>, Oleksandr Bezsonov<sup>a</sup> and Denys Yakovliev<sup>b</sup>

<sup>a</sup> Kharkiv National University of Radio, Nauky Ave. 14, Kharkiv, 61166, Ukraine

<sup>b</sup> Simon Kuznets Kharkiv National University of Economics, Nauky Ave.9, Kharkiv, 61166, Ukraine

## Abstract

A method for detecting a face and recognizing a person or a group of persons at pictures or videos that can be corrupted by non-gaussian noise using different architectures of convolutional neural networks is proposed. The technical details of building a deep learning-based face recognition system are discussed. The test results confirm the prospects of using the developed method for solving tasks of face detection and recognition.

## Keywords 1

Convolutional neural network, learning algorithm, facial boundaries, marker, measure of similarity

## 1. Introduction

Visual pattern recognition is one of the most important components of information management and processing systems, automated systems and decision-making systems. Tasks related to the classification and identification of objects, phenomena and signals, characterized by a finite set of certain properties and characteristics arise in areas such as robotics, information retrieval, monitoring and analysis of visual data, and are the subject of artificial intelligence. The main task of artificial intelligence is to build intelligent information systems that would have a level of effectiveness in solving informal problems that can be compared with human capabilities or those that exceed them.

When interacting with other people, a person's face is an important source of information. Facial expressions, gestures when talking, head movements are a convenient and natural way to convey information. The inability of the computer on the one hand to perceive, and on the other hand to reproduce the natural human ways of communication complicates the transmission and perception of information when working with a computer. To achieve the goal of computer recognition of head movements, facial expressions, it is necessary to implement stable algorithms for analysis and classification of digital images of human faces keeping in mind that they can be corrupted by noisy pixels.

Such algorithms can perform a wide range of commercial and non-commercial tasks. To date, there is no algorithm that could identify a human face with 100% accuracy. However, it is already possible to implement programs for detecting faces, which can be successfully used in education, medicine, security and safety, computer games, advertising and other areas.

The purpose of this work is the development and testing of a method for detecting faces and recognizing a person or a group of persons at pictures and videos that can be corrupted by non-gaussian noise based on the use of convolutional neural networks.

---

CMIS-2021: The Fourth International Workshop on Computer Modeling and Intelligent Systems, 27 of April, 2021, Zaporizhia, Ukraine  
EMAIL: oleg.rudenko@nure.ua (O. Rudenko); oleksandr.bezsonov@nure.ua (O. Bezsonov); denys.yakovliev@hneu.net (D. Yakovliev);  
ORCID: 0000-0003-0859-2015 (O. Rudenko); 0000-0001-6104-4275 (O. Bezsonov); 0000-0002-6369-6223 (D. Yakovliev);



© 2021 Copyright for this paper by its authors.  
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).  
CEUR Workshop Proceedings (CEUR-WS.org)

## 2. Recognition stages

It is generally assumed that full face recognition consists of four main steps:

1. Face detection
2. Preliminary processing
3. Isolation of signs
4. Classification

The first step is to detect faces in the image, regardless of scale and location. To do this, an advanced filtering procedure is often used to distinguish locations representing faces and filter them using some classifiers. It is noteworthy that all changes in displacement, scaling and rotation must be processed during the face detection phase. Note that facial expressions and changing hairstyles or smiling and frowning faces still present difficulties at the stage of pattern recognition [1].

While the accuracy and speed of face detection systems has improved, the two biggest challenges remain to some extent. Face detectors are essential to deal with large and complex variations of facial changes and effectively distinguish between faces and non-faces in unrestricted environments. In addition, large differences in face position and size in a large search space create problems that reduce detection efficiency [2]. This requires a trade-off between high detection accuracy and computational efficiency of the detection procedure.

In the next step, a system based on an anthropometric dataset predicts the approximate location of key features such as eyes, nose and mouth. The entire procedure is repeated to predict small features in relation to major features and is checked with collocation statistics to reject any misplaced features.

Highlighted anchor points are generated as a result of geometric combinations on the face image, and then the actual recognition process begins. It does this by finding a local representation of the facial appearance at each of the anchor points. The presentation scheme depends on the used approach.

Feature extraction usually occurs immediately after face detection and can be considered one of the most important steps in face recognition systems, since their effectiveness depends on the quality of the extracted features. This is because the facial landmarks are identified by a given network that determines how accurately features are represented. Traditional landmark locators are model-based, while many recent methods are based on cascade regression [3].

After feature extraction, face recognition is performed. The next stage of classification is simply performed using the general nearest neighbor technique, while more specific algorithms are developed for the preprocessing stage. This step usually consists of cropping, zooming, and aligning the detected faces. Unless sophisticated techniques are required for cropping and zooming, alignment is not an easy problem to solve. In practice, alignment consists in detecting a series of landmarks on the face (nose, eyes, etc.) and then transforming the face pictures so that the position of these landmarks is constant.

## 3. Convolutional neural networks for face recognition

Significant progress has been made in face detection and recognition through the use of convolutional neural networks (CNN).

The first CNN framework, known as LeNet-5, was developed in 1990 to classify handwritten digits by recognizing visual patterns in image pixels without the need for preprocessing [4]. In [5], a neural network used for vertical, frontal face recognition in grayscale was presented for the first time, which, although quite simple as for today's architectures, is comparable in accuracy to modern methods of that time.

Because face recognition differs from object recognition in that it requires alignment before extraction, this is reflected in the differences between CNNs designed for face recognition and those used for object recognition.

Because of the proven ability of deep neural network (DCNN)-based systems to outperform human performance in face-checking tasks, research in this area has skyrocketed.



Face detection and recognition research is currently focusing on DCNNs, which have demonstrated impressive accuracy on highly complex databases such as the WIDER FACE dataset [6] and MegaFace Challenge [7], as well as older databases such as Labeled Faces in the Wild (LFW) [8].

The growth in deep neural network research has been accompanied by the emergence of many deep learning frameworks and the development of the Caffe, TensorFlow, Torch, MXNET, and Theano frameworks that use platforms such as CUDA and libraries such as cuDNN. Finally, these frameworks can be used with a number of programming languages such as C++, Python, or Matlab.

Rapid progress in this direction has been driven by the increase in the availability of powerful GPUs and improvements in CNN's architecture for real-world applications. In addition, the development of large annotated datasets and a better understanding of nonlinear mapping between input images and class labels have contributed to increased research interest in these networks. DCNNs are very efficient due to their ability to approximate non-linear functions. Their significant disadvantage is high computational costs due to the presence of intensive convolution and nonlinear operations [9]. However, DCCN is expected to have a bright future, and is currently being developed by such large corporations as Google, Facebook and Microsoft [10].

Using CNN in face recognition tasks consists of two main stages: training and inference. Training is a global optimization process that involves identifying network parameters by observing huge datasets. Inference essentially involves deploying a trained CNN to classify observed data [11]. The training process includes minimization the loss function to determine the parameters of the network and the number of layers required, and the organization of connections between the layers.

CNN face recognition systems are characterized by the training data used to build the model, the network architecture and settings, and the type of the loss function [12]. DCNNs have the ability to learn highly discriminatory and invariant representations of objects when trained on very large datasets. Training is reduced to the application of optimization algorithms to minimize the loss function. In this case, the role of the loss function is to determine the forecast error, and its choice depends on the problem being solved (regression, classification, etc.). Different loss functions will produce different error values for an identical forecast and, thus, largely determine the network performance. To minimize the error, a backpropagation algorithm is used, the weights are adjusted (corrected) by using some algorithm.

Modern face recognition systems that use DCNN implement deep feature extraction and similarity comparison, which involves conversion of test images to deep representations and computation of Euclidean or cosine distance.

## 4. Convolutional neural network architecture

Initially, the structure of the convolutional neural network was created taking into account the structural features of some parts of the human brain that are responsible for vision. The development of such networks is based on three mechanisms:

- local perception;
- formation of layers in the form of a set of feature maps (shared weights);
- subsampling.

According to these mechanisms to build a convolutional neural network three main layers are applied: convolution, pooling (or subsampling), fully connected layer.

### 4.1. Convolution Layer

The convolution formula for the  $l$ -th ( $l=1,...,L$ ) layer of the network, which looks like [13]

$$x_{ij}^l = \sum_{a=-\infty}^{\infty} \sum_{b=-\infty}^{\infty} w_{ab}^l \cdot y_{(is-a)(js-b)}^{l-1} + b^l, \quad (1)$$

reflects the movement of the nucleus  $w^l$  from the input image or feature map for this layer  $y^{l-1}$  ... Here  $y_{ij}^{l-1} = f(x_{ij}^{l-1})$  – image after the  $(l-1)$ -th layer;  $f(\bullet)$  – used activation function;  $b^l$  – offset.  $i, j, a, b$  – indices of elements in matrices,  $s$  – the size of the convolution step.

As can be seen from (1), convolution operations are performed for each element  $i, j$  of image matrices  $x^l$ .

Convolution preserves spatial relationships between pixels.

Each convolutional layer is followed by a downsampling (subsampling) layer, or a computational layer, which serves to reduce the image dimension by local averaging the neuron output values.

## 4.2. Subsampling layer (pooling, MAX-pooling)

The subsampling layer reduces the scale of the planes by locally averaging the neuron outputs. Thus, a hierarchical organization is achieved. Subsequent layers extract more general characteristics less dependent on image distortion.

The pooling layer is described by the expression

$$x_{ij}^l = \sum_{a=-\infty}^{\infty} \sum_{b=-\infty}^{\infty} \beta^l \text{down}(w_{ab}^l \cdot y_{(is-a)(js-b)}^{l-1} + b^l), \quad (2)$$

where  $\text{down}(\bullet)$  – the pooling function. This function adds blocks of the input image, thus reducing the dimension of the output image of the layer. In this case, any output map is set by two offset parameters: multiplicative  $\beta^l$  and additive  $b^l$ .

The difference between the subsampling layer and the convolution layer is that the regions of neighboring neurons overlap, which does not happen in the subsampling layer.

The pooling layer operates independently of the slice depth of the input data and scales the volume spatially using a maximum function.

In the architecture of the convolutional network, it is generally accepted that the presence of a feature is more important than information about its location. Therefore, from several neighboring neurons in the feature map, the maximum is selected and its value is considered one neuron in the feature map of a lower dimension.

In addition to the maximum subsampling, pooling layers can perform other functions, such as an averaging subsampling or even an L2-normalized subsampling.

## 4.3. Non-linear activation layer

On these layers within the network, a non-linear activation function is applied to all input values  $f(\bullet)$  and the result is sent to the output. Thus, the activation layer does not change the size of the entrance.

Usually, due to significant positive properties, ReLU is used for hidden layers. ( $\text{ReLU}(x) = \max(0, x)$ ) and its various modifications (Leaky ReLU, Parametric ReLU, Randomized ReLU), and for fully

connected layer – SoftMax function (when solving classification problems)  $f_j^L = e^{x_j^L} \left( \sum_{i=1}^{N^L} e^{x_i^L} \right)^{-1}$  or linear (for regression problems).

## 4.4. Dropout layer

Various regularization methods are used to avoid overfitting the network.

Dropout is a simple and effective method of regularization and consists in the fact that in the process of training a network from its aggregate topology, a subnet is repeatedly randomly allocated,

i.e. some neurons are turned off from the process and the next update of the weights occurs only within the dedicated subnet. Thus, weights change only in the remaining neurons. Each neuron is dropped from the aggregate network with a certain probability, which is called the dropout ratio.

This layer reduces the time of one training epoch due to the smaller number of parameters to be optimized, and also makes it possible to better deal with network overfitting in comparison with standard regularization methods.

#### 4.5. Normalizing layer

On this layer, the standard normalization of the inputs occurs (the sample mean of their values is subtracted, and the result is divided by the root of the sample variance). Sample values are calculated taking into account the values at the inputs of this layer at previous training iterations. This approach allows you to increase the speed of network learning and improve the final result.

#### 4.6. Fully connected layer

This layer is a conventional multilayer perceptron, the purpose of which is classification. It simulates a complex nonlinear function, optimization of which improves the recognition quality.

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 conventional 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 an offset.

The difference between fully connected and convolutional layers is that the neurons of the convolutional layer

- 1) are connected only to the local input area;
- 2) can share parameters.

#### 4.7. Choice of training criterion

CNN training is an iterative process. Each iteration calculates the network outputs for one (or more) samples in the training set, and adjusts the network weights to reduce the error between the actual network output.  $(y_{i,p}^L)(i = 1, \dots, N^L)$  and the target output for a given sample  $d_{i,p}$ . Therefore, training is reduced to minimizing some functionality.

In practice, a quadratic function, cross entropy, or some combined functional is used as a criterion for the error function.

A backpropagation neural network (BPNN) is a multi-layer feedforward neural network that uses a supervised training algorithm known as an error backpropagation algorithm. Errors accumulated on the output layer are propagated back to the network to correct the weights. A conventional MP, which consists of three types of layers: input, hidden and output has no backward computation other than the operations used in training. All operations are performed in the forward direction during simulation.

### 5. Training of convolutional neural network

CNN training is similar to training of any direct propagation ANN and consists in correcting its weight parameters based on minimizing some selected loss function, which are usually used as a quadratic function or cross entropy.

To train convolutional neural networks, both the standard backpropagation method and its various modifications can be used. The derivation of the backpropagation algorithm for CNN training is discussed in sufficient detail in [14]. This method is based on the stochastic gradient descent algorithm (SGD).

In practice, the most common neural network learning algorithm is usually used, based on the gradient descent method (the backpropagation error method) and its modifications: SGD, Adam, AdaGrad, AdaDelta, etc. [15].

## 6. Training neurons of the output (fully connected) layer

To train the neurons of this layer, a gradient backpropagation algorithm is used

$$\theta(k+1) = \theta(k) - \eta \nabla_{\theta} J(\theta(k)), \quad (3)$$

where  $\theta$  – network parameters (elements of weight matrices, displacements, angles of inclination of activation functions, etc.);  $J(\theta(k))$  – objective function (loss function E or C);  $\eta$  – the parameter of the learning rate.

### 6.1. Training neurons when choosing a quadratic function E

If the quadratic function E is chosen as the network loss function, and the sigmoidal function is chosen as the activation function of neurons, then to adjust the weights matrix of the  $L$ -th layer following equations are used

$$\nabla_{w_{ki}^L} J(w_{ki}^L) = \frac{\partial E}{\partial w_{ki}^L} = \frac{\partial E}{\partial y_i^L} \frac{\partial y_i^L}{\partial x_i^L} \frac{\partial x_i^L}{\partial w_{ki}^L} = \delta_i^L \cdot \frac{\partial x_i^L}{\partial w_{ki}^L} = \delta_i^L \cdot y_k^{L-1}, \quad (4)$$

$$\forall i \in (0, \dots, N^L), \forall k \in (0, \dots, N^{L-1}),$$

where

$$\begin{aligned} \delta_i^L &= \frac{\partial E}{\partial y_i^L} \frac{\partial y_i^L}{\partial x_i^L}; \\ \frac{\partial E}{\partial y_i^L} &= y_i^L - d_i; \\ \frac{\partial y_i^L}{\partial x_i^L} &= y_i^L (1 - y_i^L); \end{aligned} \quad (5)$$

$d_i$  – required value of the  $i$ -th output.

If, however, as the activation function of neurons SoftMax is selected, then in (4) one should take  $\frac{\partial y_i^L}{\partial x_i^L} = y_i^L (1 - y_i^L)$ , and besides,  $\frac{\partial y_i^L}{\partial x_j^L} = -y_i^L y_j^L$ , because  $\frac{\partial E}{\partial x_i^L} = \sum_{j=1}^{N^L} \frac{\partial E}{\partial y_j^L} \frac{\partial y_j^L}{\partial x_i^L}$ ,  $i = (1, \dots, N^L)$ .

Similarly, one can get the procedure for adjusting the offsets for neurons of a fully connected layer

$$\frac{\partial E}{\partial b_i^L} = \frac{\partial E}{\partial y_i^L} \frac{\partial y_i^L}{\partial x_i^L} \frac{\partial x_i^L}{\partial b_i^L} = \delta_i^L \cdot \frac{\partial x_i^L}{\partial b_i^L} = \delta_i^L. \quad (6)$$

When training neurons of other (hidden,  $l$ -x,  $l = 1, \dots, L-1$ ) layers it is needed to calculate gradients

$$\begin{aligned} \frac{\partial E}{\partial y_k^l} &= \sum_{i=1}^{N^{l+1}} \delta_i^{l+1} \cdot \frac{\partial x_i^{l+1}}{\partial y_k^l} = \sum_{i=1}^{N^{l+1}} \delta_i^{l+1} \cdot w_{ki}^{l+1}, \\ \forall i &\in (1, \dots, N^L) \forall k \in (1, \dots, N^{L-1}). \end{aligned} \quad (7)$$

Here  $w_{ij}^{l+1}$  – the weight of the connection between the neuron  $i$  of the current (hidden) layer and neuron  $j$  of the next layer.

## 6.2. Training neurons on selection cross entropy function as a loss function

When choosing the cross entropy as the loss function, the training procedure for the output (fully connected) layer of the CNN will take the form

$$\theta(k+1) = \theta(k) - \eta \nabla_{\theta} C(\theta(k)), \quad (8)$$

where  $\theta = (w, b)$ .

Calculating the partial derivatives with respect to the tunable parameters, we have (here it is taken into account that for the output layer  $y_i^L = f_i^L$ )

$$\nabla_w C = \frac{\partial C}{\partial w_{ik}^L} = \frac{\partial C}{\partial y_i^L} \frac{\partial y_i^L}{\partial w_{ik}^L} = f_k^{L-1} (f_k^{L-1} - d_i); \quad (9)$$

$$\nabla_b C = \frac{\partial C}{\partial b_i^L} = \frac{\partial C}{\partial y_i^L} \frac{\partial y_i^L}{\partial b_i^L} = f_i^L - d_i; \quad (10)$$

$$\nabla_f C = \frac{\partial C}{\partial f_j^L} = -\frac{\partial}{\partial f_j^L} \left( \sum_{i=1}^n d_i \ln f_i^L \right) = -\frac{d_j}{f_j^L} = f_i^L - d_i. \quad (11)$$

When calculating the gradient, the following derivatives are used:

$$\frac{\partial f_i^L}{\partial x_j^L} = \begin{cases} \frac{e^{x_j^L}}{\sum_{k=1}^n e^{x_k^L}} - \left( \frac{e^{x_j^L}}{\sum_{k=1}^n e^{x_k^L}} \right)^2 & i = j; \\ \frac{e^{x_i^L} e^{x_j^L}}{\left( \sum_{k=1}^n e^{x_k^L} \right)^2} & i \neq j; \end{cases} \quad (12)$$

$$\frac{\partial C}{\partial x_i^L} = -\frac{\partial}{\partial x_i^L} \left( \sum_{j=1}^n d_j \ln f_i^L \right) = \sum_{j=1}^n \frac{\partial C}{\partial f_j^L} \frac{\partial f_j^L}{\partial x_i^L} = y_i^L - d_i^L. \quad (13)$$

## 6.3. Training neurons in the subsampling layer

The peculiarity of this layer is that it sits in front of both fully connected and convolutional layers. In the first case, it has the same neurons and connections as a fully connected one. Therefore  $\delta$  is calculated in the same way as in a hidden layer. If the subsampling layer is in front of convolutional layer,  $\delta$  is computed by the inverse convolution method using the rotation by  $180^\circ$  ( $rot_{180^\circ} \{\delta_{ij}^l\}$ ) [16–18].

## 6.4. Training of neurons of the convolutional layer

The backpropagation algorithm is also used to train the neurons of this layer.

Core is a filter that slides over the entire image and finds its features anywhere, i.e. provides invariance to displacements.

Formula for updating the convolution kernel  $w_{ab}^l$  has the form [16–18]

$$\frac{\partial E}{\partial w_{ab}^l} = \sum_i \sum_j \frac{\partial E}{\partial y_{ij}^l} \frac{\partial y_{ij}^l}{\partial x_{ij}^l} \cdot y_{(is-a)(js-b)}^{l-1}, \quad \forall a \in (-\infty, \dots, +\infty) \forall b \in (-\infty, \dots, +\infty). \quad (14)$$

Only one offset can be used for one feature map  $b^l$ , which is “connected” with all the elements of

this map. Accordingly, when correcting the value of this offset, all values from the map obtained during the back propagation of the error should be taken into account. In this case (when using one displacement for one feature map) we have

$$\frac{\partial E}{\partial b^l} = \sum_i \sum_j \frac{\partial E}{\partial y_{ij}^l} \frac{\partial y_{ij}^l}{\partial x_{ij}^l} \frac{\partial x_{ij}^l}{\partial b^l} = \sum_i \sum_j \frac{\partial E}{\partial y_{ij}^l} \frac{\partial y_{ij}^l}{\partial x_{ij}^l}. \quad (15)$$

As an alternative, one can take as many offsets  $b_{ij}^l$  for a separate feature map, as many elements are in this map, but in this case there will be too many displacement parameters (more than the parameters of the convolution kernels themselves). For this case  $\frac{\partial E}{\partial b_{ij}^l} = \frac{\partial E}{\partial x_{ij}^l} \dots$

To train neurons in the convolution layer, the gradient procedure uses the derivative  $\frac{\partial E}{\partial y_{ij}^{l-1}}$  calculated as follows:

$$\frac{\partial E}{\partial y_{ij}^{l-1}} = \sum_{i'} \sum_{j'} \frac{\partial E}{\partial y_{i'j'}^l} \frac{\partial y_{i'j'}^l}{\partial x_{i'j'}^l} \frac{\partial x_{i'j'}^l}{\partial y_{ij}^{l-1}} = \sum_{i'} \sum_{j'} \frac{\partial E}{\partial y_{i'j'}^l} \frac{\partial y_{i'j'}^l}{\partial x_{i'j'}^l} \cdot w_{(i's-i)(j's-j)}^l. \quad (16)$$

## 7. Histogram of oriented gradients (HOG)

Histogram of Oriented Gradients (HOG) – is a method of information evaluation of special points of the image, based on the calculation of the number of directional gradients in the space of these points. The basic idea of the algorithm is the assumption that the appearance and shape of the object in the image area can be described by the distribution of intensity gradients or the direction of the edges [19]. Such a description is carried out by dividing the image into cells and constructing histograms of directional gradients of cell pixels. The result of the algorithm is a descriptor, which includes a combination of the obtained histograms.

HOG algorithm contains following stages:

1. Gradient calculation.
2. Calculation of the histograms of the image cells.
3. Forming and normalizing descriptor blocks.
4. The final step is to classify the HOG descriptors using a training system with a teacher.

The result of the classifier's work is two object images with positive and negative weights of reference vectors. Positive weights mean that the features belong to the target class, and negative weights mean that the features belong to the background.

## 8. SSD network

The SSD method is based on such CNN architectures as Faster R-CNN and YOLO, however, the authors of [20] took into account their shortcomings, thanks to which SSD could achieve new peaks of accuracy and speed. The method is based on the Feed-forward convolutional network, which creates a finite set of limiting rectangular frames and quantitative estimates of the presence of objects of various classes in these frames, after which it produces the suppression of maximums ("Non-maximum suppression predictors"). The structure of the network is shown in Fig. 1 and can be divided into four functional parts:

1. Network input. It accepts a three-channel (color) image with a size of 300x300.
2. Backbone network. Some standard architecture for image classification is usually used as a backbone network. In this case it is VGG-16. However, the fully connected layer of VGG-16 is not used in SSD. This architecture is described in detail in [20].

3. Layer of additional features. These are convolutional layers and subsampling layers that generate feature maps of different resolution for detecting objects of different scales.

4. At this stage operations are performed to merge the outputs of various layers and suppress maximums to obtain the output of the network.

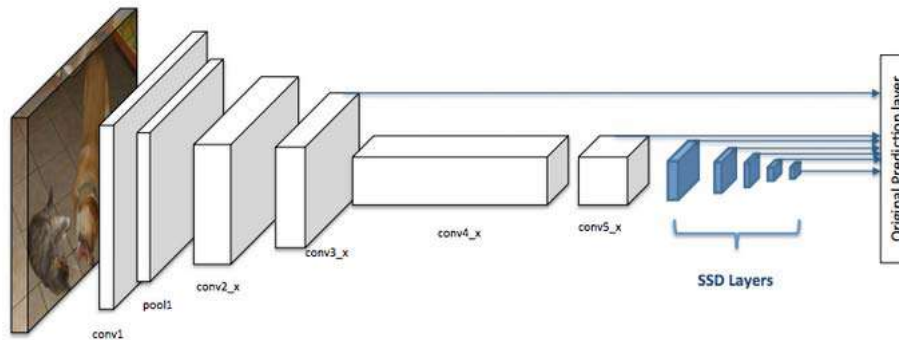


**Figure 1:** The architecture of the SSD network [20]

The SSD method has the following key features:

1. Detection of objects of various scales. Several convolutional layers of different sizes are sequentially added to the base network, which make it possible to obtain maps of signs of different resolutions and, accordingly, to predict objects at different scales (unlike, for example, YOLO, which receives only one map).

2. Convolutional predictors for detection. Each feature layer generates a finite set of assumptions about the class and position of the object using a set of convolutional filters. Each such convolutional filter has a kernel that determines the probability of belonging to the default frame offset class itself.

3. Default bounding frames. The default set of frames is linked to the cards of signs of different resolutions described in point 1: each cell of the card of signs corresponds to a frame from the set by default. For each cell of the feature map, estimates are determined according to the class of objects, as well as four offsets of the default frame relative to its initial position.

Combining these properties allowed efficiently sampling many different forms of the resulting bounding frames, which had a positive impact on the speed of the network.

## 9. MTCNN network

Of the neural network approaches in face detection, Multi-task Cascaded CNN (MTCNN) is especially significant. This network is described in details in [21].

## 10. RetinaFace Network

The architecture of the RetinaFace network [22] is shown in Fig. 2. This network is a single-stage face detector that detects faces through collaborative, supervised and self-guided multi-tasking learning. The network has several features that improve the ArcFace network's face detection performance and surpass most network architectures in the vast majority of tests for such systems:

- RetinaFace uses five additional key points on faces to dramatically improve facial recognition results;
- RetinaFace has a self-checking grid decoder branch to predict per-pixel information about 3D face shapes in parallel with existing controlled branches;
- By utilizing lightweight backbones, RetinaFace delivers performance that allows it to be run on the CPU, while most similar systems achieve such performance solely using the GPU.

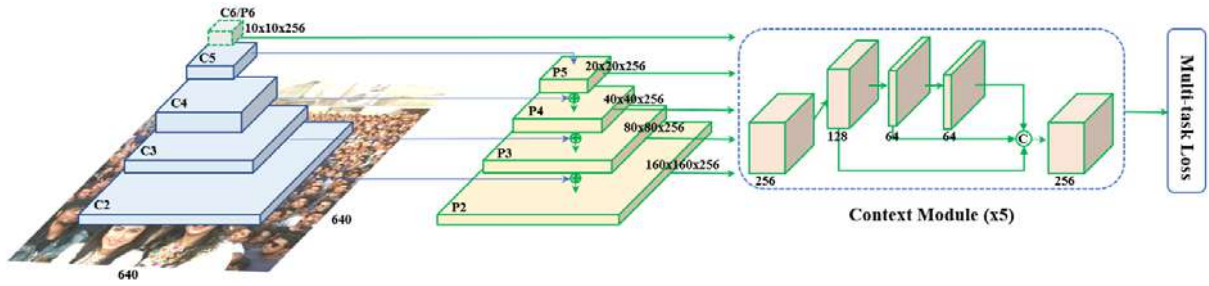


**Figure 2:** The architecture of the RetinaFace network [22]

## 11. Simulation

The face recognition system consists of the following steps:

1) Locate all the faces in the image.

It applies the location method to the given image and returns a list of bounding boxes of human faces at the image. The HOG method and CNN based methods (SSD, MTCNN, RetinaNet) are used as face detectors. HOG is less accurate than convolution networks models, but much faster on CPUs.

2) Face encoding (features extracting).

The method performs face landmark estimation and extract specific points from the found face. This will make it a lot easier to compare faces and not worry about the problem when faces turned in different directions look totally different for a vision system.

Found specific points are passed to the pre-trained CNN model, which generates a 128- or 256-dimension face encoding for each face at the image.

3) Matching the encoded faces with a database of known people.

Finally, the method compares the list of face encodings against to the candidate encoding to see if they match. Given the list of face encodings, compare them with the known face encoding and get the Euclidean distance for each comparison face. The distance indicates how similar the faces are. After that, return a list of True/False values indicating which known face encodings match the face encoding to check.

**Experiment 1.** A comparison was made between different face detectors in combination with different parameters of the RNET feature encoder. So the following detectors were used:

1. The Retina MNET model, which is a lightweight model for finding facial boundaries with fast coordinate regression. Given the found face boundaries, it returns 2d-106 face markers. A 192x192 image was fed to the network input.

2. The MTCNN model used to find facial boundaries and regression of facial markers. Given the found face boundaries, it returns 5 exact markers.

3. The SSD model, which is the fastest neural network model for object detection.

4. The RNET encoder accepts the locations of the faces, separates the face by the found coordinates. Using the resulting face landmarks (eye sockets – 2, nose tip – 1, lip edges – 2) aligns the picture to obtain a frontal view of the face. Then the result is reduced to 112x112 pixels, after which it is processed by the network. At the output of the network, we get a vector of face features with dimension 512.

To compare and identify the obtained representation of the encoder's face with the entire base of faces, a "cosine coefficient" is used – a binary similarity measure.

The results of the system operation on a test set of 500 images with various combinations of detection and recognition methods are shown in Table 1. Some selected images are presented in Fig. 3–4.

**Experiment 2.** Experiments were carried out to recognize faces in images with one face, a group of faces and in the presence of non-gaussian noise (salt-and-pepper). RetinaFace MNET network and the arc512d recognition method were used as a detector.

Some selected results are shown in Fig. 5.

The recognition accuracy for 500 test images was 94.7%, which indicates a high noise canceling ability of the method in question.

The results for other combinations of detectors and recognition methods are presented in Table 2.



**Table 1**

System performance results

| Method  | HOG    | SSD    | MTCNN  | RetinaFace |
|---------|--------|--------|--------|------------|
| fr128d  | 86.67% | 90.67% | 92.67% | 96%        |
| arc512d | 87%    | 93.33% | 95.67% | 96.67%     |

**Table 2**

System performance results with noise presents

| Method  | HOG    | SSD    | MTCNN  | RetinaFace |
|---------|--------|--------|--------|------------|
| fr128d  | 73.46% | 86.23% | 90.65% | 91.25%     |
| arc512d | 75.23% | 88.54% | 91.88% | 94.70%     |

**Figure 3:** Results of recognition of single objects

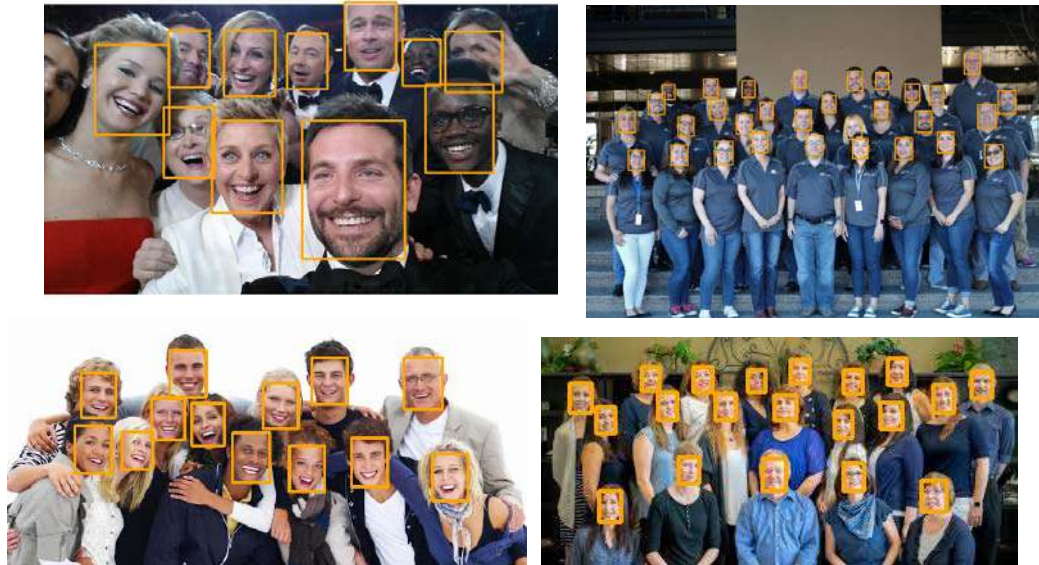


**Figure 4:** Results of recognition of a group of objects

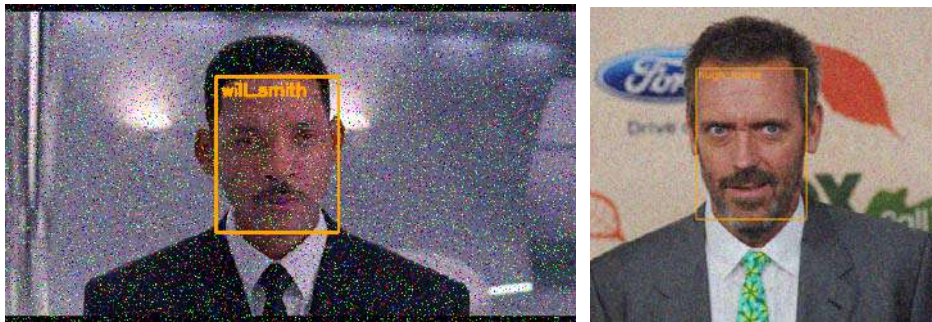


**Figure 5:** Results of recognition of noisy images

## 12. Conclusions

Analysis of the obtained results allows us to draw the following conclusions:

1. Increasing the number of more boxes leads to more accurate detection, while reducing its speed
2. The presence of MultiBox on several layers also leads to better detection, due to the fact that the detector works with several resolutions at the same time;
3. 80% of the time is required to perform computations on the VGG-16 core network. This means that with a faster and equally accurate network, the performance of an SSD can be significantly improved;
4. SSD mixes objects with similar categories (for example, animals). This is probably because cue points are common across multiple classes
5. The worst performance of the SSD network is observed for small objects, as they may not appear on all object maps. Some increase in network performance can be achieved by increasing the resolution of the input image.
6. The SSD network is quite effective for solving the problem of face detection both in the absence of interference and in the presence of non-gaussian interference.

## 13. Acknowledgements

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

## 14. References

- [1] B. Yang, J. Yan, Z. Lei, S. Li, Aggregate channel features for multi-view face detection, in IEEE International Joint Conference on Biometrics, 2014, pp. 1-8.
- [2] S. Zhang, et al., Detecting Face with Densely Connected Face Proposal Network, 2017, pp. 3-12.
- [3] R. Ranjan, et al., Deep Learning for Understanding Faces: Machines May Be Just as Good, or Better, than Humans. Signal Processing Magazine, IEEE, 2018.35 (1), p. 66-83.
- [4] Y. LeCun et al., Handwritten digit recognition with a back-propagation network, 1990.
- [5] H.A. Rowley, S. Baluja, T. Kanade, Neural network-based face detection. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 1998.20 (1), p. 23-38.
- [6] S. Yang, et al., WIDER FACE: A Face Detection Benchmark, 2015.
- [7] I. Kemelmacher, et al., The MegaFace Benchmark: 1 Million Faces for Recognition at Scale, 2016, pp. 4873-4882.
- [8] G.B. Huang, et al., Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments, in Workshop on Faces in 'Real-Life' Images: Detection, Alignment, and Recognition, Marseille, France, 2008.
- [9] S. Wu, et al., Funnel-structured cascade for multi-view face detection with alignment-awareness. Neurocomputing, v. 221, 2017, p. 138-145.
- [10] A. Krizhevsky, I. Sutskever, G. Hinton., ImageNet classification with deep convolutional neural networks. Communications of the ACM, 2017.60 (6), p. 84-90.
- [11] A. Parashar, et al., SCNN: An Accelerator for Compressed-sparse Convolutional Neural Networks. ACM SIGARCH Computer Architecture News, v.45 (2), 2017, pp. 27-40.
- [12] J. Deng, J. Guo, S. Zafeiriou. ArcFace: Additive Angular Margin Loss for Deep Face Recognition, 2018.
- [13] Z. Zhang., Derivation of Backpropagation in Convolutional Neural Network (CNN), Technical Report. University of Tennessee, Knoxville, TN, October 18, 2016. URL: <http://web.eecs.utk.edu/~zzhang61/docs/reports/2016.pdf>
- [14] Y. LeCun, Y. Bengio, Convolutional networks for images, speech, and timeseries, The Handbook of Brain Theory and Neural Networks, 1995, pp. 255-258.  
O. Rudenko, O. Bezsonov, K. Oliinyk, First-Order Optimization (Training) Algorithms in Deep Learning Proceedings of the 4th International Conference on Computational Linguistics and Intelligent Systems (COLINS 2020). Volume I: Main Conference Lviv, Ukraine, April 23-24, 2020, pp. 921-935.
- [15] Deep Learning for Image Processing Applications, Ed. Hemanth DY, Estrella VV, 2017, 273 p. URL: <http://ebooks.iospress.nl/volume/deep-learning-for-image-processing-applications>
- [16] Deep learning tutorial, Stanford University. Autoencoders. URL: <http://ufldl.stanford.edu/tutorial/unsupervised/Autoencoders/>
- [17] Convolutional network in python. Part 2. Derivation of formulas for training the model. URL: <https://habr.com/ru/company/ods/blog/344116/>
- [18] P. Viola, M.J. Jones, and D. Snow, Detecting pedestrians using patterns of motion and appearance, The 9-th ICCV, Nice, France, v.1, 2003 , pp. 734-741.
- [19] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, SSD: Single Shot MultiBox Detector, 2016. URL: <https://arxiv.org/abs/1512.02325>.
- [20] K. Simonyan, A. Zisserman, Very Deep Convolutional Networks for Large-Scale Image Recognition. URL: <https://arxiv.org/abs/1409.1556v6>
- [21] K. Zhang, Z. Zhang, Z. Li, Y. Qiao, Joint Face Detection and Alignment using Multi-task Cascaded Convolutional Networks. URL: <https://arxiv.org/pdf/1604.02878>
- [22] J. Deng, J. Guo, Y. Zhou, J. Yu, I. Kotsia, S. Zafeiriou, RetinaFace: Single-stage Dense Face Localisation in the Wild. URL: <https://arxiv.org/abs/1905.00641>

# Formal Models of Question-Answering Machine

Igor A. Chimir<sup>a</sup>, and Anatolii F. Verlan<sup>b</sup>

<sup>a</sup> Odessa State Environmental University, 15 Lvovskaya St., 65016, Odessa, Ukraine

<sup>b</sup> The Georgy Puchov Institute for Energy Modeling. National Academy of Science, 15 General Naumov St., 03164, Kiev, Ukraine

## Abstract

The article describes two models of question-answering dialogue machine: (1) model based on the idea of Mealy finite automata, and (2) model based on the idea of Petri net. Both models are problem-independent and describe question-answering dialogue process, which is independent of the subject area of the dialogue. The problem independence of the models is a consequence of a unified cognitive cycle of dialogue used in them. The unified cognitive cycle of dialogue is similar to Neisser's cyclical model of perception. Models are designed to specify a "dialogue machine" that simulates a goal-oriented behavior of the active dialogue agent when solving a problem by means of a question-answering dialogue. Implementation of the models presupposes data-driven approach when main components of the "dialogue machine" are not map in the program code but represented by data stored in the database.

## Keywords 1

Question-answering dialogue, Dialogue agent, Dialogue script, Cognitive cycle of dialogue.

## 1. Introduction

Dialogue communication is one of the phenomena of human mental activity and is a subject of study in such sciences as artificial intelligence, cognitive psychology, erotetic logic and epistemology.

The Turing intelligence test is based on a dialogue process between a human and a technical system [1]. Modern presentation of artificial intelligence is based on the concept of intelligent agents [2]. Dialogue communication between the environment and the intelligent agent determine the behavior of the latter.

In cognitive psychology, one can find theories and models that contribute to understanding the essence of dialogue. Neisser's model of perception is a model of the dialogue process if the source of the flow of sensory events is one of the dialogue partners [3].

The vehicle for the exchange of knowledge between dialogue partners is a question-answering pair, therefore erotetic logic provides an apparatus for modeling the logical structure of a dialogue [4].

Epistemology is related to understanding the essence of the dialogue process, since it studies those types of knowledge that circulate within the question-answer pair. Dialogue plays a key role in Socratic epistemology [5].

Question-answering dialogue is one of the types of dialogue that can be operated by artificial dialogue agents. In a question-answering dialogue, the information messages of the active agent represent questions, and the information messages of the reactive agent represent the answers to these questions. The concepts of "question" and "answer" should be understood not in the narrow, linguistic sense of the word, but in the broad, behavioral sense. The question is, in the general case, the interrogative stimulus of the active agent, and the answer is the reaction of the reactive agent.

A feature of the question-answering dialogue is its ability to serve as a means of solving a number of ill-formalized problems. Newell was the first to draw attention to the need to develop methods for

---

CMIS-2021: The Fourth International Workshop on Computer Modeling and Intelligent Systems, April 27, 2021, Zaporizhzhia, Ukraine

EMAIL: chimir@gmail.com (I. Chimir); a.f.verlan@gmail.com (A. Verlan)

ORCID: 0000-0001-5326-768X (I. Chimir); 0000-0002-8565-2037 (A. Verlan)



© 2020 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)



solving ill-formalized problems [6]. Newell believed that ill-formalized problems are those that had one or more of the following characteristics: problems cannot be specified numerically; goals cannot be expressed in terms of well-defined objective functions; there is no algorithmic solution to the problem; an algorithmic solution exists, but it cannot be used due to limited computer resources. In applied systems of artificial intelligence, there are systems oriented on solving ill-formalized problems. One of the most extensive classes of such systems is intelligent tutoring systems [7]. The problem of teaching and knowledge transfer is ill-formalized and can be solved by dialogue methods.

The pragmatic dimension in the modeling of question-answering dialogue processes is aimed at designing systems focused on solving ill-formalized problems and systems called "chatbots" [8].

In the field of theory and practice of dialogue processes, focused on solving ill-formalized problems, an important question is how to create, store and repeatedly reproduce dialogue methods. In a goal-oriented dialogue process, in which the solution of an ill-formalized problem is carried out, the active agent plays a key role, since it is the active agent who is the carrier of the method for solving the problem. The article is devoted to the synthesis and description of cognitive and formal problem-independent dialogue models that can generate protocols of goal-oriented question-answering dialogues. Formal models have applied value and can be used as a basis for the development of artificial dialogue agents.

## **2. Cognitive cycle of question-answering dialogue**

When developing a dialogue model, it is important that the model is adequate to the processes of perception and processing of information by a human. In the case when the basis of the formal model of dialogue is an adequate psychological model, then we can expect that the artificial dialogue agent inherits the anthropomorphic properties of the system of perception of a human.

Among the theories and models proposed by cognitive psychology and related to the dialogue process, let us dwell on the models that describe dialogue at the level of perception, which does not depend on the subject area of the dialogue and on the nature of the problem, which is solved in the dialogue process. In this sense, a psychological model of dialogue, useful in an applied aspect, should describe some unified cognitive "dialogue machine".

The dialogue process is similar to the process of perceptual human interaction with the environment, modeled by the Neisser cycle [3]. The difference lies in the fact that in the process of dialogue, the main components of the human sensory system - sight and hearing - are connected not to the "natural" environment, but to the "artificial" environment formed by streams of visual and sound sensory events generated by the opposite agent of the dialogue. Thus, in the dialogue process, the real environment is replaced by an artificial one. However, it is obvious that the perception and subsequent processing of both sensory events generated by the artificial environment (dialogue agent) and sensory events generated by the natural environment are carried out according to the same "rules and laws". A diagram illustrating the cognitive cycle of a question-answering dialogue is shown in Figure 1.

The cognitive cycle of the dialogue process, shown in Figure 1, simulates the behavior of an active agent in a dialogue with fixed roles. Fixing agent roles means prohibiting changing agent roles during a conversation. One of the agents always asks questions, and the other always answers them. For convenience, we will assume that the dialogue cycle begins from the moment when the active agent operates with a relatively small set of answers, called the set of expected answers, and the real answer of the reactive agent coincides with one of the answers from the set of expected answers of the active agent. Thus, one of the main cognitive structures that the active agent operates is the schema-answer, or mental representation of the answer. The set of expected answers is embedded in a broader cognitive framework called the active agent script. It is assumed that the script of the active agent includes all the schema-answers necessary for the given dialogue, and that the set of expected answers is some subset of answers from the script that is relevant to the current cycle of the dialogue process. After the perception and recognition of the real answer received from the reactive agent, the current set of expected answers is modified, the purpose of which is to prepare the set of expected answers for the next cycle. As a result of the modification, a new set of expected answers is formed, including the

answer expected on the next cycle. During the modification process, the current set of expected answers can be replaced in whole or in part.

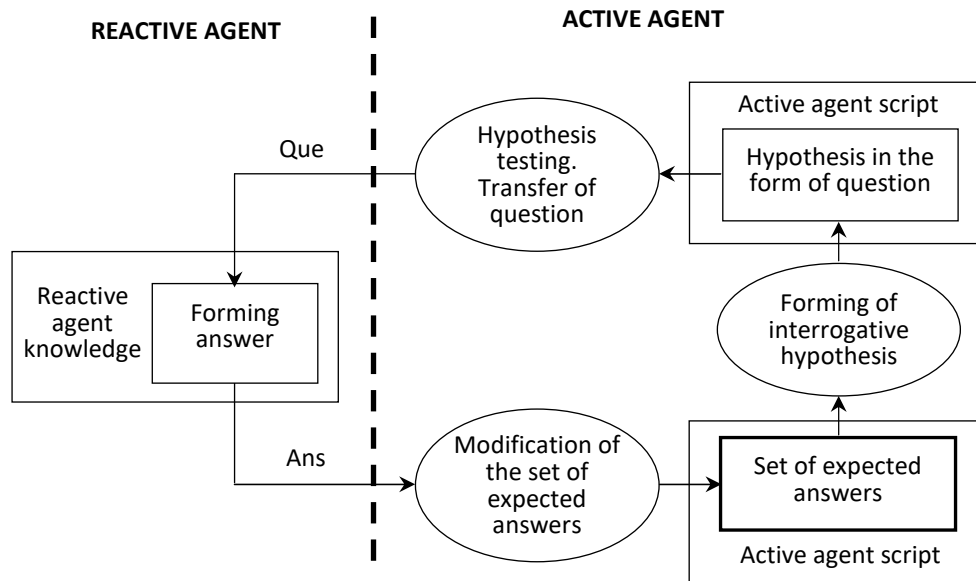


**Figure 1:** Cognitive cycle of a question-answering dialogue process

The concept of a dialogue script means that the active agent does not generate new question using some "super-algorithm", but searches for it in the memory of questions, using the method of achieving the goal of the dialogue as a method of accessing the memory of questions. Thus, the method of accessing the memory of questions is a repository of the method for achieving the goal of a dialogue, or a dialogue method for solving a problem. The concepts "memory of questions" and "dialogue method of solving a problem" are included in the scope of the concept of "script" of an active agent and detail its structure.

Introspection and analysis of real protocols of question-answering dialogue processes reveals that there is an ambiguous relationship between the current answer of the reactive agent and the subsequent question generated by the active agent. In other words, in different dialogue transactions for the same answer received from the reactive agent, the active agent can generate different questions. Therefore, when developing dialogue methods for solving problems, for modeling the noted ambiguity, it is necessary to take into account the following three principles.

**The principle of "depth of dialogue".** The principle of "depth of dialogue" means that the active agent, when forming the next question, must take into account both the perceived answer and the index of the dialogue cycle. Here, the term "index" is used synonymously with the term name, or identifier. In different dialogue cycles, different questions can be generated for the same answer received from the reactive agent.

**The principle of "history of answers".** The principle of "history of answers" means that the active agent, when forming the next question, must take into account both the perceived answer and the history of the received answers. In the same dialogue cycle, different questions can be generated for the same answer received from the reactive agent, depending on what answers were received in the previous cycles.

**The principle of "history of questions".** The principle of "history of questions" means that the active agent, when forming the next question, must take into account both the perceived answer and the history of previously formed questions. In the same dialogue cycle, different questions can be formed for the same answer, depending on what questions were formed in the previous cycles.

The dialogue cycle shown in Fig. 1 describes well the "harmonious dialogue" corresponding to the Neisser cycle for the case of routine perception. By harmonious dialogue we mean such a dialogue when both agents are satisfied with their roles and do not want to change them. However, a harmoni-

ous dialogue is not always possible. The initiator of the role change is usually the reactive agent, and the sign of the moment of the role change is the generation and transmission to the active agent of an information message with the status of a question. Therefore, one of the ways to take into account the possibility of changing roles can be to include a question detector in the list of expected answers. A question detector can be included in each set of expected answers (then agents will be able to change roles in any dialog cycle) or only in some sets of expected answers (then the roles of agents can be changed only in some predefined dialogue cycles).

Considered, further, formal problem-independent models of dialogue are models of the dialogue method of accessing the memory of questions, considered also as a dialogue method for solving a problem. These models model the behavior of an active agent of a question-answering dialogue, as the functioning of some "dialogue machine". The purpose of modeling is to identify the components of the "dialogue machine" that is invariant to the problem being solved and to the subject area of the dialogue process.

### 3. Finite-automaton model

The finite-automaton model of question-answering dialogue simulates a harmonious dialogue and is based on three assumptions: (1) the problem solved during the question-answering dialogue can be solved in a finite number of dialogue cycles (dialogue steps); (2) by the beginning of solving the problem, all questions and answers necessary for the formation of dialogue steps are determined; (3) modification of the current set of expected answers always generates a new set that is different from the current one and has a unique index.

The finite-automaton model is a variant of the Mealy automaton [9] and is described by the following formula

$$\text{DiAM} = (\text{QUE}, \text{ANS}, \text{S}, \varphi) \quad (1)$$

where DiAM denotes a dialogue access method to the memory of questions.

$$\text{QUE} = \{\text{Que}_i\}; i = 1, \dots, k \quad (2)$$

QUE – a set of indexes of questions used to solve a problem.

$$\text{ANS} = \{\text{Ans}_i\}; i = 1, \dots, l \quad (3)$$

ANS – a set of answers used to solve a problem.

$$\text{S} = \{\text{S}_i\}; i = 1, \dots, m \quad (4)$$

S – a set of states of waiting for an answer. The number of elements of the set S is equal to the total number of dialogue steps.

$$\varphi : \text{S} \times \text{ANS} \rightarrow \text{S} \times \text{QUE} \quad (5)$$

$\varphi$  – step function that determines the index of the new question and the index of the next step, depending on the current answer and the index of the current step.

When performing a step, an automaton that simulates a question-answering dialogue process performs iterations consisting of the following sequence of actions: (1) perceived the current answer of the reactive agent; (2) the current answer of the reactive agent is recognized; (3) the index of the next question of the active agent is determined.

If the set of answers ANS is not structured in any way, this means that in order to recognize the current answer of the reactive agent, the dialog access method must, at each step, operate with the entire set of answers ANS. For large scripts, the cardinality of this set can be significant. In order to reduce the cardinality of the set of answers that DiAM operates with at each step, and to make the model practically realizable, we introduce into the model a set of answers recognizable at the  $i$ -th step. The set of recognizable answers contains only those answers that are expected at the  $i$ -th step, are necessary for the implementation of the dialogue method and, therefore, must be recognized. All other answers will be referred to the class of answers that are not recognizable at the  $i$ -th step. We will use the following

notation:  $R^i$  – is the set of answers recognizable at the  $i$ -th step;  $NR^i$  – is the set of answers not recognizable at the  $i$ -th step.

The set of answers that are not recognizable at the  $i$ -th step includes all answers belonging to the set  $ANS$  and not belonging to the set  $R^i$ , i.e.

$$NR^i = ANS - R^i \quad (6)$$

However, the set of answers that are not recognizable at the  $i$ -th step must be considered wider and understood as the set of any conceivable answers of a reactive agent that do not belong to the set  $R^i$ , i.e.

$$NR^i = U - R^i \quad (7)$$

where  $U$  – is the set of any responses that the reactive dialogue agent can ever generate.

It is natural to assume that the appearance at the  $i$ -th step of any answer belonging to  $NR^i$  leads to the fact that DiAM generates the same index of the next question. Otherwise, it would mean that the answers from the  $NR^i$  belong to the class of recognizable answers. Therefore, in the step function  $\phi$  for each step of the dialogue, the entire class of unrecognizable answers will be modeled with one unrecognizable answer.

The above reasoning allows us to return to the cycle in Figure. 1, and clarify the structure of the set of expected responses. At each step of the dialogue, the set of answers that the active agent expects can include responses from the following three classes: (1) the class of recognizable answers; (2) the class of unrecognizable answers (modeled by one answer); (3) the class of question detectors (to determine the need for a role reversal).

Let us consider an example that will allow, firstly, to assess how useful the finite-automaton model of the question-answer dialogue is when constructing a dialogue access method to the memory of questions, and, secondly, will provide a transition to the subsequent model in the form of a Petri net. The example includes several steps.

Step 1. The active agent generates a question with index  $Que_1$  and expects to receive the following answers:

$Ans_1$  – request for a hard copy of a question;

$Ans_3$  – answer followed by a question with an index  $Que_3$ ;

$Ans_4$  – answer followed by a question with an index  $Que_4$ ;

$Ans_5$  – unrecognizable answer (any answer other than  $Ans_1$ ,  $Ans_3$ ,  $Ans_4$ ). Answer  $Ans_5$  means that it is necessary to generate a question with  $Que_2$  index.

Step 2. The active agent generates a question with index  $Que_2$ , which is a reformulated question with index  $Que_1$ . In this step, the active agent expects to receive the same answers as in the first step, as well as an  $Ans_2$  answer, which is a request to return to step 1.

DiAM is described using the following formulas

$$ANS = \{Ans_1, Ans_2, Ans_3, Ans_4, Ans_5\} \quad (8)$$

$$QUE = \{Que_1, Que_2, Que_3, Que_4\} \quad (9)$$

$$S = \{S_1, S_2, S_3, S_4\} \quad (10)$$

$$\phi : \left. \begin{array}{l} (S_1, Ans_1) \rightarrow (S_1, Que_1), \\ (S_1, Ans_3) \rightarrow (S_3, Que_3), \\ (S_1, Ans_4) \rightarrow (S_4, Que_4), \\ (S_1, Ans_5) \rightarrow (S_1, Que_2), \\ (S_2, Ans_1) \rightarrow (S_2, Que_2), \\ (S_2, Ans_2) \rightarrow (S_1, Que_1), \\ (S_2, Ans_3) \rightarrow (S_3, Que_3), \\ (S_2, Ans_4) \rightarrow (S_4, Que_4), \\ (S_1, Ans_5) \rightarrow (S_2, Que_2), \end{array} \right\} \quad (11)$$



The step function  $\phi$ , represented by mappings (11), on the one hand, describes the “logic” of access to the memory of questions, and on the other, the “logic” of the dialogue method for solving a certain problem. The step function can be presented in tabular form. The tabular presentation of the step function is convenient in that it allows you to easily switch to a graphical representation of the question-answering dialogue in the form of a state diagram. An example of a graphical representation of a question-answering dialogue in the form of a state diagram is shown in figure 2.

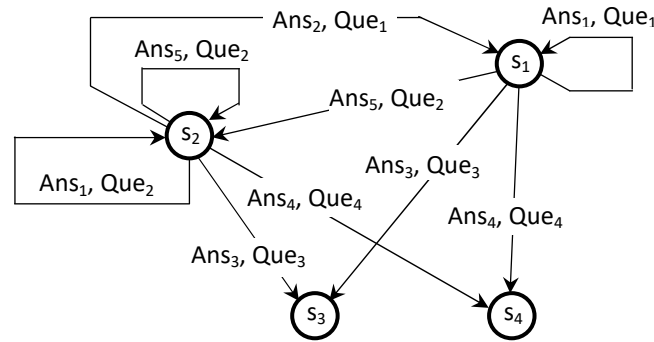


**Figure 2:** Question-answering dialogue state diagram for the example described in the text

In the state diagram shown in Figure 2, the set of vertices corresponds to the set of steps of the question-answering dialogue (or to the set of stable states), and each arc corresponds to one of the possible options for the development of the dialogue, determined by the answer of the reactive agent. Arcs are marked by pairs  $Ans_i, Que_i$ . Where  $Ans_i$  – one of the expected answers;  $Que_i$  – question index.

As noted earlier, the step function can be considered as a way of describing the "logic" of the dialogue problem solving method, and in this sense, a state diagram, an example of which is shown in Figure 2 can be viewed as a diagram describing the behavior of an active agent in the process of solving a problem.

#### 4. Model in the form of a Petri net

The finite-automaton model of question-answering dialogue has its own sphere of applicability and can be used to specify a number of applied dialogue systems. However, the finite-automaton model has disadvantages that appear at the stage of its computer implementation.

An effective approach to the computer implementation of question-answer dialog systems is the datalogical approach, which, if applied to the design of the dialog access method DiAM, involves the implementation of the step function not in the form of a program code, but in the form of a mapping into a certain database. A decisive advantage of the datalogical approach is the ability to easily modify the dialogue method using full-screen editing tools. Thus, the datalogical approach excludes the process of compiling the source code when modifying the system and allows the dialogue method to be modified by its author.

From the standpoint of the datalogical approach, the formalism adopted to describe the dialogue method must be identified with the conceptual model of some database. As noted in [10], the problem of the conceptual database model is associated with such a representation of the model, which, on the one hand, most naturally reflects the subject area (in our case, the dialogue method of accessing the memory of questions), and, on the other hand, can be supported by computer means. In other words, a formalism is needed that can be easily transformed into a data schema. The experience of designing dialogue applications allows us to conclude that the noted property is inherent not in the state diagram of the dialogue access method (an example of which is shown in Figure 2), but in the graph modeling the dialogue access method in the form of a Petri net [11].

The Petri model of the dialogue method, like the automaton model, is based on three assumptions: (1) the problem solved in the dialogue process can be solved in a finite number of dialogue steps; (2) by the beginning of the solution of the problem, all questions and answers necessary for its solution are determined; (3) modification of the current set of expected answers of the reactive agent, in any case, generates a set of answers that is different from the current one and, therefore, has a unique index. We describe the Petri-model of the dialogue method with the following four:

$$\text{DiAM} = (\text{QUE}, \text{ANS}, \text{NextQue}, \text{NextAns}) \quad (12)$$

$$\text{QUE} = \{\text{Que}_i\}; i = 1, \dots, n \quad (13)$$

where QUE – a collection of question indices made up of elements of the set (2), which allows the presence of several instances of the same element [11].

$$\text{ANS} = \{\text{Ans}_i\}; i = 1, \dots, p \quad (14)$$

ANS – a collection of answers made up of elements of the set (3).

$$\text{NextQue} : \text{ANS} \rightarrow \text{QUE} \quad (15)$$

NextQue – function of subsequent question indices. Since the active agent generates only one question at each step, NextQue determines the index of a single question for each answer.

$$\text{NextAns} : \text{QUE} \rightarrow \text{ANS} \quad (16)$$

NextAns – function of subsequent answers, which map the collection of question indices into the collection of answers.

To illustrate the Petri-model, let us present the example, described earlier, by formulas (12) – (16).

$$\text{ANS} = \{\text{Ans}_1, \text{Ans}_2, \text{Ans}_3, \text{Ans}_4, \text{Ans}_5\} \quad (17)$$

$$\text{QUE} = \{\text{Que}_1, \text{Que}_2, \text{Que}_3, \text{Que}_4\} \quad (18)$$

$$\left. \begin{aligned} \text{NextAns}(\text{Que}_1) &= \{\text{Ans}_1, \text{Ans}_3, \text{Ans}_4, \text{Ans}_5\} \\ \text{NextAns}(\text{Que}_2) &= \{\text{Ans}_1, \text{Ans}_2, \text{Ans}_3, \text{Ans}_4, \text{Ans}_5\} \\ \text{NextAns}(\text{Que}_3) &= \{\} \\ \text{NextAns}(\text{Que}_4) &= \{\} \end{aligned} \right\} \quad (19)$$

$$\left. \begin{aligned} \text{NextQue}(\text{Ans}_1) &= \{\text{Que}_1\} \\ \text{NextQue}(\text{Ans}_2) &= \{\text{Que}_1\} \\ \text{NextQue}(\text{Ans}_3) &= \{\text{Que}_3\} \\ \text{NextQue}(\text{Ans}_4) &= \{\text{Que}_4\} \\ \text{NextQue}(\text{Ans}_5) &= \{\text{Que}_2\} \end{aligned} \right\} \quad (20)$$

Formulas (19) and (20) can be used to define NextAns and NextQue functions in the form of tables. The right-hand side of the i-th line of the NextAns function specifies the set of reactive agent answers expected at the i-th step.

$$\text{NextAns}(\text{Que}_i) = R^i \cup NR^i \quad (21)$$

The right-hand side of the i-th line of the NextQue function consists of one element of the collection Que, which is a consequence of the limitation inherent in the question-answering dialogue, which is that for each answer from the set of expected answers, the function of indexes of subsequent questions determines the index of one single question.

In the graphical interpretation of the Petri-model of the dialogue method, the collection of indexes of the questions Que and the collections of answers Ans are represented by sets of places (circles) and transitions (rectangles), respectively. The set of answers  $R^i \cup NR^i$  expected at the i-th step is repre-

sented by a set of transitions that are incident to the  $i$ -th place and are connected to it by outgoing arcs.

For the Petri-model, a step is a natural "building block" of a dialogue access method and for the  $i$ -th step can be defined as a set

$$Que_i, (R^i \cup NR^i) \quad (22)$$

Figure 3 shows the graph of the Petri-model of the dialogue method, corresponding to the example described in the text.

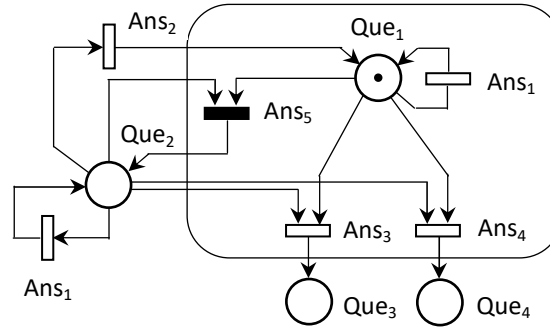


**Figure 3:** The Petri-model graph for the example described in the text. The black transition simulates all unrecognizable answers

When interpreting the Petri-model graphically, a dialogue step is a place and a set of transitions connected to it by outgoing arcs. In Figure 3 highlights the elements included in the scope of the concept of "dialogue step".

The Petri-model graph is marked with a single token marking the current active step. The initial marking, in this case, should be understood as indicating the first step from which the implementation of the dialogue method begins. The condition for firing a transition is the coincidence of the answer encoded by this transition and the answer received from the reactive agent. In other words, a transition is fired and "skips" a token if an answer is received from the reactive agent, which coincides with the answer encoded by this transition.

The theory of Petri nets provides a convenient apparatus for tracking the dynamics of a simulated process, consisting of marking places and conditions for firing transitions. Since, in the Petri-model of the question-answering dialogue, there is a single token, the marking can be represented by a vector, the number of components of which is equal to the number of elements of the collection of question indices – the number  $n$ .

$$\mu = (\mu_1, \mu_2, \dots, \mu_n) \quad (23)$$

Each component of the vector  $\mu$  takes values on the two-element set  $\{0, 1\}$ .  $\mu_i = 0$  if the token is absent in the  $i$ -th place corresponding to the  $i$ -th element of the collection  $Que$ .  $\mu_i = 1$  if the token is in the  $i$ -th place corresponding to the  $i$ -th element of the collection  $Que$ .

The question-answering dialogue process is accompanied by the movement of the token across the Petri-model without destroying or multiplying it. Therefore, at any moment of the dialogue,  $\sum \mu_i = 1$  takes place. Thus, in the process of a question-answering dialogue, the dialogue method is characterized by a continuously changing vector  $\mu$ , which indicates the current, active step of the dialogue. For the Petri-model shown in Figure 3, the vector  $\mu$  indicates step number one and has a value  $\mu = (1, 0, 0, 0)$ . To move a token to the next place, it is necessary that one of the transitions of the step fires.

Representation of the dialogue method in the form of a Petri net (formulas 12 – 16) models its spatial structure and does not reflect the dynamics of the dialogue process, which is set, in the general case, by the cognitive cycle of the question-answering dialogue process shown in Figure 1. Let us supplement the structural description of the dialogue method with components that simulate the dynamics of the dialogue process. For this purpose, let us consider how, within the framework of the proposed Petri-model, the cycle of the question-answer dialogue process is implemented, based on the

assumption that the **NextAns** and **NextQue** functions are presented in tables.

It can be assumed that a separate row in the **NextAns** function table corresponds to one step of the dialogue. Each row starts with the index of the question generated at the given step, followed by a list of answers associated with that step. In this case, the **NextAns** function table consists of  $n$  rows, where  $n$  is the number of elements in the collection of question indices. The number of columns in the **NextAns** function table is determined by the maximum possible number of answers expected from the reactive agent during the dialogue. Since the set of answers associated with a step is individual for each step and is determined by the number of answers expected at a given step, some of the cells in the **NextAns** table remain empty.

The **NextQue** function table has a simpler structure. It consists of  $p$  rows, where  $p$  is the number of elements in the collection of answers. Each row in the **NextQue** function table contains two elements: the expected answer and the associated follow-up question.

Let's add a unified DiMC process to the model, which simulates the cyclically repeating dynamics of the dialogue

$$\text{DiMC}(\mu, \text{NextAns}, \text{NextQue}) \quad (24)$$

The unified DiMC process controls the transition from the previous step of the dialogue to the next, which is equivalent to firing one of the transitions of the step and moving the token. The unification of the DiMC process means that for the implementation of any step of the dialogue it is necessary to perform the same sequence of actions, and that this sequence of actions does not depend on the subject area of the dialogue process. We describe a separate iteration of the DiMC cycle as follows.

1. The marking vector  $\mu$  defines the currently active step of the dialogue process, which corresponds to the row number in the **NextAns** function table.
2. The index of the current question is determined and transferred to the memory of questions. Using the question specifications in the memory of questions, the active agent generates and passes the current question to the reactive agent. The active agent enters the waiting state for the answer of the reactive agent.
3. The active agent perceives the current answer of the reactive agent.
4. The current answer of the reactive agent is processed. The perceived answer is sequentially compared with all the answers expected at this step, stored in the cells of the selected row of the **NextAns** function table. For a matched answer, a **NextQue** function table row is determined.
5. Using the table of the **NextQue** function, the component of the vector  $\mu$  is searched for, which determines the next step of the dialogue. The value of this component changes from zero to one, and the process returns to item 1.

Among the actions iteratively performed by the unified DiMC process, there are two actions that require more detailed consideration and clarification: (1) generating the current question; (2) processing the current answer. Since DiMC is a unified process that processes any step "uniformly", both question generation and response processing are handled by DiMC using the same rules for any dialogue step in any dialogue. These rules define a certain standard model for the functioning of an active dialogue agent. In other words, DiMC generates a question and processes the answer in a certain standard way. It is clear that even with a fairly extensive standard, there may be cases where standard tools are not enough. For example, when generating a question, in the case when it is presented non-verbally, it is impossible to take into account all the diversity of non-verbal presentation of information, and when processing the answer, specific numerical processing may be required. Therefore, in addition to the basic DiMC process, it is rational to include in the model "external" processes in relation to DiMC. These processes must be called from DiMC and return control to DiMC when finished. Let's call such processes process-daemons and add two classes of process-daemons to the model

QueDemon and AnsDemon

Process-daemons carry out "additional processing" of a question or answer at any step of the dialogue process in the case when standard DiMC tools are not enough [12].

Let's consider the sphere of applicability of the Petri-model of the question-answering dialogue. By the sphere of applicability we mean the division of problems solved in the dialogue process into two classes: (1) easily implemented within the Petri-model and (2) difficult to implement within the Petri-model.

Problems that are difficult to be implemented within the framework of the Petri-model include, for example, problems that are solved when conducting psycho-diagnostic tests. As a rule, at the first stage of psycho-diagnostic testing (in the process of dialogue interaction with the subject), a primary data set is formed, which is then used to form a conclusion about the intellectual abilities or the structure of the subject's intelligence. An example of a psycho-diagnostic test is the Raven's progressive matrix technique used to test intellectual abilities [13]. The dialogue process that implements the Raven's progressive matrix technique has two specific features. Firstly, at each step of the dialogue, the reactive agent (subject of testing) is presented with stimuli-questions in a non-verbal form. Secondly, the sequence of stimuli-questions transmitted to the reactive agent is formed by the active agent without taking into account the answers of the reactive agent.

We will describe a simple psycho-diagnostic test as follows. The first step generates a  $Que_1$  question that expects one of the answers  $Ans_1$  or  $Ans_2$ . In the second step, regardless of the answer received, a  $Que_2$  question is generated, for which one of the answers  $Ans_3$  or  $Ans_4$  is expecting. The conclusion is determined depending on the combination of the answers received. The graph of the Petri-model of the dialog method that implements the described test is shown in Figure 4.

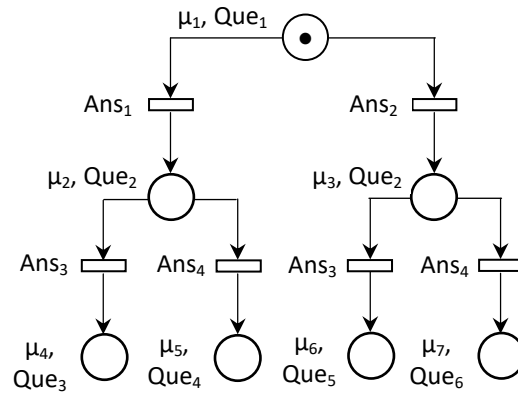


**Figure 4:** Petri-model of a "difficult" dialogue method

At the bottom of Fig. 4 are target places corresponding to the end of the dialogue and the formation of a conclusion. The difficulty of implementing the Petri-model is manifested in the rapid increase in the number of steps. If a answers are expected at each step, and subsequent questions do not depend on previous answers, then the total number of steps is determined by the formula

$$n = a^0 + a^1 + a^3 + \dots + a^b \quad (25)$$

where  $b$  is the number of questions. The Petri-model for Raven's progressive matrix technique consists of  $2^{61}$  steps. The tree-like shape of the Petri-model and a large number of steps are determined by the fact that the model takes into account the combination of answers by forming a unique trajectory of the token in the network. A way to combat this deficiency is to include a "history of answers" memory into the model. The presence of such memory means taking into account the principle of the history of answers formulated in Section 1, which means that the dialogue method, when forming the next question, should take into account both the answer just received and the previously perceived answers. To simplify the presentation, we will assume that the model includes the memory  $Mem$ , which stores only one answer obtained at the previous step of the dialogue. Thus, the contents of  $Mem$ : (1) are automatically updated when dialogue go to the next step; (2) DiMC takes into account the contents of  $Mem$  when determining the index of the next question.

Fig. 5 shows the graph of the Petri-model with memory  $Mem$  storing the answer obtained in the previous step. The inclusion in the model the memory of the previous answer does not require the formation of a unique trajectory for each conclusion. The target positions remain unique, and the trajectories leading to the target positions have a common part.

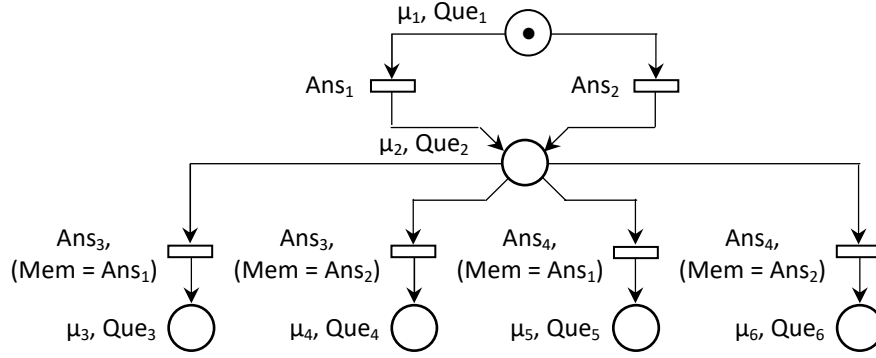


**Figure 5:** Petri-model in the presence of memory of the previous answer Mem

## 5. Conclusion

The article describes two formal models of a question-answer dialogue: (1) a finite-automaton model and (2) a model in the form of a Petri net. The models can be used to design a "dialogue machine" that simulates the purposeful behavior of an artificial active dialogue agent, which does not depend on the subject area of the dialogue. The functioning of the dialogue machine is based on the implementation of a unified cognitive cycle of dialogue. The question-answering dialogue model in the form of a Petri net has a wider sphere of applicability than the finite-automaton model. The advantage of this model is that it simulates a typical question-answering process loop using the unified DiMC process, and also allows custom processing of questions and answers using the unique QueDemon and AnsDemon processes. The model in the form of a Petri net also allows taking into account the history of the answers of the reactive agent when forming the next question.

It is supposed to direct further research to develop the main ideas of the described models, but to start not from mathematical structures in the form of a Mealy machine or a Petri net, but from the idea of representing the dialogue method in the form of a network structure *consisting of nodes of different types*. A priori, it is assumed that the network structure is mapped to a relational database. In such a datalogical and network model, the following should be essential: (1) two of the three principles for organizing a question-answering dialogue formulated in Section 1 should be taken into account: the principle of the "depth of dialogue" and the principle of the "history of answers" of a reactive agent; (2) a datalogical approach should be developed and, therefore, a dialogue method for solving a problem, as well as questions of an active agent, should be represented by data stored in a database.

## 6. References

- [1] A. M. Turing, Computing machinery and intelligence, *Mind*, 49, (1950) 433-460.
- [2] S. Russell, P. Norvig, *Artificial Intelligence: A Modern Approach*, 2<sup>nd</sup> ed., Prentice Hall, New Jersey, 2003.
- [3] U. Neisser, *Cognition and Reality*, W.H. Freeman and Company, San Francisco, 1976.
- [4] N. Belnap, T. Steel, *The Logic of Questions and Answers*, Yale University Press, New Haven and London, 1976.
- [5] J. Hintikka, *Socratic Epistemology: Explorations of Knowledge Seeking by Questions*, Cambridge University Press, Cambridge, 2007.
- [6] A. Newell, Remarks on the relationship between artificial intelligence and cognitive psychology, in: R. Banerji and J.D. Mesarovich (Eds.), *Theoretical Approaches to Non-Numerical Problem Solving*, Springer-Verlag, New York, NY, 1970, pp. 363-399.
- [7] E. Wenger, *Artificial Intelligence and Tutoring Systems. Computational and Cognitive Approaches to the Communication of Knowledge*, Morgan Kaufman, Los Altos, California, 1987.
- [8] Zhou, L., Gao, J., Li, D., and Shum, H.-Y., The design and implementation of Xiaolce, an empathetic social chatbot, *Computational Linguistics* 46(1) (2020), 53-93.

- [9] Lawson, Mark V, Finite automata, CRC Press, Taylor & Francis Group, 2004.
- [10] Graham A. A., Framework for Enterprise Data Architecture, 2<sup>nd</sup> edition, Koios Associates Ltd., 2012.
- [11] J. Peterson, Petri Net Theory and the Modeling of Systems, Prentice-Hall, Englewood Cliffs, New Jersey, 1981.
- [12] I. Chmyr, Dialogue of Partners as a Method of Non-Formal Problem Solving, in Maddy D. Brouwer-Janse and Thomas Harrington (Eds.), Human-Machine Communication for Educational Systems Design, NATO ASI Series F129, Springer-Verlag, Berlin, 1994, pp. 222 – 228.
- [13] J. Raven, Mental tests used in genetic studies: The performance of related individuals on tests mainly educative and mainly reproductive, Master's thesis, University of London, London, 1936.

# Software for statistical processing and modeling of a set of synchronously registered cardio signals of different physical nature

Serhii Lupenko<sup>a</sup>, Iaroslav Lytvynenko<sup>a</sup>, Andrii Sverstiuk<sup>b</sup>, Andrii Horkunenko<sup>b</sup> and Borys Shelestovskyi<sup>a</sup>

<sup>a</sup> Ternopil Ivan Puluj National Technical University, Ruska Street 56, Ternopil, 46000, Ukraine

<sup>b</sup> I. Horbachevsky Ternopil National Medical University, Maidan Voli 1, Ternopil, 46000, Ukraine

## Abstract

It has been developed the software complex, that allows to perform mutual statistical processing of synchronously registered cardiosignals on the basis of the vector model of the cyclic rhythmically connected stochastic processes.

The developed software provides the ability to form a discrete rhythm function, depending on the number of cycles and zones of the cardio signal to be modeled. The developed set of programs allows to perform statistical processing of synchronously registered cardiosignals of different physical nature and to obtain estimates of mathematical expectation, dispersion, autocorrelation and mutual correlation functions taking into account the rhythm function and the period. The software also allows you to simulate cardio signals, using the loaded input data of the cardiosignal, rhythm functions, the number of cycles and zones of the signal to be modeled.

## Keywords 1

Synchronously registered cardiosignals, statistic processing methods, software complex

## 1. Introduction

Cardiovascular diseases in the world are a serious problem. The medical-social burden of the circulatory system diseases, especially exacerbated by the rapid spread of COVID-19 and is that they significantly affect the duration and life quality, the loss economic potential indicators of countries. That is why the fight against diseases of the circulatory system at the present stage is a priority problem of modern medicine [1, 2]. Therefore, the development of modern technologies for instrumental diagnosis of the cardiovascular system (CVS) pathology in the early disease stages, monitoring the effectiveness of preventive and curative measures, as well as monitoring the vital functions of the body in critical conditions is an urgent problem of modern medicine and technology [3, 4].

A promising direction to overcome this problem is to improve existing methods of instrumental diagnosis of CVS by developing new diagnostic and prognostic features based on the choosing adequate mathematical analysis methods of synchronously registered signals. To date, a significant number of functional methods for studying the state of the CVS, which uses the synchronous registration of several identical or different in origin cardiosignals. One of the main ones is the synchronous registration of the electrocardiogram (ECG) in different leads [5] and in standardized formats [6].

---

CMIS-2021: The Fourth International Workshop on Computer Modeling and Intelligent Systems, April 27, 2021, Zaporizhzhia, Ukraine  
EMAIL: lupenko.san@gmail.com (S. Lupenko); iaroslav.lytvynenko@gmail.com (I. Lytvynenko); sverstyuk@tdmu.edu.ua (A. Sverstiuk); horkunenkoab@tdmu.edu.ua (A. Horkunenko); kaf\_vm@tu.edu.ua (B. Shelestovskyi)  
ORCID: 0000-0002-6559-0721 (S. Lupenko); 0000-0001-7311-4103 (I. Lytvynenko); 0000-0001-8644-0776 (A. Sverstiuk); 0000-0002-2021-006X (A. Horkunenko); 0000-0002-5606-6290 (B. Shelestovskyi)



© 2021 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)



When monitoring the vital functions of the body, which occurs during the provision of emergency and urgent medical care, synchronously recorded key indicators of the body: ECG, curves of arterial and central venous pressure, external respiration, oxygen and carbon dioxide in the blood and others, depending on specific clinical situation and capabilities of the applied equipment [7, 8]. Mechanical devices to support blood circulation in advanced heart failure are considered in the work [9].

However, and in this situation the diagnostic analysis is conducted on each signal in particular which is compared with age norm or magnitude which is defined and corrected by the doctor. An urgent problem today is the improvement of existing methods of instrumental diagnosis of the cardiovascular system by developing a software package for the analysis of cardio signals.

A smart system for monitoring and predicting heart disease, which is based on in-depth training, is presented in the work [10]. The article [11] presents cardio diagnostic systems that use artificial intelligence and machine learning, as it concerns the health of the cardiovascular system. Cyclic analysis between meteorological rhythms of respiratory and cardiovascular diseases is considered in [12]. Diseases of the circulatory system under the influence of ionizing radiation are considered in this work [13].

Algorithms and methods of cardiac signal processing are created on the basis of their adequate mathematical models. Thus, it can be stated that today the methods of compatible automated processing of the set of synchronously registered cardiosignals (SRCS) have not received significant development. This is due to the insufficient level of development of unifying ideas in the construction of mathematical models of different types of cardiosignals and diversity of methods for their processing. The development of other highly informative diagnostic methods also had a significant impact. However, it can be stated that the development of optimal ways to analyze SRCS or vital parameters of the organism with access to new integrated indicators and empirical confirmation of their higher diagnostic value, compared to existing ones, can raise to a new level the nature of diagnostic process and health management treatment.

Despite significant advances in mathematical modeling of cyclic heart signals, their existing models do not take into account the following facts. For example, new diagnostic features for cardio diagnostic systems based on the model of electrocardiographic signals in the form of a vector of cyclic random processes are considered in [14].

The above arguments indicate the relevance of software for statistical processing and modeling of cardiosignals based on a mathematical model in the form of a vector of synchronously registered cardio signals for the needs of automated cardiodynamics.

The effectiveness of modern cardiodynamic systems largely depends on the hardware and software components on which they are based. The use of various processing methods included in the software significantly expands the functionality and increases the reliability of human heart diagnosis [15]. While the methods are based on mathematical models that establish the possibility and specifics of processing. In particular, the simultaneous processing of cardiosignals of different physical nature can be performed only if their mathematical models are in some way correlated and have a similar structure.

## **2. Analysis of recent researches**

A stochastic approach to their simulation is mainly used in the direction of cardiac signal processing. Thus, for this purpose, use the mathematical models of cyclostationary signals [16-18], modeling diagnostic features [19-20], analytical approaches for myocardial fibrillation signals [21], modeling the pulse signal by wave-shape function [22]. An interesting approach to the description of cardio signals is the mathematical model presented in [23]. It allows to take into account both morphological features of cardiac signals and their rhythmic structure. It is important that this mathematical model takes into account the common rhythm of synchronously recorded cyclic cardio signals of different physical nature, which do not allow to implement other mathematical models using a stochastic approach. In order to be able to apply the proposed mathematical model, a necessary condition is the availability of information about the characteristic elements of cardio signals (cycles and zones, such as zones P, Q, R, S, T - electrocardiogram). It is possible to obtain tau information about the characteristic elements of cyclic cardiac signals by applying their segmentation

methods, for example, the methods presented in [24], other methods of analysis of cardio signals are presented in [25-27].

### 3. Objectives

The work is devoted to the creation and application of software for statistical processing and modeling of synchronously registered cardiac signals as components of cardio diagnostic systems.

The importance of these studies lies in the possibility of using a common rhythm in the modelling and static processing of different in nature cyclic signals but formed by a single source.

### 4. Method

According to paper [22], we give a definition of the vector of cyclic rhythmically connected stochastic processes.

Definition 1. If there is a function such as  $T(t, n)$ , which satisfies the conditions of the rhythm function that finitely measurable vectors  $\{\xi_{i_1}(\omega, t_1), \xi_{i_2}(\omega, t_2), \dots, \xi_{i_k}(\omega, t_k)\}$  and  $\{\xi_{i_1}(\omega, t_1 + T(t_1, n)), \xi_{i_2}(\omega, t_2 + T(t_2, n)), \dots, \xi_{i_k}(\omega, t_k + T(t_k, n))\}$   $n \in \mathbf{Z}, i_1, \dots, i_k = \overline{1, N}$ , where  $\{t_1, \dots, t_k\}$  - multiple separability of the vector  $\Theta_N(\omega, t)$ , for all the integers  $k \geq 1$  are stochastic equivalent in the broadest sense, we will call the vector  $\Theta_N(\omega, t)$  of cyclic stochastic processes  $\left\{ \xi_i(\omega, t), i = \overline{1, N}, \omega \in \Omega, t \in \mathbf{W} \right\}$  as the vector of strictly rhythmically connected stochastic processes and the processes as strictly rhythmically connected.

Area of definition  $\mathbf{W}$  vector of cyclic rhythmically connected stochastic processes can be as ordered discretely  $\mathbf{W} = \mathbf{R} = \{t_{ml} \in \mathbf{R}, m \in \mathbf{Z}, l = \overline{1, L}\}$  or continuous  $\mathbf{W} = \mathbf{R}$  set of real numbers. In the case of discrete domain definition  $\mathbf{W} = \mathbf{D}$  for its elements if  $m_2 > m_1$ , or if  $m_2 = m_1$ , a  $l_2 > l_1$ , in other cases  $t_{m_1 l_1} > t_{m_2 l_2}; m_1, m_2 \in \mathbf{Z}, l_1, l_2 \in \overline{1, L}$  there is a type of linear ordering:  $t_{m_1 l_1} < t_{m_2 l_2}$ . Moreover  $0 < t_{m, l+1} - t_{m, l} < \infty$ .

The rhythm function  $T(t, n)$  determines the law of changing the time intervals between the single-phase values of the vector of cyclic rhythmically connected stochastic processes. The function of the rhythm satisfies such conditions according to the theorem which is proved in the paper:

- a)  $T(t, n) > 0$ , if  $n > 0$  ( $T(t, 1) < \infty$ );
  - b)  $T(t, n) = 0$ , if  $n = 0$ ;
  - c)  $T(t, n) < 0$ , if  $n < 0, t \in \mathbf{W}$ ;
- (1)

for any  $t_1 \in \mathbf{W}$  and  $t_2 \in \mathbf{W}$ , for which  $t_1 < t_2$ , for function  $T(t, n)$  should be performed a strict inequality:

$$T(t_1, n) + t_1 < T(t_2, n) + t_2, \forall n \in \mathbf{Z}; \quad (2)$$

function  $T(t, n)$  is the smallest modulo ( $|T(t, n)| \leq |T_\gamma(t, n)|$ ) among all such functions  $\{T_\gamma(t, n), \gamma \in \Gamma\}$ , which satisfy (1) and (2).

In the partial case, if the rhythm function is  $T(t, n) = n \cdot T$  ( $T > 0, n \in \mathbf{Z}$ ), we will call the vector  $\Theta_N(\omega, t)$  as the vector  $T$  - periodically connected stochastic processes.

Let us consider the properties of some probabilistic characteristics of the vector  $\Theta_N(\omega, t)$  cyclic rhythmically connected stochastic processes. So, for its compatible  $k$  - measurable distribution function takes place equation:

$$F_{k_{\xi_{i_1} \dots \xi_{i_k}}} (x_1, \dots, x_k; t_1, \dots, t_k) =$$

$$= F_{k_{\xi_{i_1} \dots \xi_{i_k}}} (x_1, \dots, x_k; t_1 + T(t_1, n), \dots, t_k + T(t_k, n)), n \in \mathbf{Z}, i_1, \dots, i_k = \overline{1, N}, t_1, \dots, t_k \in \mathbf{W} \quad (3)$$

Combined central moments function of order  $p = \sum_{j=1}^k R_j :$

$$r_{p_{\xi_{i_1} \dots \xi_{i_k}}} (t_1, \dots, t_k) = \mathbf{M}\{(\xi_{i_1}(\omega, t_1) - m_{\xi_{i_1}}(t_1))^{R_1} \cdot \dots \cdot (\xi_{i_p}(\omega, t_k) - m_{\xi_{i_k}}(t_k))^{R_k}\} =$$

$$= r_{p_{\xi_{i_1} \dots \xi_{i_k}}} (t_1 + T(t_1, n), \dots, t_k + T(t_k, n)), t_1, t_2, \dots, t_k \in \mathbf{W}, i_1, \dots, i_k = \overline{1, N}, n \in \mathbf{Z}. \quad (4)$$

## 5. Results and Discussion

It was made the series of experiments on processing of the cardio signals of the same and different physical origin which were investigated for the purpose of approving of the greater effectiveness of the simultaneous processing of synchronously registered cardio signals (SRCS) based on the model of the vector of cyclic rhythmically connected stochastic processes in comparison with the well-known method of their processing.

As the example it is shown on the Figure 1 the realizations of SRCS electrocardiosignal (ECS) and phonocardiosignal (PCS) and on the figures 2-4 are represented the results of a comparative analysis.

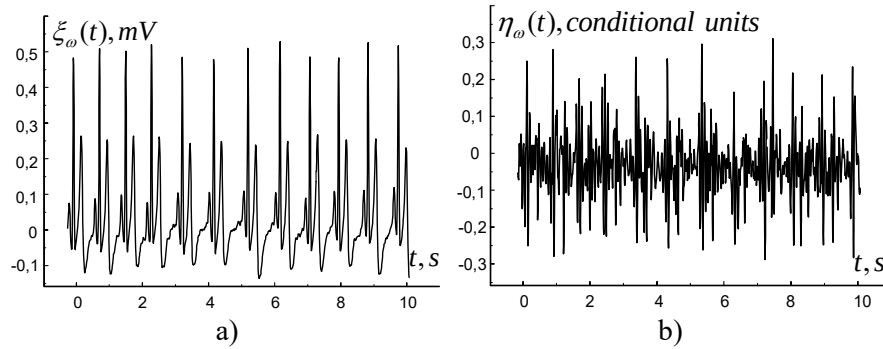


**Figure 1:** The realizations of SRCS: (a) electrocardiogram, (b) phonocardiogram.

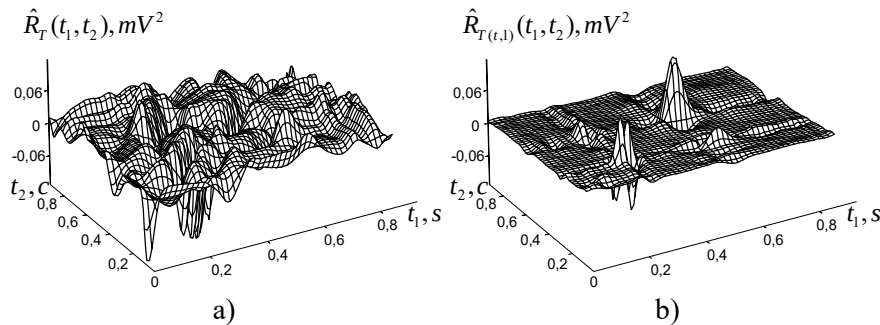


**Figure 2:** The realization graphs of statistical estimations of the autocorrelation function of the ECS while its processing on the basis of: (a) the vector of periodically connected stochastic processes; (b) the vector of cyclic rhythmically connected stochastic processes.

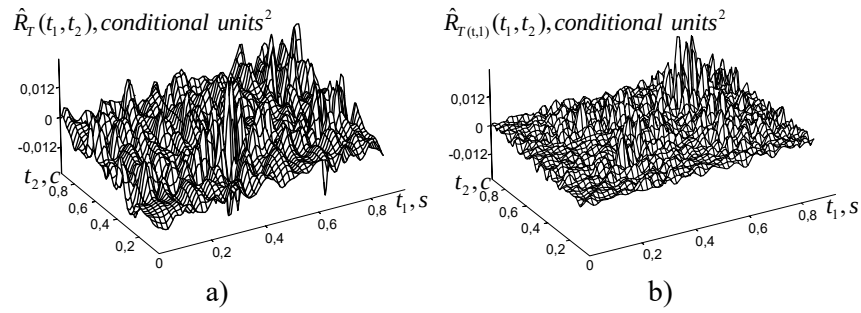


**Figure 3:** The realization graphs of statistical estimations of the autocorrelation function of PCS while its processing on the basis of: (a) the vector of periodically connected stochastic processes; (b) the vector of cyclic rhythmically connected stochastic processes.

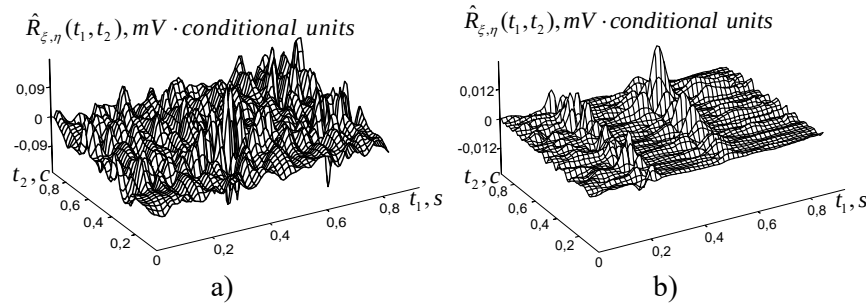


**Figure 4:** The realization graphs of statistical estimations of the mutual correlation function of the ECS and PCS while its processing on the basis of: (a) the vector of periodically connected stochastic processes; (b) the vector of cyclic rhythmically connected stochastic processes.

It was set in the result of the comparative analysis of the statistical processing of sets of SRCS which was made that the method of statistical processing of the analyzed cardiosignals on the basis of the vector of cyclic rhythmically connected stochastic processes significantly reduces the negative effect of "blurring" of statistical estimations of mutual correlation functions as a part of set of synchronous cyclic signals of a heart that is strongly-pronounced in results of statistical processing of the cardiac signals which were investigated on the basis of the vector of periodically connected stochastic processes. It is proved by the fact that the new method of compatible statistical processing takes into account the variability of the SRCS rhythm, in contrast to the well-known methods.

## 5.1 Software operating principle

It was developed the software complex which allows to perform SRCS statistical processing and its structural scheme is shown in Figure 5.

In this complex was implemented: block of SRCS formation, SRCS filtering block, trend removal block, SRCS segmentation block, block for determining the SRCS rhythm function, statistical analysis block of the rhythm function, valuation block of statistical estimates of SRCS, statistical analysis block of SRCS, dimension minimization block of the vector of informative features, diagnostic block.

The exported synchronously registered cardio signals were used from PhysioNet in accordance with the article [28] in the SRCS implementation generation unit. Pre-processing units are designed to filter and extract trends from SRCS. The segmentation unit is used to obtain the time-time structure of SRCS to determine their rhythmic functions. Blocks of statistical processing are intended for the normalization of statistical estimations, the analysis of investigated cardiac signals. The block of minimization of dimensionality of a vector of informative signs is necessary for minimization of their quantity representing norm or a certain pathology of the cardiovascular system. The diagnostic unit allows you to diagnose the cardiovascular system using SRCS (Fig. 5).

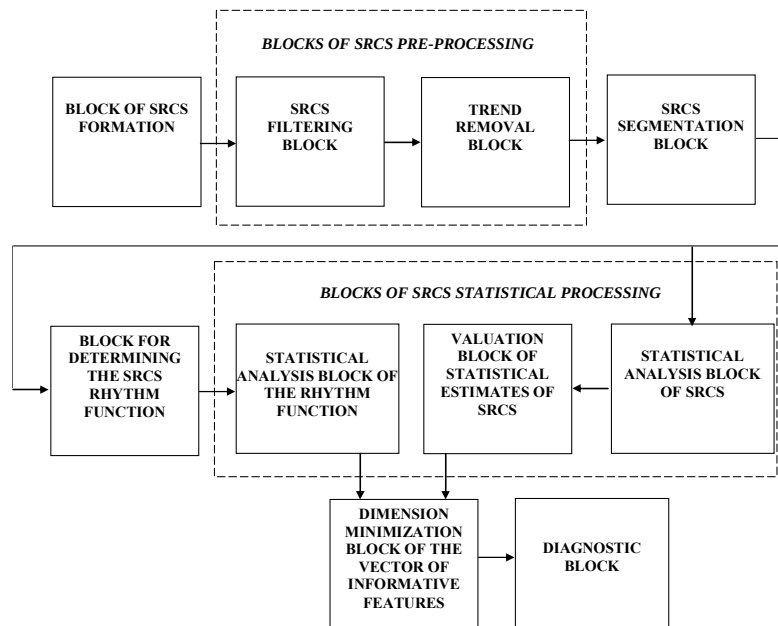


**Figure 5:** Software complex for statistical processing of synchronously registered cardiosignals.

For the compatible statistical processing of SRCS (electrical, magnetic, acoustic nature) the module “Compatible statistical processing of SRCS” is used, the structural-functional scheme of which is shown in fig. 6.

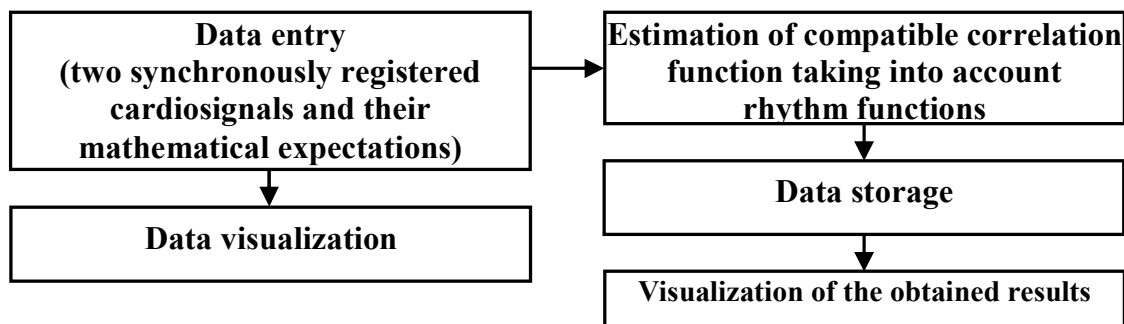


**Figure 6:** Structural-functional module scheme of the program "Compatible statistical processing SRCS".

The input data for compatible statistical processing are oversampled values of signals and their estimates of mathematical expectation. The result of joint statistical processing is a cross-correlation function obtained taking into account the rhythm function (Fig. 6).

To simulate cardiosignals, a simulation unit SRCS was developed taking into account the rhythm function, the structural-functional scheme of which is shown in Fig. 7.

In order to simulate cardio signals taking into account the rhythm function, it is first necessary to open the input data of the cardio signal and rhythm function, set the number of cycles and signal zones, load the input cycle (Fig. 7).

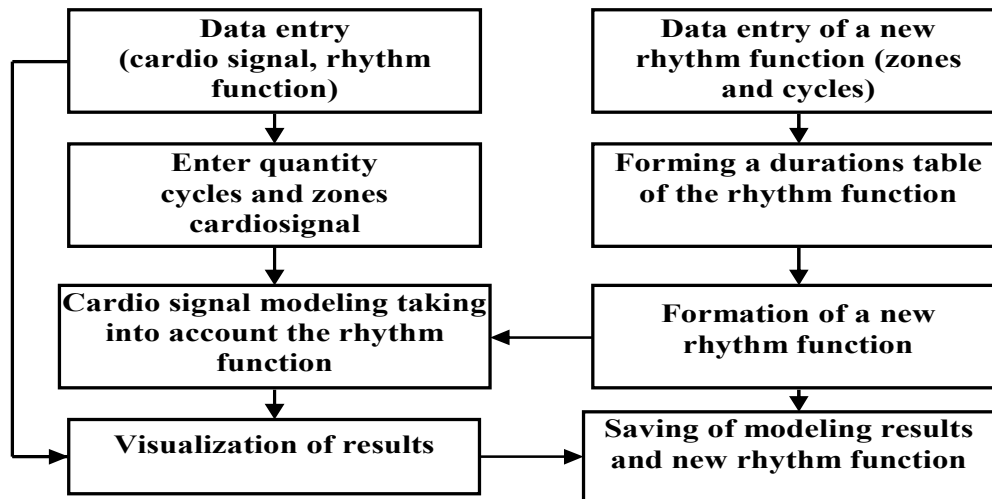


**Figure 7:** Structural-functional scheme of the SRCS simulation unit taking into account the rhythm function.

## 5.2 Software design

In Fig. 8 the general view of program interfaces for data input (cardiosignals, rhythm functions) and their visualization is given.

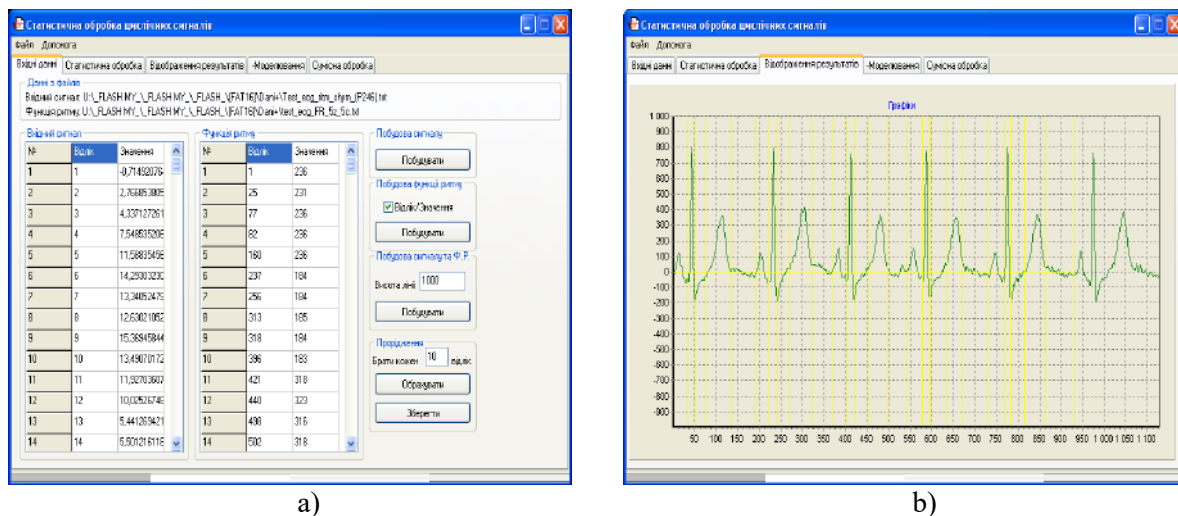
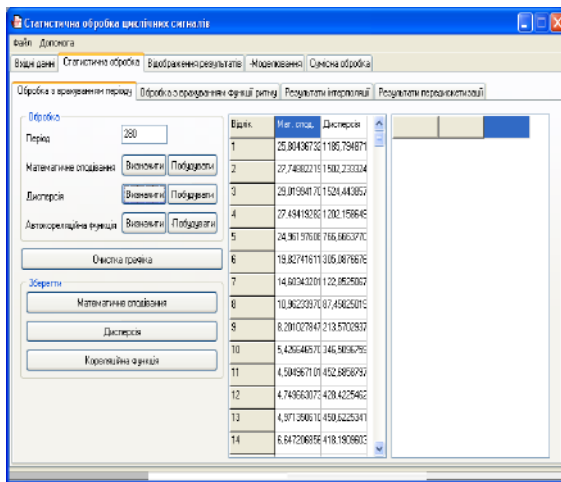


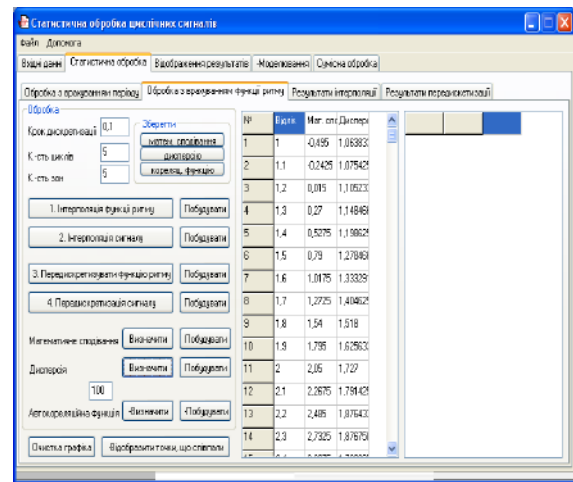
**Figure 8:** General view of the program interfaces for data entry (a) and their visualization (b).

Two methods were used in the program for statistical processing of SRCS: the method of averaging the values of signal realization taking into account the period and the method of averaging the values of signal realization taking into account the rhythm function. An example of the program interface for statistical processing taking into account the period is shown in Fig. 9.

The method of averaging the values of signal realization taking into account the rhythm function, in comparison with known methods, significantly reduces the negative effect of "blurring" of statistical estimates of probabilistic characteristics of SRCS, which allows to increase the accuracy, reliability and informative of complex automated diagnosis of the human cardiovascular system by synchronous cyclic heart signals (Fig. 9).



a)



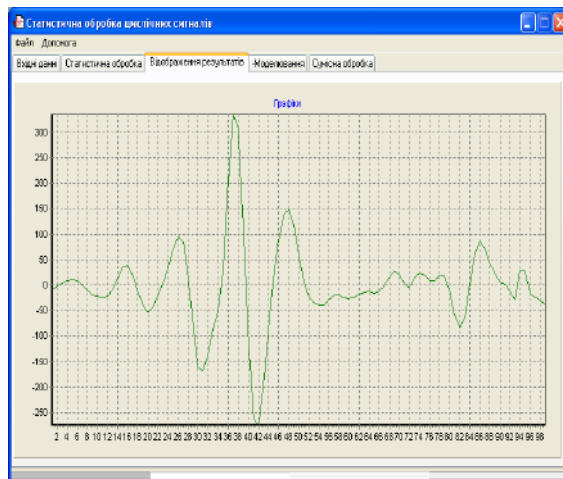
b)

**Figure 9:** General view of the program interfaces for statistical processing through period (a) and taking into account the rhythm function (b).

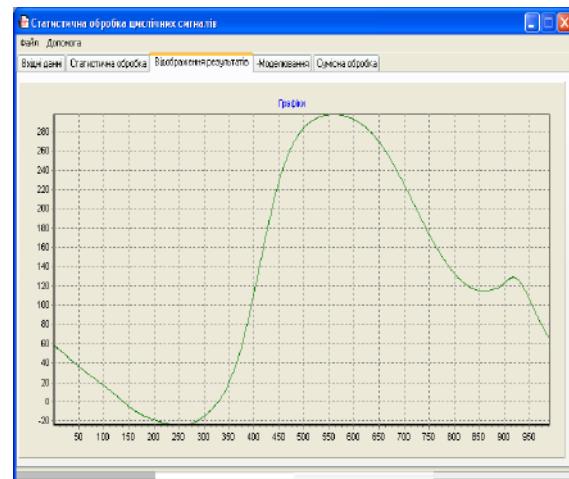
During statistical processing, taking into account the rhythm function, it is necessary to specify additional parameters, namely: the sampling step of the signal and rhythm function (required during resampling of signal and rhythm function), as well as the number of heartbeat cycles and the zones number (per cycle) of registered signal implementation. Performing consistent steps: rhythm function interpolation; signal interpolation; rhythm function resampling; signal oversampling we obtain the necessary data for statistical processing, taking into account the rhythm function.

In Fig. 9 shows that as a result of statistical processing we can obtain estimates of the following probabilistic characteristics: mathematical expectation, variance and autocorrelation function. All received data can be saved in text files if necessary.

The general view of the program interfaces for visualization of the estimation of the mathematical expectation of the phonocardiogram (a) and the integrated rheogram (b), by the method of statistical processing taking into account the rhythm function, is given in Fig. 10.



a)



b)

**Figure 10:** General view of the program interfaces for estimation visualization of the mathematical expectation of the phonocardiogram (a) and the integrated rheogram (b).

The input data for compatible statistical processing are oversampled values of signals and their estimates of mathematical expectation. The result of compatible statistical processing is a mutual correlation function obtained taking into account the rhythm function.

An example of the program interface for entering synchronously registered data (a) and their compatible processing (b) is shown in Fig. 11.

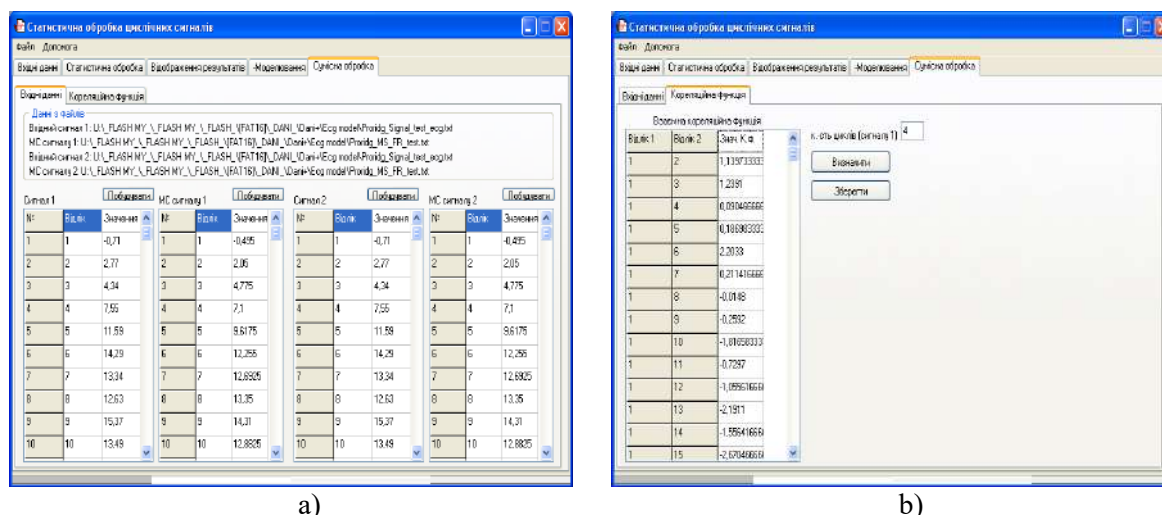


**Figure 11:** General program interface view for entering synchronously registered data (a) and their compatible processing (b).

In order to simulate cardiosignals taking into account the rhythm function, it is first necessary to open the input data of the cardiosignal and rhythm function, set the number of cycles and signal zones, load the input cycle. The program interfaces are used for input data (Fig. 12 (a)) and for visualization of the input cycle (Fig. 12 (b)).

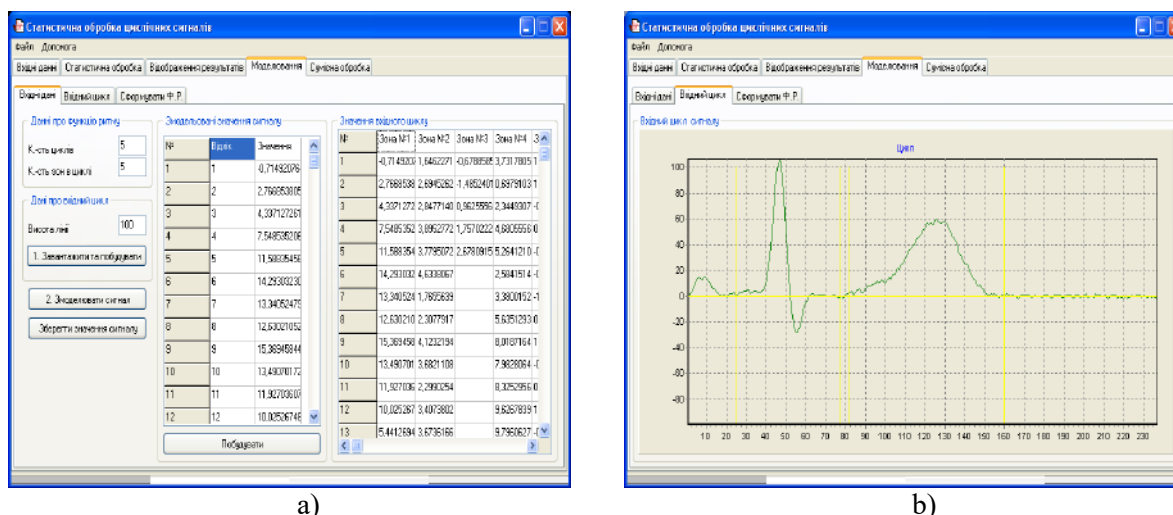


**Figure 12:** General view of the interface of the program of input data input (a) and visualization of the input cycle (b).

The developed software provides the ability to form a discrete rhythm function, depending on the number of cycles and zones of the cardiosignal to be modeled. The program interfaces for forming the rhythm function and visualization of the modeled cardio signal are shown, respectively, in Fig. 13 (a, b).



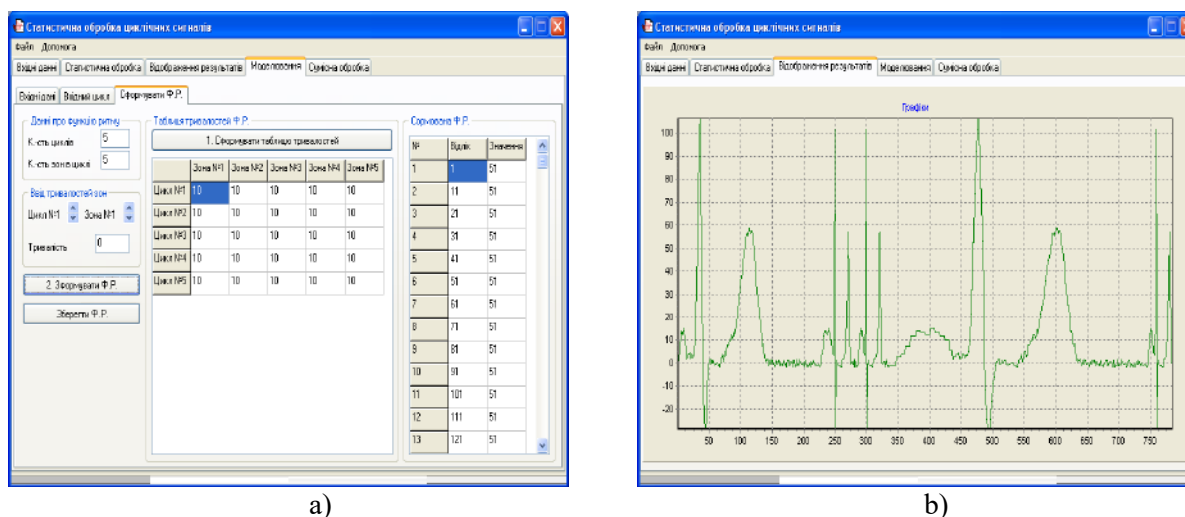


**Figure 13:** General view of the program interface for forming the rhythm function (a) and the modeled cardiosignal generated by the rhythm function (b).

The developed set of programs makes it possible to perform statistical processing of SRCS of different physical nature and obtain estimates of mathematical expectation, dispersion, autocorrelation and mutual correlation functions (taking into account the rhythm function and taking into account the period). The software also allows you to simulate cardio signals. The developed software provides the ability to form a discrete rhythm function, depending on the number of cycles and zones of the cardio signal to be modeled.

## 6. Conclusions

A set of programs has been developed that allows for statistical processing of cardiac signals of different physical origin and to obtain estimates of mathematical expectation, variance, autocorrelation and cross-correlation (taking into account the function of rhythm and period). The software allows you to simulate the implementation of cardio signals of different physical nature, taking into account their constant rhythm (taking into account the period), variable (taking into account the rhythm function) and the common rhythm. The modeling method includes information obtained at the stage of statistical processing of cardiac signals, in particular, morphological features are taken into account on the basis of estimates of mathematical expectation and variance of cardiac cycle cycles.

The developed software provides the possibility of forming a discrete rhythm function depending on the number of cycles and zones of the simulated cardio signal.

This software can be used as a component of diagnostic systems of the human heart.

## 7. References

- [1] Anastasia P. Nesterova, Eugene A. Klimov, Maria Zharkova, Sergey Sozin, Vladimir Sobolev, Natalia V. Ivanikova, Maria Shkrob, Anton Yuryev, Diseases of the circulatory system, in: Disease Pathways, 1st. ed., Elsevier, 2020, pp. 327-390. doi:10.1016/C2018-0-00586-1.
- [2] Jer-Hwa Hsu, I-Chia Chien, Ching-Heng Lin, Increased risk of ischemic heart disease in patients with bipolar disorder: A population-based study, *Journal of Affective Disorders* 281 (2021), 721-726, doi:10.1016/j.jad.2020.11.083.
- [3] Jun Yang, Wenjing Xiao, Huimin Lu, Ahmed Barnawi, Wireless high-frequency NLOS monitoring system for heart disease combined with hospital and home, *Future Generation Computer Systems* 110 (2020) 772-780. doi:10.1016/j.future.2019.11.001.

- [4] M.T. Donofrio, A.J. Moon-Grady, L.K. Hornberger, J.A. Copel, M.S. Sklansky, A. Abuhamad, et al. Diagnosis and treatment of fetal cardiac disease: a scientific statement from the American heart association *Circulation* 129 (2014) 2183-2242. doi:10.1161/01.cir.0000437597.44550.5d.
- [5] Margherita Padeletti, Giuseppe Bagliani, Roberto De Ponti, Fabio M. Leonelli, Emanuela T. Locati, Surface Electrocardiogram Recording: Baseline 12-lead and Ambulatory Electrocardiogram Monitoring, *Cardiac Electrophysiology Clinics* 11 (2019) 189-201, doi:10.1016/j.ccep.2019.01.004.
- [6] Raymond R. Bond, Dewar D. Finlay, Chris D. Nugent, George Moore, A review of ECG storage formats, *International Journal of Medical Informatics* 80 (2011) 681-697, doi:10.1016/j.ijmedinf.2011.06.008.
- [7] Gail Baura, Electrocardiograph design lab, in *Medical Device Technologies*, 2d. ed., Academic Press, 2020, pp. 545-548. doi:10.1016/B978-0-12-811984-6.00022-2.
- [8] Abhinav Nair, Alec Vishnevsky, Daniel R. Frisch, Atrial fibrillation or unstable angina? Utilization of a mobile electrocardiographic device to diagnose acute coronary syndrome, *Cardiovascular Digital Health Journal* 1 (2020) 52-54. doi:10.1016/j.cvdhj.2020.06.001.
- [9] Jefferson L. Vieira, Hector O. Ventura, Mandeep R. Mehra, Mechanical circulatory support devices in advanced heart failure: 2020 and beyond, *Progress in Cardiovascular Diseases* 63 (2020) 630-639. doi:10.1016/j.pcad.2020.09.003.
- [10] Farman Ali, Shaker El-Sappagh, S.M. Riazul Islam, Daehan Kwak, Amjad Ali, Muhammad Imran, Kyung-Sup Kwak, A smart healthcare monitoring system for heart disease prediction based on ensemble deep learning and feature fusion, *Information Fusion* 63 (2020) 208-222, doi:10.1016/j.inffus.2020.06.008.
- [11] Arman Kilic, Artificial Intelligence and Machine Learning in Cardiovascular Health Care, *The Annals of Thoracic Surgery* 109 (2020) 1323-1329. doi: 10.1016/j.athoracsur.2019.09.042.
- [12] Pan Ma, Shigong Wang, Ji Zhou, Tanshi Li, Xingang Fan, Jin Fan, Siyi Wang, Meteorological rhythms of respiratory and circulatory diseases revealed by Harmonic Analysis, *Heliyon* 6 (2020). doi:10.1016/j.heliyon.2020.e04034.
- [13] Soile Tapio, Mark P. Little, Jan Christian Kaiser, Nathalie Impens, Nobuyuki Hamada, Alexandros G. Georgakilas, David Simar, Sisko Salomaa, Ionizing radiation-induced circulatory and metabolic diseases, *Environment International* 146 (2021). doi:10.1016/j.envint.2020.106235.
- [14] Martsenyuk V.P., Sverstiuk A.S., Klos-Witkowska A., Horkunenکو A.B., Rajba S. Vector of Diagnostic Features in the Form of Decomposition Coefficients of Statistical Estimates Using a Cyclic Random Process Model of Cardiosignal, *The 10th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications*, volume 1 of IDAACS. Metz, France, 2019, pp. 298–303. doi: 10.1109/IDAACS.2019.8924398
- [15] Edward B. Panganiban, Arnold C. Paglinawan, Wen Yaw Chung, Gilbert Lance S. Paa, ECG diagnostic support system (EDSS): A deep learning neural network based classification system for detecting ECG abnormal rhythms from a low-powered wearable biosensors, *Sensing and Bio-Sensing Research* 31 (2021), 100398, <https://doi.org/10.1016/j.sbsr.2021.100398>.
- [16] W. A. Gardner, Cyclostationarity: Half a century of research, *Signal Processing* 86 (2006) 639–697. doi:10.1016/j.sigpro.2005.06.016.
- [17] W. A. Gardner Exploitation of cyclostationarity for identifying the Volterra kernels of non-linear systems, *IEEE Transactions on Information Theory* 39 (1993) 535–542. doi:10.1109/18.212283.
- [18] W. A. Gardner Fraction-of-time probability for time-series that exhibit cyclostationarity, *Signal Processing* 23 (1991) 273–292. doi:10.1016/0165-1684(91)90005-4.
- [19] W. A. Gardner, C. M. Spooner Higher-order cyclostationarity, in: *International Symposium on Information Theory and Applications*, ISITA '90, Honolulu, HI, 1990, pp. 355–358.
- [20] W.A. Gardner, T.L. Archer Exploitation of cyclostationarity for identifying the Volterra kernels of non-linear systems, *IEEE Transactions on Information Theory* 39 (1993) 535-542. doi:10.1109/18.212283.
- [21] W.A. Gardner, C.M. Spooner The cumulant theory of cyclostationary time-series. II. Development and applications, *IEEE Transactions on Signal Processing* 42 (1994) 3409-3429. doi:10.1109/78.340776.

- [22] S. Lupenko, N. Lutsyk, Y. Lapusta, Cyclic linear random process as a mathematical model of cyclic signals, *Acta mechanica et automatica* 9 (2015) 219–224. doi:10.1515/ama-2015-0035.
- [23] A. S. Sverstyuk, Argumentation and verification of the mathematical model of synchronously registered cardiosignals using the vector of cyclic rhythmically connected stochastic processes, *Measuring and Computing Devices in Technological Processes* 1 (2009) 143–147. doi:10.1515/ama-2015-0035.
- [24] I. Lytvynenko, A. Horkunenko, O. Kuchvara, Y. Palaniza, Methods of processing cyclic signals in automated cardiagnostic complexes, in: *Proceedings of the 1st International Workshop on Information-Communication Technologies and Embedded Systems*, Mykolaiv, Ukraine, 2019, pp. 116-127.
- [25] S.-W. Chen, H.-C. Chen, H.-L. Chan, A real-time QRS detection method based on moving-averaging incorporating with wavelet denoising, *Computer Methods and Programs in Biomedicine* 82 (2006) 187–195. doi.org/10.1016/j.cmpb.2005.11.012.
- [26] V. S. Chouhan, S. S. Mehta, N. S. Lingayat, Delineation of QRS-complex, P and T-wave in 12-lead ECG, *IJCSNS International Journal of Computer Science and Network Security* 8 (2008) 185–190.
- [27] Bommel J.H. van, Past and future research goals for computerized ECG processing, in: *Comput. ECG Anal.: Towards Stand. Proc. IFIP - IMIA Work Cont.*, Amsterdam e.a., 1986, pp. 367-381.
- [28] A.L. Goldberger, L.A. Amaral, L. Glass, J.M. Hausdorff, P.Ch. Ivanov, R.G. Mark, J.E. Mietus, G.B. Moody, C.-K. Peng, H.E. Stanley. *PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals*. *Circulation* 101(23):e215-e220 2000 (June 13). ULR : <http://circ.ahajournals.org/content/101/23/e215.full>.

# Methodology for Performance Analysis of Distributed Knowledge-Based Systems

Nataliia Kulykovska<sup>a</sup>, Artur Timenko<sup>a</sup>, Svitlana Hrushko<sup>a</sup> and Stepan Skrupsky<sup>a</sup>

<sup>a</sup> "Zaporizhzhia Polytechnic" National University, Zhukovsky str., 64, Zaporizhzhia, 69063, Ukraine

## Abstract

In this paper, the performance of distributed knowledge-based systems is the time spent by an application to execute messaging functions. To analyze the performance of distributed knowledge-based systems, direct performance measurement with tracing of web service function calls is used. The main advantage of this approach is the preservation of complete information about the interaction of processes in the trace, which makes it possible to analyze the performance of web services in detail. The identification of performance problems based on the analysis of the collected data is determined by the selected criterion. The paper proposes an expert methodology for solving the performance problem. The scientific novelty of the work lies in the presentation of the model of the problem of the performance of web services, as a set of actions of processes, certain combinations of which, under certain conditions, can lead to an identifiable performance problem.

## Keywords 1

Performance, analysis, web service, knowledge engineering, distributed knowledge-based systems, methodology

## 1. Introduction

The development of distributed knowledge-based systems requires research in the field of knowledge representation models, methods for describing them, and improving the performance of web services. Knowledge engineering in distributed knowledge-based systems implies the use of an ontological representation of knowledge. Thus, an object is formed that has a logical programmatic descriptive environment for further automatic machine processing [1].

The performance of distributed systems is a rather complex and multifaceted concept [2]. In [3] it is noted that under the metric of measuring performance, such diverse indicators as total operating time, parallel efficiency, scalability, memory requirements, bandwidth, latency, development cost, etc. can be considered.

In this paper, the performance of distributed knowledge-based systems refers to the time spent by an application to perform messaging functions.

Performance analysis is understood as the study of the values of service performance metrics [4]. In [5] the following goals of performance analysis are listed:

- determination of system performance under certain conditions;
- study of the system performance when varying parameters;
- comparison of performance of different systems;
- performance debugging - identifying the reasons why the system's performance does not meet expectations.

---

CMIS-2021: The Fourth International Workshop on Computer Modeling and Intelligent Systems, April 27, 2021, Zaporizhzhia, Ukraine  
EMAIL: natalya.gontar@gmail.com (N. Kulykovska); timenko.artur@gmail.com (A. Timenko); grushko\_ss@i.ua (S. Hrushko); sskrupsky@gmail.com (S. Skrupsky)  
ORCID: 0000-0003-4691-5102 (N. Kulykovska); 0000-0002-7871-4543 (A. Timenko); 0000-0002-0064-408X (S. Hrushko); 0000-0002-9437-9095 (S. Skrupsky)



© 2020 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

- performance tuning – i.e. determining the values of system parameters that provide the best overall performance.

## 2. Methods for analyzing and measuring the performance of distributed systems

Currently, there are three main methods for analyzing the performance of distributed systems [5]:

1. Analytical modeling [6]. This method provides a mathematical description of the system under consideration and, in comparison with other methods, it often provides the least detailed information about the system. But a simple analytic model can provide insight into the general behavior of a system or its components and what details need to be learned by other methods.
2. Simulation [7]. This method involves writing a simulator to simulate important characteristics of the original system. The advantage of this method is that the simulator can be easily modified to study certain aspects of the original system. The disadvantages include the inability to imitate every detail of the original system. As a result, simplifying assumptions must be made in order to have a reasonable execution time for the simulator, but this limits the accuracy of this method.
3. Measuring the performance of an existing system [8, 9]. In most cases, this method is preferred because no simplifying assumptions need to be made - the real system is being measured. On the other hand, this method is not very flexible, since provides information only about the system that it is measuring and for a specific set of its parameters. While it is possible to investigate performance for a few specific data parameters, it is generally difficult to do so.

When analyzing performance based on the method of measuring the performance of an existing system, there are three main methods of measuring performance [5, 10]:

- Event-driven method. This method collects performance indicators when a specific event occurs. In the simplest case, a global counter is used to count the number of events that have occurred. This method is best suited for events where the frequency of occurrence is not very high. otherwise, the behavior of the program changes.
- Tracing. This method is similar to the previous one, but in addition to recording the event that has occurred, some information about the state of the system is also used. The disadvantages of this method include increased memory requirements for storing the collected information, which can lead to an even greater change in the behavior of the program.
- Sampling. In contrast to the event-driven method, this method stores system state and performance metrics at fixed time intervals. As a result, the measurement overhead does not depend on the timing of the events. However, using this method, information is not collected about all the events that have occurred. The measurement result is the statistics of the program operation.

To analyze the performance of distributed knowledge-based systems, direct performance measurement with tracing of web service function calls is most often used. The main advantage of this approach is the preservation of complete information about the interaction of processes in the trace, which makes it possible to analyze the performance of web services in detail.

Improving the performance of distributed knowledge-based systems by directly measuring performance includes solving the following problems:

1. Choosing a criterion (metric) for evaluating the performance of a web service.
2. Collect performance data.
3. Visualization of performance data.
4. Identify performance problems.
5. Finding ways to fix performance problems.
6. Troubleshoot performance problems.

As noted above, the metric of measuring performance can be considered such diverse indicators as total execution time, parallel efficiency, scalability, memory requirements, bandwidth, etc.

Collecting performance data consists of tracing the calls to the web service functions. Depending on the selected criterion, one or several launches of the web service with representative data and certain parameters are performed on a particular computing system (or different systems).

The visualization of the collected performance data is usually done when identifying performance problems and finding ways to fix them will be done “manually”. There are various tools for

visualizing performance data, for example, space-time diagrams, function statistics, and others [11, 12].

Apache JMeter, HP Loadrunner, Microsoft Visual Studio, TSUNG are used for performance testing [13, 14, 15, 16].

Apache JMeter is a load testing tool developed by the Apache Software Foundation. The advantages of the tool include: conducting tests with complex logic and correlation of dynamic parameters; performance testing of web applications, including APIs, web services, and database connectivity; the ability to reproduce the behavior of several users with parallel threads.

HP LoadRunner has a distributed architecture (VuGen, Controller, Load Generator, Analysis). Provides a set of tools for working with various protocols. Provides detailed logging of the actions of each virtual user.

Microsoft Visual Studio provides the following advantages: storage of test results in MS SQL Server database and their convenient structuring; no restrictions on the number of virtual users with a license.

TSUNG - support for HTTP, WebDAV, SOAP, PostgreSQL, MySQL, LDAP, XMPP; convenience in testing API requests; load distribution on the cluster from client machines.

This review reviewed existing web service testing systems to improve system performance. Table 1 shows a comparative analysis of the considered systems.

**Table 1.**

Comparison of performance testing tools

| The challenge of increasing productivity | Apache JMeter | HP Loadrunner | Microsoft Visual Studio | TSUNG |
|------------------------------------------|---------------|---------------|-------------------------|-------|
| 1. Choice of metrics                     |               |               |                         |       |
| 2. Collecting performance data           | +             | +             | +                       | +     |
| 3. Performance data visualization        | +             | +             |                         |       |
| 4. Identifying performance problems      |               | +             |                         | +     |
| 5. Finding ways to fix problems          | +             |               |                         |       |
| 6. Troubleshooting                       |               |               |                         |       |

### 3. Models and methods of knowledge representation for performance analysis

A web service in a distributed knowledge-based system (Fig. 1) is considered as a set of interacting processes:

$$PR = \{pr_1, pr_2, \dots, pr_N\}. \quad (1)$$

The interaction of processes is carried out using actions initiated processes in a certain sequence:

$$PR_i \Rightarrow A_i = \langle a_{ij} \rangle; i = 1, 2, \dots, N; j = 1, 2, \dots, M_i. \quad (2)$$

Each action of the process consists in calling the functions of the web service:

$$F = \{f_k\}, k = 1, 2, \dots, K. \quad (3)$$

Each function is characterized by input and output arguments:

$$\begin{aligned}
f_k &\Rightarrow FA_k^{IN} = (fa_{k1}^{IN}, \dots, fa_{kL_k^{IN}}^{IN}); k = 1, \dots, K \\
f_k &\Rightarrow FA_k^{OUT} = (fa_{k1}^{OUT}, \dots, fa_{kL_k^{OUT}}^{OUT}); k = 1, \dots, K
\end{aligned} \tag{4}$$

Thus, each process action is characterized by a web service function, the values of its arguments when called and terminated:

$$\begin{aligned}
a_{ij} &= \langle f_k, FAval_k^{IN}, FAval_k^{OUT} \rangle; i = 1, 2, \dots, N; j = 1, 2, \dots, M_i; f_k \in F \\
FAval_k^{IN} &= (av_{k1}^{IN}, av_{k2}^{IN}, \dots, av_{kL_k^{IN}}^{IN}); k = 1, \dots, K \\
FAval_k^{OUT} &= (av_{k1}^{OUT}, av_{k2}^{OUT}, \dots, av_{kL_k^{OUT}}^{OUT}); k = 1, \dots, K
\end{aligned} \tag{5}$$

There are several messaging models that are defined by the syntax and semantics of function  $F$ .

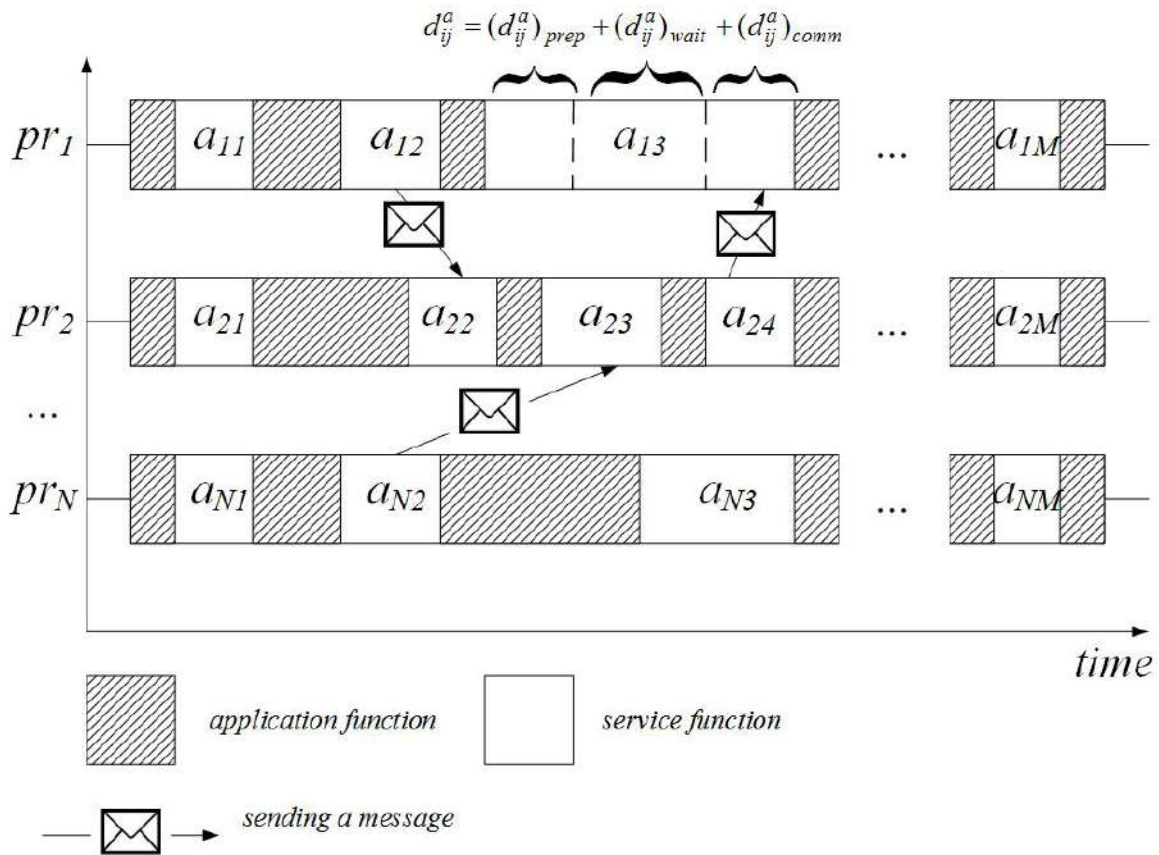


**Figure 1:** Model messaging web services

The duration of actions to call a web service consists of the following components:

$$d_{ij}^a = (d_{ij}^a)_{prep} + (d_{ij}^a)_{wait} + (d_{ij}^a)_{comm}, \tag{6}$$

where  $(d_{ij}^a)_{prep}$  - is the duration of data preparation for transmission / reception of messages;

$(d_{ij}^a)_{wait}$  - the duration of waiting for the web service to be ready;

$(d_{ij}^a)_{comm}$  - duration of transmission / reception of messages over the network.

#### 4. Identifying performance problems

The identification of performance problems based on the analysis of the collected data is determined by the selected criterion [17]. The search for ways to eliminate the identified performance problems can be carried out at different levels: the application level, the level of the web service

functions, the level of the computing system, etc. [18, 19]. To fix problems at the application level, you need to change the source code of the service.

The process of performing the listed tasks can be repeated (for example, if there are several performance problems) until the required performance is achieved. However, it is recommended that you start by fixing the problems that will yield the most results, so productivity tools should provide information about the extent to which performance problems affect the overall running time of the application.

Thus, the work proposes an expert methodology for solving the set task, including the following stages:

- systematization of typical productivity problems and establishment of the reasons for their occurrence;
- development of recommendations to eliminate these causes;
- a description of the identified performance problems, rules for determining the causes of their occurrence and recommendations for their elimination in the knowledge base.

Events are recorded in the track when calling action functions. A simple event is represented as:

$$e = \langle f, et, EPval, t, d, cs \rangle, \quad (7)$$

where  $f \in F$  - action function;

$et$  - the type of event that sets its parameters  $et \Rightarrow EP = (ep_1, \dots, ep_K)$ ;

$EPval = (v_1, \dots, v_K)$  - event parameter values;

$t$  - time of the event;

$d$  - event duration;

$cs$  - service code.

The following action tracing rules must be set to execute tracing:

$$a = \langle f, FAval^{IN}, FAval^{OUT} \rangle \xrightarrow{rule} e = \langle f, et, EPval, t, d, pr, cs \rangle, \quad (8)$$

$$rule = \langle f, et, EPtemp \rangle$$

where  $f \in F$  - function that triggers the tracer to generate a rule event;

$et$  - the type of event that sets its parameters  $et \Rightarrow EP = (ep_1, \dots, ep_K)$ ;

$EPtemp = \langle ep_i, Expr_i, kind_i \rangle; i = 1, \dots, N$  - description of event parameters and a method of obtaining their values, where:

$ep_i$  - event parameter;

$Expr_i : FAval^{IN} \times FAval^{OUT} \rightarrow v_i$  - function for calculating the parameter value  $v_i$  based on the arguments of the called function;

$kind_i \in \{in, out\}$  - character of parameter ( $in$  - input, i.e. calculated only on the basis of the input arguments of the function;  $out$  - output, i.e. formed on the basis of the output arguments of the function or input and output).

The values of the  $t$ ,  $d$ ,  $pr$  and  $cs$  parameters of the event are generated by the tracer.

Trace rules are specified for the subset of functions whose call actions cause the described performance problem. This provides selective tracing of activities. And the parameters of the events logged in the trace contain only the information that is relevant for identifying the described performance problem.

The work uses the following model of the problem of web services performance:

$$pb = \langle pd, dur, TrRulesAnRules, REC(A^{INFO}) \rangle, \quad (9)$$

where  $pd$  - verbal description of the problem;

$dur$  - the duration of the appearance of this problem;

$TrRules$  - rules for tracing the action of the problem;



*AnRules* - rules for recognizing the problem;

*REC* - recommendations for its elimination;

$A^{INFO} = \langle f, t, d, pr_i, cs_i \rangle$  - a description of the actions that led to the problem.

The rule for identifying a performance problem is represented as:

$$\begin{aligned} ce = \langle cet, CEPval, ME \rangle &\xrightarrow{pbrule} pb = \langle pd, dur, REC(A^{INFO}) \rangle \\ pbrule = \langle cet, \varphi, pd, RECtemp, L_{dur} \rangle \end{aligned}, \quad (10)$$

where *cet* – the type of compound event as a contender for the identified performance problem;

*CEPval* – the condition for the occurrence of the problem (if the value of the function is 1, then the problem exists and not otherwise);

*pd* – a verbal description of the problem determined by the rule;

*RECtemp* : *CEPval* x *REC* – recommendation template to fix the problem;

*Ldur* – function for calculating the duration of the performance problem.

It is concluded that there is a performance problem *pb*, when the conditions of the rule are met. The software environment recursively extracts simple events from the trace and generates information about a set of actions based on them.

The rules for identifying performance problems are described using a sequence of ontological axioms that have the syntax:

```
declare problem for <event name>
when
< logical expression >
parameters(
name = "< problem name >",
description = "< description of the problem >",
advice = "< recommendation >",
duration = < duration expression >);
```

where *declare problem for* is a description of the rule for identifying a performance problem for an applicant event *cet*;

*when* – condition for the occurrence of the problem (as a logical expression over the parameters of a composite event);

*name* – name of the problem;

*description* – verbal description of the *pb* problem;

*advice* – recommendation template for its resolution *RECtemp* ;

*duration* is an expression for calculating the duration of the manifestation of this problem *L<sub>dur</sub>* .

The analysis flowchart is illustrated in Fig. 2. The initial data for the analysis is the service message trace, formed as a result of the trace and containing simple events, and the analysis result is a set of identified performance problems. The analysis is carried out in two stages:

1. Constructing compound events as candidates for specific types of performance problems.
2. Identify performance problems from multiple constructed composite events.

For a practical solution to the problem of increasing productivity, special systems are being developed that allow evaluating the performance of web services and identifying the most "heavy" parts of the code (performance problems). Based on these data, the expert can make a decision to optimize the program code in order to reduce losses in these areas. A common disadvantage of such systems is that the task of identifying performance problems remains quite difficult due to the large volume of information being processed and the complexity of the links that generate individual problems. In addition, the performance problems identified by the expert and the decisions made to eliminate them in such systems do not persist, while most of the problems are, as a rule, of a fairly typical nature and can manifest themselves in other parallel applications. The practical value of this

methodology is in the development of a system capable of accumulating expert knowledge on identifying typical performance problems of distributed knowledge-based systems and ways to eliminate them.

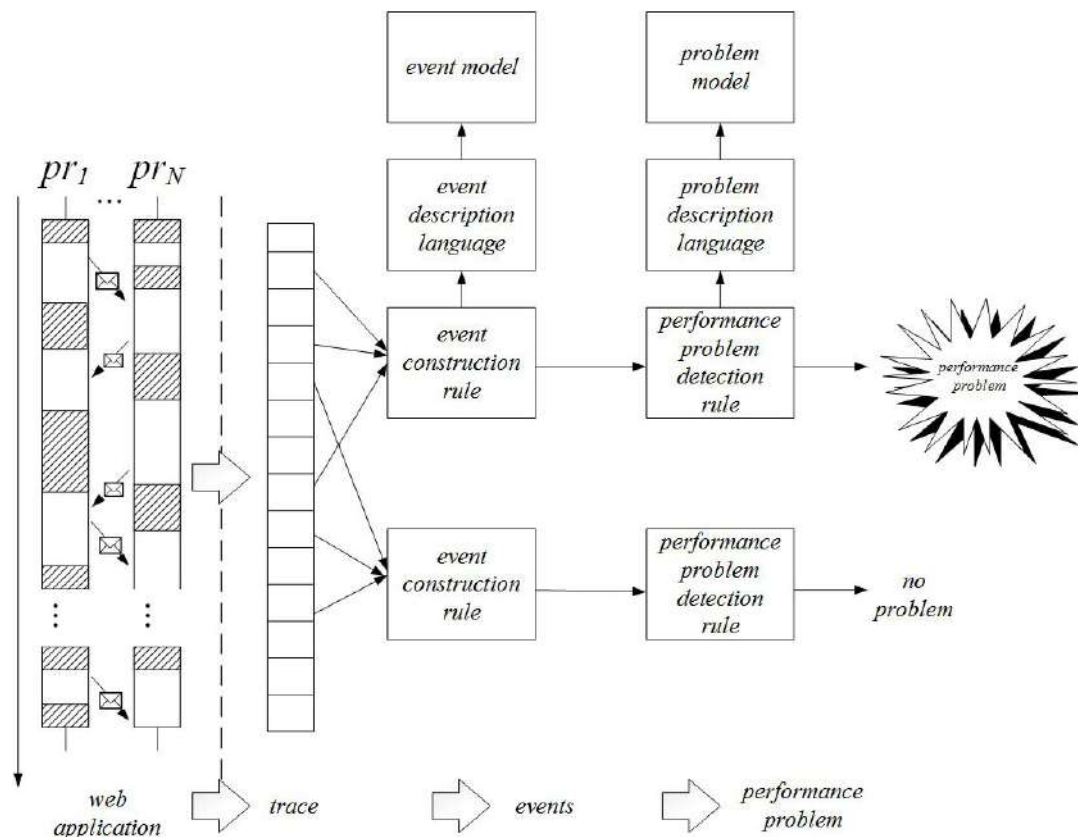


**Figure 2:** Performance analysis flow chart

All stages of performance analysis are implemented in a software environment for working with ontologies and the knowledge base of distributed knowledge-based systems. Descriptions of typical performance problems were included in the knowledge base of the system (Table 2).

**Table 2**

Performance challenges for distributed knowledge-based systems

| Problem type                         | Description                                                                                                     |
|--------------------------------------|-----------------------------------------------------------------------------------------------------------------|
| The problem of knowledge engineering | - service ontology error<br>- search error                                                                      |
| Bandwidth problem                    | - the limit of service operations has been exceeded<br>- the time of using the service has been exceeded        |
| The concurrency problem              | - late message sending<br>- late message reception<br>- sync error                                              |
| Failure problem                      | - memory access delays<br>- errors in the service<br>- errors in incoming parameters<br>- network / port errors |
| The resource problem                 | - not enough RAM<br>- insufficient physical memory                                                              |

|                                      |                                            |
|--------------------------------------|--------------------------------------------|
| Code problem                         | - program code errors<br>- output errors   |
| The problem of knowledge engineering | - service ontology error<br>- search error |

The general scheme of the system operation is shown in Fig. 3. Working with the system consists of two stages: preparatory (training stage) and application stage.

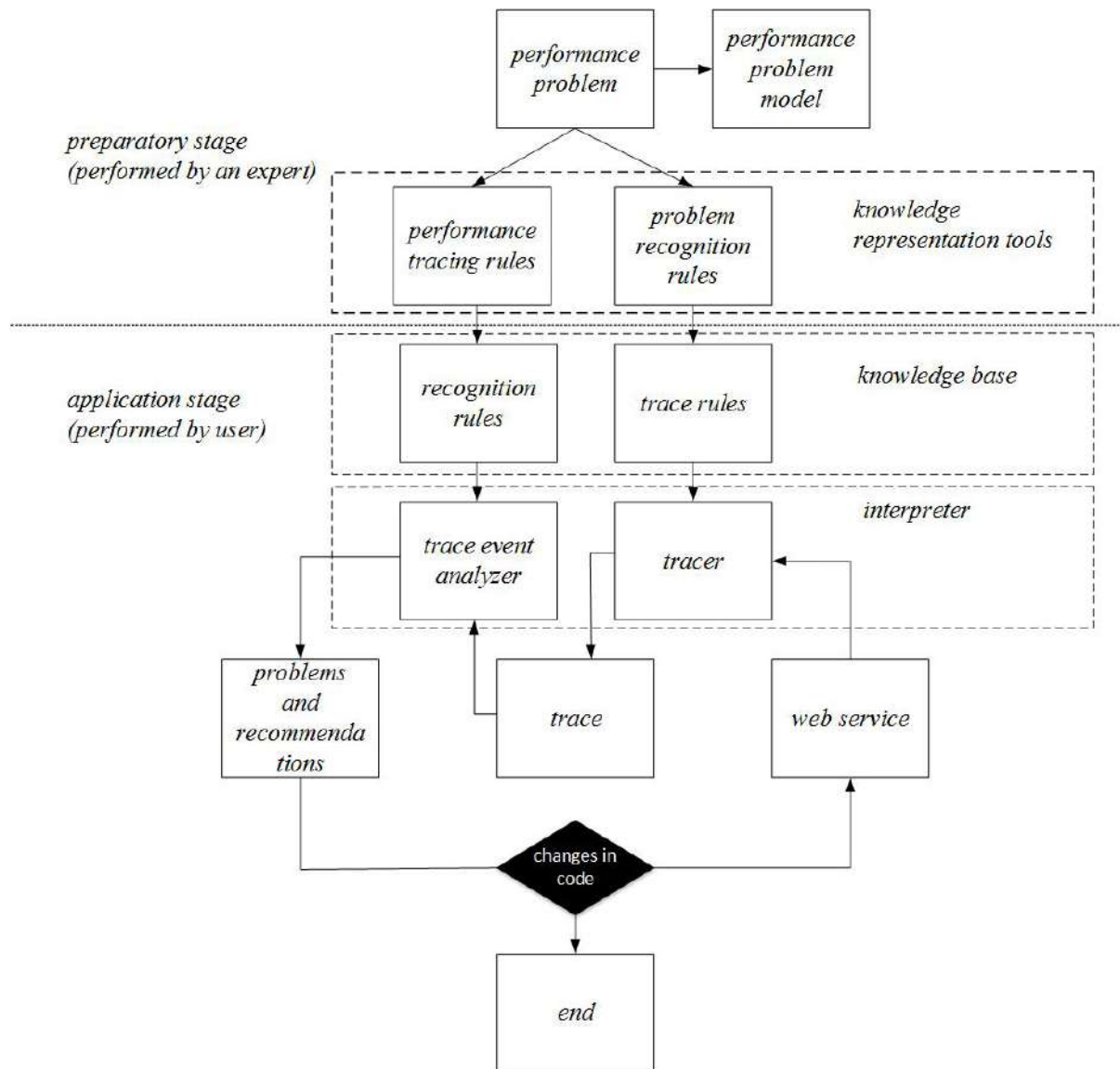


**Figure 3:** System operation diagram

The preparatory stage is performed by an expert after identifying the next performance problem. At this stage:

1. The expert describes the actions of the processes that (or a combination of which) can lead to the emergence of the identified problem. Actions are described in the form of a description of the functions of web services, the use of which is associated with the execution of these actions and the rules for tracing these functions.
2. The expert describes the rules for tracing actions and the rules for recognizing the identified problem.
3. The rules described by the expert are added to the knowledge base of the system.

The application phase is performed by the user. At this stage:

1. The system tracer is launched along with the application.
2. During the execution of the application, the tracer, guided by the tracing rules, generates a trace in which all the necessary actions of the application processes are recorded in the form of a sequence of events.
3. After executing the application, the trace analyzer, guided by the rules for recognizing known problems, determines the presence of problems, generates a list of them and recommendations for their elimination.

## 5. Conclusions

The concept of performance of distributed knowledge-based systems is given, the goals of performance analysis are listed, and it is noted that of a sufficiently large number of known performance criteria (metrics), the execution time of messaging exchange functions is most often used as the main indicator of overhead loss of web services.

This article discusses existing methods for analyzing the performance of distributed systems and methods for measuring performance. It is noted that for web services, direct performance measurement with tracing of web service function calls is most often used. The scientific novelty of the work lies in the presentation of the model of the performance problem of web services as a set of process actions, certain combinations of which, under certain conditions, can lead to an identifiable performance problem.

An overview of existing tools for testing the performance of web services is given. It is noted that most modern tools are based on visualizing performance data, and the rest of the tasks (identifying problems, finding ways and fixing problems) are solved manually by the user. Currently, systems are being developed that allow solving problems of identifying performance problems and finding ways to eliminate them in an automatic mode. It is necessary to set a fixed set of problems that they can analyze, as well as add new problems, without reprogramming the source code. The practical value of this methodology is in the development of a system capable of accumulating expert knowledge on identifying typical performance problems of distributed knowledge-based systems and ways to eliminate them.

All stages of the performance analysis methodology are implemented in a software environment for working with ontologies and a knowledge base of distributed knowledge-based systems. Descriptions of typical performance problems were included in the knowledge base of the system. Thus, the proposed model for representing knowledge in the form of an ontology capable of accumulating expert knowledge.

## 6. References

- [1] N. Kulykovska, S. Skrupsky, T. Diachuk. A model of semantic web service in a distributed computer system. CMIS-2020 Computer Modeling and Intelligent Systems: Proceedings of the Third International Workshop on Computer Modeling and Intelligent Systems (CMIS-2020) CEUR-WS.org, online. (2020) 338-351.
- [2] B. Gregg. Systems Performance: Enterprise and the Cloud. First edition, Upper Saddle River (New Jersey): Prentice Hall, 2013.
- [3] N. Herbst [et al.]. Quantifying Cloud Performance and Dependability: Taxonomy, Metric Design, and Emerging Challenges. ACM Trans. Model. Perform. Eval. Comput. Syst. 3, 4, Article 19 (2018) 36 pages. doi: 10.1145/3236332.
- [4] A. Balalaie, A. Heydarnoori, P. Jamshidi. Microservices architecture enables DevOps: Migration to a cloud-native architecture. IEEE Softw. 33, 3 (2016) 42-52.
- [5] D. Lilja. Measuring computer performance: A Practitioner's Guide, Cambridge: Cambridge University Press, 2000.
- [6] A. Grama. Introduction to Parallel Computing, 2nd edition, Boston: Addison-Wesley, 2003.

- [7] W. Serrai, A. Abdelli, L. Mokdad, Y. Hammal. Towards an efficient and a more accurate web service selection using MCDM methods. *Journal of Computational Science*. Volume 22 (2017) 253-267. doi:10.1016/j.jocs.2017.05.024.
- [8] L. Li, Allen D. Malony. Automatic Performance Diagnosis of Parallel Computations with Compositional Models. *Proceedings of the IEEE International Parallel and Distributed Processing Symposium*, Long Beach, CA, USA (2007) 1-8. doi: 10.1109/IPDPS.2007.370401.
- [9] S. Tsafarakis, Th. Kokotas, A. Pantouvakis. A multiple criteria approach for airline passenger satisfaction measurement and service quality improvement. *Journal of Air Transport Management*. Volume 68 (2018) 61-75. doi:10.1016/j.jairtraman.2017.09.010.
- [10] R. Jain. *The Art of Computer Systems Performance Analysis: Techniques for Experimental Design, Measurement, Simulation and Modeling*, New York: Wiley-Interscience (1991).
- [11] A. Ghasempour. Internet of Things in Smart Grid: Architecture, Applications, Services, Key Technologies, and Challenges *Inventions* 4, no. 1: 22 (2019). doi:10.3390/inventions4010022.
- [12] A. Sunardi, A. Suharjito. MVC Architecture: A Comparative Study Between Laravel Framework and Slim Framework in Freelancer Project Monitoring System Web Based. *Procedia Computer Science*. Volume 157 (2019) 134-141. doi:10.1016/j.procs.2019.08.150.
- [13] Apache JMeter™, 2021. URL: <https://jmeter.apache.org/>
- [14] LoadRunner Professional. Simplified project-based performance load testing solution to quickly identify abnormal application behavior, 2021. URL: <https://www.microfocus.com/en-us/products/loadrunner-professional/overview>
- [15] Visual Studio, 2021. URL: <https://visualstudio.microsoft.com/ru/>
- [16] Tsung 1.7.0 released, 2021. URL: <http://tsung.erlang-projects.org/>
- [17] P. McMullan. An Expert System Approach to Data Distribution and Distribution. *Proceedings of the 5th International Conference on Parallel Computing Technologies*, St. Petersburg (1999) 487-488.
- [18] C. Seragiotto [et al.] On Using Aksum for Semi-Automatically Searching of Performance Problems in Parallel and Distributed Programs. *Proceedings of the Eleventh Euromicro Conference on Parallel, Distributed and Network-Based Processing* (2003) 385-392.
- [19] S. Subbotin, A. Oliinyk, V. Levashenko, E. Zaitseva. Diagnostic rule mining based on artificial immune system for a case of uneven distribution of classes in sample. *Communications - Scientific Letters of the University of Zilina* Vol. 18(3) (2016) 3–11.

# Modern Systems for Processing and Analyzing GEO-data Based On OLAP Technology

Pavel Lukashevich<sup>a</sup>, Alexei Belotserkovsky<sup>a</sup> and Hrachya Astsatryan<sup>b</sup>

<sup>a</sup> *United Institute of Informatics Problems of the National Academy of Sciences of Belarus, 6 Surganova str. Minsk 220012, Republic of Belarus*

<sup>b</sup> *Institute for Informatics and Automation Problems of the National Academy of Sciences of the Republic of Armenia, 1 P. Sevak str. Yerevan 0014, Republic of Armenia*

## Abstract

In Belarus, the demand for applications that are focused on geodata processing has been significantly increasing recently. Although Belarus has its satellite BKA, there is also access to satellites of the Kanopus-V constellation, and highly relevant open data from meteorological satellites, only a few specialists of the corresponding profile can use them. To meet the wide demand for geodata analysis and expand the scope of their application, it was decided to study the modern experience of market stakeholders based on On-Line Analytical Processing (OLAP) technologies. The paper describes the currently popular approach to processing satellite images based on OLAP and Data Cube technology. The existing approaches from major market players such as Google Earth engine and Copernicus DIAS are compared with the alternative to develop and host an own service based on the Open Data Cube. All pros and cons are considered, the cost is compared, the possibilities of their coding, and the flexibility of the tools provided are also described here.

## Keywords 1

Satellite Image, Image Processing, On-Line Analytical Processing, GIS, Google Earth Engine, Copernicus Data and Information Access Services, Open Data Cube.

## 1. Introduction

Earth observations (EO), which include both in-situ and satellite data, provide robust monitoring to tackle environmental challenges, becoming critical because of the continuous pressure on natural resources. Remote sensing EO accurate and reliable data is a compulsory component of the environmental monitoring systems. Remote sensing open access data repositories offer precious, timely, and accurate remotely sensed EO information, such as American Landsat or European Sentinel satellites. The increasing complexities of practical tasks and applications, based on the operational analysis of remote sensing data, require new generations of remote sensing satellites providing more and more data of various types. Several new data services have emerged worldwide with broad potential to significantly impact on crucial environmental, economic and social issues at their local, regional, and global levels [1].

There is also a great demand in Belarus for operational geodata analytics applications. Since 2012 high-resolution remote sensing satellite BKA is operated in Belarus, providing high-precision capabilities to obtain remote sensing data focusing on the territory of Belarus [2] (operator GIS.by). Besides, it also supports receiving images from Russian Kanopus-V satellites [3]. There are other services for providing data that are relevant and unique for Belarus and its surroundings (see Table 1).

---

CMIS-2021: The Fourth International Workshop on Computer Modeling and Intelligent Systems, April 27, 2021, Zaporizhzhia, Ukraine  
EMAIL: pavel.lukashevich@newman.bas-net.by (P.Lukashevich); alex.belot@gmail.com (A.Belotserkovsky); hrach@sci.am (H.Atsatryan)  
ORCID: 0000-0001-9138-545X (P.Lukashevich); 0000-0002-8544-8554 (A.Belotserkovsky); 0000-0001-8872-6620 (H.Atsatryan)



© 2020 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

At present, these data are widely used in various expert systems, which are developed mainly for large government customers, such as the Ministry of Natural Resources, Forestry, Cadastral Services, agro-technical enterprises, the Ministry of Emergency Situations, and other major stakeholders [4].

**Table 1**

List of services

| Name                                                                      | Description                                                                                                                                                   | URL                                                           |
|---------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------|
| A Geoportal<br>"MeteoEye"                                                 | Provides highly relevant meteorological data received from various meteorological satellites to predict the weather and prevent emergencies and forest fires. | <a href="https://meteoeye.gis.by">https://meteoeye.gis.by</a> |
| Geoportal BelPSKHAGI                                                      | Provides ultra-high-resolution data received by manned and unmanned aircraft. This data is mainly used to update digital maps and cadastral needs.            | <a href="https://www.dzz.by">https://www.dzz.by</a>           |
| Geoportal of the Land<br>Information System of<br>the Republic of Belarus | Maintains its own set of high-resolution images for cadastral needs.                                                                                          | <a href="http://gismap.by">http://gismap.by</a>               |

The applications extract and process valuable information from remote sensing data to address societal and scientific challenges in the country and beyond. However, the high threshold for data processing specialists and infrastructure complexities to process the data limited the widespread usage to dive into the applied analysis of remote sensing data. Remote sensing data are highly heterogeneous in terms of the acquisition, digital representation formats, spatial resolution, etc. Besides, even today, many archived data from foreign and domestic satellites are used insufficiently, despite their wide availability.

The solution to this problem is complicated for countries with developed economies and even more difficult for developing countries interested in using EO. It is technically or financially unmanageable for many researchers to use their means to obtain and process data locally to solve applied problems taking into account the data size, data pre-processing and processing difficulties, and data storage and sharing.

Fortunately, data management and data analysis challenges are possible overcome with the support of relatively new infrastructural approaches to organizing data storage and processing. The article aims to present the modern systems for processing and analyzing GEO-data based on OLAP (Online Analytical Processing) technology, and the capabilities and experience of Belarussian and Armenian communities actively developing new paradigms and services for representing and processing remote sensing data. The remaining content of the paper is organized as follows. The local unique data in Belarus is given in Section 2. The overview of OLAP technology can be found in Section 3 and coding OLAP systems in Section 4. Finally, the Belarussian Data Cube system in Section 5 and the conclusions in Section 6.

## 2. Local unique data in Belarus

Since July 22, 2012, Belarus has its own high-resolution remote sensing satellite BKA operated by "Geoinformation Systems".

BKA, equipped with a panchromatic imaging system (PAN), was developed by order of the National Academy of Sciences of Belarus. PAN allows obtaining black-and-white images with a resolution of 2.1 m, and a multispectral imaging system (MUL) for acquiring images with a resolution of 10.5 m in four spectral ranges). The satellite has a mass of 400 kg, an orbit with an altitude of about 500 - 520 km. Also, "Geoinformation Systems" has access to Russian EO satellites Kanopus-V. Currently, these satellites form one constellation.

At present, GIS provides the most relevant high-resolution data on the territory of Belarus. These data are unique for Belarus and have a higher resolution and relevance than competitive platforms,

where high-resolution data is limited at all for our region at all, or is updated extremely rarely (once a year or worse).

In addition to space imagery in Belarus, remote sensing data are having been collecting by drones or manned aircrafts (such data are represented in geoportal BelPSKHAGI [5]). The resolution of such images is up to 0.05 meters per pixel. Unfortunately, the update frequency is not enough.

Currently, there are plans to launch a new generation satellite with a higher resolution - 0.35 m in pan chrome and 1.4 m in multispectral range. In addition to high-resolution data, the GIS operator has developed and maintains a new service called "MeteoEye" [6], which allows users to obtain operational data from a variety of available meteorological satellites: AQUA (AIRS, MODIS), MetOp-A (AVHRR, AMSU, HIRS, IASI, MHS), MetOp-B (AVHRR, AMSU, HIRS, IASI, MHS), MetOp-C (AVHRR, AMSU, IASI, MHS), NOAA18 (AVHRR, AMSU), NOAA19 (AVHRR, AMSU, HIRS, MHS), NOAA20 (ATMS, CRIS, VIIRS), Suomi NPP (ATMS, CRIS, VIIRS), TERRA (MODIS), Feng-Yun 3D (MERSI, HIRAS, MWHS, MWTS, MWRI).

MeteoEye data are received on Belarus and neighboring territory close to real-time mode (as the satellites fly over the receiving stations).

All these collected data may be used to: create and update topographic and navigation maps, control over land use and agricultural production, monitor land reclamation facilities, monitor changes in the forest fund, control forest inventory and manage reforestation, control constructions, control mining, control environment and manage hydrometeorology.

As already mentioned, there is a demand in the scientific community for the widespread use of these data, and it is also supported by the state's strategic plans for the informatization of society. The suggested solution is to apply of already proven approaches based on OLAP technologies.

### **3. Overview of OLAP technology**

The growth of the number of EO data sources and the improvement of their qualitative characteristics have been stimulating the research and development of science and technology specialists to simplify access to multidimensional data, their interactive online processing and analysis. Approaches to remote sensing data processing based on the OLAP methodology mean online preparation and delivery of aggregated information based on extensive data arrays structured according to the multidimensional principle. Currently, it has become widespread.

OLAP approaches led to developing a specialized high-performance database for storing, processing, and issuing large volumes of raster information RasDaMan (raster data manager), funded by the European Union.

In May 2018, an international team of specialists formulated a list of additions to the ISO SQL standard to provide the ability to operate with multi-dimensional raster data (datacube functionality). In June 2019, the ISO / IEC 9075-15: 2019 SQL multi-dimensional arrays (SQL / MDA) standard, expands the storage, processing, and delivery of multi-dimensional raster data through SQL database queries, was officially approved.

Based on OLAP approaches and SQL / MDA standards, nowadays, new services for storing, processing, and providing remote sensing data are widely developed worldwide. For example, the European Commission has launched an initiative to facilitate access to Copernicus data and information services. In addition to the classical approaches to data distribution services, a new conception of access was proposed. It is based on the DIAS (Data and Information Access Services) approach, which is designed to improve the methods of obtaining and processing Copernicus EO data. Based on this methodology, five large cloud platforms were created: CREODIAS, MUNDI, ONDA, SOBLOO, and WEKEO.

Other global market players have similar services [7], such as Google Earth Engine (GEE) [8], Earth on Amazon Web Services (EAWS) [9], PDGS-DataCube [10], Earth System Data Lab (ESDL) [11], and others [12].

Australian Geoscience DC (AGDC), renamed Digital Earth Australia (DEA) [13], was the first successful attempt, making the entire continent's geographical datasets available to researchers and



policy-makers [14]. Now several other countries have a national-scale DC, including Switzerland [15], Colombia [16], or China [17].

The Armenian DC [18] contains US Landsat and European Sentinel analysis ready data over the territory of Armenia. The full coverage of Armenia includes 11 Sentinel (38TLL, 38TML, 38TNL, 38TLK, 38TMK, 38TNK, 38SMJ, 38SNJ, 38SPJ, 38SNH, 38SPH) and 9 Landsat (171031, 170031, 169031, 171032, 170032, 169032, 168032, 169033, 168033) scenes. The Armenian DC is used to address diverse challenges and scientific questions in Armenia. For instance, to improve the hourly air temperature prediction for up to 24 hours in the Ararat valley based on machine learning methods and approaches by utilizing the EO data received from several meteorological stations and the large satellite analysis-ready datasets at different frequencies and resolutions [19].

Lessons [20] learned from design and implementation of AGDC underpin Chinese DC (CDC) based on the new Open Geospatial Consortium (OGC) Discrete Global Grids System (DGGS) standard and cloud computing technologies and Colombian DC. The Committee of EO Systems (CEOS) has a vision, that more over 20 countries will develop and realize their Data-Cube infrastructure by 2022 [21].

In addition to data processing convenience, DC is almost the only way to process Big Data. After all, for example, the European Union's Copernicus program generates up to 12 terabytes of data every day. Such large scale data offers various opportunities for climate change analysis, land monitoring, marine and coastal monitoring, atmospheric monitoring, human safety, and emergency and disaster management. However, self-loading and storing this data brings some sophisticated logistical challenges. DC answers these questions as users no longer need to download bulky files from multiple sources and process them locally. Instead, DC platforms provide massive cloud storage of satellite data, acting as a single point of access to data, allowing users to independently develop and deploy new topical applications in the cloud, including very complex ones that require large volumes and deficiency of good processing capacities for Big Data processing.

Big Data and cloud computing enable Earth scientists and developers to create web-accessible frameworks and platforms to store, retrieve and analyze Big Earth Data efficiently [22]. In the Earth science domain, scientists rely on a series of data models, frameworks, and initiatives to ensure heterogeneous data sharing and analysis.

DC is now considered a promising technology to perform time-series analyses of significant satellite Analysis Ready data-sets like Landsat and Sentinel [23]. There are several operational DC initiatives, covering various spatial scales and storing different data, using a wide range of infrastructures and software implementations – such as GEE, EAWS, PDGS, ESDL, and others – is a new paradigm aiming to meet the Big Earth Data challenge as a new approach to store, organize, manage and analyze EO data [24].

However, it is essential to distinguish Data Cubes for S&E (science and education) and commercial cloud-based processing facilities, such as DIAS, GEE, or EAWS. Cloud-based EO platforms commonly provide free and open access to global EO datasets (available datasets are growing daily) together with powerful space and time analysis tools supporting different programming languages (e.g., JavaScript, Python, and R). Recently, these online platforms were transformed into an environment where the user community works with satellite EO data. They remove most of the burden for data preparation, yield rapid results and foster a community of contributors, which is growing fast. However, it brings a hard dependency for users to work with a commercial platform, so with well-known challenges [25].

## 4. Coding OLAP systems

The main focus of this section is to present best practices for displaying and analyzing programs based on the well-known Google Earth Engine. The other services, which are Mundi (<https://jupyterhub.mundiwebservices.com>), PDGS Data Cube (<https://jupyter.pdgs.eo.esa.int>), Terrascope (<https://notebooks.terrascope.be>) and so on, will have a very similar IPython-based interface and Jupyter Notebook. One of the Google Earth Engine differences is that it provides standard interfaces for Python analogs and a JavaScript API, which is much more convenient when

embedding programs directly into a web application. The OLAP is widely used by the Armenian and Belarussian communities, such as the web-based interactive visualization and analytical platform for weather data in Armenia by integrating the three existing infrastructures for observational data, numerical weather prediction, and satellite image processing [26]. Another example is to apply OLAP system to develop a data quality alerting model for Big Data analytics [27].

One may practice coding JavaScript API at <https://code.earthengine.google.com>. Google Earth Engine JavaScript API code sample, development environment interface and results are presented in Figures 1 and 2.

```
var dem = ee.Image('USGS/SRTMGL1_003');
var xy = ee.Geometry.Point([86.9250, 27.9881])
var elev = dem.sample(xy, 30).first().get('elevation').getInfo()
print('Mount Everest elevation (m): ' + elev) //OUTPUT: 8729

// Set visualization parameters.
Map.setCenter(86.9250, 27.9881, 4);
var vis_params = { min: 0, max: 8729,
  palette: ['006633', 'E5FFCC', '662A00', 'D8D8D8', 'F5F5F5']}
Map.addLayer(dem, vis_params, 'Mount Everest DEM');
```

**Figure 1: Google Earth Engine: JavaScript API – Code Sample**

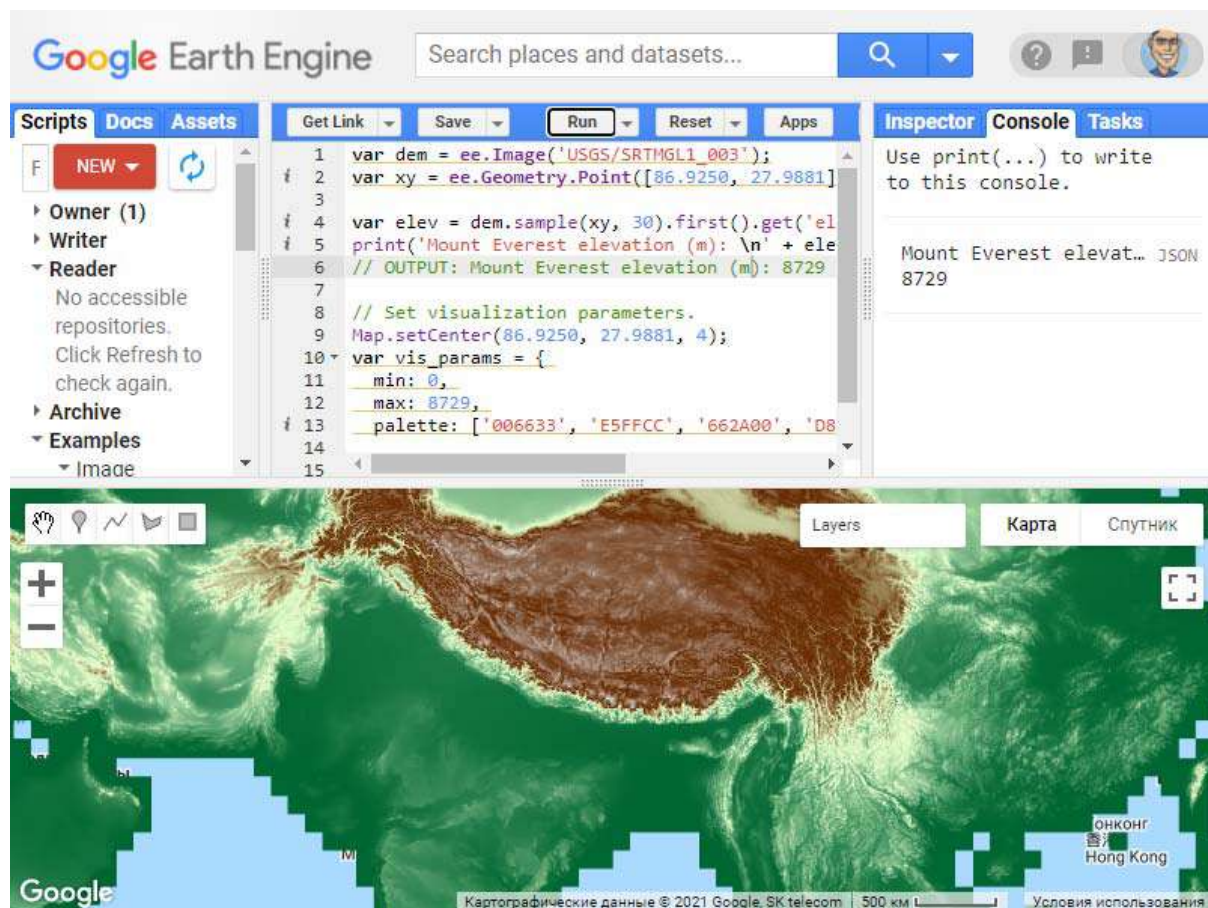


**Figure 2: Google Earth Engine: JavaScript API – Development Environment Interface and Results**

Also, Google Earth Engine provides access to its tools through the Python API. One may use the Python API from the Google Cloud Platform resources and through the Google Colab Platform (<https://colab.research.google.com> can be used for free).

Google Earth Engine Python API code samples are presented in Figures 3 and 5. The result of image generating from geodata, Google Colab development environment interface and result of map overlay are presented in Figures 4 and 6.

```

import ee
ee.Authenticate() # Trigger the authentication flow.
ee.Initialize()   # Initialize the library.

# Print the elevation of Mount Everest.
dem = ee.Image('USGS/SRTMGL1_003')
xy = ee.Geometry.Point([86.9250, 27.9881])

elev = dem.sample(xy, 30).first().get('elevation').getInfo()
print('Mount Everest elevation (m):', elev) # OUTPUT: 8729

# Import the Image function from the IPython.display module.
from IPython.display import Image
ROI = ee.Geometry.Rectangle([85.9250, 26.9881, 87.9250, 28.9881])
# Display a thumbnail of global elevation.
Image(url = dem.updateMask(dem.gt(0))
      .getThumbURL({'min': 0, 'max': 8729, 'dimensions': 512,
                    'palette': ['0000FF', '00FF00', 'FF0000'],
                    'region': ROI}))

```

**Figure 3:** Google Earth Engine: Python API Code Sample – Data Analysis and Image Composing

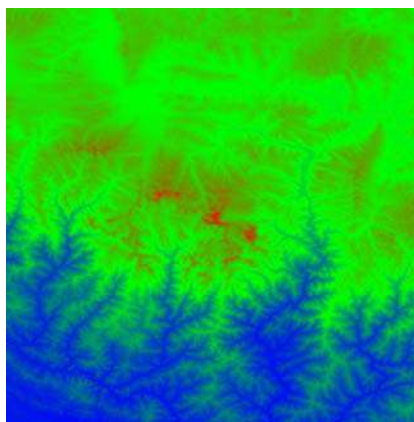


**Figure 4:** Google Earth Engine: Python API – Result of Generating of Image from raw Geodata

```

import ee
ee.Authenticate() # Trigger the authentication flow.
ee.Initialize()   # Initialize the library.

# Import the Folium library.
import folium

# Define a method for displaying Earth Engine image tiles to folium map.
def add_ee_layer(self, ee_image_object, vis_params, name):
    map_id_dict = ee.Image(ee_image_object).getMapId(vis_params)
    folium.raster_layers.TileLayer(
        tiles = map_id_dict['tile_fetcher'].url_format,
        name = name,
        overlay = True,
        control = True
    ).add_to(self)

# Add EE drawing method to folium.
folium.Map.add_ee_layer = add_ee_layer

# Set visualization parameters.
vis_params = {'min': 0, 'max': 4000,
              'palette': ['0000FF', '00FF00', 'FF0000']}

# Create a folium map object.
my_map = folium.Map(location=[27.9881, 86.9250], zoom_start=4)
my_map.add_ee_layer(dem.updateMask(dem.gt(0)), vis_params, 'Everest DEM')
my_map.add_child(folium.LayerControl())

# Display the map.
display(my_map)

```

**Figure 5:** Google Earth Engine: Python API Code Sample – Map Overlay

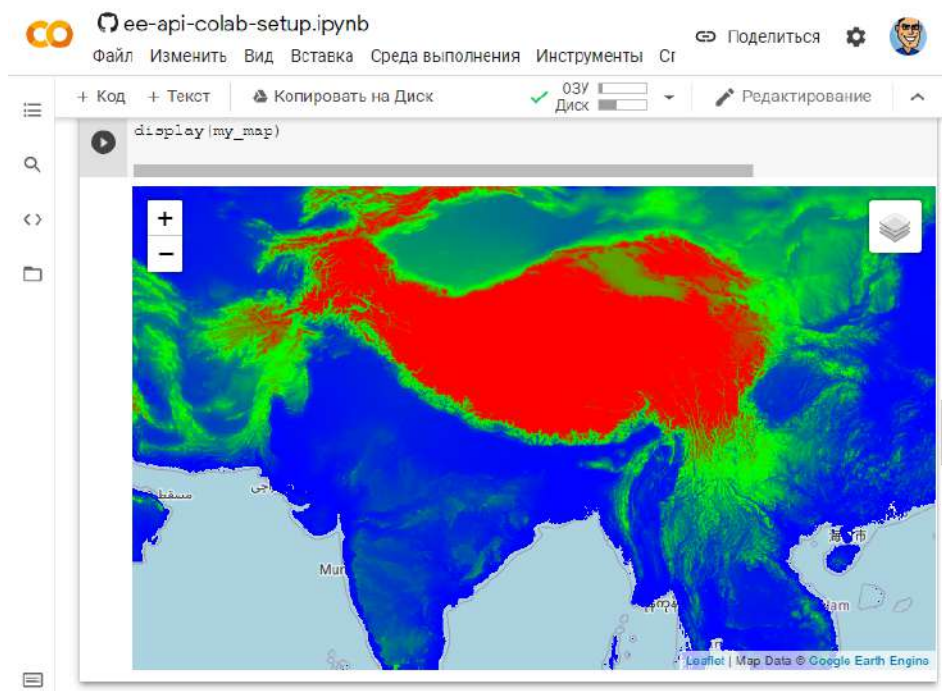


**Figure 6:** Google Earth Engine: Python API – Development Environment and Resulting Map Overlay

## 5. Belarusian Data Cube system

As seen from the previous section, modern EO information systems based on DC technology, such as Google Earth Engine, systems similar to Copernicus DIAS, and others, look quite attractive. Many of them even provide limited free access and a convenient development environment.

Unfortunately, the full functionality of such platforms for third countries (non-EU and not associated countries) is possible only under a commercial license or eligible at all. We provide a price assessment of using such systems.

As a basis we have taken the maximum configuration (LARGE) of the Mundi service from Copernicus DIAS (<https://mundiwebservices.com/offer>) – vCPU: 32, RAM: 256 GB, Storage: 80 TB. Such a configuration should be sufficient for the simultaneous work of several specialists and the background calculation of a very limited number of tasks.

The exact configuration for the Mundi service (<https://mundiwebservices.com/offer>) will be around 3000 EUR / month, similar to other Copernicus DIAS services. Google Cloud resources (<https://cloud.google.com/products/calculator>) look a bit costly – approx. 4000 USD / month, but AWS services (<https://calculator.aws>) are even higher – the minimum price for such a configuration is 4837 USD per month. The actual cost of servicing such a system on cloud platforms will be increased since the intensive use of the system presupposes the active transfer of large geodata, which is also charged.

But the cost issue is not the last one. Only data optimized to provide the best coverage for the European and North American regions are available for processing and analysis on the platforms listed. In most cases, there is no up-to-date high-resolution data for our region, and low-resolution weather data comes with a significant delay from 1 to 9 days. So such a set of data does not always provide the necessary accuracy and relevance to solve domestic scientific and practical problems in Belarus.

To order and receive additional high-resolution datasets for our region is not always possible through the platform. It may be available under require extra agreements and financial costs. Besides, Belarus already has its datasets optimized for current scientific needs (see section 2). The use of these data on platform resources will require downloading them to a remote resource (which is tens of gigabytes per day), payment for their storage (tens of terabytes), and their manual preprocessing and



combining with geodata presented on that platforms. It negates all the advantages of the cloud platform and the DC approach.

Therefore, the only option is to develop our DC system to build a reliable national system, which will increase the availability of our geodata, value and usability. When creating our own regional "BYCube" system, the main emphasis will be on receiving and downloading highly relevant data from various meteorological satellites (receiving delay up to 3 hours), and high-resolution EO data from the national satellite BKA and other satellites of the constellation.

Figure 7 shows a conceptual infrastructural scheme of the regional system "BYCube". We are deliberately tied to the Belarusian Earth Remote Sensing System (elements enclosed in the first rectangle, excluding meteorological satellites), which is hosted by the Unitary Enterprise "Geographic Information Systems". For us, as a parent organization, this means the fastest possible automatic retrieving, preprocessing, and uploading of highly relevant data from a variety of meteorological satellites, as well as high-resolution remote sensing data from the national satellite BKA at the organizational level. On the other hand, the BASNET academic network is a provider of computing services and connectivity for state science and education, and primarily for scientists of the National Academy of Sciences of Belarus. BASNET is the national research and educational network of the GÉANT Association and has the necessary capacity to host and further support the "BYCube".

It is assumed a lot of work on the study and testing of the technologies available at the present stage to implement a DC that allow on-demand scaling of the computing infrastructure and expanding its functionality using a deep learning algorithm for analyzing EO data.

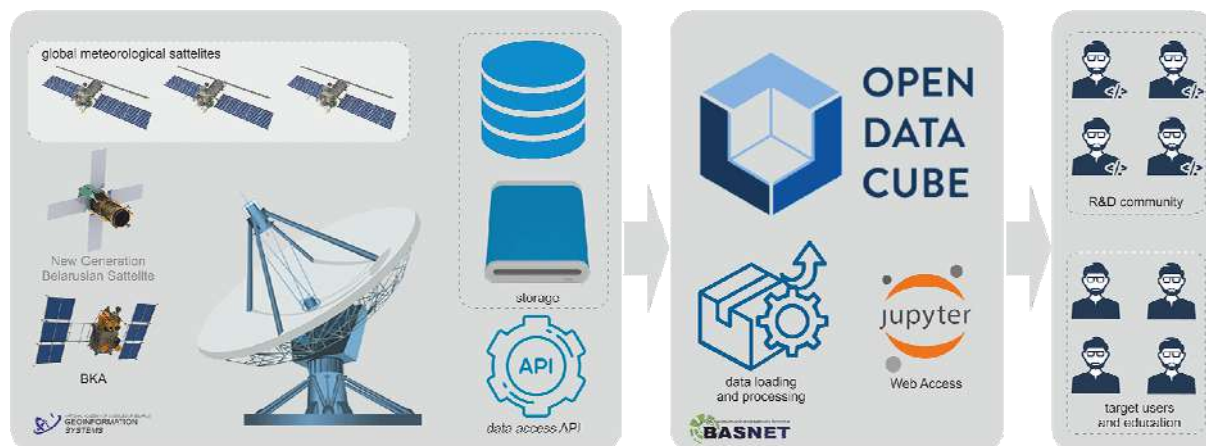


**Figure 7:** Conceptual infrastructural scheme of the regional system "BYCube"

## 6. Conclusions

The implementation of an online computing infrastructure based on the DC technology for the needs of Belarusian researchers and institutions of higher education will significantly increase the value of the received EO data in Belarus due to:

- the minimization of time and special knowledge required for access and preparation of satellite data;
- accumulation of free and open satellite data in Belarus;
- providing easy access to open-source software solutions for remote sensing data processing, which are promoted thanks to the contribution of the community;
- using consistent data structures that allow code, tools, and algorithms to be shared;
- joint use of several heterogeneous sets of remote sensing data in complex analysis algorithms.

The first demonstration applications on the developed platform are planned utilities for the operational determination of heat anomaly maps, temperature distribution maps, cloud masks, and snow masks based on data from meteorological satellites. These data can be used both independently and together with high-resolution EO data. In the future, it is planned to scale BYCube by increasing

the applied algorithmic repository by active users from S&T community, which in turn will contribute to the formation of scientific potential and new specialized personnel in the academic environment capable of quickly solving applied problems of EO.

## 7. Acknowledgements

The work is supported by the EaPConnect project funded by European Union within the EU4Digital initiative under Grant Agreement number ENI / 2019 / 407-452. Provide scheme will be elaborated in the frame of state activity “To create a software package for collecting, storing and processing remote sensing data based on technology “Data Cube”, providing remote access to up-to-date multi-satellite remote sensing data for the Republic of Belarus, their operational centralized processing and analysis for solving scientific, technical and educational problems” subprogramme 6 “Exploration and use of outer space in peaceful purposes” of the State programme “Science-intensive technologies and equipment” for 2021 - 2025.

## 8. References

- [1] L. Friedl, Benefits Assessment of Applied Earth Science, in: M. Onoda, O. R. Young (Eds.), Satellite Earth Observations and Their Impact on Society and Policy, Springer, Singapore Singapore, 2017, pp. 73–79, ISBN 978-981-10-3713-9.
- [2] Belarusian spacecraft (BKA), Research Center for Earth Operative Monitoring. URL: <http://gis.by/en/tech/bka>.
- [3] Kanopus-V, ESA Earth Observation. URL: <https://earth.esa.int/web/eoportal/satellite-missions/k/kanopus-v-1>.
- [4] S. A. Zolotoy, A. V. Kosilo, A. A. Stavrov, I. B. Strashko, Software solutions in the field of geoinformatics, in: Proceedings of the National Academy of Sciences of Belarus, physical-technical series 4 (2016) 113–120.  
URL: <https://vestift.belnauka.by/jour/article/download/281/279>.
- [5] BelPSKHAGI Geoportal. URL: <https://www.dzz.by>.
- [6] MeteoEye Geoportal. URL: <https://meteoeye.gis.by>.
- [7] K. Anderson, B. Ryan, W. Sonntag, A. Kavvada, L. Friedl, Earth observation in service of the 2030 Agenda for Sustainable Development, Geo-Spat. Inf. Sci. 20 (2017) 77–96.
- [8] Google Earth Engine, A planetary-scale platform for Earth science data & analysis. URL: <https://earthengine.google.com>.
- [9] Earth on AWS, Build planetary-scale applications in the cloud with open geospatial data. URL: <https://aws.amazon.com/earth>.
- [10] PDGS-DataCube. URL: <http://eodatacube.eu>.
- [11] The Earth System Data Lab (ESDL), Earth in a Box. URL: <https://www.earthsystemdatalab.net>.
- [12] A. Lewis, L. Lymburner, M. B. J. Purss, B. Brooke, B. Evans, A. Ip, A. G. Dekker, J. R. Irons, S. Minchin, N. Mueller, Rapid, high-resolution detection of environmental change over continental scales from satellite data – The Earth Observation Data Cube, Int. J. Digit. Earth 9 (2016) 106–111.
- [13] Digital Earth Australia (DEA). URL: <https://www.ga.gov.au/dea>.
- [14] A. Gore, The Digital Earth, Australian Surveyor, 43 (1998) 89–91.
- [15] EO for monitoring the environment of Switzerland in space and time, Swiss Data Cube (SDC). URL: <https://www.swissdatacube.org>.
- [16] C. Ariza-Porras, G. Bravo, M. Villamizar, A. Moreno, H. Castro, G. Galindo, E. Cabera, S. Valbuena, P. Lozano, CDCol: A Geoscience Data Cube that Meets Colombian Needs, in: A. Solano, H. Ordoñez (Eds.), Proceedings of the Advances in Computing, Springer International Publishing, Cham Switzerland, 2017, pp. 87–99.
- [17] X. Yao, Y. Liu, Q. Cao, J. Li, R. Huang, R. Woodcock, M. Paget, J. Wang, G. Li, China Data Cube (CDC) for Big Earth Observation Data: Lessons Learned from the Design and Implementation, in: Proceedings of the 2018 International Workshop on Big Geospatial Data and Data Science (BGDDS), Wuhan China, 22–23 September 2018, pp. 1–3.

- [18] S. Asmaryan, et al. Paving the Way towards an Armenian Data Cube, *Earth Observation Data Cubes*, 4 (2019) 117. URL: <https://www.mdpi.com/2306-5729/4/3/117>.
- [19] H. Astsatryan, H. Grigoryan, A. Poghosyan, et al. Air temperature forecasting using artificial neural network for Ararat valley, *Earth Sci. Inform.*, 2021. URL: <https://doi.org/10.1007/s12145-021-00583-9>. doi:10.1007/s12145-021-00583-9.
- [20] Open Data Cube Manual, Open Data Cube Documentation. URL: <https://datacube-core.readthedocs.io/en/latest>.
- [21] S. R. Rizvi, B. Killough, A. Cherry, S. Gowda, The Ceos Data Cube Portal: A User-Friendly, Open Source Software Solution for the Distribution, Exploration, Analysis, and Visualization of Analysis Ready Data, in: *Proceedings of the IGARSS 2018–2018 IEEE International Geoscience and Remote Sensing Symposium*, Valencia Spain, 22–27 July 2018, pp. 8639–8642.
- [22] H. Guo (Ed.), *Manual of Digital Earth*, Springer Nature, 2020. URL: <https://doi.org/10.1007/978-981-32-9915-3>. doi: 10.1007/978-981-32-9915-3.
- [23] Earth Observation: Copernicus Sentinel Satellite Data-Open Access, ESA Earth Observation. URL: <http://open.esa.int/copernicus-sentinel-satellite-data>.
- [24] An Open Source Geospatial Data Management & Analysis Platform, Open Data Cube. URL: <https://www.opendatacube.org>.
- [25] G. Giuliani, J. Masò, P. Mazzetti, S. Nativi, A. Zabala, Paving the way to increased interoperability of Earth Observations Data Cubes, *Earth Observations Data Cubes*, 4 (2019) 113, JRC117467, ISSN 2306-5729.
- [26] H. Astsatryan, H. Grigoryan, E. Gyulgyulyan, A. Hakobyan, A. Kocharyan, W. Narsisian, V. Sahakyan, Y. Shoukouryan, R. Abrahamyan, Z. Petrosyan, J. Aligon, *Weather Data Visualization and Analytical Platform. Scalable Computing: Practice and Experience* (2018).
- [27] E. Gyulgyulyan, and J. Aligon, F. Ravat, H. Astsatryan, Data Quality Alerting Model for Big Data Analytics, *International Workshop on Qualitative Aspects of User-Centered Analytics - European Conference on Advances in Databases and Information Systems (QAUCA@ADBIS 2019)*, Bled, Slovenia.

# Towards Designing of Recommendation System for Recruiting of Software Development Teams

Karina Melnyk, Natalia Borysova, Zoia Kochuieva and Dina Huliieva

National Technical University "Kharkiv Polytechnic Institute", Kirpichova, 2, Kharkiv, 61002, Ukraine

## Abstract

The problem of recruiting software development teams based on the choice of potential candidates is considered. An analytical review of methods for creating recommendation systems has been conducted. A suitable similarity measure based on the type of input data of the domain area has been chosen. The recruiting task of software development teams in an IT-company is formalized. The functional model of business-process of recruiting of software development teams has been developed. The method of resolving the recruiting task has been proposed. The algorithm for solving the issue is given. The proposed microservice architecture of the information system, which has been used for solving the problem of automation of the personnel selection process, has been improved at the expense of creating and adding a new microservice to solve the recruiting task of software development teams. The architecture of recommendation system in the form of deployment diagram and database model are developed. The team building process according to technical appraisal is demonstrated. The evaluation process of the quality of the team building is described. The set of quality metrics and the convolution criterion are proposed.

## Keywords 1

Recommendation system, team building, software development teams, content-based filtering approach, similarity measure, quality assessment

## 1. Introduction

Recommender systems (RS) long ago have taken their rightful place in various spheres of human life and activity. They are used to provide the user with goods, services, information, etc. The conceptual framework of such actions are most relevant search requests, the analysis of user preferences, the analysis of the goods, services, information, etc. However, the potential of recommendation systems in the field of project management has not yet been fully disclosed. There is no doubt that now there are many software products, which allow automating some of the routine actions of a project-manager (PM) and head resource manager (HR). Some of them, such as Applicant Tracking Systems, are used to search for candidates for vacant company positions [1, 2], others are aimed directly at the project management process when a software development team has already been formed [3, 4]. Nevertheless, the recruiting task of teams remains without support of RS.

Today, team building problems are mostly solved by HR-managers manually. The research [1] proposes a solution to the task of personnel selecting for team building according to the technical appraisal. As a result, a list of candidates with suitable technical skills for each vacant position is formed without taking into account psychological characteristics. Further, HR and PM conduct face-to-face interviews with each potential applicant to assess the candidate's communication, leadership and personal qualities. However, it is a time-consuming process. Therefore, some HR-managers use

---

CMIS-2021: The 4th International Workshop on Computer Modeling and Intelligent Systems, April 27, 2021, Zaporizhzhia, Ukraine  
EMAIL: karina.v.melnyk@gmail.com (K. Melnyk); borysova.n.v@gmail.com (N. Borysova); aliseiko@gmail.com (Z. Kochuieva);  
dgulieva@ukr.net (D. Huliieva)  
ORCID: 0000-0001-9642-5414 (K. Melnyk); 0000-0002-8834-2536 (N. Borysova); 0000-0002-4300-3370 (Z. Kochuieva);  
0000-0001-8310-745X (D. Huliieva);



© 2021 Copyright for this paper by its authors.  
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).  
CEUR Workshop Proceedings (CEUR-WS.org)



psychological techniques for recruiting software development teams as additional stage of appraisal. This approach allows realizing the intellectual and creative potential of employees in accordance with the strategic goals of the IT-company. Many authors have researched the problem of team building based on psychotypes, for example, Robert Hogan, Bruce Tuckman, Dr. Meredith Belbin, D. Keirsey [5-7]. The task of automation of the process of recruiting software development team based on their members' psychological compatibility is highly relevant. There are many benefits of this approach. Firstly, it will speed up the decision-making process on this issue, because a computer takes much less time to process large amounts of information than a human. Secondly, automating this process will free up the PM's and HR-manager's time for other tasks. Therefore, the main point of this research is dedicated to the development of a recommendation system with particular attention to the function of the recruiting of software development team based on their psychological compatibility.

## 2. Formal problem statement

Let's consider the formulation of the recruiting task of software development teams. Input data for this task is information about the psychotypes of potential candidates. Designate  $C = \{C_1, \dots, C_n\}$  as a set of potential candidates with suitable technical skills for the software development teams from employees and non-employees. The set has formed as result of resolving the personnel selection task according to [1]. The psychological portrait of a potential candidate is a set of competencies that was built based on the model for superior performance developed by L. M. Spencer, such as result orientation, analytical thinking, search for information, etc. [8].

Let  $T = \{T_1, \dots, T_m\}$  is a set of possible teams. Therefore, the task of team building based on psychotypes is a mapping of one set to another  $f: C \rightarrow T$ . If cardinality of the  $\|T\| = 1$ , then the candidate can belong to several variations of one team; if cardinality of the  $\|T\| \geq 2$ , then the candidate can belong to several variations of every team. The final decision depends from the goals of the IT-company. For instance, the team  $T_m$  is more important for the IT-company compared to the team  $T_{m-2}$ , that means the best candidate, who fits both teams, has to belong to the team  $T_m$ .

So, the purpose of the article is to resolve the recruiting task of software development teams using the recommendation system.

To achieve the goal of the given research, it is necessary to:

- conduct analytical review of methods for creating recommendation systems;
- choose a method for recommendation system for team building;
- develop a method for the recruiting task of software development teams;
- conduct numerical researches;
- assess the proposed approach of team building.

## 3. Literature review

There are some traditional approaches to creating of RS [9, 10]:

- content-based filtering approach;
- collaborative filtering approach;
- hybrid approach.

Content-based filtering approach provides recommendations based on comparing information about the user and the object of recommendations. Firstly, a profile of users and recommendation objects are built using the same criteria and the similarity measure of these profiles is estimated; then the closest objects to the user's profile are selected for the recommendation.

Collaborative filtering approach provides recommendations based on comparing information about the preferences of different users with different characteristics and makes recommendations to new users with similar characteristics. The characteristics of recommendation objects are not taken into account in this approach.

Hybrid approach combines both of the previous approaches, trying to use their advantages and neutralize disadvantages.

To solve the given problem, it is efficient to use a content-based filtering approach. A fully formed profile of the employee is compared with the profile of the “ideal” team member for this project or with a set of requirements for a vacant position. Such problem statement moreover allows avoiding the disadvantages of the content-based filtering approach: the limited content analysis and over-specialization, because the profile of the “ideal” team member and real candidates should not contain any gaps.

An analysis of various sources of information has shown the methods used for content-based filtering have been divided into two main groups [9, 10]. The first group of methods calculates different similarity measures between the considered objects. The second group of methods is data mining methods, which involves the use of different machine learning methods. The choice of the method in this case depends on the domain area, the available statistical material and the power of the computing system. Input data for solving the task of recruiting of software development teams is the psychological portrait of some potential candidate to appropriate job. To resolve the given task, it is necessary to compare “ideal” portrait and information about a real candidate. It means to calculate the similarity between two objects. Thus, the next stage is choosing the similarity metric corresponding to the input data of the considered domain. There are many such metrics [11-12]. To find a degree of similarity between quantitative data, there are many metrics: the Euclidean measure, the Manhattan measure or City block distance, the Mahalanobis measure, the Chebyshev measure, and others. The presented metrics can also be used for interval data. For qualitative information, we can use the Hemming similarity measure, the measure of Rogers-Tanimoto, Rao coefficient. The distance between objects described by binary or dichotomous data can be measured using the Pearson metric or square contingency coefficient, the Jacquard coefficient, and the metrics used for qualitative data: the Hamming measure and the Rogers-Tanimoto measure. For mixed data, it is recommended to use Gower’s coefficient, Zhuravlev’s metric, Voronin’s similarity measure, Mirkin’s metric.

According to the L. M. Spencer’s model, each characteristic from the psychological portrait of some potential candidate is measured from one to ten, depending on the degree of manifestation of a particular characteristic. Therefore, it is proposed to calculate only the similarity measure for quantitative data. Let’s consider the most commonly used metrics [11-12], where  $C_k = \{x_1, \dots, x_n\}$  is a  $k$ -th candidate with psychological portrait  $\{x_1, \dots, x_n\}$  and  $C_{ideal} = \{x_1', \dots, x_n'\}$  is a “ideal” portrait accordingly:

1. Euclidean distance. It is the most common distance function, that represents geometric distance in multidimensional space:

$$\rho(C_k, C_{ideal}) = \sqrt{\sum_{i=1}^n (x_i - x_i')^2} \quad (1)$$

2. Squared Euclidean distance. It is used to give more weight to objects that are more distant from each other:

$$\rho(C_k, C_{ideal}) = \sum_{i=1}^n (x_i - x_i')^2 \quad (2)$$

3. Manhattan distance. The distance between two points measured along axes at right angles. For this measure, the influence of individual large differences (outliers) decreases (since they are not squared):

$$\rho(C_k, C_{ideal}) = \sum_{i=1}^n |x_i - x_i'| \quad (3)$$

4. Chebyshev distance. This distance can be useful when we need to define two objects as “different” if they differ in any one coordinate:

$$\rho(C_k, C_{ideal}) = \max_i |x_i - x_i'| \quad (4)$$

An analytical review of the proposed similarity metrics (1)-(4) from the entire arsenal of existing metrics and features of the domain area allows choosing the Euclidean distance as a measure for this research.

## 4. Materials and methods

### 4.1. Formalization of the team building process

The task of recruiting of software development team for the particular project is a multi-criteria task with a large number of restrictions and requirements. Any client always wants to receive the end-product as soon as possible with maximum benefit and with limited resources, despite the requirements of the stakeholders for the product can be contradictory. That is why the project planning and recruiting task of team for this project are very important components of the successful implementation of a quality product. To implement an effective business process in order to achieve the maximum result in team recruiting, it is necessary to create a model of this process. There are many techniques for modelling different business process, such as methodology IDEF0, BPMN Specification, Flowcharts diagrams, Data Flow diagram, notation eEPC. To formalize given business process, it is proposed to use Business Process Modeling Notation (BPMN). The BPMN-diagrams allow describing business processes in a unified standardized language that is understandable to all stakeholders, regardless of their level of technical knowledge. Let's consider the process of team recruiting in more detail way (Fig. 1).

The team recruiting process begins with the formation of a set of requirements that apply to the qualitative and quantitative composition of the team. In order to draw up a list of requirements for potential candidates for the software development team, it is necessary to hold a kickoff meeting. Project Scope and Project charter should be formed during this meeting. These documents are the basis not only for team requirements, but also for software requirement specification. The requirements to technical skill for the competence profiles of IT-specialists for this project are drawn up according to the wishes of the PO and the existing working conditions in the IT-company. The next step is the recruiting process of potential members of software development teams. The HR-manager firstly checks the resumes of those employees who are currently in the personnel reserve.

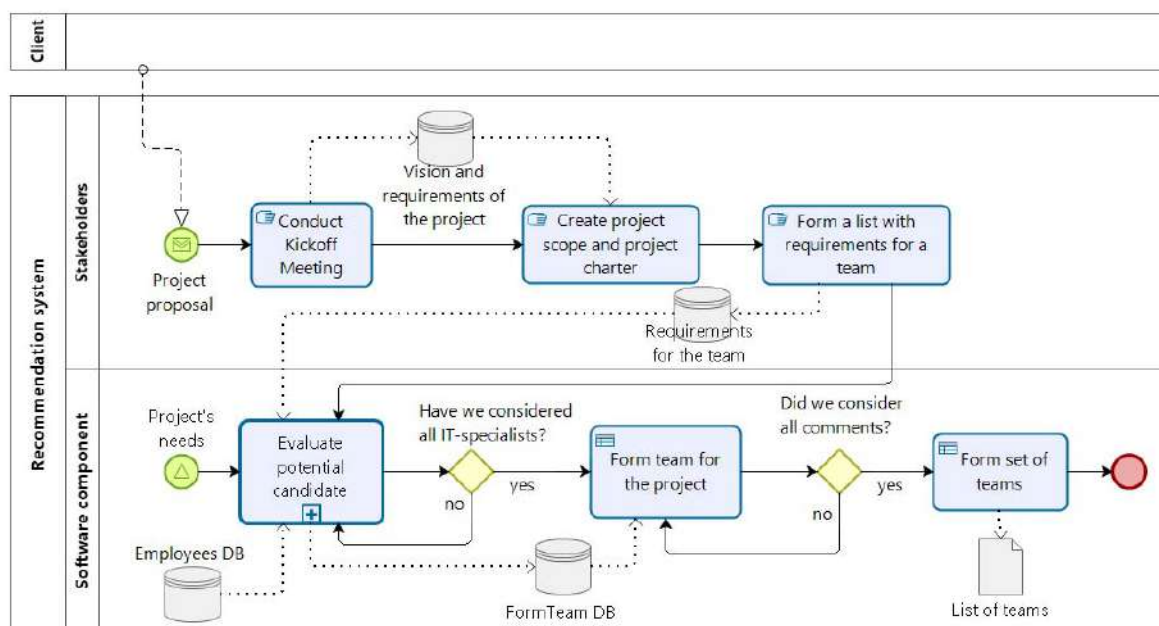


**Figure 1:** The functional model of the team building process

Next step is checking the employees who are currently included in projects from clients. Sometimes there is a need for a particular IT-specialist who has been already allocated to another project. In this case, this employee can be hired like a consultant. If a proper candidate were not find in the company, HR-manager would search for potential candidates in the Internet. The process of evaluating an IT-specialist can also be initiated due to the immediate needs of the project, for example, in case of dismissal or illness of one of the team members. The task of evaluating potential candidates for the software development teams is considered in [1, 13-15]. As a result of this stage,

potential candidates are screened and set of candidates are formed for each position in the project team. The next step is to solve the problem of team recruiting for new project. Depending on the needs of the PO, it could be recruited one or more teams with different team members.

Let's consider the task of team recruiting in more detailed way.

## 4.2. The method of resolving the recruiting task of software development teams

The team building task begin with defining the team structure. The management of the IT-company assign the PM for a particular project. The PM and the Product Owner (PO) further determine the number of people in the team and the distribution of roles between them according to the project budget.

Let's denote  $R$  as a set of different roles in a team, for example,  $r_i$  – development leader,  $r_k$  – back-end developer,  $r_l$  – quality assurance engineer. Then  $R_i, i \in R$  is the number of team members of  $r_i$ -th role in the team. So, all values of  $R_i$  are the input data from PM and PO.

The result of solving the personnel recruiting task problem is group of sets with potential candidates for each role in the project team [1]. If  $C = \{C_1, \dots, C_n\}$  is a set of potential candidates for the software development teams from employees and non-employees, then  $C_i$  – is a set of potential candidates for the  $r_i$ -th role in the team.

Let denote  $K$  as set of competencies for the psychological portrait of an IT-specialist, then  $x_k$  – is a value of the  $k$ -th competence. This variable has view  $x_k^{pi}$  for a particular  $p$ -th ( $p \in C_i$ ) specialist of  $r_i$ -th role in the team.

The algorithm for solving the recruiting task of software development teams is a step-by-step filling out the template of the hierarchical structure of the team based on proper candidates according to the results of calculating the similarity measure. Therefore, it is necessary to compute the similarity metric between potential candidates. Designate  $s_{ij}^{pq}$  as the similarity measure of  $p$ -th and  $q$ -th potential candidates between the  $r_i$ -th and the  $r_j$ -th roles in the team,  $p \in C_i, q \in C_j$ . If  $i = j$ , then  $s_{ii}^{pq}$  is a similarity measure between people with similar role in the team. The Euclidean measure as a similarity measure between candidates has been proposed in this study. It means the similarity measure can be found according to (1) as follows:

$$s_{ij}^{pq} = \sqrt{\sum_{k \in K} (x_k^{pi} - x_k^{qj})^2}, p \in C_i, q \in C_j. \quad (5)$$

Let's consider the common method of team building for a project under sufficient financial, human and time resources.

Step 1. The choosing of PM.

The Project Manager is the most important and responsible specialist for the project, who is the connecting link between the client and the team members. Therefore, he/she is the basis for the selection of the rest of the team. The formation of the team is carried out according to the principle of choosing the candidates who are closest to the Project Manager in terms of psychological characteristics.

Step 2. Appointment of specialists of the next management level: Development Leader, Quality Assurance Leader, Architect.

There are three variants:

- There are proper IT-professionals who are immediately assigned to selected roles.
- This option provides several potential specialists for one position. At the same time, the PM is ready to work with any of them because of the past positive mutual experience and communication. In this case, there is no need to calculate the similarity measure. The decision about the appointment to the position is made after the possible teams have been selected and analyzed.

- If employees and non-employees apply for the position, then the choice is based on calculating the similarity measure between the PM and possible candidates using the formula (5). The obtained set of values of the metric  $s_{ij}^{pq}$  is sorted in ascending order. The best candidate is determined as follows:

$$candidate_{best} = \min_{C_j} \{s_{ij}^{pq}\} \quad (6)$$

where  $p$  is the Project Manager in this case,  $q$  is any person from the set  $C_j$ .

In the face of uncertainty, three lists of candidates are formed: the first list corresponds to the group of people applying for the Development Leader role, where the top names from the list belong to people with the highest priority, the second list being candidates for Quality Assurance Leader, and the third one for Architect.

Step 3. Assessment of the set  $C_k$  ( $C_k \subset C$ ).

Let's consider the set of potential candidates  $C_k$  for the  $r_k$ -th role in the team. All following actions depend on the results of comparing the cardinality of the set with the number of required employees:

$$action = \begin{cases} |C_k| < R_k \Rightarrow go\ to\ HR - manager \\ |C_k| == R_k \Rightarrow C_k \subset \forall T_d (d = \overline{1, m}) \\ |C_k| > R_k \Rightarrow go\ to\ step\ 4 \end{cases} \quad (7)$$

The number of potential candidates is not enough to fill the vacancy in the first case, so it is necessary to return to the previous stage to solve the problem of personnel selection. The second option describes a situation when the candidates' set of  $r_k$ -th role is a part of all variations of the future team. The third case implies the resolving of the recruiting task.

Step 4. Recruitment of specialists for the  $r_k$ -th role in the team.

Similarly to step 2, the similarity measure between the top-level manager (Development Leader, Quality Assurance Leader, Architect) and the corresponding candidates from the set  $C_k$  is calculated using formula (5). The choice is also carried out from a ranked list according to the formula (6). If there is a need to select several candidates, then they are selected with the smallest values of the similarity measure.

Step 5. Checking the team.

If the team is not formed, then go to step 3. If the team is fully completed, then go to step 2 to form an alternative team. If all possible options are considered, then RS has to provide a list of possible teams for evaluation and analysis to PM.

Thus, the method of forming of project's team based on the psychological portrait of candidates has been considered.

### 4.3. Architectural solution of the recruiting task of software development teams

The software design always starts from the defining of the functional and non-functional requirements. Let us look more closely at the functional requirements for RS. The main user of RS is PM. He or she has the ability to view the lists of candidates for each vacant position in the project. These lists are formed by IS after solving the task of personnel selection [1]. PM can view the information about each candidate available in the system; can find a group of people according to proximity based on psychological characteristics. There is a possibility to assign a Development Leader, a Quality Assurance Leader and an Architect. The RS allows recruiting or assigning candidates to teams for the selected Development Leader, Quality Assurance Leader and Architect. The PM can assign a particular team from the list to a project. The RS can create a report with list of generated teams. Non-functional requirements include an intuitive and user-friendly interface, reliability of data transfer and storage, usability, and high performance. An important requirement is

the fact that the forming task of the software development teams uses the results of solving the recruiting task [1]. Thus, RS has to coordinate its actions with IS from [1] or be a component of it.

The research [1] has shown the feasibility of using the microservice architecture as a distributed system architecture based on the analysis of existing IS for the solving various tasks related to the assessment and selection of personnel. This approach shows how to build the software as a set of small, asynchronously interacting services, which implements one of the tasks from the list of ones, which are formulated in the analytical review of the subject area and solved using the developed models. The architecture of IS from [1] consists of several microservices:

1. Microservice Employee – this component allows processing information about employees in the IT-company.
2. Microservice Candidate serves for finding and monitoring potential candidates for vacant positions.
3. Microservice Estimation allows assessing the competencies of employees and candidates.
4. Microservice Identify – a database that contains authentication information about each user of the system.

To solve the problem of forming the software development team, it is necessary to add other microservice, which will form teams with potential candidates. Microservice architecture allows using different technologies to solve each problem. Therefore, the following architecture of the microservice for the recruiting task has been proposed. It is presented in the form of a deployment diagram (Fig. 2).

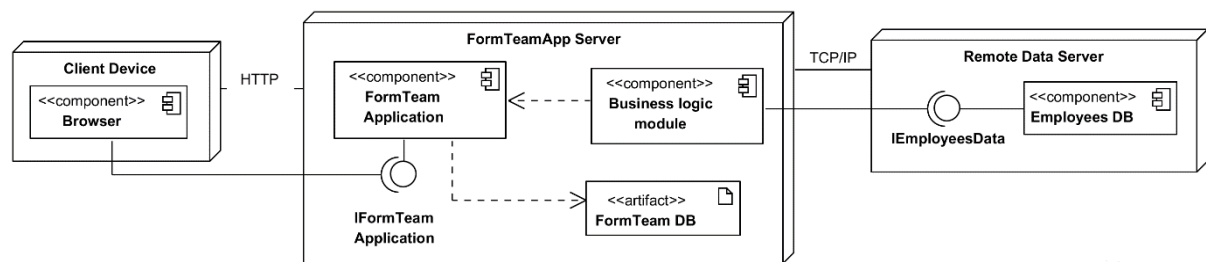


**Figure 2:** Deployment diagram for the software component

A three-tier architecture to deploy the software component has been chosen. The advantages of the chosen architecture are scalability, high level of security and reliability, low requirements for network speed between clients and the application server, and low requirements for performance and technical characteristics of clients. Let's consider the presented architecture in more detail.

The Client node includes the following components: Web browser and JQuery – JavaScript framework. The FormTeamApp Server node is an Internet Information Services server, which allow the application to work. There are many components on this node: FormTeam Application is a web application, Business logic module is a component for implementing the logic of work between domain objects, FormTeam DB is a component that contains information about formed teams with potential candidates, IFormTeam Application is an application interface. The Remote Data Server node is a system database location node that uses the SQL Server database. This node deploys the database, which is used by the IIS. The FormTeam DB database for storing, changing and processing information about the subject area, that is, about employees, potential candidates, projects, etc., was developed. The model of this base is represented using the IDEF1x notation (Fig. 3).

Database consists of 13 entities:

- the entity «Employees» contains information about all employees of the company;
- the entity «PMs» contains information about all employees of the company in the position of PM; these employees are placed in a separate table, because of the company's management should be able to choose particular PM from the list to manage a new project;
- the entity «IT\_specialists» contains information about all employees in the positions of IT-specialists, except for project managers; these employees are placed in a separate table, because there are many employees who are not involved in the implementation of IT-projects;
- the entity «Candidates» contains information about employees and non-employees who took a part in a job interviews;

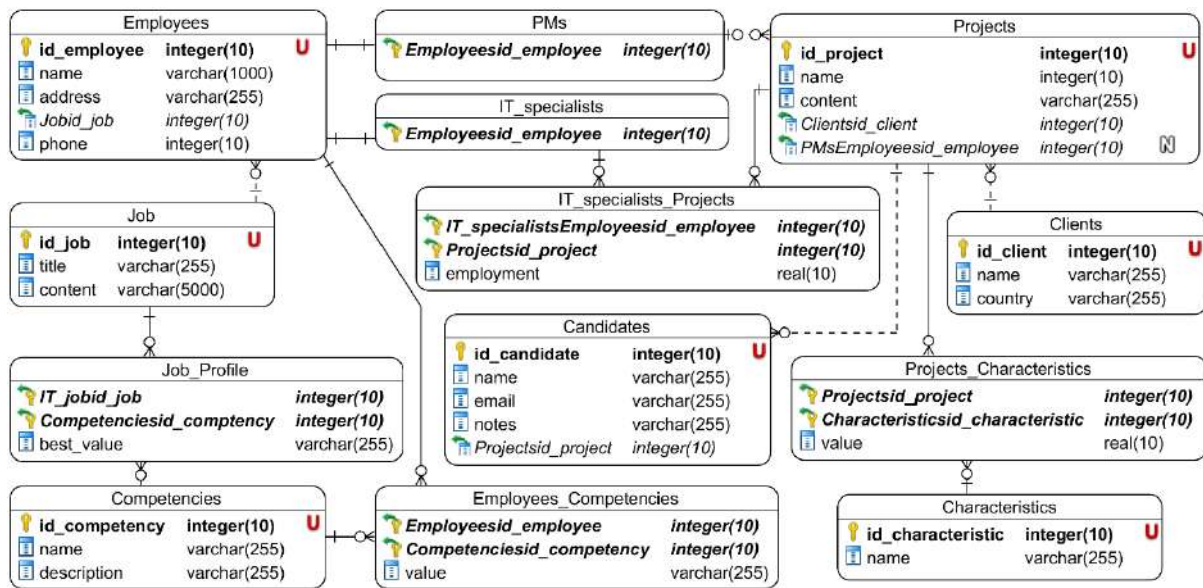


**Figure 3: Model of database**

- the entity «Job» contains information about all positions in the IT-company;
- the entity «Competencies» contains the competencies from job competency profiles;
- the entity «Projects» contains information about all projects of the IT-company;
- the entity «Clients» contains information about all clients of the IT-company including the country and the time zone;
- the entity «Characteristics» contains information about all characteristics of the projects;
- the associative entity «Job\_Profile» shows the competencies corresponding to a particular position; it contains a description of the ideal profile of each position;
- the associative entity «IT\_specialists\_Projects» shows IT-specialists who were involved or are currently working on the company's projects;
- the associative entity «Projects\_Characteristics» shows the characteristics corresponding to each project of the company;
- the associative entity «Employees\_Competencies» shows the levels of competence of the employees.

Thus, the RS architecture has been presented to solve the task of forming of software development teams.

#### 4.4. Results and Discussion

Let's consider the using of the developed RS for recruiting of software development teams. One foreign company has ordered a software from an IT-company. The names of the candidates, name of project and IT-company have been hidden. Management of the IT-company has assigned particular PM to this project. According to the proposed method of the team building, first step is done. Project Manager and HR-manager with Product Owner have chosen the following structure of the team from 10 IT-specialists: Project Manager (1 person), Development Leader (1 person), Quality Assurance Leader (1 person), Architect (1 person), Middle Front-end Developer (2 persons), Middle Back-end Developer (2 persons), Middle QA Engineer (1 person) and Junior QA Engineer (1 person). Next step is an appointment of specialists of the next management level: Development Leader, Quality Assurance Leader, and Architect. The input data is the information about employees of the IT-company and non-employees with suitable technical appraisal. The result of solving the task of selecting potential candidates according to [1] is following: five people claim for the vacant position of Development Leader in the project; four people claim for two vacant positions of Middle Front-end Developers, and three people for two vacant positions of Middle Back-end Developers. Each from group of five Development Leaders can be assigned on this role. So, final decision depends from

different variants of possible teams. The Table 1 presents the input data of potential candidates based on the L. M. Spencer model for superior performance.

**Table 1**

The psychological picture of candidates to software development team

| Employees          | Focus on achieving results | Analytical thinking | Conceptual thinking | Customer service orientation | Development of others | Influence | Information search | Leader skills | Teamwork and cooperation | Self-confidence |
|--------------------|----------------------------|---------------------|---------------------|------------------------------|-----------------------|-----------|--------------------|---------------|--------------------------|-----------------|
| DevLead 1          | 8                          | 10                  | 7                   | 8                            | 9                     | 8         | 7                  | 8             | 9                        | 10              |
| DevLead 2          | 9                          | 9                   | 6                   | 9                            | 8                     | 7         | 6                  | 9             | 8                        | 7               |
| DevLead 3          | 8                          | 6                   | 10                  | 9                            | 8                     | 7         | 8                  | 9             | 5                        | 6               |
| DevLead 4          | 6                          | 8                   | 8                   | 6                            | 7                     | 8         | 9                  | 10            | 6                        | 7               |
| DevLead 5          | 8                          | 6                   | 7                   | 7                            | 6                     | 7         | 8                  | 9             | 7                        | 8               |
| MidFront-end Dev 1 | 7                          | 8                   | 9                   | 8                            | 7                     | 9         | 9                  | 9             | 10                       | 9               |
| MidFront-end Dev 2 | 8                          | 7                   | 9                   | 9                            | 8                     | 7         | 8                  | 9             | 8                        | 9               |
| MidFront-end Dev 3 | 9                          | 8                   | 9                   | 10                           | 6                     | 8         | 7                  | 9             | 8                        | 9               |
| MidFront-end Dev 4 | 6                          | 5                   | 8                   | 9                            | 7                     | 6         | 8                  | 9             | 8                        | 7               |
| MidBack-end Dev 1  | 10                         | 9                   | 6                   | 8                            | 7                     | 5         | 6                  | 5             | 9                        | 8               |
| MidBack-end Dev 2  | 9                          | 8                   | 6                   | 9                            | 8                     | 7         | 6                  | 8             | 9                        | 9               |
| MidBack-end Dev 3  | 9                          | 10                  | 9                   | 7                            | 8                     | 10        | 10                 | 9             | 8                        | 7               |

Next step is calculating the matching between candidates. The results of calculating the similarity measure for five employees from the Development Leaders' list (designated DevLead 1-5) are presented in the Table 2. It is the base for forming the development teams. The best candidates for the team are highlighted in bold.

**Table 2**

An example of calculating the similarity measure for recruiting of software development teams

| Employees from second and third lists | DevLead 1          | DevLead 2          | DevLead 3          | DevLead 4          | DevLead 5          |
|---------------------------------------|--------------------|--------------------|--------------------|--------------------|--------------------|
| MidFront-end Dev 1                    | <b>4,582575695</b> | 6,082762530        | 6,855654600        | <b>5,291502622</b> | 5,099019514        |
| MidFront-end Dev 2                    | <b>4,472135955</b> | <b>4,690415760</b> | <b>4,472135955</b> | <b>5,196152423</b> | <b>3,872983346</b> |
| MidFront-end Dev 3                    | 5,000000000        | <b>4,582575695</b> | 5,567764363        | 6,324555320        | 4,690415760        |
| MidFront-end Dev 4                    | 7,141428429        | 5,916079783        | <b>4,582575695</b> | <b>5,291502622</b> | <b>3,741657387</b> |
| MidBack-end Dev 1                     | 5,744562647        | <b>5,000000000</b> | 8,660254038        | 8,831760866        | 6,633249581        |
| MidBack-end Dev 2                     | <b>3,316624790</b> | <b>2,645751311</b> | <b>7,141428429</b> | <b>7,071067812</b> | <b>4,898979486</b> |
| MidBack-end Dev 3                     | <b>5,567764363</b> | 6,244997998        | <b>6,708203932</b> | <b>5,099019514</b> | <b>6,324555320</b> |

The analysis of the Table 2 allows seeing three candidates from the second list instead of the two required one for DevLead 4. The reason is equal similarity measure of two candidates. Therefore, RS has formed two alternative teams for this employee. Thus, the result of using the RS is six alternative software development subteams.

The rest of the team is formed in a similar way: the similarity measures between the Quality Assurance Leader and his subordinates are calculated, and the Architect is selected. The RS forms a set of recommendations. It is a framework for the PM, who forms the final structure of the team for the project.



For checking the adequacy of working of RS with the opinion of experts, it is necessary to choose special metrics. The Recall and Precision metrics are inapplicable to assess the efficiency of the work of the developed RS, since the RS does not produce irrelevant responses. The metrics will always be equal to 100%. However, there is another useful metric. It is the *AP* metric – Average Precision. It takes into account the order of the recommendations of proposed software teams. It can be used for evaluating the efficiency of RS if the recommendation task is considered as a ranking task [16]. *AP* is calculated using the following formula:

$$AP@N = \frac{1}{m} \sum_{k=1}^N (P(k) \text{ if } k^{\text{th}} \text{ item was relevant}) = \frac{1}{m} \sum_{k=1}^N P(k) \cdot rel(k), \quad (8)$$

where  $rel(k)$  is an indicator that shows when  $k^{\text{th}}$  item was relevant ( $rel(k)=1$ ) or not ( $rel(k)=0$ );  $N$  is a number of recommendations made by RS;  $m$  is the number of relevant items from list of recommendations made by RS;  $P(k)$  is Precision at cutoff  $k$ , that is simply the precision calculated by considering only the subset of RS' recommendations from rank 1 through  $k$ .

Let's consider an example of calculating the *AP*-metric. The RS formed six alternative development subteams and arranged them in a certain order. An expert arranged formed six teams by RS in a certain order as well. The results of the allocation of teams are presented in the Table 3.

**Table 3**  
Results of the allocation of software development teams

| Number in the list | Expert | RS     | Correct or not |
|--------------------|--------|--------|----------------|
| 1                  | Team 5 | Team 5 | Correct        |
| 2                  | Team 1 | Team 1 | Correct        |
| 3                  | Team 4 | Team 4 | Correct        |
| 4                  | Team 2 | Team 3 | Incorrect      |
| 5                  | Team 3 | Team 2 | Incorrect      |
| 6                  | Team 6 | Team 6 | Correct        |

Thus, the metric *AP* according to (8) is equal to:

$$AP@6 = \frac{1}{6} \left( \frac{1}{1} + \frac{2}{2} + \frac{3}{3} + \frac{3}{4} + \frac{3}{5} + \frac{4}{6} \right) = 0,84.$$

The obtained value characterizes the adequacy of the RS operation, since the order of the formed teams reflects the priority of the command. If the importance of the commands is approximately equal, the expert can arrange in a different order than the recommendations from the RS. Thus, the obtained results indicate the results of the RS operation reflect the stated requirements.

In addition, it is possible to assess the quality of the RS work based on a set of characteristics, which shows the degree of achieving the goal.

To assess the quality of the forming process of the development teams, it is necessary to match the results of work of the HR-manager and PM before applying the RS and after. Process quality is a set of process indicators that reflects the degree of achievement of certain goals [15, 17]. To assess the quality of team building process, the following criteria or features can be used:

- $f_1$  – difficulty level of data identification;
- $f_2$  – difficulty level of processing data;
- $f_3$  – time for data processing;
- $f_4$  – level of objectivity in team building;
- $f_5$  – level of adaptation to changing conditions;
- $f_6$  – degree of consideration of psychological compatibility.

Each feature should be evaluated by 7-point scale, where 1 is the minimal value for metrics and 7 is the maximum accordingly. To obtain the numerical values of the criteria, the experts have evaluated the complexity of solving the given task manually and using RS according to the proposed

scale. In this study, a weighted average quality metric  $PQ$  is proposed in [15] as a convolution criterion, where  $f_u'$  is a normalized value of  $f_u$ -th metric and  $w_u$  is a weight of it:

$$PQ = \frac{\sum_{u=1}^6 f_u' w_u}{\sum_{u=1}^6 w_u} \quad (9)$$

In order to normalize the selected criteria, it is necessary to determine  $f_u^+$  – set of criteria that make a positive contribution to the quality assessment, and  $f_u^-$  – negative criteria. Normalization is carried out according to the formulas (10) and (11) for negative and positive features, respectively:

$$f_u' = 1 - \frac{f_u^-}{6} \quad (10)$$

$$f_u' = \frac{f_u^+}{6} \quad (11)$$

Analysis of the domain area shows that an increasing of the values of features  $f_1 - f_3$  negatively affects the quality of the process of forming the project team, and the criteria  $f_4 - f_6$  are positive. The result of the assessment of the quality of forming the software development teams by (9)-(11) is shown below (Table 4).

**Table 4**  
The quality assessment

| Criterion | Expert | RS | Weight of criterion | Normalized values |      |
|-----------|--------|----|---------------------|-------------------|------|
|           |        |    |                     | Expert            | RS   |
| $f_1$     | 6      | 1  | 0,4                 | 0,14              | 0,86 |
| $f_2$     | 5      | 1  | 0,8                 | 0,29              | 0,86 |
| $f_3$     | 7      | 2  | 0,9                 | 0,01              | 0,71 |
| $f_4$     | 4      | 5  | 0,5                 | 0,57              | 0,71 |
| $f_5$     | 6      | 3  | 0,4                 | 0,86              | 0,43 |
| $f_6$     | 2      | 7  | 0,9                 | 0,29              | 1,00 |
| $PQ$      |        |    |                     | 0,30              | 0,79 |

The results from the Table 4 allows to notice that the quality of the process of forming the project's team based on the use of RS has increased by 49%, which confirms the research hypothesis stated earlier. Thus, the decision-making process of recruiting of software development teams can be improved using recommendations that allow taking into account the psychological profile of potential candidates.

## 5. Conclusion

In the course of this study, the approach for automation of the performance of one of the main functional responsibilities of a PM was presented. It is the formation of software development teams according to recommendation from the RS. Based on a comprehensive analysis of the domain area for the implementation of the RS, the content-based filtering approach and Euclidean distance were chosen. This measure has been used for calculation the proximity of candidates according to the psychological portrait. The algorithm for solving the task of forming of software development teams and a functional model of this business process are presented. The architecture of the RS was proposed and the database model was provided.

The scientific novelty of the given research is to improve the process of recruiting of software development teams using the proposed method, which allows taking into account the psychological compatibility of team members. The similarity measures between candidates are calculated. It allows taking into account results of calculations for forming the development team. Since the Recall and

Precision metrics are inapplicable for assessing the effectiveness of the developed recommendation system, the Average Precision metric was calculated. It allows evaluating the ranking order of the items in the recommendation list. The assessment of the quality of the RS work according to the indicators proposed by the authors is calculated. Numerical studies have shown the possibility of using the proposed method in real conditions in IT-companies to improve the efficiency of management decisions.

## 6. Acknowledgements

This paper and the research behind it would not have been possible without the support of Vitalii Hlushko – Senior Solution Architect, GlobalLogic, Kharkiv, Ukraine. We thank Vitalii Hlushko for assistance with development of proposed method and for comments that greatly improved the paper.

## 7. References

- [1] K. V. Melnyk, N. V. Borysova, Rozrobka informatsiinoi systemy dlia vyrishennia zadachi avtomatyzatsii protsesu pidboru personalu, in: V. Vychuzhanin (Ed.), Aktualni problemy informatsiinykh system i tekhnolohii, NU "OMA", Odesa, 2020, pp. 129-142.
- [2] The Top 12 Best Applicant Tracking Systems (ATS) – February 2021, 2021. URL: <https://www.selectsoftwarereviews.com/buyer-guide/applicant-tracking-systems>.
- [3] A. Stepanov, 10 Sistem dlya Upravleniya Proektami: Moshhno, Nadezhno, Effektivno, 2021. URL: <https://blog.ganttpro.com/ru/sistemy-instrumenty-servisy-upravlenie-proektami/>.
- [4] The Best Project Management Software for 2021, 2021. URL: <https://www.pcmag.com/picks/the-best-project-management-software>.
- [5] R. Hogan, Personality and the Fate of Organizations, 1st. ed., Routledge, London, 2006.
- [6] B. Tuckman, The Forming, Storming, Norming & Performing Team Development Model, 2021. URL: <https://www.coachingcultureatwork.com/bruce-tuckman-team-development-model/>.
- [7] D. Keirse, Personology, Prometheus Nemesis Book Company, Del Mar, 2010.
- [8] L. M. Spencer, S. M. Spencer, Competence at Work: Models for Superior Performance, Wiley India Pvt. Limited, Delhi, 2008.
- [9] Ch. Belhekar, Recommendation Systems – A walk through, 2020. URL: <https://medium.com/@chaitanyarb619/recommendation-systems-a-walk-through-33587fecc195>.
- [10] F. O. Isinkaye, Y. O. Folajimi, B. A. Ojokoh, Recommendation systems: Principles, methods and evaluation, Egyptian Informatics Journal, 16.3 (2015) 261–273. doi:10.1016/j.eij.2015.06.005.
- [11] S. Prabhakaran, Mahalanobis Distance – Understanding the math with examples (python), 2019. URL: <https://www.machinelearningplus.com/statistics/mahalanobis-distance/>.
- [12] B. S. Everitt, S. Landau, M. Leese, D. Stahl, Cluster Analysis, 5th. ed., Wiley, Chichester, 2011.
- [13] K. Melnyk, N. Borysova, Integrated Technology for Personnel Assessment Based on the Competencies Model, in: Proceedings of the 9th International Conference "Information Control Systems & Technologies", ICTS-2020, Odessa, 2020, pp. 343-357. doi:10.13140/RG.2.2.26024.60169.
- [14] N. V. Borysova, K. V. Melnyk, I. V. Olifenko, Tekhnolohiia pidboru komandy it-fakhivtsiv dlia vykonannia proektu, in: Proceedings of the V International Scientific-Practical Conference "Information Technologies in Education, Science and Technology", ITEST-2020, ChSTU, Cherkasy, 2020, pp. 40–41.
- [15] K. V. Melnyk, N. V. Borysova, V. I. Melnyk, Information system development for teaching staff performance evaluation, in: S.V. Kotlik (Ed.), Informatsiini tekhnolohii ta avtomatyzatsiia, Astropynt, Odesa, 2020, pp. 46-54.
- [16] S. Sawtelle, Mean Average Precision (MAP) For Recommender Systems, 2016. URL: <http://sdsawtelle.github.io/blog/output/mean-average-precision-MAP-for-recommender-systems.html>.
- [17] What is process quality?, 2021. URL: <http://www.taskmanagementguide.com/glossary/what-is-process-quality.php>.

# On Keystroke Pattern Variability in Virtual Desktop Infrastructure

Ievgeniia Kuzminykh<sup>a,b</sup>, Bogdan Ghita<sup>c</sup> and Alexandr Silonosov<sup>d</sup>

<sup>a</sup> King's College London, WC2R 2ND, UK; [ievgeniia.kuzminykh@kcl.ac.uk](mailto:ievgeniia.kuzminykh@kcl.ac.uk)

<sup>b</sup> Kharkiv National University of Radio Electronics, Kharkov, 61000, Ukraine

<sup>c</sup> University of Plymouth, PL4 8AA, UK

<sup>d</sup> Blekinge Institute of Technology, Karlskrona, SE-371 79 Sweden

## Abstract

The paper investigates the impact of network traffic and network characteristics on the effectiveness of keystroke dynamics for continuous and one-time authentication in the remote desktop applications and virtual desktop infrastructure. The context presented in the paper contributes to the area of biometric authentication systems by identifying virtual desktop environment as a specific application domain that uses a keystroke pattern for user verification and identifying at what extent the context influences on permanence and the immunity of the keystroke data which was never studied before. The results showed that the keystroke pattern is not affected by network latency, but standard deviation, jitter, and packet loss have a significant combined impact onto it. It can also be concluded that RDP packets are not prioritized during transmission over the network, and therefore, we can say that any other competing traffic is likely to render keystroke dynamics unusable for continuous authentication during remote access.

## Keywords

Biometric pattern, continuous authentication, Euclidean distance, keystroke dynamics, remote display protocol, RDP

## 1. Introduction

Information Security represents a key factor to support the critical information infrastructures across all business sectors. While password-based access control of encompassing systems remains the de-facto option for protecting such systems, previous research demonstrated that such an approach is insufficient; in addition, many examples of successful attacks, including fishing, identity theft, brute force, coupled with weak password policies confirmed the inherent weakness of password-based authentication. In order to overcome this weakness and enhance the authentication process, the industry regulators recommend the adoption of Multi Factor Authentication (MFA). On an MFA-based access system, the user is authenticated using two or more methods that combine knowledge-, token-, or biometric-based inputs. While the first two options have been extensively enhanced through security education and respectively hardware hardening, the biometric alternatives are still subject to significant research due to the complexity and variability of the inputs. The two main classes of biometric inputs are physiological (retina, face, fingerprint, or eye scan) and behavioral (gait, voice, signature, typed keystrokes or mouse movements dynamics). Unlike the other options, biometric authentication requires no additional hardware or memorizing of complex patterns, which makes it the preferred choice for additional security provisioning. From an accuracy perspective, physiological authentication is indeed the better option but, given its inherent intrusiveness for sample acquisition makes it less attractive for users. In contrast, behavioral authentication is subject to variability and therefore typically leads to poorer accuracy, but it is ideal for continuous authentication and, as it draws data from the system interaction, it also has a higher user acceptance.

CMIS-2021: The Fourth International Workshop on Computer Modeling and Intelligent Systems, April 27, 2021, Zaporizhzhia, Ukraine  
EMAIL: [yevheniia.kuzminykh@nure.ua](mailto:yevheniia.kuzminykh@nure.ua) (Ie. Kuzminykh); [bogdan.ghita@plymouth.ac.uk](mailto:bogdan.ghita@plymouth.ac.uk) (B. Ghita); [asilonos@gmail.com](mailto:asilonos@gmail.com) (A. Silonosov)  
ORCID: 0000-0001-6917-4234 (Ie. Kuzminykh); 0000-0002-1788-547X (B. Ghita)



© 2020 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

Given the level and variability of system and applications, the most common interaction option and user data input remains text typing. From an MFA perspective, keystroke dynamics would be the implicit option for continuous authentication, in spite of its inherent variability and attrition, as user behavior may change over time and depends on age, therefore requiring frequent updates of the associated biometric user template.

Biometric authentication has been a highly popular area of research over the past decade; a number of studies demonstrated that keystroke dynamics provides a unique discriminator for individual users through a number of two defining features and their related variables: the duration of a button press and release and the delay between next button press depending on the distance between buttons on a keyboard.

From the intrusiveness perspective, there are two types of user input when performing keystroke dynamics: fixed- and free-text. In the fixed-text authentication, a user needs to type a predefined phrase in order to have their biometric characteristics collected and evaluated, and subsequently be allowed to access system. In free-text authentication, the user may type any text, including command keys, hotkeys, arrows, which makes the method ideal for continuous authentication (CA), to be run in the background, while the user is interacting with the system. Recent studies propose a hybrid approach for improving Free-text approach by using a subset of typed text as input for analysis. Free-text authentication is considered as supplemental security control since it helps to identify the change of legitimate user, particularly when an open session or an unlocked desktop are being intercepted by an intruder. Free-text authentication is reliable since it is hard to forge it. An attacker needs to have access to computer and availability to capture user behavior for a long period to collect data. In this context, one of the factors that further affects the user behavior for a free-text keystroke template is the computer input device, due to its layout and mechanical characteristics. As a result, keystroke biometrics accuracy is affected when collected from different devices and therefore it is preferred in scenarios where the user is regularly performing data input on the same computer, in areas such as healthcare or finance.

Prior studies exhaustively evaluated the strength of keystroke dynamics authentication and proposed a number of improvements focusing on the free-text authentication methods. Given the de-facto description of computer systems, research focused exclusively on system-based, local authentication, while the impact of remote interaction on free-text authentication has not been considered. In this study, we investigate how the variability of network characteristics, delay in particular, influence the calculation of user biometric keystroke dynamics patterns. We will perform an experiment using the most commonly used remote display protocol Microsoft RDP to determine how changes in the network load and resulting congestion influence the keystroke dynamics pattern, in particular, the Euclidean distance value.

Following this introduction, section 2 of the paper provides a brief overview of research on keystroke dynamics authentication system and protocols. Section 3 outlines the methodology for data acquisition and processing. Section 4 provides a comparative analysis and comparison of the results, to benchmark the actual impact of network path changes on the resulting biometric profile. Based on this evaluation, Section 5 draws a discussion and further opportunities for research and Section 6 concludes the paper.

## **2. Literature review**

Review of state-of-the-art in network protocols that implements remote display or so called thin client functionality shows that network latency is one of the key factor characteristic that affects remote display performance. Simoens provided in [1] an in-depth analysis of network latency and its impact of user input events. In [2], the authors show that high network latency affects temporal characteristics of the provided service and thus human perception. Long and Gutwin in their study [3] covers temporal aspect of user input actions in network context and differentiate between local lag versus network lag, which is also exist.

As can be inferred from their required functionality, one of the main performance indicators of remote display infrastructure is their ability to transmit commands and render the screen in a near-real-time fashion; as a result, low latency is critical to their success. A number of studies [4, 5]

benchmarked how network latency impacts on the perceived performance in multiplayer Internet games, more specifically how delays in interaction affect the movement, location, and actions of a player.

Another subject of our research is biometric pattern based on the user keystroke dynamics. In this area slightly more works but any work related to the application of the keystroke for remote access.

In [6] authors acquired keystroke types of the users, built biometric template called enrolment vector, and then tested five matching methods presented by different researchers by comparing enrolment vector with test vector. Experiment included only password typing template, the position of the keys on the keyboard was not taking into account as well as any network or environmental conditions. All results were obtained only by manipulation of the datasets that were collected from users who typed several times password for experiment.

Miguel Lizarraga in [7] collected keystroke dynamics from set of the users, created template for each user and then tested similarity of template and sample. The method of verification of the similarity is developed by authors and based on mean and standard deviation of each feature. The features extracted were three time-features of pressed keys (down-down, down-up, and up-down times). They tested the efficiency of the verification algorithm for different situations: when user is legitimate and when user is impostor.

In this work [8] the timing values from 10 users for typing of the password/username pair were collected and presented. The users should login with the same credentials. The flight and hold time for each user were analysed and showed the difference of the user's biometric behaviour, some users typed text faster, some slower. The further analysis of the obtained data with timing characteristics was not provided.

In [9] the authors investigated the feasibility of using keystrokes for creating cryptography key for files stored remotely in the cloud as alternative to the algorithmic key generation. The key based on the biometry provides both authentication and integrity of data. The work showed the reliability of creating a 256-bit key with biometric pattern. The results showed very low FAR parameter for keystroke-based cryptography key.

Another work devoted to the authentication in the cloud [10] uses keystroke activity as one of the input parameters for creating user template align with other host-based characteristic and network flow-based features. Totally, authors used 36 features to form user profile. But only two features were used to describe keystroke activity of the user: the key press-down time and the time interval between two pressed keys.

Gunetti et.al. [11] and Monroe et.al. [12] analysed non-static biometric technique based on user free-text typing dynamics. It is an attempt to consider using of keystroke dynamics for continuous authentication. The user's typing profiles were obtained using of digraph-specific measures, and in [11] Imposter Pass Rate (IPR) and False Alarm Rates (FAR) have been calculated. This technique provides useful information for user identification and authentication even when a long time has passed since user profiles were formed.

In [13] the authors conducted static manipulation with the dataset of collected keystroke dynamics. The time related features were extracted from the dataset, VKF is calculated based on the typing speed. Using Genetic Algorithm and Support Vector Machines (SVM) algorithms the feature subset selection was done from the reduced. Feature reduction rate and FAR were calculated using the MATLAB.

Bajaj et.al. [14] and Yvonne J. et.al. [15] used timing characteristics of the keystroke as pressing time, dwell time and total time of password entering. Then statistical method is used to calculate the mean time. If value is above the threshold value, then access can be granted, otherwise access is denied. In both papers the users were required to enter only password for collecting biometric pattern.

Literature review shows that only few works analysed free-text keystroke dynamics, and there are no studies researching how characteristics of virtual desktop infrastructures affects free-text based authentication. Moreover, most of the researchers used extraction of timing information from the raw data for forming user profile, but some researchers rely only on two parameters, some use 3-4 timing parameters to describe keystroke activity of the user. Only one work proposes using additional parameter of key location on the keyboard for user profile that we will use in our research. The use of four timing features with adjacency class for keys to authenticate users is novel. Therefore, we conduct an experiment to investigate the impact of network and host characteristics on the user

biometric profile based on keystroke dynamics and to derive the threshold for delay variation beyond which the underlying behavioural biometric algorithms become ineffective.

### 3. Methodology

Null hypothesis is that network latency has no effect on consistency of behaviour pattern that is calculated based on keystroke dynamics of user input. The first part of the methodology is to develop input rendering engine for thin client, then implement algorithms to calculate biometric pattern from real input keystroke dynamics in a VDI session, and then calculate distance for user reference and current biometric pattern. Current biometric pattern will be obtained under different network environmental conditions.

Thin client as part of VDI will capture input events from locally attached keyboard, transfer them to remote desktop over a network protocols such TCP/IP and then replay or render input events on a remote display to simulate, user input. The infrastructure of the remote session is showed in Figure 1.

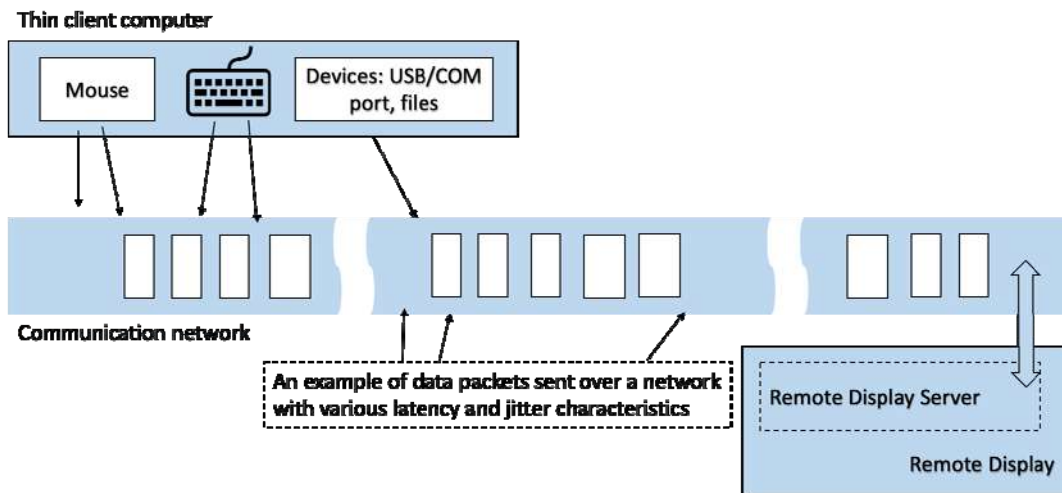


**Figure 1:** Typed keystrokes, mouse clicks and devices data are sent to Remote Display in a raw format in different TCP packets. Screen image and sound are sent back to thin client

We need to experiment with network protocol, network load, running processes on the remote host and biometric pattern which is calculated from keystroke dynamics by an algorithm. The objective is to study how network and host characteristics affect the keystroke dynamics user pattern.

#### 3.1. Network generator

For the purpose of experimenting with network load the review of current state-of-art techniques to generate network workload and measure network performance in VDI scenario has been made. Using literature review four methods of network workload generation were identified and presented in Table 1. The table reflects main characteristics of network simulation and tools that were used for traffic generation. Most of the papers run experiments in testbed that includes network environment close to VDI, where client computer accesses a service or data on a server via heterogeneous network. The results showed that the network latency is the most popular parameter for network simulation experiments, and many researchers [18, 20, 22] used D-ITG software to simulate controlled network traffic. Several research [19,22] used OpenFlow protocol to retrieve network latency and loss rate from network equipment. We will use D-ITG tool as traffic generator in our experiment.

### 3.2. Dataset

The subject is represented by a dataset on natural human-computer interaction [11] that includes keystrokes dynamics of 39 users and three typing patterns of each user resulting into a total of 122 typed keystrokes to be used as input data for creating user profile. This reference profile will be compared with current user profile that is captured under different network and host performance. The dataset contains keystroke-timing information from three different passwords, free-text questions (500 chars) and the transcript of Steve Jobs' Commencement Speech split into two parts. Each user performed the typing test in two separate laboratory sessions, with each session taking about one hour and containing approximately ten thousand keystrokes.

**Table 1**

Research on using the traffic generator tools

| Ref  | Latency   | Workload              | Speed,<br>Mbs | Delay,<br>ms | Packet<br>loss, % | VDI | Year |
|------|-----------|-----------------------|---------------|--------------|-------------------|-----|------|
| [17] | DiffProbe | DiffProbe             | 1-100         | -            | 0.15-1.68         | Y   | 2010 |
| [18] | D-ITG     | D-ITG                 | 1             | 1            | 1                 | Y   | 2012 |
| [19] | OpenFlow  | OpenFlow              | 100           | 10-20        | 1                 | Y   | 2014 |
|      |           | OpenNet-Mon,<br>NetEm |               |              |                   |     |      |
| [20] | Yaz       | D-ITG                 | 7             | 300-1000     | n/a               | N   | 2018 |
| [21] | NorNet    | Netem, D-ITG          | 10-20         | 25-400       | 1-5               | N   | 2018 |
| [22] | OpenFlow  | D-ITG                 | 30            | 5-100        | n/a               | N   | 2018 |
| [23] | Emulab    | D-ITG                 | 4             | 30           | 1                 | N   | 2019 |

### 3.3. Feature Extraction

In order to verify the impact of external impairments on the effectiveness of keystroke authentication, we selected a recognised, successful method and applied it to both the raw initial dataset and the new dataset, based on the received keystroke events. We followed the studies [7,16] and used digraph latency for feature extraction. Features were computed for each key-pair using two main values, specifically the press time (P) and the release time (R) of each key in milliseconds. Four keystroke features were extracted from each key-pair for each user as shown in Figure 2:

1. *Press-Release Time*: Hold time for a key that define the time the key remains pressed.
2. *Press-Press Time*: Time interval between pressing two successive keys.
3. *Release-Release Time*: The interval between releasing two successive keys.
4. *Release-Press Time*: The interval between releasing and pressing two successive keys. This value could be positive or negative according to when keystroke was pressed, before or after previous key was released. That is why we will use absolute value of this feature.



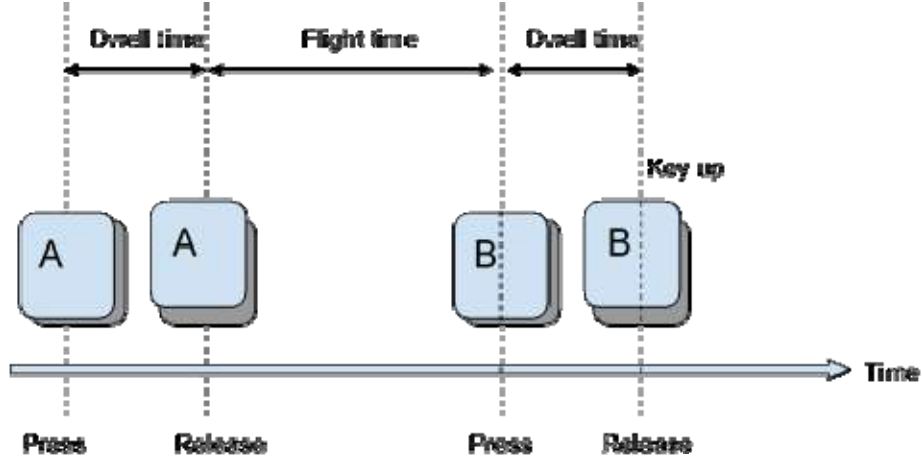


**Figure 2:** Timing features for the key pair

It is apparent only three of the four parameters are independent but, given there was no post-processing applied to the data, the full set of four was maintained for completeness.

Before feature extraction, the dataset was ordered by grouping records from dataset according to the keystroke value  $c_i$  that first pressed and then released. The adjacency class was introduced as an input to set the level of adjacency between two keys on the keyboard and allows to decrease the deviation of the time values during feature extraction. Each row was labelled by the class parameter that corresponds to one of the five adjacency classes and related to key-pair. This ordered dataset was used for the extraction of the timing features, and the reference profile of a user is created for a particular typing session, consisting of mean time characteristics for 121 digraphs (corresponds to 122 characters).

The difference in the assessment of the two profiles, the reference profile  $P_{ref}$  and current  $P_t$  is defined in Equation (1) and could be measured using different classification algorithms as statistical algorithms and machine learning algorithms [11, 16, 24-27]. Few of the common statistical measures are k-nearest neighbour, Euclidean distance, Manhattan distance, Mahalanobis distance, Bayesian distance, statistical t-test, and degree of disorder. Among ML algorithms the common algorithms include Neural Networks, SVM, Regression, Decision Tree, Random Forest, KNN, K-Means, Naive Bayes.

$$distance = |P_{ref} - P_t| / P_{ref} \quad (1)$$

There are two approaches to calculate distance with several features for the authentication purpose. First way is to calculate distance for summative vector with all features that corresponds to the user profile, another way is to calculate distance value for each feature separately. The determination of the threshold is an important issue in the methodology. The analysis of the real collected keystroke data needs to be done to determine thresholds. The works [7, 16] showed that a user's feature with a higher variation demands a lower threshold, while a feature with a lower variation demands a higher threshold. So, the threshold for each feature in each account is obtained based on its standard deviation. Decision about an optimal threshold for each feature based on the a priori knowledge of authentication system administrator and usually the task of determination of the threshold is simple linear function maxima problem. In our work we do not consider decision-making part of the matching phase, the user's identification and authentication is out of scope. We only used an Euclidean distance value as indicator for changes that happened with keystroke pattern under certain conditions with variety of network and host characteristics.

### 3.4. Experiment setup

For the experiment, a network consisting of two routers, two switches and the client and server side was constructed. On the client side:

- Keystroke dynamic was simulated according to one of the profiles
- Remote access to the system was initiated using one of the RD protocols.

Remote server software was installed on the server side, the pattern recognition algorithms program was launched and the received data was compared with the user reference model that was obtained in advance and stored on the server.

The network latency was simulated between the client and the server using the D-ITG traffic generator simulator. We can control and change this variable of latency during an experiment by changing parameter C (intensity of traffic, packets/sec) and c (packet size, bytes) in D-ITG. Since keystroke dynamics is transferred over the network then changing of traffic delays will predetermine the temporal characteristics of the keystroke dynamics and as a result will affect a pattern. The dependent variable is a biometric pattern.

Three types of the experiment were conducted performing the efficiency of keystroke dynamics:

- Under increasing network latency
- During transmission the files from remote desktop to the host
- During watching the video on the remote desktop.

In the first experiment, the transmission channel was loaded with UDP traffic, which was generated using D-ITG, there were two streams of the same intensity, the channel capacity was reduced to 1mbps to obtain high delays in early iterations. To fix the behaviour of the system in the normal state, baseline for the network of characteristics was measured in the absence of load. Further, the network load increased due to an increase in the -C parameter in D-ITG until the packet loss began to exceed 1%.

In the second experiment, the communication channel was not loaded with generated traffic, but the keystroke data prompted for authentication was accompanied by copying the file from remote desktop to the client machine.

In the third experiment, the biometrics authentication process was accompanied by watching a fullHD video on the youtube channel on remote desktop.

## 4. Results

Process of the performing continuous authentication based on the keystroke dynamics was implemented under different conditions. The difference between reference biometric user profile and current obtained during experiment was estimated by its distance value that shows similarity of two patterns, and summary is showed in the Table 1.

**Table 1**  
Results of the experiment

| Experiment             | 1     | 2       | 3     | 4                         | 5   | 6            | 7              |
|------------------------|-------|---------|-------|---------------------------|-----|--------------|----------------|
| C parameter            | 0     | 10..100 | 110   | 110,                      | 115 |              |                |
| D-ITG [pkts/s]         |       |         |       |                           |     |              |                |
| Additional actions     |       |         |       | Mouse moves on RDP window |     | Copying file | Watching video |
| Av. Euclidean distance | 1.068 | 1.07    | 1.074 | 1.77                      | 3.2 | 2.12         | 3              |
| Av. packet loss, %     | 0     | 0       | 0     | 0.14                      | 3.6 | n/a          | n/a            |
| Av.delay, ms           | <1    | <10     | 10.5  | 64                        | 370 | n/a          | n/a            |

As shown by the first three experiments, the impact of low or medium congestion is minimal on the resulting keystroke dynamics parameters, as the Euclidian distance of experiments 2 and 3 is comparable to the baseline obtained in experiment 1. In the case of experiment 3, the measured average bitrate was 901 kb/s, which is close to the bottleneck bandwidth for the testbed, still with minimal impact on the Euclidian distance between the original vector and the one collected at the RDP server. Although the channel utilisation was rather high, there was no packet loss recorded and the average delay (10.5ms) and jitter (3.2ms) were also low. In order to see the behaviour of the

channel under heavy congestion, the data flows were further increased to an average bitrate of 916kb/s, leading to a bottleneck link utilisation of 91.7% and subsequently increasing the average delay (370ms); given the additional traffic was CBR, the jitter varied only slightly, increasing to 3.7ms.

The last three experiments involved a combination of RDP-based interaction, RDP-based traffic, and bandwidth limitations. Experiment 5 was similar to the original conditions, running through a limited 1Mbps bandwidth, but included additional activity on the remote display, more specifically random fast window movements for an open application. This affected slightly the average delay, increasing it to 64 ms, but had a slightly more significant impact on jitter (7ms) and packet loss, which increased to 0.14%. While none of the individual parameters increased significantly, the resulting Euclidian distance varied by a significant factor of 1.77. Experiments 6 and 7 also investigated the impact of additional RDP activities onto the keystroke timings but using a non-restricted 100Mbps link. For experiment 6, a text file was transmitted from the remote machine to the client machine via RDP, while experiment 7 generated additional traffic by streaming a video from the remote machine to the thin client. The video chosen for experiment 7 was an HD 1080p video, with bitrate varying between 4000-8000kbps; while this is indeed the bitrate between the YouTube server and the RDP server, optimised for the bandwidth between the two endpoints, not the same can be said about the link between the RDP server and the client, which uses non-optimised video communication between the them. As a result, it is likely that the down-streaming from the server to the client is higher in terms of bitrate.

## 5. Discussion

The results showed that increasing the channel load with UDP traffic to a certain level does not affect the biometric pattern. This could be explained by the fact that UDP traffic is non-bursty, smooth, and the standard deviation and jitter values are relatively small [28, 29]. But when the channel utilization goes beyond 0.9 (as it is the case for experiments 3-5) the throttling algorithm is starting to work causing a packet loss. This affects the keystroke pattern and the similarity coefficient changes from 1 to 1.7 even with small packet loss values of 0.15-0.5%.

The keystroke dynamics pattern is also affected by the value of standard deviation and jitter, as demonstrated where the transmission of RDP keyboard events is concurrent with events from the VDI. In this case, the average delay was within normal limits but the deviation in delay values was almost two times greater (100 msec). From this, we can conclude that such an effect will be caused by applications running on TCP, since TCP traffic is characterized as bursty, greedy, and is more resistant to delays, but has large jitter values due to its nature [30,31].

When the network load approaches bandwidth the network latency increases to an unacceptable 370msec (ITU-T recommendation for UDP traffic network latency is less than 150ms) and packet loss occurs. This leads to a modification of the timing information required for user current biometric pattern, as a result, the similarity coefficient has a very large distance value which means that the user is most likely not to be identified.

Additional actions such as copying and watching videos on the remote display also significantly affect the keystroke pattern which means that biometrics dynamics are transmitted with the same priority as the rest of the traffic and are served in a queue. From this we can conclude that continuous authentication will be not effective in this situation.

Not prioritizing of RDP messages that carry biometric information can be explained by the fact related to the functioning/processing of keyboard events. Keyboard messages often have a low priority over other operations such as file management [32]. This is evident when typing in a word processor. If there is any disk activity when typing, the text being written is held in a buffer until such time that the processor is free to deal with it (causing the display to pause and then the 7 characters appearing). In order to achieve the high accuracy of timing between keystrokes, the keyboard driver and timing device need a high priority in order not to be affected by other activities.

## 6. Conclusion

This paper presents assessment of the applicability of keystroke dynamics for continuous and one-time authentication in the remote desktop applications and VDI. The combination of four features (Press-Release, PP, RR, RP times) were collected to form a user profile that was object of the experiments. This profile was compared with the current user profiles obtained during experiments under increasing network latency, during transmission the files from remote desktop to the host and during watching the video on the remote desktop. The similarity of two profiles was assessed using Euclidean distance.

We found that the pattern is not affected by network latency that approved our null hypothesis; but by the values of its standard deviation, jitter, and packet loss. This means that working with applications that require TCP transmission on the remote desktop will have a greater effect on the pattern than UDP based ones. It can also be concluded that data transmitted via RDP is not priority traffic which means that as of today, it is impossible to use keystroke dynamic based pattern for continuous authentication for remote access. For one-time authentication at the beginning of the session it is possible to use keystroke pattern when there is no other running application or open windows that can affect the pattern.

But for future, that needs to be reconsidered and modified the RDP protocol in order to take into account the prioritizing of the traffic that is necessary for transmission of biometric information. It is very important in the situation we are faced now, during coronavirus, caused social isolation and needs to work remotely. The biometric identification and authentication based on the keystroke dynamics could be great solution for control the remote access.

## 7. References

- [1] P. Simoens, B. Vankeirsbilck, L. Deboosere, F. A. Ali, F. D. Turck, B. Dhoedt, and P. Demeester, Upstream bandwidth optimization of thin client protocols through latency-aware adaptive user event buffering, *Int. J. Communication Systems* 24 (2011): 666–690.
- [2] Md Liakat Ali, John V Monaco, Charles C Tappert, and Meikang Qiu, Keystroke biometric systems for user authentication, *Journal of Signal Processing Systems*, 86(2-3):175–190, 2017
- [3] M. Long and C. Gutwin, Characterizing and modeling the effects of local latency on game performance and experience, in: *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play*, 2018, pp. 285–297.
- [4] D. Stuckel, C. Gutwin, The effects of local lag on tightly-coupled interaction in distributed groupware, in: *Proceedings of ACM conference on Computer, CSCW '08*, 2008, pp. 447–456.
- [5] K. Almeroth, J. Nelson, On the impact of delay on real-time multiplayer games, in: *Proceedings of 12th International Workshop on Network and Operating Systems Support for Digital Audio and Video (NOSSDAV'02)*, 2002, pp. 23-29.
- [6] R. Giot, M. El-Abed and C. Rosenberger, Keystroke dynamics authentication for collaborative systems, in *Proceedings of 2009 International Symposium on Collaborative Technologies and Systems*, 2009, pp. 172-179.
- [7] Livia C. F. Araújo, Luiz H. R. Sucupira Jr., Miguel G. Lizárraga, Lee L. Ling, Andjoão B. T. Yabu-Uti, "User Authentication Through Typing Biometrics Features," *IEEE Transactions on Signal Processing* 53(2) (2005): 851-855.
- [8] P. V. Pawar, Static User Authentication through Typing Behavior, *International Journal of Latest Trends in Engineering and Technology* 4(1) (2014): 144-148.
- [9] L. Abed, N. Clarke, B. Ghita, A. Alruban, Securing Cloud Storage by Transparent Biometric Cryptography, in: J.L. Lanet, C. Toma (Eds.), *Innovative Security Solutions for Information Technology and Communications*, volume 11359 of LNCS, Springer, Cham, 2018, pp.97-108. doi: 10.1007/978-3-030-12942-2\_9.
- [10] W. Liu, A. S. Uluagac and R. Beyah, MACA: A privacy-preserving multi-factor cloud authentication system utilizing big data, in *Proceedings of IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, 2014, pp. 518-523.

- [11] D. Gunetti, C. Picardi, and G. Ruffo, Keystroke Analysis of Different Languages: A Case Study, in: A.F. Famili, J.N. Kok, J.M. Peña, A. Siebes, A. Feelders (Eds.), *Advances in Intelligent Data Analysis*, volume 3646 of *Lecture Notes in Computer Science*, Springer, Berlin, Heidelberg, 2005, pp. 133-144. doi:10.1007/11552253\_13.
- [12] F. Monrose, A. Rubin, Keystroke dynamics as a biometric for authentication, *Future Generation Computer Systems* 16 (2000): 351-359.
- [13] K. Senathipathi, K. Batri, Keystroke Dynamics Based on Human Authentication System using Genetic Algorithm, *European Journal of Scientific Research* 82(3) (2012): 446-459.
- [14] S. Bajaj, S. Kaur, Typing Speed Analysis of Human for Password Protection (Based on Keystrokes Dynamics), *International Journal of Innovative Technology and Exploring Engineering* 3 (2) (2013): 88-91.
- [15] Y.J. Stark, M.R. Kellas-Dicks. Incorporating False Reject Data into Templates for User Authentication. US Patent US8533486B1, 2013.
- [16] A. Alsultan, K. Warwick, H. Wei, Improving the Performance of Free-text Keystroke Dynamics Authentication by Fusion, *Applied Soft Computing*, 70 (2018): 1024-1033.
- [17] K. Partha, and C. Dovrolis, Diffprobe: detecting ISP service discrimination, in: *Proceedings of IEEE INFOCOM*, San Diego, CA, USA, 2010, pp. 1-9, doi: 10.1109/INFOCOM.2010.5461983.
- [18] A. Botta, A. Dainotti, A. Pescapè, A tool for the generation of realistic network workload for emerging networking scenarios, *Computer Networks (Elsevier)* 56 (15) (2012): 3531-3547.
- [19] N. L. M. van Adrichem, C. Doerr and F. A. Kuipers, Opennetmon: Network monitoring in openflow software-defined networks, in *IEEE Network Operations and Management Symposium (NOMS)*, Krakow, Poland, 2014, pp. 1-8, doi: 10.1109/NOMS.2014.6838228.
- [20] G. Aceto, F. Palumbo, V. Persico and A. Pescapè, Available Bandwidth vs. Achievable Throughput Measurements in 4G Mobile Networks, in: *14th International Conference on Network and Service Management (CNSM)*, Rome, Italy, 2018, pp. 125-133.
- [21] S. Ferlin, S. Kucera, H. Claussen and Ö. Alay, MPTCP Meets FEC: Supporting Latency-Sensitive Applications Over Heterogeneous Networks, *IEEE/ACM Transactions on Networking*, 26 (5) (2018): 2005-2018, doi: 10.1109/TNET.2018.2864192.
- [22] H. Tahaei, R. B. Salleh, M. F. Ab Razak, K. Ko and N. B. Anuar, "Cost effective network flow measurement for software defined networks: A distributed controller scenario." *IEEE Access* 6 (2018): 5182-5198.
- [23] V. Vu Anh, and B. Walker, Redundant Multipath-TCP Scheduling with Desired Packet Latency, in: *Proceedings of the 14th Workshop on Challenged Networks*, Oct 2019, pp. 7-19.
- [24] P. Kang, S. Cho, Keystroke dynamics-based user authentication using long and free text strings from various input devices, *Inf. Sci.*, 308 (2015): 72-93.
- [25] J. Huang, D. Hou, S. Schuckers, T. Law and A. Sherwin, Benchmarking keystroke authentication algorithms, in: *IEEE Workshop on Information Forensics and Security (WIFS)*, Rennes, France, 2017, pp. 1-6, doi: 10.1109/WIFS.2017.8267670.
- [26] A. Carlsson, I. Kuzminykh, R. Franksson, A. Liljegren, Measuring a LoRa Network: Performance, Possibilities and Limitations, in: O.Galinina, S.Andreev, S. Balandin, Y. Koucheryavy (Eds.), *Internet of Things, Smart Spaces, and Next Generation Networks and Systems*, volume 11118 of *LNCS*, Springer, Cham, 2018.
- [27] I. Kuzminykh, D. Shevchuk, S.Shiaeles, B. Ghita, Audio Interval Retrieval Using Convolutional Neural Networks. In: O. Galinina, S. Andreev, S. Balandin, Y. Koucheryavy (Eds.), *Internet of Things, Smart Spaces, and Next Generation Networks and Systems, NEW2AN 2020, ruSMART 2020*, volume 12525 of *LNCS*, Springer, Cham, 2020.
- [28] T. Bakhshi and B. Ghita, Traffic Profiling: Evaluating Stability in Multi-device User Environments, in: *Proceedings of 30th International Conference on Advanced Information Networking and Applications Workshops (WAINA)*, Crans-Montana, Switzerland, 2016, pp. 731-736, doi: 10.1109/WAINA.2016.8.

- [29] T. Lebedenko, M. Ievdokymenko, A.S. Ali, Research of Influence Flow Characteristics to Network Routers Queues Utilization, in: Proceedings of 1st International Conference Advanced Information and Communication Technologies, Lviv, Ukraine, 2016, pp.111-112.
- [30] H.Oudah, B.Ghita, T. Bakhshi, A Novel Features Set for Internet Traffic Classification using Burstiness, in: Proceedings of the 5th International Conference on Information Systems Security and Privacy, Prague, Czech Republic, 2019, pp. 397-404, doi:10.5220/0007384203970404.
- [31] I. Kuzminykh, B. Ghita, A. Silonosov, Impact of Network and Host Characteristics on the Keystroke Pattern in Remote Desktop Sessions, arXiv preprint arXiv:2012.03577, 2020.
- [32] R. Shorrock, D. J. Atkinson, S. S. Dlay, Biometric verification of computer users with probabilistic and cascade forward neural networks, URL: <https://pdfs.semanticscholar.org/9825/90e785721f5dae5e53183b97c028a996544e.pdf>.

# Methodology for Evaluation the Performance Indicators of the Ergatic Information System Functioning

Tetiana Vakaliuk<sup>a</sup>, Ihor Pilkevych<sup>b</sup>, Andrii Tokar<sup>b</sup> and Roman Loboda<sup>b</sup>

<sup>a</sup> Zhytomyr Polytechnic State University, Chudnivska str., 103, Zhitomir, 10005, Ukraine

<sup>b</sup> Zhytomyr military institute named after S.P. Korolov, Mira ave., 22, Zhitomir, 10004, Ukraine

## Abstract

The paper proposes a generalized methodology for evaluation the performance indicators of the ergatic information system, which is based on an algorithmic model of the operator's activity and takes into account its reliability. For this purpose, the process of operator activity is represented in the form probabilistic graph, which is a graphical representation of individual operations and allows to describe the algorithm mathematically, to evaluate its hourly and probabilistic indicators using the rules of transformation graph-scheme. Proceeding from the purpose of the ergatic information system, the tasks that it solves, the indicators are selected the effectiveness of its operation. The order of their calculation is given. The influence exerted by the operator reliability on the efficiency performance of the ergatic information system is taken into account by means a coefficient which characterizes the operator's ability to work. By the example radio-monitoring system calculations indicators efficiency of its functioning are carried out.

## Keywords <sup>1</sup>

Coefficient the reduction of operator reliability, ergatic system, operator reliability, performance indicators, system functioning performance.

## 1. Introduction

The efficiency of functioning within any ergatic information system depends on the reliability of its components: technical means and operator. The modern science allows us to establish successfully the laws of occurrence failure occurrence of devices and methods to forecast them, to find methods to improve reliability during their design and the following manufacturing, as well as methods and techniques to maintain reliability during their storage and operation. At the same time, the individual nature and high variability of human psychological, physiological and professional capabilities and characteristics, its sensitivity to the influence exerted by external and internal environment factors complicates the processes related to analysis, forecasting and improvement of human-operator reliability. As a consequence, due to human errors as a result of her insufficient training, unfavorable psychological factors, fatigue occur the majority share in all the accidents and accidents in different branches of activity [1-3]. Under such conditions, the issue relating to the evaluation indicators of the effectiveness at functioning ergatic information system is topical. Under such system we will understand the interaction of technical means and the operator, whose activity is aimed at the timely identification and processing significant volume of information flows, rapid analysis of the received data, development an optimum decision concerning the further actions. Examples such activities are the actions of an air traffic controller, operator unmanned aircraft complex, operator of a radio monitoring post and the like.

---

CMIS-2021: The Fourth International Workshop on Computer Modeling and Intelligent Systems, April 27, 2021, Zaporizhzhia, Ukraine  
EMAIL: tetianavakaliuk@gmail.com (T. Vakaliuk); igor.pilkevich@meta.ua (I. Pilkevych); tapir@i.ua (A. Tokar);  
romaloboda0704@gmail.com (R. Loboda)  
ORCID: 0000-0001-6825-4697 (T. Vakaliuk); 0000-0001-5064-3272 (I. Pilkevych); 0000-0001-7534-2820 (A. Tokar); 0000-0003-4010-0252 (R. Loboda)



© 2021 Copyright for this paper by its authors.  
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).  
CEUR Workshop Proceedings (CEUR-WS.org)

To prevent a failure in the execution of tasks, it is necessary to evaluate the performance indicators of the information system in the process of its operation and compare their values with the limiting value. Based on this, there is an urgent need to calculate these indicators in real time, taking into account the peculiarities of the operator's activities in specific conditions, in order to prevent the deterioration of the quality of tasks performed by the system as a whole. To do this, we will formalize and simulate the process of the system functioning, as a result of which we will obtain the dependence of the performance indicators of the information system on the operator's reliability. When constructing a model, we will determine its qualitative composition and numerical parameters.

## **2. Review of the literature**

Sufficiently well-developed approaches to the evaluation efficiency of systems at the stage of their design [4, 5]. Estimation a degree of growth probability human error depending on changes of environmental conditions using prediction model is described in [3], but its influence on the efficiency the system is not shown. Impact that system reliability and task uncertainty have on unmanned aerial vehicle operator performance during patrolling and target recognition tasks is investigated in [6], but the inverse effect is not explored. The issue concerning quantitative estimation probability of human error on the basis the qualitative analysis human factor and task context with additional consideration of factors forming performance is devoted to the research work [7]. In research [8] measurement from the operator's reliability level is carried out using Human Error Assessment and Reduction Technique (HEART) on the example of 3 male operators using computer numerical control milling with work experience > 3 months, operator age between 18 and 39 years. Method HEART is proposed to be applied to analyze human realibility in production by is done using Hierarchical Task Analysis in research [9]. In [10] the HEART method is used in order to quantify the human failure event for probabilistic safety assessment of a TRIGA Mark II reactor, situated at the Malaysia Nuclear Agency, Malaysia. This approach requires the selection values nominal probabilities of human error and does not take into account the time factor. A mutual dependence between machine degradation and human error in man-machine systems are studied in research [11], but the influence of other factors on the efficiency of the system functioning is not taken into account. In [12], it is proposed to evaluate the efficiency of the system using some quantitative parameters such as servicing time and costs. This approach does not take into account the degree of decrease in the reliability of the system user. The issue of professional suitability at the stage of selection of candidates for training pilots of the United States Air Force MQ-1 Predator is investigated in [13]. The study of changes in the efficiency of the MQ-1 Predator depending on the reliability of the operator in this work is not performed. A number of works are devoted to assessing and improving the reliability and efficiency of systems functioning, but at the same time, no attention is paid to the influence of operator reliability on the performance indicators of the system as a whole [14-16].

Analysis shows that quite successfully solved the issue concerning the evaluation of the effectiveness systems at the stage their design. At the same time, during the assessment and improvement of the reliability and efficiency of the systems functioning, attention is not paid to the influence of the operator's reliability on the performance of the system as a whole. In the direction of evaluation efficiency systems in their functioning studies conducted limited to the assessment a person's reliability without calculating the performance measures of the system as a whole, in addition is not taken into account the factor changes in the probability or human errors over time.

So, the purpose the work is to develop a methodology for evaluation the performance indicators of the ergatic information system, taking into account the reliability of the operator.

## **3. Research methods**

The functioning of the ergatic information system will be considered as a process of timely identification, receipt and processing of information flows by the operator, rapid analysis of the data obtained, the development of optimal solutions for further action. In this case, we will assume that the technical means of the information system work faultlessly at a given time interval. Under such conditions, the effectiveness of the functioning of the ergatic information system will be determined



by the timeliness and validity of decision-making by the operator. The probability of a certain event is used as performance indicators in conditions of uncertainty [3, 6–8]. Therefore, as indicators of efficiency of functioning of ergatic information system we use probability of timely decision making –  $P_t$  and probability of error-free decision making –  $P_d$  by operator of information system.

### 3.1. Method for evaluation the probability of timely decision-making

The probability of timely decision-making by the operator of information system –  $P_t$  characterizes the probability of event  $T_{min} \leq P_o \leq T_a$ , that is

$$P_t = P(T_{min} \leq P_o \leq T_a), \quad (1)$$

where  $T_{min}$  – is the minimum possible time of the operator's decision;  $T_a$  – is the maximum allowable time of the operator's decision;  $T_o$  – time is the operator's decision.

According to probability theory [17]

$$P(T_{min} \leq P_o \leq T_a) = F(T_a) - F(T_{min}), \quad (2)$$

where  $F(T_a)$  – is the time distribution function  $T_a$ ;  $F(T_{min})$  – is the time distribution function  $T_{min}$ .

Due to the influence of a large number and random variables on the information system operator (increase or redistribution efforts from information processing bodies when the situation changes; constant influence on the operator the results from the information flows processing; failures in the regulation process and reporting the data received, decrease in personnel training, deterioration of communication channels quality, quality level of information base used; lack as well as inadequate resources to implement processing in case of complicated situation on the information system operator's side) there is every reason to believe that the time for making a decision by the operator is distributed according to the normal law [17].

In such a case the density of the distribution function of the operator's decision-making time will be equal to

$$f(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t-m)^2}{2\sigma^2}}. \quad (3)$$

From expression (3) we obtain the time distribution functions  $T_{min}$  and  $T_a$

$$F(T_{min}) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{T_{min}} e^{-\frac{(T_{min}-m)^2}{2\sigma^2}} dT_{min}, \quad (4)$$

$$F(T_a) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{T_a} e^{-\frac{(T_a-m)^2}{2\sigma^2}} dT_a, \quad (5)$$

where  $\sigma$  – is the standard deviation of the operator's decision time;  $m$  – is the mathematical expectation of the operator's decision time.

Expressing the distribution functions (4) and (5) through the standard distribution function (6)

$$\Phi(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{(x)^2}{2}} dx, \quad (6)$$

using expressions (1) and (2) we obtain

$$P_t = P(T_{min} \leq P_o \leq T_a) = \Phi\left(\frac{T_a - m}{\sigma}\right) - \Phi\left(\frac{T_{min} - m}{\sigma}\right). \quad (7)$$

Over time, as a result for various reasons, when an operator makes a decision, and hence the mathematical expectation time of making a decision will increase and is determined by the dependence  $m = m + \Delta t$ , where  $\Delta t$  – is the value of the increase in the average time of the operator's decision due to a decrease in its reliability

$$\Delta t = (T_{max} - m)K_R, \quad (8)$$

where  $K_R$  – is a coefficient characterizing the reduction of operator reliability. In the optimal state operator –  $K_R = 0$ , in a critical decrease in the reliability operators  $K_R = 1$ . Procedure for calculation of  $K_R$  is given in [18];  $T_{max}$  – time for decision making at a low level of reliability the operator of the information system, i.e. when  $K_R = 1$ .

Taking into account (7), (8), we obtain an expression for the evaluation of the probability of timely decision making by the information system operator at any point of time

$$\begin{aligned} P_t &= \Phi\left(\frac{T_a - (m + \Delta t)}{\sigma}\right) - \Phi\left(\frac{T_{min} - (m + \Delta t)}{\sigma}\right) = \\ &= \Phi\left(\frac{T_a - T_{max}K_R - m(1 - K_R)}{\sigma}\right) - \Phi\left(\frac{T_{min} - T_{max}K_R - m(1 - K_R)}{\sigma}\right). \end{aligned} \quad (9)$$

To calculate the probability of timely decision-making by the operator at an arbitrary moment using expression (9), it is necessary to know the minimum possible decision-making time –  $T_{min}$ , the maximum possible decision-making time by the operator –  $T_a$ , average decision-making time by the operator in optimal conditions –  $m$ , decision-making time at a low level operator reliability –  $T_{max}$ .

For definition of mentioned parameters, we will use method on the base an algorithm probability graph [19, 20], which is based on analysis of algorithm of operator's activity. For the effective usage algorithmic models, it is necessary to obtain reliable data on the elements of operator activity, i.e. to analyze the functioning process in a particular information system. As a result, to establish the time, probabilistic characteristics and accuracy indicators of the information system operator's activity, based on the data given in the reference literature and their clarification, taking into account various factors (operator's state, ergonomic characteristics of the system etc.).

For example, let us calculate the probability of a timely decision by the operator of the radio monitoring system. For this, an analysis of the functioning of such a system based on a radio receiver was carried out. As a result, based on the list of operations and logical conditions that were obtained by the expert method, a structural diagram of the activity of the operator of the radio monitoring information system was drawn up (Figure 1). A list of operations of the operator's activity algorithm has been compiled and the time indicators of each operation are given –  $T_{mini}$ ,  $T_{ai}$ ,  $m_i$ ,  $T_{max_i}$  (Table 1).

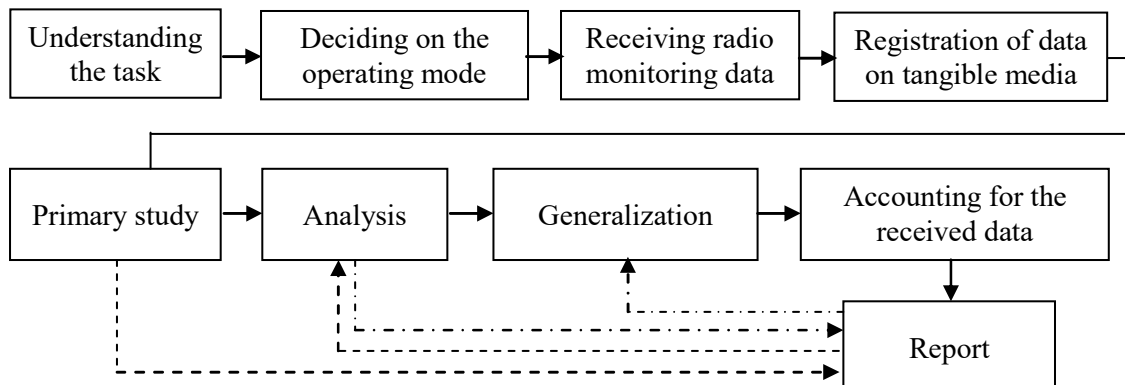


**Figure 1:** Block diagram of the activity of the operator of the radio monitoring information system

**Table 1**

Operator activity operations and their timing

| No<br>operation | Operation name                                                 | $T_{min_i}, c$ | $T_{a_i}, c$ | $m_i, c$ | $T_{max_i}, c$ |
|-----------------|----------------------------------------------------------------|----------------|--------------|----------|----------------|
| 1               | Understanding the task                                         | 25             | 50           | 30       | 60             |
| 2               | Deciding on the operating mode                                 | 1,2            | 3            | 1,8      | 5              |
| 3               | Switching the receiver to radio monitoring mode                | 5              | 10           | 7        | 15             |
| 4               | Switching the receiver into search mode                        | 5              | 10           | 7        | 15             |
| 5               | Signal processing decision making                              | 1,5            | 3            | 2        | 4              |
| 6               | Selection of the signal of interest                            | 3              | 6            | 4        | 8              |
| 7               | Pre-latching (signal acquisition)                              | 5              | 10           | 7        | 12             |
| 8               | Preliminary analysis of the information message                | 10             | 25           | 15       | 32             |
| 9               | Determination of signal novelty                                | 5              | 25           | 10       | 32             |
| 10              | Sending a direction finding command                            | 2,3            | 4,3          | 2,8      | 6              |
| 11              | Getting results direction finding                              | 2              | 5            | 3        | 7              |
| 12              | Assignment of bearing to a specific source                     | 5              | 10           | 7        | 16             |
| 13              | Establishing the main features of the source                   | 60             | 120          | 80       | 130            |
| 14              | Source identification                                          | 1,5            | 3            | 2        | 5              |
| 15              | Registration of data on tangible media                         | 50             | 110          | 70       | 120            |
| 16              | Familiarization with data content                              | 60             | 160          | 85       | 174            |
| 17              | Highlighting known and unknown elements                        | 10             | 30           | 17       | 50             |
| 18              | Determining the importance of a new element                    | 3              | 7            | 4        | 11             |
| 19              | Determination of urgency                                       | 3              | 7            | 4        | 11             |
| 20              | Report                                                         | 10             | 20           | 11       | 30             |
| 21              | Validity informations assessment                               | 30             | 70           | 40       | 90             |
| 22              | Determining the informations characteristics                   | 40             | 80           | 50       | 110            |
| 23              | Comprehensive data exploration, comparison with available data | 120            | 240          | 150      | 280            |
| 24              | Identifying changes in condition and character                 | 5              | 15           | 8        | 17             |
| 25              | Formation of conclusions                                       | 130            | 270          | 170      | 340            |
| 26              | Formation of information documents                             | 180            | 400          | 220      | 420            |
| 27              | Thematic systematization                                       | 25             | 45           | 30       | 70             |
| 28              | Accounting for the received data                               | 120            | 250          | 170      | 310            |
| 29              | Results report                                                 | 10             | 20           | 11       | 30             |

Time indices  $T_{min}$ ,  $T_a$ ,  $m$ ,  $T_{max}$  we will receive on the basis of time indices individual operations  $T_{min_i}$ ,  $T_{a_i}$ ,  $m_i$ ,  $T_{max_i}$  (Table 1) by representing the algorithm and using the rules of its transformation [19, 20].

Probabilistic oriented graph, corresponding to the algorithm of radio monitoring system functioning is shown in Figure 2. It is a graphical representation of individual operations and allows to describe the algorithm mathematically, to estimate the time and probabilistic indicators.

Each vertex of the graph corresponds to an operation according to the numbers from Table 1 and a 4-dimensional vector  $V_i$ , whose components are  $T_{min_i}$ ,  $T_{a_i}$ ,  $m_i$ ,  $T_{max_i}$ . The set graph vertices and its branches form the graph-scheme of the algorithm. To obtain time indicators in the algorithm we use the rules of graph-scheme transformation [20].

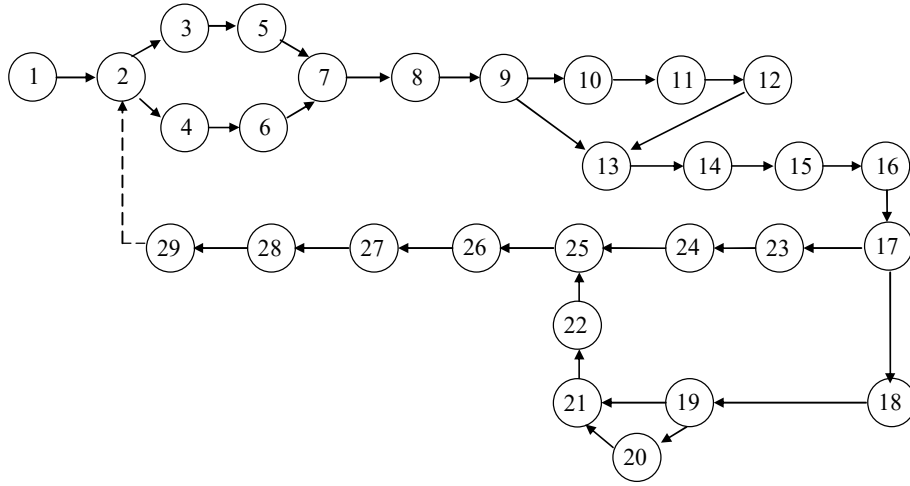


**Figure 2:** Probabilistic directed graph

1. Combining paths without branches.

The path graph without branches has the form (Figure 3):

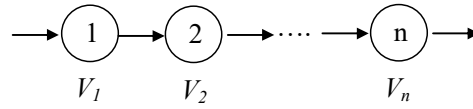


**Figure 3:** Graph without branches

The concatenation operation turns a branch without branches into an equivalent branch with one vertex (Figure 4).

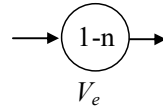


**Figure 4:** Equivalent branch with one vertex

The components of the vector  $V_e$  are found by the expressions:

$$\begin{aligned}
 T_{min_e} &= \sum_{i=1}^n T_{min_i} \\
 T_{a_e} &= \sum_{i=1}^n T_{a_i} \\
 m_e &= \sum_{i=1}^n m_i \\
 T_{max_e} &= \sum_{i=1}^n T_{max_i}
 \end{aligned} \tag{10}$$

2. Combining paths with a branching.

The graph of a path with a branching has the form (Figure 5):

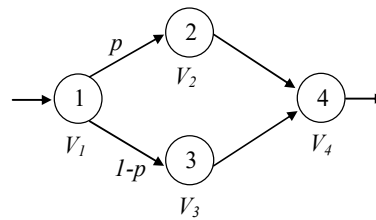


**Figure 5:** Graph of a path with a branching

The concurrent paths joining operation turns a branch with a branching into an equivalent branch with one vertex with vector  $V_e$  whose components are found by the expressions:

$$\begin{aligned} T_{min_e} &= T_{min_1} + pT_{min_2} + (1-p)T_{min_3} + T_{min_4} ; \\ T_{a_e} &= T_{a_1} + pT_{a_2} + (1-p)T_{a_3} + T_{a_4} ; \\ m_e &= m_1 + pm_2 + (1-p)m_3 + m_4 ; \\ T_{max_e} &= T_{max_1} + pT_{max_2} + (1-p)T_{max_3} + T_{max_4} , \end{aligned} \quad (11)$$

where  $p$  – is the probability of transition along a branch, taken as 0,5 for all branches.

According to the given rules let us join paths without branches into graph (see Figure. 2) (1, 2), (3, 5), (4, 6), (7, 8, 9), (10, 11, 12), (13, 14, 15, 16, 17), (18, 19), (21, 22), (23, 24), (25, 26, 27, 28, 29). As the result we obtain a probabilistic oriented graph presented in Figure 6.

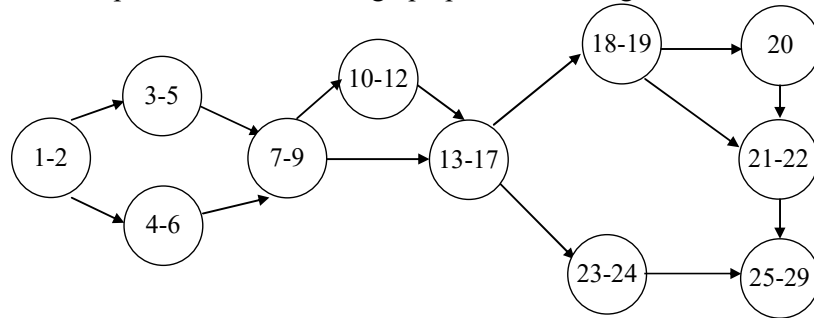


**Figure 6:** Result of consecutive branch merging

Using expressions (10), we obtain equivalent time characteristics of graph vertices (Table 2).

**Table 2**

Time characteristics after combining

| № operation | $T_{min_i}, c$ | $T_{a_i}, c$ | $m_i, c$ | $T_{max_i}, c$ |
|-------------|----------------|--------------|----------|----------------|
| 1-2         | 26,2           | 53           | 31,8     | 65             |
| 3-5         | 6,5            | 13           | 9        | 19             |
| 4-6         | 8              | 16           | 11       | 23             |
| 7-9         | 20             | 60           | 32       | 76             |
| 10-12       | 9,3            | 19,3         | 12,8     | 29             |
| 13-17       | 131,5          | 313          | 184      | 359            |
| 18-19       | 6              | 14           | 8        | 22             |
| 20          | 10             | 20           | 11       | 30             |
| 21-22       | 70             | 150          | 90       | 200            |
| 23-24       | 125            | 255          | 158      | 297            |
| 25-29       | 465            | 925          | 601      | 1170           |

To combine paths with branching, accept the same probability to go through each branch of the logical operator is  $p = 0,5$ . As a result combining paths with branching (1-2, 3-5, 4-6, 7-9) and (18-19, 20, 21-22) the graph-scheme will look like (Figure 7).

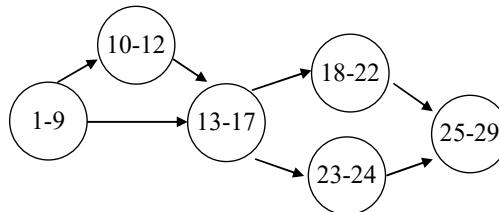


**Figure 7:** Result of consecutive branch merging

Using expressions (11), we obtain equivalent time characteristics of graph vertices (Table 3).

**Table 3**

Time characteristics after combining

| No operation | $T_{min_i}, c$ | $T_{a_i}, c$ | $m_i, c$ | $T_{max_i}, c$ |
|--------------|----------------|--------------|----------|----------------|
| 1-9          | 53,3           | 127,2        | 73,6     | 161,6          |
| 10-12        | 9,3            | 19,3         | 12,8     | 29             |
| 13-17        | 131,5          | 313          | 184      | 359            |
| 18-22        | 81             | 174          | 103,5    | 237            |
| 23-24        | 125            | 255          | 158      | 297            |
| 25-29        | 465            | 925          | 601      | 1170           |

Combine the paths with branching (1-9, 10-12, 13-17) (Figure 8) and obtain the equivalent time characteristics of the vertices of the graph (Table 4).

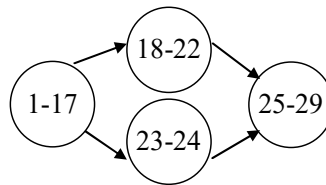


**Figure 8:** Result of consecutive branch merging

**Table 4**

Time characteristics after combining

| No operation | $T_{min_i}, c$ | $T_{a_i}, c$ | $m_i, c$ | $T_{max_i}, c$ |
|--------------|----------------|--------------|----------|----------------|
| 1-17         | 187,59         | 446          | 261,4    | 529,3          |
| 18-22        | 81             | 174          | 103,5    | 237            |
| 23-24        | 125            | 255          | 158      | 297            |
| 25-29        | 465            | 925          | 601      | 1170           |

After combining paths with branching (1-17, 18-22, 23-24, 25-29) (Figure 9), we obtain equivalent time characteristics of the equivalent graph vertices (Table 5).

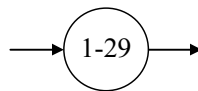


**Figure 9:** Result of consecutive branch merging

**Table 5**

Time characteristics after combining

| No operation | $T_{min_i}, c$ | $T_{a_i}, c$ | $m_i, c$ | $T_{max_i}, c$ |
|--------------|----------------|--------------|----------|----------------|
| 1-29         | 818            | 1626         | 1072     | 1996           |

Obtained time indices algorithm activity of the radio monitoring system operator gives the opportunity to find the probabilities of timely decision making by the operator at any point in time by expression (9). Taking into account that in solving the problem the operator has normal law distribution, the average deviation in time of taking the decision by the operator will be found by expression

$$\sigma = \frac{(m - T_{min})}{3} \quad (12)$$

The obtained value of the probability of a timely decision-making by the operator must be compared with the admissible value. The permissible value can be obtained in further research in relation to a specific task performed by the system.

So, the method for evaluation the probability of timely decision-making by the information system operator will consist of the following steps:

1. Make a probabilistic oriented graph of the functioning of the ergatic information system.
2. Determine the timing of individual operations –  $T_{min_i}$ ,  $T_{a_i}$ ,  $m_i$ ,  $T_{max_i}$ .
3. Perform transformations of the graph using the rules of combining paths without branches and with branches and calculate the time indices –  $T_{min}$ ,  $T_a$ ,  $m$ ,  $T_{max}$ .
4. Calculate the average deviation of the operator's decision time by expression (12).
5. Calculate the probability of timely decision-making by the operator of the information system by expression (9).

### 3.2. Method for evaluation the probability of error-free decision-making

To calculate the probability of error-free decision making by an information system operator we will use a method based on information analysis of decision making processes [21, 22]. The basis of the informational approach is the fact that the most essential factor, which influences the error-free decision making, is the volume of initial information. Accordingly, for any complex system, an increase in the amount of initial information leads to an increase in the probability of error-free decision making by an operator. Taking into account loss of initial information due to decrease of operator's reliability, let us write the expression for probability of error-free decision –  $P_{d_i}$  of the  $i$ -th operation:

$$P_{d_i} = 1 - Y_0 e^{-\frac{(I_{in} - I_l)}{K_i}}, \quad (13)$$

where  $I_{in}$  – amount of initial information;  $I_l$  – amount of information lost;  $Y_0$  – initial uncertainty of the decision taken characterizes the psychological confidence of the operator (at the optimal state of the operator  $Y_0 = 1$ );  $K_i$  – coefficient that characterizes the value an information and affects the growth rate of probability depending on the amount of processed information by the operator. At maximum value of the information  $K_i$  tends to 0, at minimum –  $K_i = 1$ . The value of this coefficient will be found on the assumption that as the number of possible states of the system, and consequently the number of variants of decisions  $n$ , grows, the value of the information by the expression

$$K_i = \frac{I}{n}. \quad (14)$$

The amount of initial information about some system is equal to the entropy of this system  $I_{in} = N$ , and at equal-probability  $n$  states of the system [17]

$$I_{in} = \log n. \quad (15)$$

Over time, due to the impact of various reasons, part the original information will be lost, the value of the loss will depend on the level efficiency of the operator and determined by the introduced coefficient decrease in the reliability of the operator –  $K_R$ . The amount is lost information we will find by the expression.

$$I_l = I_{in} \cdot K_R. \quad (16)$$

The level of reliability will also affect the level in psychological confidence, that is, the coefficient  $Y_0$  will decrease in accordance with the value of  $K_R$ .

With the decrease of the level a reliability operators value of information will decrease. To take this into account let's write expression (14) in the form:

$$K_i = \frac{K_R}{n}. \quad (17)$$

Given expressions (13), (15), (16), (17), and assuming that the operator solution is completely indeterminate is  $Y_0=1$ , we obtain an expression for determining the probability of error-free solution at the  $i$ -th operation at an arbitrary time moment:

$$P_{d_i} = 1 - K_R e^{-\frac{n_i(1-K_R)\log n_i}{K_R}}. \quad (18)$$

To carry out calculations on the example the radio monitoring system, by expert polling was established the number from the  $i$ -th operation, determining the number of states –  $n$  (Table 6, Table 7). In the table the numbers of operations correspond to the numbers from Table 1.

**Table 6**

Time characteristics after combining

| No operation                 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8  | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|------------------------------|---|---|---|---|---|---|---|----|---|----|----|----|----|----|----|
| Number of possible solutions | 3 | 2 | 2 | 2 | 2 | 4 | 4 | 13 | 4 | 4  | 4  | 4  | 14 | 7  | 9  |

**Table 7**

Time characteristics after combining

| No operation                 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 |
|------------------------------|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| Number of possible solutions | 9  | 16 | 16 | 16 | 2  | 16 | 10 | 9  | 16 | 2  | 2  | 4  | 5  | 2  |

To obtain the probability for error-free decision by the operator of the radio-monitoring system, was used probabilistic directed graph, represented on Figure 2. In the process of transforming the initial graph-scheme to the equivalent one, with one vertex, the union of the same paths was carried out, as in determining the time characteristics, taking into account the rules for transforming the graph [20]. It was taken into account that when merging paths without branching (Figure 10)

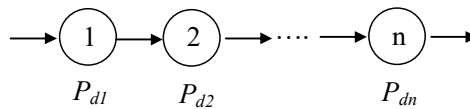


**Figure 10:** Paths without branching

probability characteristics are determined by the expression [20]

$$P_d = \prod_{i=1}^n P_{d_i}. \quad (19)$$

When combining tracks with branching (Figure 11)

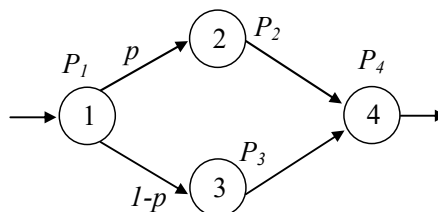


**Figure 11:** Paths without branching

probability characteristics are determined by the expression



$$P_d = P_t P_4 (p P_2 + (1 - p) P_3). \quad (20)$$

So, the method for evaluation the probability of error-free decision-making by the information system operator will consist of the following steps:

1. Make a probabilistic oriented graph for functioning of an ergatic information system.
2. Determine the number of possible solutions of the  $i$ -th operation determining the number of states –  $n$ .
3. Calculate the probability for error-free decision by the operator when performing the  $i$ -th operation at an arbitrary moment of time by expression (18).
4. To carry out transformations to the graph using the rules of combining paths without branching and with branching and to carry out calculation –  $P_d$  by expressions (19), (20).

In order to verify the proposed methods for evaluation the performance of the ergatic information system, the probabilities involved in the timely decision-making and the probability of error-free decision-making by the operator information system for the range of changes in the coefficient  $K_R = [0 - 1]$  were calculated. The results are shown in Table 8, for which the corresponding dependencies were built (Figure 12).

**Table 8**  
Calculation results

| $K_R$ | $P_t$ | $P_d$ |
|-------|-------|-------|
| 0     | 1,00  | 1,00  |
| 0,1   | 0,99  | 0,99  |
| 0,2   | 0,99  | 0,99  |
| 0,3   | 0,99  | 0,99  |
| 0,4   | 0,94  | 0,97  |
| 0,5   | 0,81  | 0,85  |
| 0,6   | 0,58  | 0,50  |
| 0,7   | 0,28  | 0,13  |
| 0,8   | 0,05  | 0,02  |
| 0,9   | 0,01  | 0,01  |
| 1     | 0,00  | 0,00  |

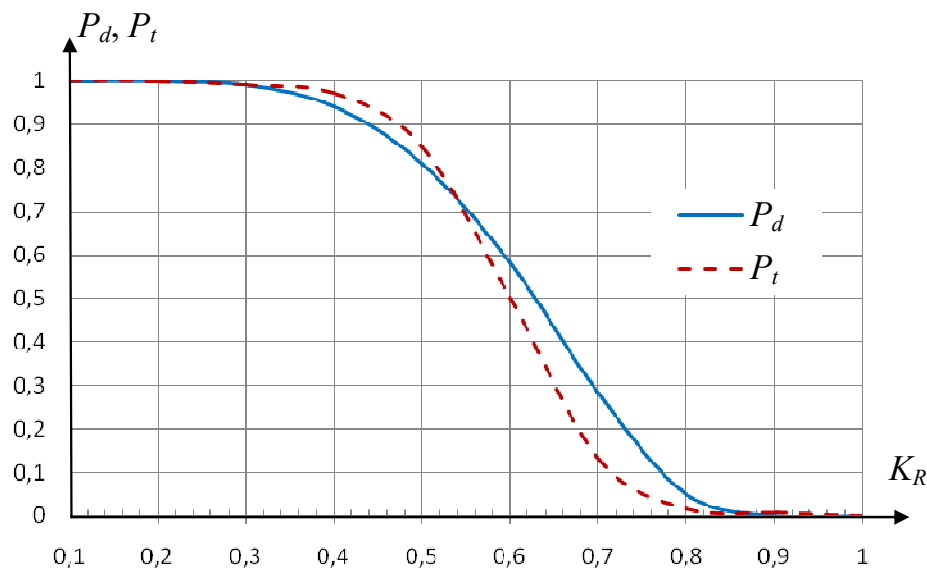


**Figure 12:** Dependence of indicators on  $K_R$

Analysis of the data obtained shows that at the initial stage of reducing the operator's reliability (up to  $K_R = 0.4$ ), the performance indicators of the system functioning practically do not change. This can be explained by the inclusion of human compensatory mechanisms. After a certain value, a sharp decrease in the performance indicators of the system functioning to critical values is observed.

## 4. Conclusions

The developed methods constitute the content of the generalized methodology for evaluation the performance indicators for ergatic information system, based on an algorithmic model of the operator's activity. The methods take into account the reduction factor of operator reliability, made it possible to obtain an estimate of probabilities for timely decision-making and probability of error-free decision-making by the information system operator in real time while on duty.

## 5. References

- [1] T. Kletz, P. Amyotte, Accidents said to be due to human error, What Went Wrong?, 6th. ed. Butterworth-Heinemann. (2019) 53–172. doi:10.1016/B978-0-12-810539-9.00007-0.
- [2] B. Strauch, Investigating Human Error: Incidents, Accidents, and Complex Systems, 1st. ed., Routledge (2018) doi:10.4324/9781315251851.
- [3] Y. Tan, D. Feng, H. Shen, Research for Unmanned Aerial Vehicle components reliability evaluation model considering the influences of human factors, 3rd International Conference on Mechanical, Electronic and Information Technology Engineering., ICMITE 2017, vol. 139, article number 00221. doi:10.1051/mateconf/201713900221.
- [4] Z. Mian, D. Mavris, Mission effectiveness quantification and assessment of Micro Autonomous Systems and Technologies, IEEE Systems and Information Engineering Design Symposium (2013) doi:10.1109/SIEDS.2013.6549514.
- [5] Haifeng Zhu, Fundamental Models for Missions of Engineered Systems, Systems Conference (SysCon) IEEE International. (2019) 1–6. doi:10.1109/syscon.2019.8836714.
- [6] D. Liu, M. Jaramillo, D. Vincenzi, The effects of system reliability and task uncertainty on autonomous unmanned aerial vehicle operator performance under high time pressure. Hum. Factors Ergon. Vol. 25, № 5. (2015) 515–522. doi:10.1002/hfm.20565.
- [7] X. Pan, Z. Wu, Performance shaping factors in the human error probability modification of human reliability analysis, International Journal of Occupational Safety and Ergonomics, Vol. 26. (2020) 538–550, doi:10.1080/10803548.2018.1498655.
- [8] Y. Rejeki, E. Achiraeniwati, A. Wanda, Measurement of operator reliability level using the Human Error Assessment and Reduction Technique (HEART) method. IOP Conference Series: Materials Science and Engineering, Vol. 830, 2020, pages 032095, doi:10.1088/1757-899X/830/3/032095.
- [9] D. Safitri, R. Astriaty, N. Rizani, Human Reliability Assessment dengan Metode Human Error Assessment and Reduction Technique pada Operator Stasiun Shroud PT.X, Jurnal Rekayasa Sistem Industri Vol. 4, No. 1, (2015) 1–7. doi:10.26593/jrsi.v4i1.1388.1-7.
- [10] H. Ahmad, M. Faizal, M. Mazleha, Research reactor operator performance based on the human error assessment and reduction technique (HEART), in IOP Conference Series: Materials Science and Engineering, Vol. 785, International Nuclear Science Technology and Engineering Conference 2019 (iNuSTEC2019) 29–31 October 2019, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia. doi:10.1088/1757-899x/785/1/012048.
- [11] H. Che, S. Zeng, J. Guo, Reliability assessment of man-machine systems subject to mutually dependent machine degradation and human errors. Reliab Eng Syst, Vol. 190. (2019) 106504. doi:10.1016/j.ress.2019.106504.
- [12] Y. Fan, Study of a new Analysis Method of Risk Priority Number Based on FMEA, SAE Technical Paper 2015-01-0136. (2015) doi:10.4271/2015-01-0136.
- [13] W. Chappelle, T. Goodman, J. Swearingen, W. Thompson. A Preliminary Investigation into Cognitive Aptitudes Predictive of Overall MQ-1 Predator Pilot Qualification Training

- Performance. Wright-Patterson AFB (OH): U.S. Air Force School of Aerospace Medicine, 2015. Technical Report AFRL-SA-WP-SR-2015-0025. 14 p.
- [14] A. Noroozian; A. Sayyah, A new hybrid reliability allocation method, based on developed ME-OWA method, Birnbaum importance measure and fuzzy DEMATEL technique, *International Journal of Critical Infrastructures (IJCIS)*, Vol. 16, No. 2. (2020) 162–185. doi:10.1504/IJCIS.2020.107269.
  - [15] D. Kwon, A. Lucero, Optimized reliability test design to reduce uncertainties in reliability assessment in: 62nd. Electronic Components and Technology Conference, San Diego, CA, USA. (2012) 1726–1730, doi: 10.1109/ECTC.2012.6249070.
  - [16] J. Lee, M. Mitici, An integrated assessment of safety and efficiency of aircraft maintenance strategies using agent-based modelling and stochastic Petri nets, *Reliability Engineering & System Safety*, Vol. 202, 107052 (2020) doi:10.1016/j.ress.2020.107052.
  - [17] Werner L. Probability Theory. De Gruyter Textbook (2016) 395 p. doi:10.1515/9783110466195.
  - [18] I. Korobiichuk, A. Tokar, Y. Danik, V. Katuha, in: Evaluation methods for the ergatic system reliability operator. *Automation 2019: progress in automation, robotics and measurement techniques*, Vol. 920 (2020) 560–570. doi:10.1007/978-3-030-13273-6\_52.
  - [19] L. Ying, G. Wen-xiang, A plan recognition algorithm based on the probabilistic goal graph, in: 2011 International Conference on Network Computing and Information Security, Guilin, 359–362. doi:10.1109/ncis.2011.79.
  - [20] Graph Theory, in: *Probabilistic Networks and Expert Systems*. Information Science and Statistics. Springer, New York, NY (1999) doi:10.1007/0-387-22630-3\_4.
  - [21] M. Bar Eli, H. Plessner, M. Raab, Theories of Decision Making, in *Judgement, Decision Making and Success in Sport* (2011) 27–38. doi: 10.1002/9781119977032.ch3.
  - [22] J. Mishra, D. Allen, A. Pearman, Information seeking, use, and decision making, *Journal of the Association for Information Science and Technology*, vol. 66(4). (2015) 662–673. doi: 10.1002/asi.23204.

# Simple Autonomous Security System Based On Arduino UNO Platform and Fingerprint Scanner Module: A Study Case

Volodymyr Rusyn<sup>a</sup>, Sergey Subbotin<sup>b</sup> and Aceng Sambas<sup>c</sup>

<sup>a</sup> Yuriy Fedkovych Chernivtsi National University, Kotsybynsky str., 2, Chernivtsi, 58012, Ukraine

<sup>b</sup> National university "Zaporizhzhia polytechnic", Zhukovsky str., 64, Zaporizhzhia, 69063, Ukraine

<sup>c</sup> Universitas Muhammadiyah Tasikmalaya, Jl. Tamansari No. KM 2,5, Mulyasari, Kec. Tamansari, Tasikmalaya, Jawa Barat, 46196, Indonesia

## Abstract

In this paper, we presented a simple autonomous security system. This security system based on the Arduino Uno and fingerprint scanner module. The Arduino was connected to a computer through the USB port and programmed using a language similar to C++. Programming code (sketch) was uploaded into Arduino using program software ArduinoIDE. Connection scheme and experimental results are presented and can be used for education.

## Keywords 1

Fingerprint scanner, security system, Arduino

## 1. Introduction

In the modern world, biometric information protection is one of the most effective methods for its preservation. Thanks to the process of authenticating a person, that is, comparing his characteristics with the characteristics previously entered into the system, it is possible to determine with maximum accuracy whether a given person has access to the requested information or still does not. Nowadays, everyone cares about how to protect their data as efficiently as possible. In an effort to secure personal information, mankind has been developing more and more security systems for many years. Biometric protection turned out to be one of the most effective such systems.

Biometric security technologies are becoming more and more everyday attributes of the ordinary person's life. In recent years, they have become extremely widespread in the production of mobile technology: fingerprint scanners, face geometry scanners, and voice recognizers are built into smartphones [1-7]. All this makes the life of an ordinary user much more convenient, quieter and safer. Precisely the same devices are now very often equipped with personal computers, smart homes and many other computerized equipment.

Biometric systems are also quite popular among enterprises, where the entrance to the territory or access to information is open to a large, but limited number of people. At the checkpoints of such enterprises, specialized equipment is usually installed that allows you to identify "your" person through the iris of the eye, the geometry of the hand, face or through the drawing of papillary lines on the pads of the fingers, and sometimes all together.

The introduction of biometrics opens up a sea of different possibilities in banking, making life much easier not only for the bank, but also for its customers. By introducing biometric technologies, bank managers want to reduce the number of data thefts, cases of fraud and greatly simplify the procedure for a client's work with an ATM.

In this paper, we proposed a simple autonomous security system that based on the Arduino Uno and fingerprint scanner module that can be used for education.

---

CMIS-2021: The Fourth International Workshop on Computer Modeling and Intelligent Systems, April 27, 2021, Zaporizhzhia, Ukraine

EMAIL: rusyn\_v@ukr.net (V. Rusyn); subbotin@zntu.edu.ua (S. Subbotin); acenx.bts@gmail.com (A. Sambas)

ORCID: 0000-0001-6219-1031 (V. Rusyn); 0000-0001-5814-8268 (S. Subbotin); 0000-0002-1623-0770 (A. Sambas)



© 2021 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

## 2. ARDUINO board

The Arduino can be compared to a computer motherboard, which also has a processor under the cooler. The motherboard with the processor performs the same functions as the Arduino Uno. Other devices are connected to it: the display, disk drives, hard drives, and all other peripherals. You can also connect many different devices [8].

The uniqueness of the Arduino is that now even an inexperienced radio amateur will be able to create various complex digital devices, without going into theory. Previously, for example, this required programmers, deep knowledge of digital electronics, programming. Arduino was created so that almost any user could create their own interesting and useful device.

Fig. 1 shows Arduino Uno that we used for our project.



**Figure 1:** Arduino Uno board

Arduino Uno controller built on ATmega328. The platform has 16 MHz crystal oscillator, 6 analog inputs, 14 digital inputs / outputs (6 of which can be used as PWM outputs), a USB connector, an ICSP connector, a power connector and a reset button. For work, we must connect the platform to the computer via a USB cable, or apply power using battery or AC/DC adapter.

The Arduino Uno board has 3 ways to connect power on board: via USB, via an external power connector and via the Vin connector, brought out to one of the combs on the side. The platform has a built-in stabilizer on board, which allows not only to automatically select the power source, but also to equalize the current to stable 5 Volts, which the controller needs to work.

External power can be supplied either directly from the USB port of the computer, or from any AC / DC power supply via the power connector or USB.

The board has several pins that allow powering connected sensors, sensors and actuators from it. All of these findings are labeled:

- Vin – power input, used to receive power from an external source. Through the data output, only power is supplied to the board, it is impossible to receive power from there for external devices. It is recommended to apply a voltage in the range from 7 V to 20 V to the Vin input, in order to avoid overheating and burning of the built-in stabilizer.
- 5 V – a five-volt voltage source for powering external devices. When the board receives power from any other source (USB, power connector or Vin), you can always get a stable voltage of 5 Volts on this pin. It can be displayed on a breadboard or fed directly to the desired device.

- 3V3 – source of 3.3 Volt voltage for powering external devices. Works on the same principle as the 5 V pin. From this leg, you can also output the voltage to the breadboard, or apply it directly to the required sensor / sensor.
- GND – pin for ground connection. Needed to create a closed circuit when connected to Vin, 5 V or 3V3 pins. In all cases, the GND leg must be output as a minus, otherwise the circuit will not be closed and the power (both external and internal) will not be supplied.

The Arduino Uno platform has an ATmega328 microcontroller on board, which has Flash, SRAM and EEPROM memory:

- FLASH – 32 kB (0.5 kB is used for bootloader storage SRAM (RAM).
- 2 kB EEPROM.
- 1 kB (available with EEPROM library).

### 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 Uno program is written, compiled and uploaded on the Arduino board as depicted in Fig. 2.

The Arduino can be connected to a computer through the USB port and programmed using a language similar to C++.

In the first stage, we must control the fingerprint sensor module with the Arduino. It is by using the Adafruit library for this sensor.

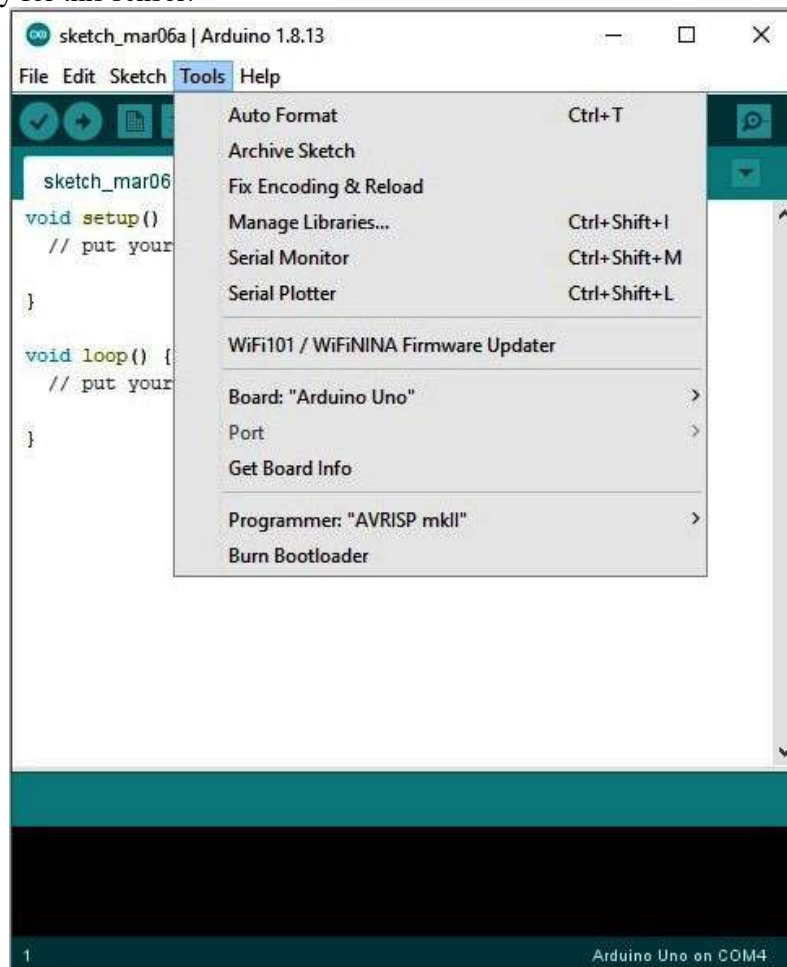


Figure 2: Arduino IDE

Having the fingerprint sensor module wired to the Arduino, we must follow the next steps to enroll a new fingerprint. Make sure you have installed the Adafruit Fingerprint Sensor library previously (Fig. 3).



**Figure 3:** Enroll a new fingerprint

You will be asked to place the same finger twice on the scanner. If you get the "Prints matched!" message, the fingerprint was successfully stored. If not, we repeat the process, until was succeed (Fig. 4).

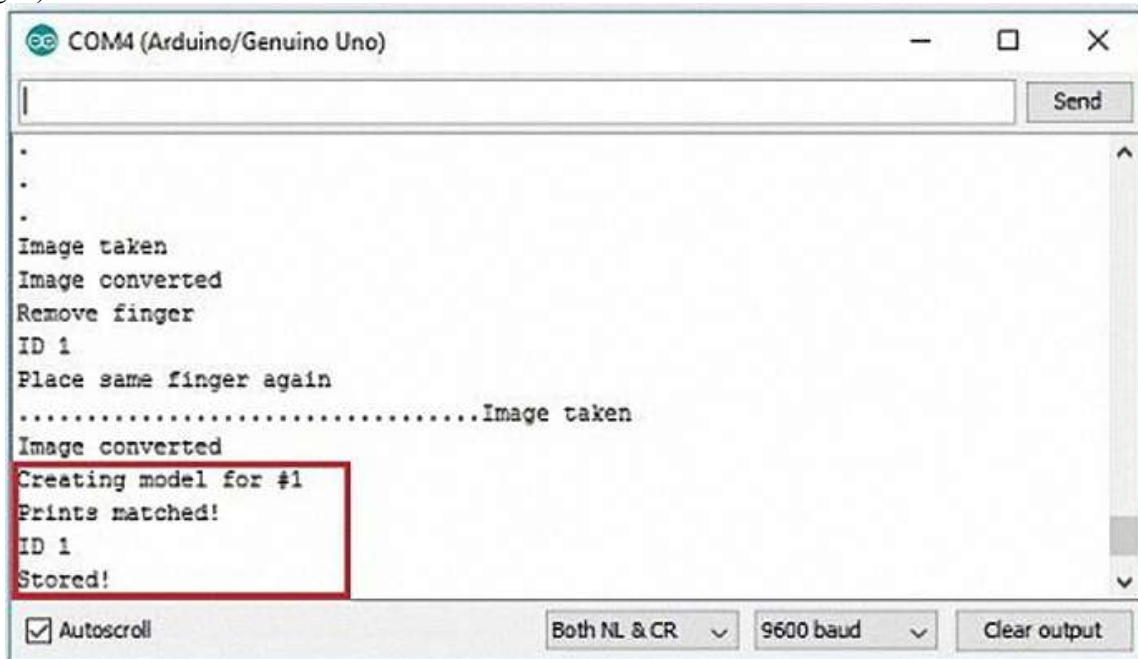


**Figure 4:** Interface that demonstrate that a fingerprint was successfully stored

Now, we should have several fingerprints saved on different IDs (Fig. 5).



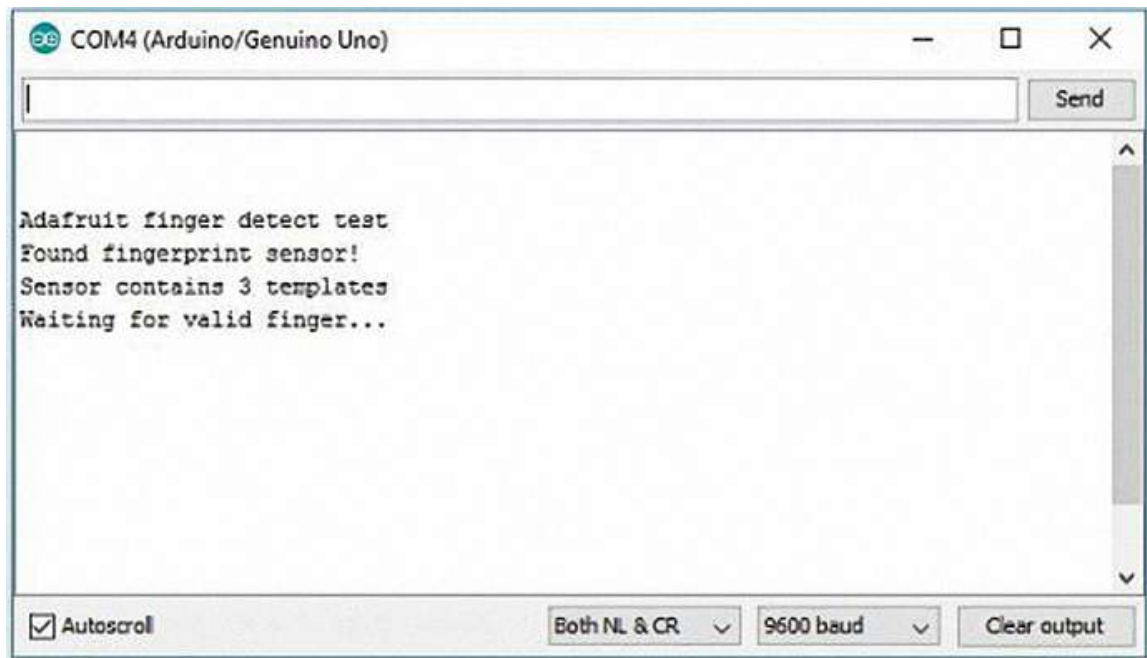


**Figure 5:** Interface that demonstrate that a sensor contains three templates

Next stage: place the finger on the scan for identified . On the serial monitor, we can see the ID user that matches the fingerprint. It also shows the confidence – the higher the confidence, the similar the fingerprint is with the stored fingerprint (Fig. 6).

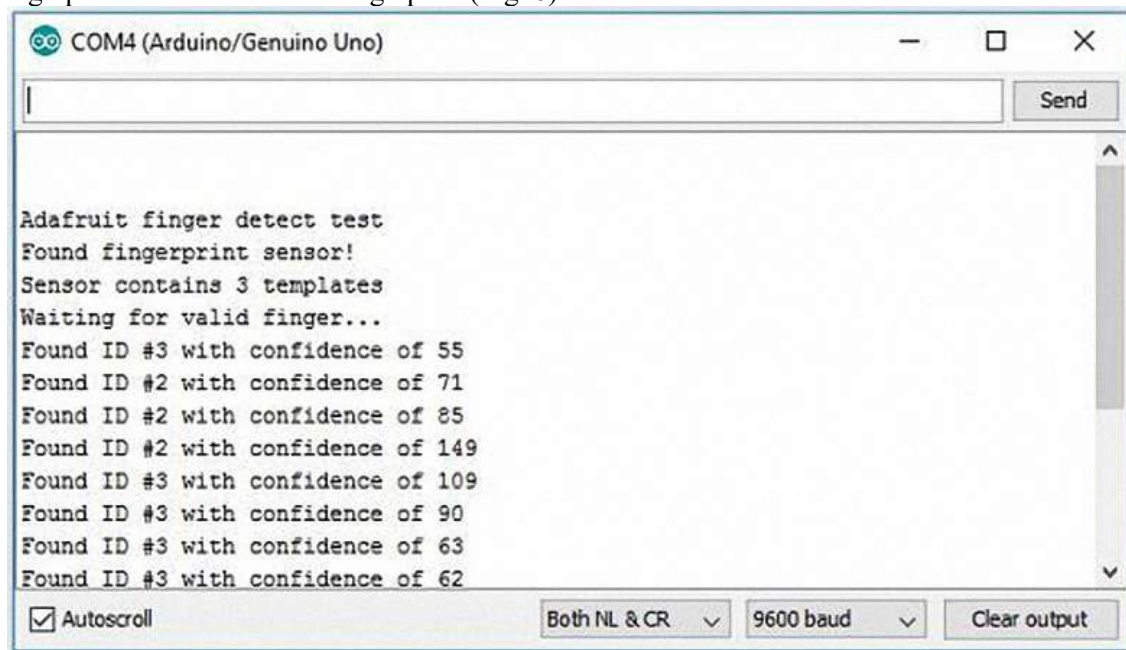


**Figure 6:** Interface that demonstrate ID with confidence



The programming code:

```
#include <Wire.h>
#include <Adafruit_GFX.h>
#include <Adafruit_SSD1306.h>
#define OLED_RESET 4
Adafruit_SSD1306 display(OLED_RESET);

#include <Adafruit_Fingerprint.h>
#include <SoftwareSerial.h>
SoftwareSerial mySerial(2, 3);

Adafruit_Fingerprint finger =
Adafruit_Fingerprint(&mySerial);
int fingerprintID = 0;
String IDname;

void setup(){
  Serial.begin(9600);
  finger.begin(57600);

  if (finger.verifyPassword()) {
    Serial.println("Fingerprint sensor found!");
  }
  else {
    Serial.println("Fingerprint sensor did not find");
    while (1) { delay(1); }
  }

  Wire.begin();
  display.begin(SSD1306_SWITCHCAPVCC, 0x3C);
  displayMainScreen();
}

void loop(){
  displayMainScreen();
  fingerprintID = getFingerprintIDez();
  delay(50);
  if(fingerprintID == 1 || fingerprintID == 2 ||
fingerprintID == 3)
  {
    IDname = "Volodymyr";
    displayUserGreeting(IDname);
  }
  else if(fingerprintID == 2){
    IDname = "Sergey";
    displayUserGreeting(IDname);
  }
  else if(fingerprintID == 3){
    IDname = "Aceng";
    displayUserGreeting(IDname);
  }
}
```

```

int getFingerprintIDez()
{
    uint8_t p = finger.getImage();
    if (p != FINGERPRINT_OK) return -1;

    p = finger.image2Tz();
    if (p != FINGERPRINT_OK) return -1;

    p = finger.fingerFastSearch();
    if (p != FINGERPRINT_OK) return -1;

    Serial.print("Found ID #");
    Serial.print(finger.fingerID);
    Serial.print(" with confidence of ");
    Serial.println(finger.confidence);
    return finger.fingerID;
}

void displayMainScreen()
{
    display.clearDisplay();
    display.setTextSize(1);
    display.setTextColor(WHITE);
    display.setCursor(7,5);
    display.println("Waiting fingerprint");
    display.setTextSize(1);
    display.setTextColor(WHITE);
    display.setCursor(52,20);
    display.println("...");
    display.display();
    delay(2000);
}

void displayUserGreeting(String Name){
    display.clearDisplay();
    display.setTextColor(WHITE);
    display.setTextSize(2);
    display.setCursor(0,0);
    display.print("Welcome");
    display.setCursor(0,15);
    display.print(Name);
    display.display();
    delay(5000);
    fingerprintID = 0;
}

```

## 4. Experimental realization

Fig. 7 shows fingerprint sensor module that made fingerprint recognition more accessible and easy to add to different projects. This means that is super easy to make fingerprint collection, search, registration and comparison. This module work with any microcontroller or system with TTL serial and come with FLASH memory to store the fingerprints. This fingerprint sensor module can be added to security systems, door locks, time attendance systems, and much more.

Specifications of the fingerprint sensor module:

- Voltage supply: DC 3.6 to 6.0 V.
- Current supply: < 120mA.
- Backlight color: green.
- Interface: UART.
- Bad rate: 9600.
- Safety level: five (from low to high: 1,2,3,4,5).
- False Accept Rate (FAR): < 0.001 % (security level 3).
- False Reject Rate (FRR): < 1.0 % (security level 3).
- Able to store 127 different fingerprints.



**Figure 7:** Fingerprint sensor module

Fig. 8 shows pins of the fingerprint sensor module.

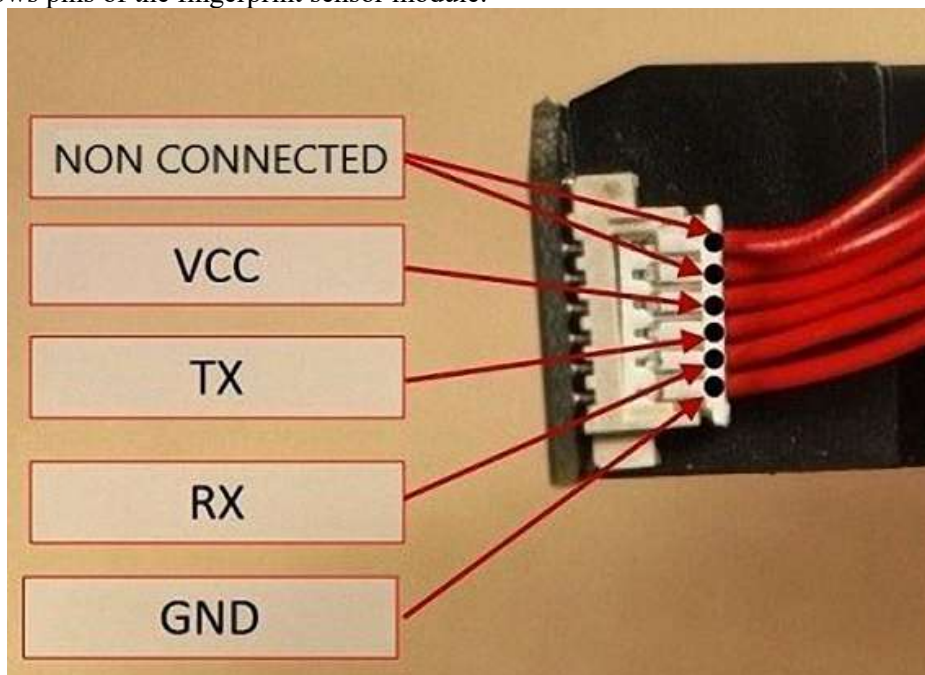


**Figure 8:** Pins of the fingerprint sensor module

The fingerprint sensor module used in this project came with really thin wires, so soldering breadboard-friendly wires was needed. We recommend using different colors according to the pin function.

In our case:

- DNC – white wires.
- VCC – red wire.
- TX – blue wire.
- RX – green wire.
- GND – black wire.

The connection scheme is quite simple and it's rather difficult to make a mistake (Fig. 9).

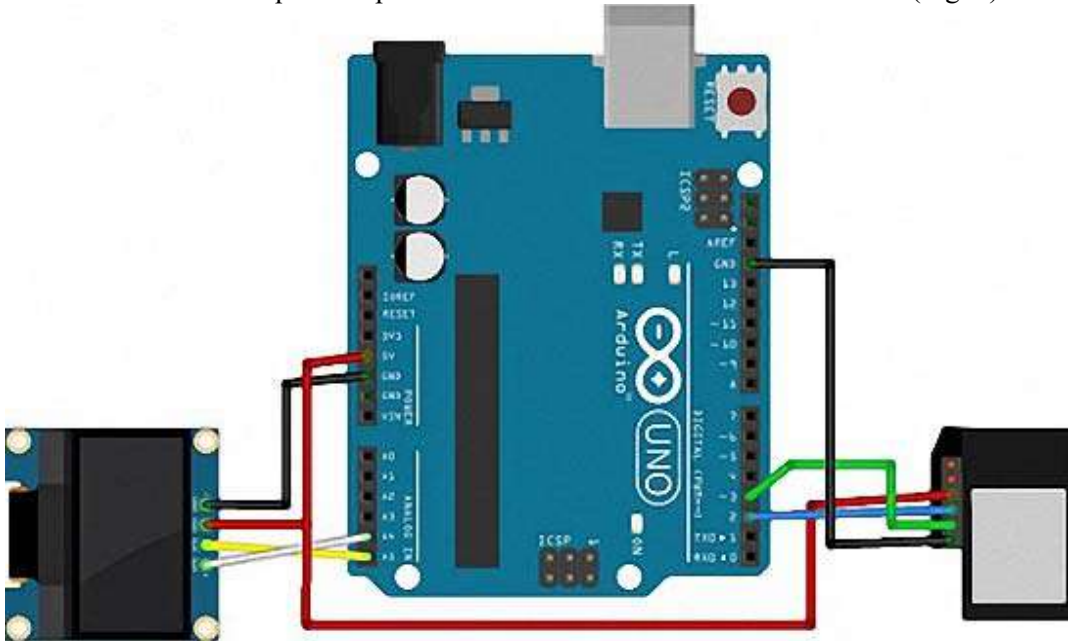


**Figure 9:** The connection scheme

Fig. 10 shows connection scheme for experimental demonstration of the simple autonomous security system.

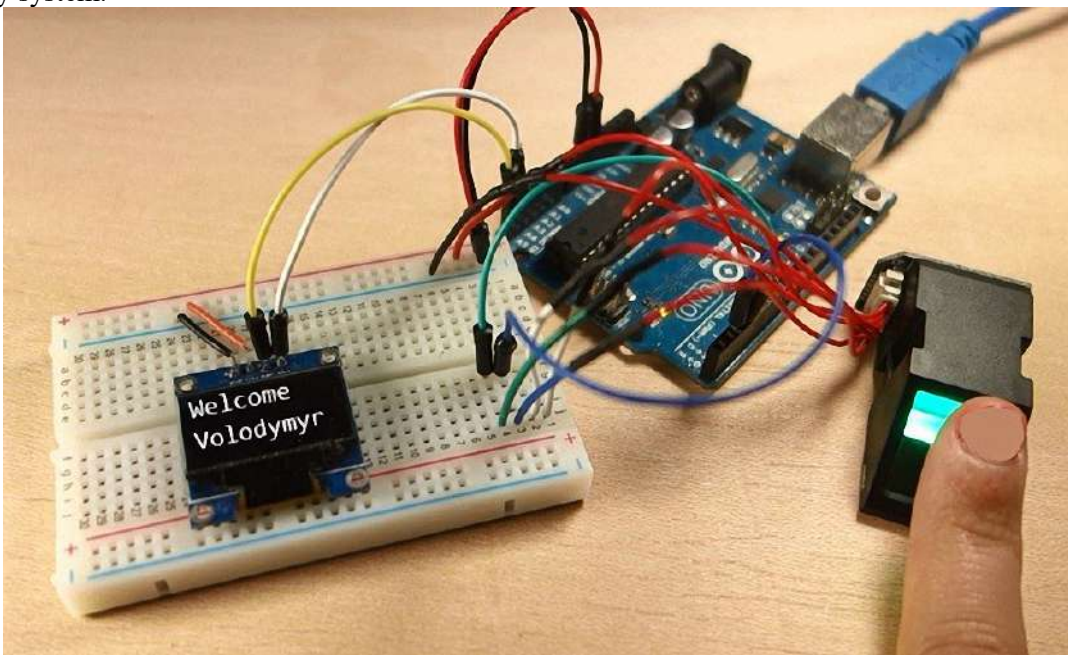


**Figure 10:** Connection scheme for experimental demonstration

## 5. Conclusions

The simple autonomous security system that can be used for individual or different commercial security as example personal data, based on the fingerprint scanner module and Arduino Uno.

The Arduino was connected to a computer through the USB port and programmed using a language similar to C++. Programming code (sketch) was uploaded into Arduino using program software ArduinoIDE. Connection scheme and experimental results are also presented.

## 6. References

- [1] Y. Elmir, O. Ghazaoui, F. Boukenni, Multimodal Biometrics System's Resistance to noise (Fingerprint and Voice). CEUR Workshop Proceedings 942 (2012) 25-28.
- [2] S. Xia, Y. Liu, G. Yuan, M. Zhu, Z. Wang, Indoor fingerprint positioning based on Wi-Fi: an overview, ISPRS Int. J. Geo-Inf. 6 (2017) 135.
- [3] H. Suining, S. H. Gary Chan, Wi-Fi Fingerprint-Based Indoor Positioning: Recent Advances and Comparisons, IEEE Commun. Surv Tut, 18 (2016) 466-490.
- [4] G. Li, C. Busch and B. Yang, "A novel approach used for measuring fingerprint orientation of arch fingerprint," 2014 37th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), Opatija, 2014, pp. 1309-1314. doi: 10.1109/MIPRO.2014.6859770
- [5] S. P. Sandip and P. H. Zope, "Selective review of fingerprint enhancement, classification and matching techniques," 2015 IEEE Bombay Section Symposium (IBSS), Mumbai, 2015, pp. 1-6. doi: 10.1109/IBSS.2015.7456656
- [6] G. Caso, L. De Nardis, On the Applicability of Multi-wall Multi-floor Propagation Models to WiFi Fingerprinting Indoor Positioning, Future Access Enablers for Ubiquitous and Intelligent Infrastructures: Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, Springer, 2015.
- [7] Tomas Trainys, Algimantas Venčkauskas. Encryption Keys Generation Based on Bio-Cryptography Finger Vein Method. CEUR Workshop Proceedings 2145 (2018) 106-111.
- [8] Rusyn V, Subbotin S., Sambas A., Analysis and Experimental Realization of the Logistic Map Using Arduino Pro Mini. CEUR Workshop Proceedings 2608 (2020) 300-310.

# An Approach to Development of Interactive Adaptive Software Tool to Support Data Analysis Activity

Dmytro Orlovskiy, Andrii Kopp and Ivan Bilous

National Technical University “Kharkiv Polytechnic Institute”, Kyrpychova str. 2, Kharkiv, 61001, Ukraine

## Abstract

In the recent decades, databases have been used in any field of human activity to keep valuable data about ongoing processes. Large amounts of data stored in enterprise databases are useless without having a specialized software tool for data discovery or querying. Most business users that make data-driven decisions usually do not have special training and experience in database querying using special formal languages. Existing solutions are based on “query wizards” and database query forms that require knowledge of a database schema and inconvenient for users without special training. Proposed approach is based on the content-based filtering of already executed queries by usage frequency and similarity criteria in order to suggest relevant queries that may be re-used. It is a baseline of the interactive adaptive system for data analysis, which design and development is outlined in this study. Software prototype was demonstrated and its usage was discussed. Conclusions were made and future research was formulated.

## Keywords 1

Database Query, Adaptive Query Interface, Data Analysis, Interactive Software Tool

## 1. Introduction

Nowadays enterprise databases contain extremely large collections of transactional and aggregated analytical data records. Such data volumes contain descriptions of hundreds or even thousands of data entities, described by multiple various attributes of different data types. Without having interactive and flexible querying tools, such data collections are basically useless, since no data could be obtained for analytical processing, visualization, and decision making. Existing querying languages, such as SQL (Structured Query Language), LINQ (Language Integrated Query), DAX (Data Analysis eXpressions), or XQuery, need special training and experience in software engineering, database design, and database administration. Besides querying languages, it is required to know the database schema structure to apply these query languages. Such data preparation tasks may distract data analysts from their primary activities, which include interpretation and communication of data analysis results (e.g. using simple spreadsheets or data visualizations based on descriptive, or predictive models [1]) to stakeholders. Also improper usage of database querying languages and lack of database schema knowledge, may mislead data analysts and block problems solving.

Modern database management systems are accompanied with so-called “query wizards” intended to simplify data querying and do not require knowledge of SQL or other data querying languages. These software components are considered as user-friendly tools that could be used by non-technical users to build analytical reports and glean insights. On practice most data analysis problems require integration of enterprise databases or other corporate data sources with a specialized software tool or service. There are Business Intelligence (BI) tools focused on visual reports design (e.g. Power BI or QlikView) [2], data science methods and models provided by a standalone software tool (e.g. RStudio or Jupyter) [3] or libraries and frameworks (e.g. Pandas or Matplotlib) [4]. The first step for data

CMIS-2021: The Fourth International Workshop on Computer Modeling and Intelligent Systems, April 27, 2021, Zaporizhzhia, Ukraine  
EMAIL: orlovskiy.dm@gmail.com (D. Orlovskiy); kopp93@gmail.com (A. Kopp); ivanbilous2000@gmail.com (I. Bilous)  
ORCID: 0000-0002-8261-2988 (D. Orlovskiy); 0000-0002-3189-5623 (A. Kopp); 0000-0001-8110-786X (I. Bilous)



© 2020 Copyright for this paper by its authors.  
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).  
CEUR Workshop Proceedings (CEUR-WS.org)



visualization, or even for simple data presentation using spreadsheets, is data querying. Since query languages require skills and experience to be applied, and existing “query wizards” are limited in their functionality and mostly are not understandable for end users, the problem of development of a software tool for data querying becomes relevant. This software tool should provide interactive adaptive user interface, which can be used by such users as: data analysts (if they for some reason do not have special training in database management systems), business analysts, top managers, or board members.

Therefore, this study proposes an approach to development of interactive adaptive software tool to support data analysis activity that is considered as the research objective. Research subject includes the approach to development of interactive adaptive software tool. This study aims to decrease complexity of data analysis process by supporting related activities with the interactive adaptive software tool.

This paper consists of six sections. First section introduces research problem and is already outlined. Second section demonstrates literature review and problem relevance. In third section a formal problem statement is given. Fourth section includes description of a recommending procedure, elicited software requirements, a database structure, and a system design. Software prototype usage and its discussion is outlined in fifth section. Sixth section includes conclusions and a future research statement.

## **2. Literature Review**

Analysis of the state-of-the-art has demonstrated that considered research topic is already covered by several studies. Tang et al. in [5] proposed a database query form (DQF) interface used to generate query forms based on captured user’s preferences. In these DQF preferences are gathered to rank query form components in order to assist users in making decisions. Query form generation process outlined in [5] is iterative and is guided by users: among proposed lists of components users may choose desired ones in order to include into the query form. At each iteration users can submit their forms and adjust DQF until they are satisfied with the obtained response [5].

The term DQF introduced in [5] then appeared in multiple research studies. Authors of [6] propose a DQF system based on user interactions captured in order to adapt questions submitted through query forms. Proposed procedure is also iterative, which means that form could be dynamically refilled till the user satisfies with query results [6]. Paper [7] proposes a survey on DQF approach and reviews its core concepts: query form interfaces, interface components ranking metrics, and estimation of ranking score. The DQF survey also surveyed in [8], where four functional modules of DQF software systems are considered: query form enrichment module, query execution module, customizable query form, and query recommendation module. It is also concluded in all of the considered survey papers that dynamic database querying approach results in higher success rate and easier usage in compare to static database querying approach [6, 7, 8].

Minor contributions to DQF proposed in [5] were made by authors of [9] and [10], while substantial contribution was made by Dagade and Bhonsle, who proposed search optimization for dynamic query form approach or SODQF, which is based on keywords in addition to user feedbacks, as the idea for further development [11]. The keywords based DQF is proposed in [12], which authors also consider queries as items for the collaborative filtering approach in order to provide recommendations. There are also many papers devoted to DQF study, such as [13, 14, 15, 16, 17, 18, 19]. All of this papers are based on previously developed concept of DQF, while proposing some adjustments and improvements to the core idea. For example, in [13] the focus is made on elaboration of differences in DQF application to SQL and NoSQL databases. In [14] authors propose to develop methods to capture user preferences to replace direct feedbacks. Authors of [15] propose an alternative to DQF called auto query forms (AQF) that take queries written in human understandable language and translate them into SQL queries using natural language processing (NLP) methods. Performance measurement of DQF approach was made in paper [16], authors of [17] elaborated the algorithm for DQF generation, and query form formalization was made in [18].

While previously mentioned papers considered mostly SQL baseline and relational databases, in study [19] authors also consider DQF with iterative search and query enhancements by users

feedbacks and responses, but with the use of document-oriented NoSQL database management systems based on JSON (JavaScript Object Notation) structures like MongoDB [19]. Usage of DQF approach for NoSQL databases also considered in research [20], where obtained results of performance comparison between SQL based and NoSQL database management systems demonstrate prevalence of relational databases in terms of querying performance.

Fig. 1 demonstrates rapid increase of DQF research topic popularity after it was mentioned for the first time in 2013 by authors of [5]. However, after 5 years of presumably insufficient results or lack of practical implementations, it does not seem really popular (see Fig. 1). Publications statistics outlined in Fig. 1 was received from Google Scholar as the most comprehensive index of research papers.

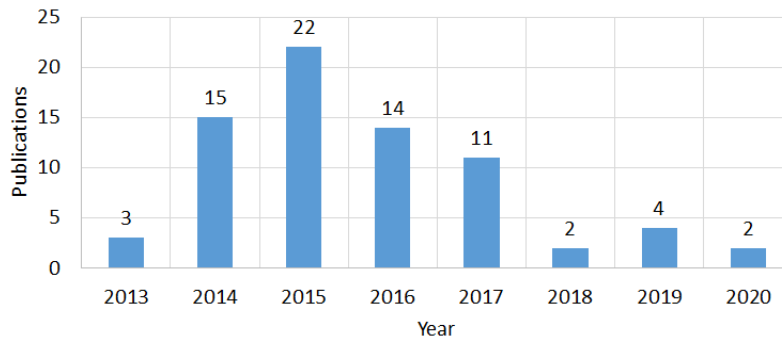


**Figure 1:** Research interest to DQF

The problems of DQF technique mentioned in two other survey and review papers [21, 22] related to ranking of suggested queries or form components, high complexity of formal querying languages for end users, and inconvenience when working with large database schemas. Considering such problems remaining after years of research and unfair lack of interest for such promising, by our opinion, research topic, the DQF approach should be elaborated in this study in order to develop interactive adaptive software tool to support data analysis activity.

### 3. Formal Problem Statement

The interactive adaptive software system for data analysis is supposed to be utilized by two types of users: administrator and analyst. Administrator should be responsible for data sources connections and configuration, analysts profiles creation, and management of their permissions. Analyst can run queries to data sources with possibility of further visualization of obtained results. Administrator should be able to use the same functionalities that analyst can use.

When using this system, administrator should configure a workspace for analysts. The configuration process starts with analysis of business requirements received from managers or other stakeholders who are responsible for data-driven decision making. Then, administrator should add required data sources using their properties. Usually these properties are server name, database name, login and password. At the next stage administrator should prepare queries (e.g. using stored procedures on the server side) and provide additional information on their parameters. Then administrator creates user profiles for analysts and grants permissions: which data sources certain analysts could access and which queries they could execute. Querying tools are essential for the proposed system. That is why prepared queries, presented as stored procedures, should be described with the metadata sufficient for their displaying as graphical user interface (GUI) forms.

Using provided GUI analyst could query connected data sources without knowledge of SQL or other database languages. Using different interface components user could provide parameters for attributes demonstration, records filtering, and other stored procedure features. When preparing queries analysts may use instructions of stakeholders and use required data source, query, and corresponding parameters. Formed using GUI query then executed and received data are displayed on the screen ready for further processing and visualization. The system of adaptive interfaces includes recommending techniques for analytical queries. System should observe user activity and capture



statistics of queries execution. Using this statistics analysts may receive recommendations of queries the most similar to queries executed before. Recommendations should be formulated for each data source (e.g. stored procedure). Structural model of such system is demonstrated on the IDEF0 diagram below (Fig. 2).

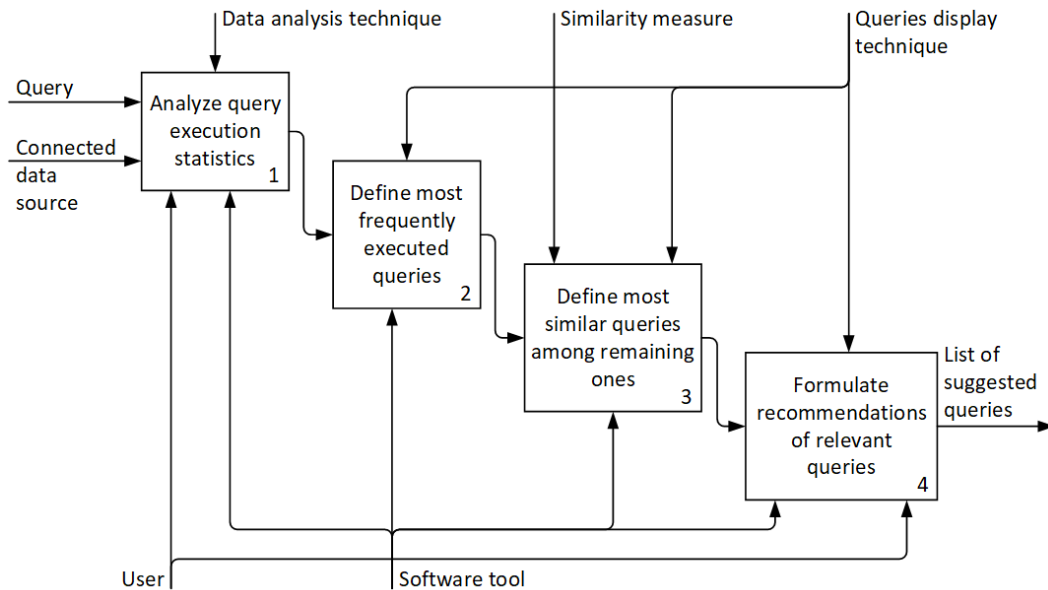


**Figure 2:** Structural model of the analytical queries recommending system

Recommending algorithm should analyze query execution statistics and define frequencies of query execution in order to detect relevant queries. Three most frequently used queries should be compared to remaining queries in order to detect two most similar queries for each of the most frequently used ones. Therefore, there will be proposed nine most relevant queries for data analytics. Such number of recommended queries (9 items) originates from the statement that human perception allows to take into account  $7 \pm 2$  objects at once. Recommending mechanisms provide interactive querying system that may adapt to user's preferences. Formulated suggestions are optional and not mandatory to follow.

## 4. Proposed Approach

### 4.1. Recommending Procedure

Proposed approach is based on the content-based filtering approach. Users get recommendations of those queries, which are similar to the queries executed by them before. The underlying idea considers capturing statistics of queries execution by users and therefore it becomes possible to formulate the list of frequently executed queries for each user by each data source.

Let us briefly describe proposed recommending procedure:

- At first, three most frequent queries should be chosen.
- Then for each of the most frequent queries, selected at first step, should be chosen two most similar to them queries within corresponding data sources. Similarity between two queries could be calculated using analysis of their SQL code.
- As the result, the list of nine recommended queries should be formulated.

Most of recommending systems that support content-based filtering approach use keyword matching or vector space model (VSM) with simple TF-IDF weighting [23].

VSM is the algebraic representation of document collection using vectors that belong to the vector space common for the whole document collection. TF-IDF is the statistical indicator used to evaluate importance of terms (words) within the context of a document that belongs to the document collection (corpus). Weight (importance) of the term is directly proportional to the frequency of such term use in

a document and inversely proportional to the frequency of such term use in remaining documents of the collection [23].

The usage of the considered recommending procedure is demonstrated in Fig. 3.

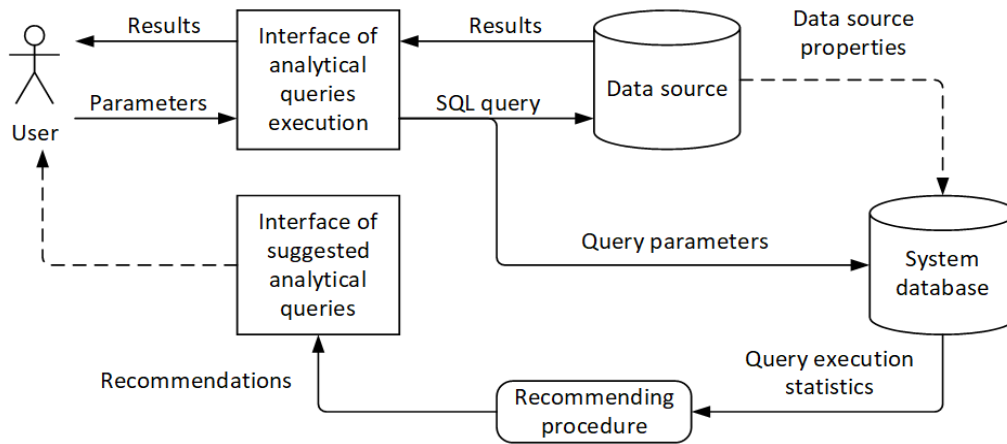


**Figure 3:** Usage of the recommending procedure

According to the proposed recommending procedure, SQL statements that correspond to queries are considered as documents. Document collection includes the set of all SQL statements accessible to the user within a corresponding data source.

Therefore, formal description of the proposed recommending procedure is based on the pair:

$$\langle D, T \rangle, \quad (1)$$

where:

- $D = \{d_1, d_2, \dots, d_m\}$  is the collection of SQL statements (documents),  $d_j$ ,  $j = \overline{1, m}$  is the SQL statement (document) represented as the vector of length  $n$ , and  $m$  is the number of SQL statements (documents) within the collection [23];
- $T = \{t_1, t_2, \dots, t_n\}$  is the general vocabulary of the document collection, which contains  $n$  terms (pairs of parameter names and values of SQL statements that call stored procedures) [23].

Each document (SQL statement)  $d_j$ ,  $j = \overline{1, m}$  of the collection defined in (1) could be represented as the vector of length  $n$ :

$$d_j = \{w_{1j}, w_{2j}, \dots, w_{nj}\}, \quad (2)$$

where  $w_{kj}$  is the weight of the term (represented by the pair of parameter name and value of the SQL statement that calls a stored procedure)  $t_k$ ,  $k = \overline{1, n}$  in the document  $d_j$ ,  $j = \overline{1, m}$ .

Documents representation using VSM allows weighting terms and detecting similarity of document vectors.

TF-IDF weighting approach is based on the following empirical assumptions [23, 24]:

- Rare terms are not less relevant than frequent terms (so called IDF assumption).
- Multiple occurrences of a term in the document make it less relevant than single occurrence of a term in the document (so called TF assumption).
- Documents of large size do not have advantage over small documents (so called normalization assumption).

Hence, terms that occur in one document, but rarely occur in remaining documents, are more often relevant to the document's topic. Also normalization of result vectors of weights allow avoiding large documents' advantage. These assumptions are demonstrated by the TF-IDF function [24]:

$$\text{TF-IDF}(t_k, d_j) = \frac{n_j}{\sum_{l=1}^n n_l} \cdot \log \frac{m}{n_k}, \quad (3)$$

where:

- $m$  is the number of documents (SQL statements) within the collection;
- $n_k$  is the number of documents within the collection of SQL statements, in which the term  $t_k$ ,  $k = \overline{1, n}$  occurs at least once;
- $n_j$  is the number of times the term  $t_k$ ,  $k = \overline{1, n}$  occurs in the document  $d_j$ ,  $j = \overline{1, m}$  (2);
- $n_l$  is the number of times each term  $t_l$ ,  $l = \overline{1, n}$  occurs in the document  $d_j$ ,  $j = \overline{1, m}$  (2).

In order to obtain weights within the interval  $[0,1]$  and documents represented using vectors of the same length, TF-IDF measures obtained using (3) should be normalized [23, 24, 25]:

$$w_{kj} = \text{TF-IDF}(t_k, d_j) \cdot \left( \sqrt{\sum_{s=1}^{|T|} \text{TF-IDF}(t_k, d_s)^2} \right)^{-1}, k = \overline{1, n}, j = \overline{1, m}. \quad (4)$$

Cosine similarity measure could be used to calculate similarity between two vectors of normalized TF-IDF measures (4) [23]:

$$\text{sim}(d_i, d_j) = \sum_{k=1}^n (w_{ki} \cdot w_{kj}) \cdot \left( \sqrt{\sum_{k=1}^n w_{ki}^2} \cdot \sqrt{\sum_{k=1}^n w_{kj}^2} \right)^{-1}, i \neq j, i = \overline{1, m}, j = \overline{1, m}. \quad (5)$$

Besides cosine similarity, there may be used other measures, such as Jaccard similarity [26] or even adjusted cosine-based similarity measures, such as soft cosine similarity [27].

Using equation (5), the most frequently used queries should be compared to remaining queries that belong to the collection of SQL statements  $D$ . For three of the most frequently used queries, remaining queries with the highest similarity scores are then selected in order to provide recommendations within respective data source. In this study we consider data sources as stored procedures also known as executable routines hosted on a database server. Unlike related papers that consider both SQL and NoSQL databases, in this study we focus on relational database management systems (DBMS) only, since among the most widely used and popular DBMS at are still SQL-based systems, such as Oracle, MySQL, Microsoft SQL Server, or PostgreSQL [28]. User-defined stored procedures are created by DBA for encapsulation, security, and performance reasons. On practice stored procedures implement parameterized queries on the server side to provide business logic for various heterogeneous clients. According to the proposed approach, stored procedures should be used as data sources for adaptive interactive querying tools. For each of these data sources (stored procedures) the set of its calls exists. The SQL query to call a stored procedure includes its name and the list or parameters. Sample SQL query that calls the stored procedure to obtain the list of students by the year of admission and country of origin is shown below:

$$\text{EXEC GetStudents @Year = '2020', @Country = 'Turkey';} \quad (6)$$

For example, such stored procedure is called ten times with different parameters. Table 1 shows data gathered using these calls: documents collection  $D = \{d_1, d_2, \dots, d_{10}\}$  of ten SQL queries used to call the stored procedure with different parameters (used as terms in pairs with their names).

**Table 1**

Sample documents collection

| Document | Year | Country  | $t_1$ | $t_2$ | $t_3$ | $t_4$ | $t_5$ | $t_6$ | $t_7$ | $t_8$ |
|----------|------|----------|-------|-------|-------|-------|-------|-------|-------|-------|
| 1        | 2018 | Morocco  | 1     | 0     | 0     | 0     | 1     | 0     | 0     | 0     |
| 2        | 2019 | Turkey   | 0     | 1     | 0     | 0     | 0     | 1     | 0     | 0     |
| 3        | 2019 | Turkey   | 0     | 1     | 0     | 0     | 0     | 1     | 0     | 0     |
| 4        | 2018 | Turkey   | 0     | 1     | 0     | 0     | 1     | 0     | 0     | 0     |
| 5        | 2020 | Algeria  | 0     | 0     | 1     | 0     | 0     | 0     | 1     | 0     |
| 6        | 2017 | Turkey   | 0     | 1     | 0     | 0     | 0     | 0     | 0     | 1     |
| 7        | 2017 | Morocco  | 1     | 0     | 0     | 0     | 0     | 0     | 0     | 1     |
| 8        | 2020 | Pakistan | 0     | 0     | 0     | 1     | 0     | 0     | 1     | 0     |
| 9        | 2020 | Turkey   | 0     | 1     | 0     | 0     | 0     | 0     | 1     | 0     |
| 10       | 2018 | Morocco  | 1     | 0     | 0     | 0     | 1     | 0     | 0     | 0     |

Therefore, e.g. documents  $d_1$  and  $d_4$  could be represented using the following vectors:

$$d_1 = \{0.71, 0, 0, 0, 0.71, 0, 0, 0\}, d_4 = \{0, 0.5, 0, 0, 0.87, 0, 0, 0\}. \quad (7)$$

These vectors were obtained using equations (3) and (4), while similarity between  $d_1$  and  $d_4$  could be calculated using (5) applied to the vectors of document weights (7),  $\text{sim}(d_1, d_4) = 0.61$ .

There is also another document exists, which similarity is greater than zero when compare it to the  $d_1$  document, it is  $d_7$ :

$$d_1 = \{0.71, 0, 0, 0, 0.71, 0, 0, 0\}, d_7 = \{0.6, 0, 0, 0, 0, 0, 0, 0.8\}. \quad (8)$$

Similarity between  $d_1$  and  $d_7$  calculated using (5) applied to the vectors of document weights (8) is  $\text{sim}(d_1, d_7) = 0.42$ . Therefore, there are two queries that should be recommended as relevant to (6):

EXEC GetStudents @Year = '2018', @Country = 'Turkey'; (9)

EXEC GetStudents @Year = '2017', @Country = 'Morocco';

Obtained queries (9) are similar to (6) by one of the year or country parameter values.

The workflow of recommending procedure is demonstrated in Fig. 4.

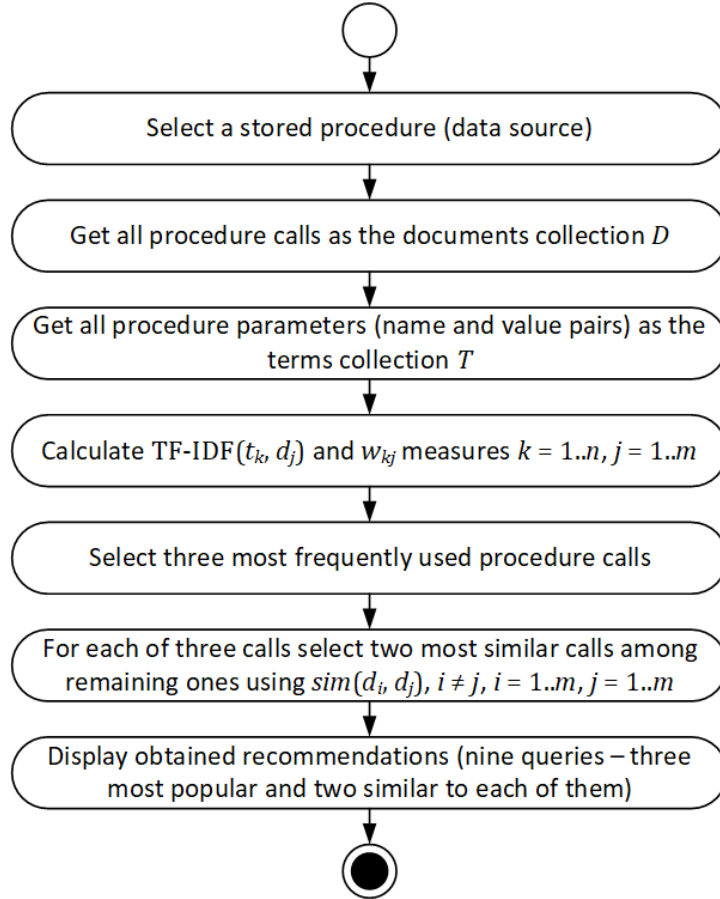


**Figure 4:** Workflow of the recommending procedure

Demonstrated workflow is the baseline of the interactive adaptive data analytics system proposed in this paper. Software requirements, database schema, and system design are outlined in following sub-section. Prototype demonstration and discussion are given in section 5.

## 4.2. Software Requirements

There are two kinds of users that may interact with the system of adaptive interactive interfaces:

- Database administrators (DBA).
- Data analysts (DA).

Preliminary domain analysis has led to the functional requirements (FR) for the system under design demonstrated in Table 2.

**Table 2**  
Elicited functional requirements

| User       | Requirement | Details                                                                                                 |
|------------|-------------|---------------------------------------------------------------------------------------------------------|
| DBA        | FR1         | Manages data source connections (connects, disconnects, and modifies connection properties)             |
|            | FR2         | Manages analytical queries and corresponding stored procedures (creates new and modifies existing ones) |
|            | FR3         | Manages DA profiles (creates, modifies, and disables) and profile privileges for data access            |
| DBA,<br>DA | FR4         | Runs available analytical queries within granted data sources                                           |
|            | FR5         | Searches over, visualizes, or exports received datasets to Microsoft Excel spreadsheets                 |
|            | FR6         | Receives recommendations and suggestions for analytical queries                                         |
|            | FR7         | Searches over available data sources, analytical queries, and visualization tools                       |

As it is outlined in Table 2 above, DBA should be able to access the same functionality accessible to DA. Functional requirements are demonstrated on SysML requirements diagram [29] below (Fig. 5), which clarifies user roles, granted permissions, and responsibility areas.

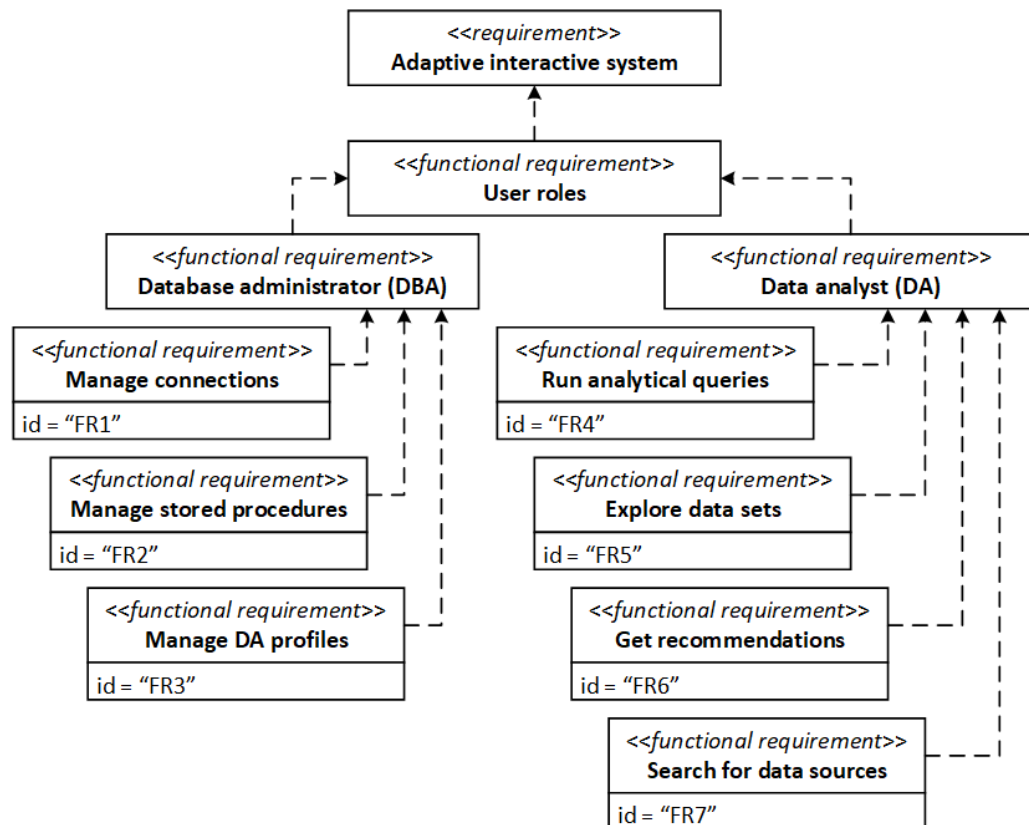


**Figure 5:** Functional requirements diagram

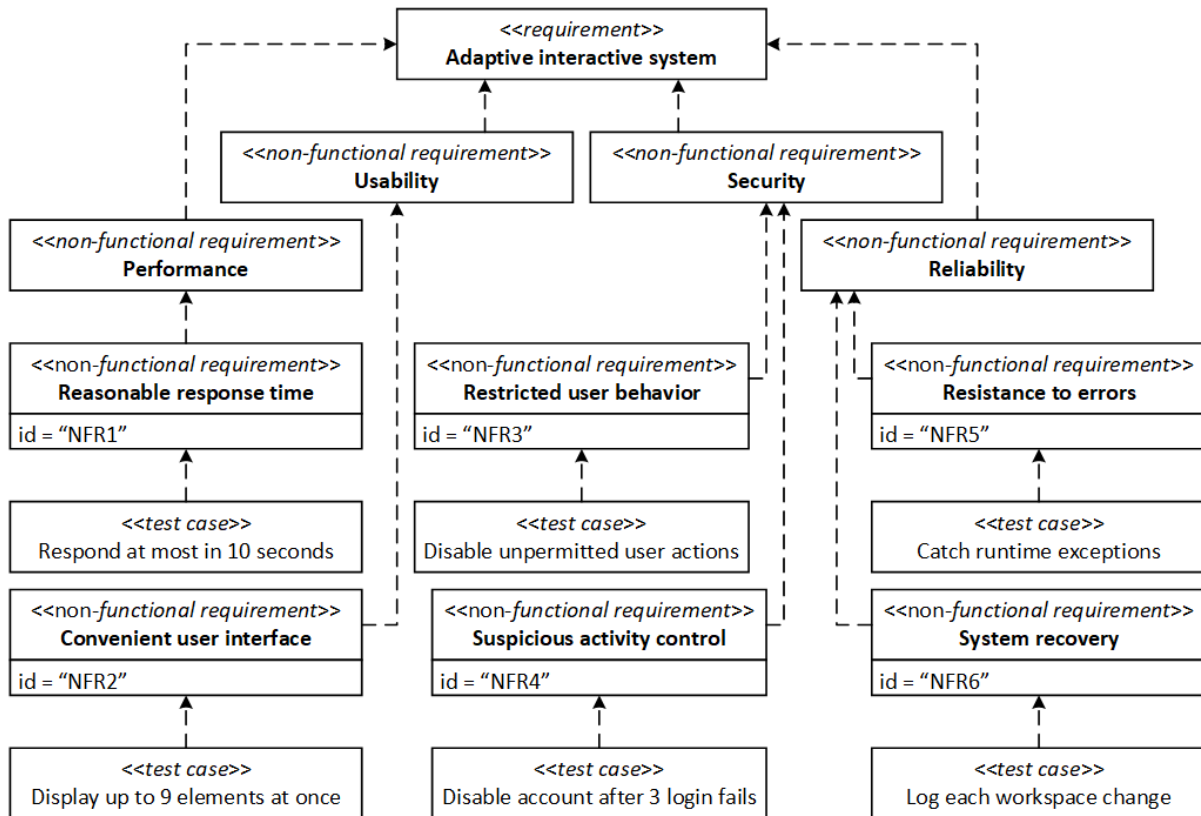
Also there were formulated non-functional requirements for the system of adaptive interactive tool for data analysis activities. Elicited non-functional requirements (NFR) include performance, usability, security, and reliability requirements (Table 3).

**Table 3**

Elicited non-functional requirements

| Attribute   | Requirement | Details                                                                                                               |
|-------------|-------------|-----------------------------------------------------------------------------------------------------------------------|
| Performance | NFR1        | System should respond in a reasonable time even for complex queries and large data sets                               |
| Usability   | NFR2        | Graphical user interface should be friendly and convenient for end users without knowledge of SQL and database schema |
| Security    | NFR3        | User's actions should be restricted according to their privileges and responsibility areas (DBA or DA)                |
| Reliability | NFR4        | Users' accounts should be disabled if suspicious activity detected                                                    |
|             | NFR5        | System should be resistant to errors and user's actions should not damage connected data sources                      |
|             | NFR6        | States of user's workspace should be periodically saved for recovery purposes                                         |

For each of NFR outlined in Table 3 should be given specific formal or even numerical measures to verify future software system. Non-functional requirements are demonstrated on SysML requirements diagram [29] below (Fig. 6), which clarifies measurable quality attributes.

**Figure 6:** Non-functional requirements diagram

Recommending system should have its own database to store user profiles and privileges, connected data sources, and configured analytical queries.

### 4.3. Database Structure

Besides storing user profiles, data sources, and queries, the database should contain captured usage statistics to utilize it for recommending purposes.

The data model, which describes database tables and relationships, is shown in Fig. 7.

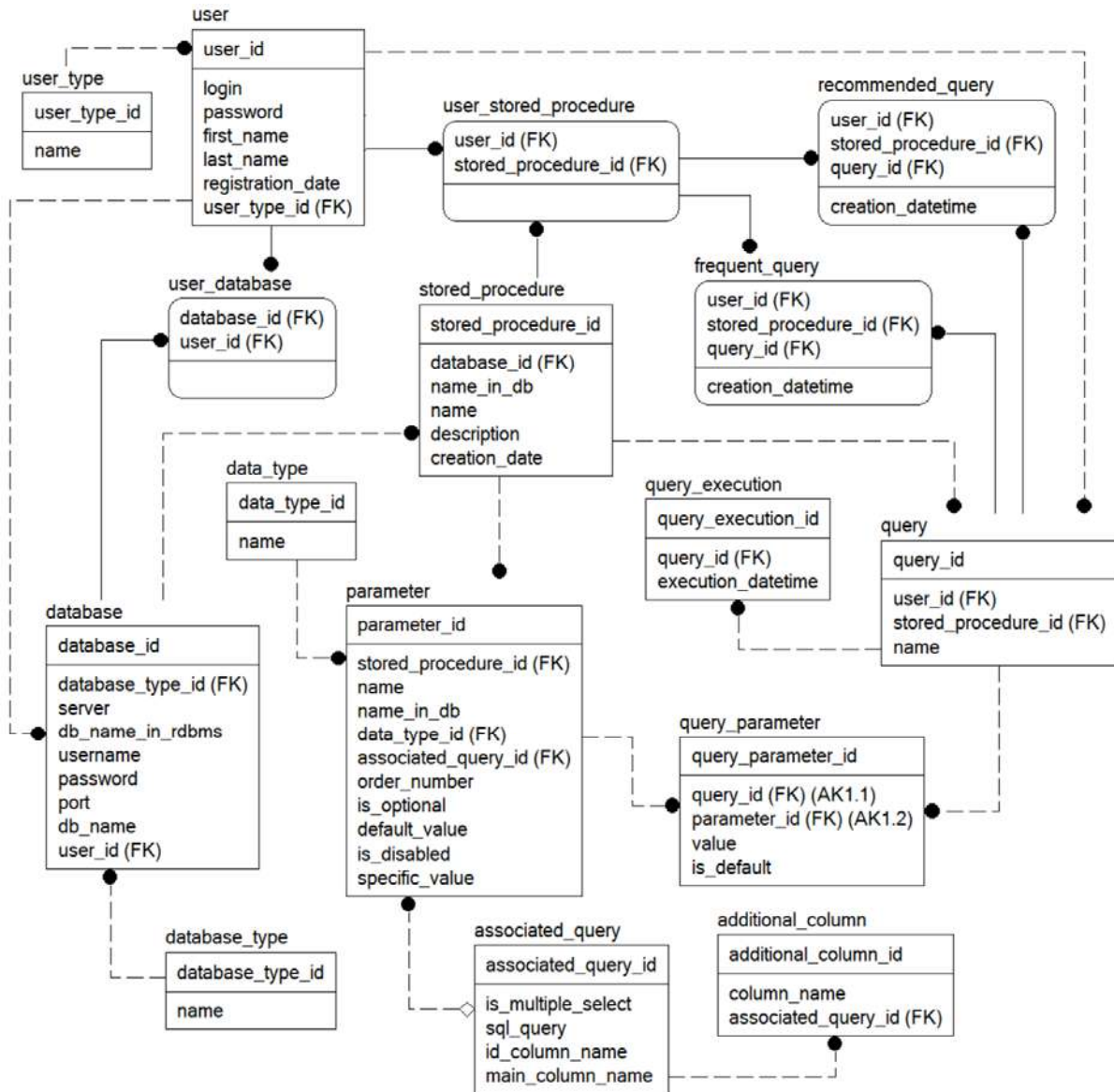


**Figure 7:** Database structure

Demonstrated data model includes entities for generic user, as well as for administrators and analysts with their own attributes. Administrators attach databases of different types assigned to analysts, stored procedures, and given recommendations. Stored procedures are described by parameter names and data types. Attributes of datasets produced by stored procedures are also considered. Stored procedures are assigned to analysts according to permissions configured by administrators. Each query run is stored in order to capture usage statistics for recommending purposes.

#### 4.4. System Design

Let us start system design description with the dynamic models based on CMMN (Case Management Model and Notation) graphical diagrams [30]. When user selects desired data source from the dropdown list with the search feature, respective stored procedures appear in the left side of application's window, while recommended analytical queries appear in the right side of the window. Administrators can access all attached databases and stored procedures used as data sources for



analytical queries, while analysts can access only databases and stored procedures according to their privileges. Both lists could be filtered by name of a database or stored procedure respectively. After certain query is executed, a modal window is displayed to provide parameters of a stored procedure (to filter produced dataset by rows) and choose desired attributes (to filter produced dataset by columns). After query is executed, result set of records is demonstrated. Recommended queries already have relevant parameters and attributes and do not need to be additionally configured.

Case management notation is the novel technique for description of ad-hoc processes [30], including ad-hoc analytical data querying. The CMMN model that describes query execution using the interactive adaptive software system for data analysis is shown in Fig. 8.

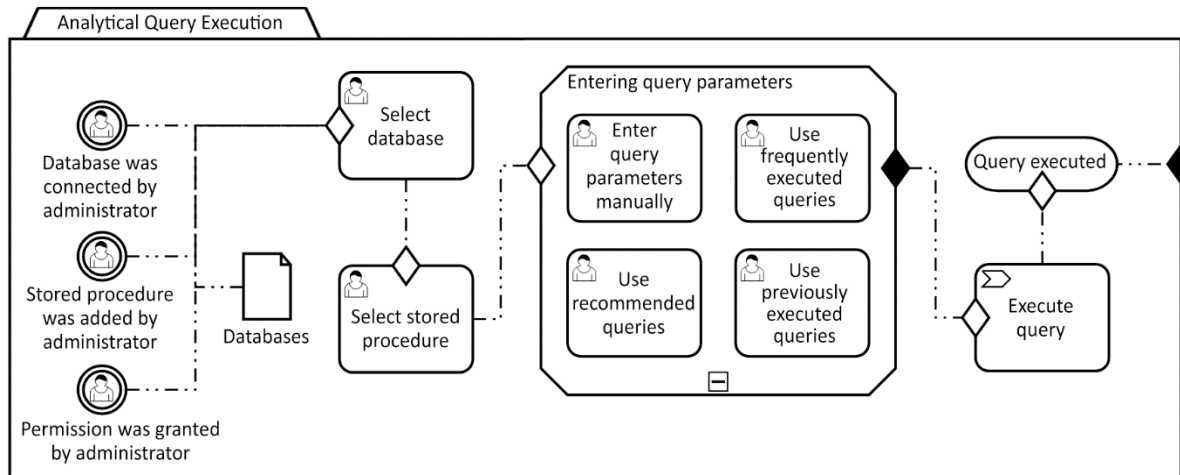


**Figure 8:** Case management model of the interactive adaptive system for data analytics

User interface structure of analytical querying could be described using the IFML (Interaction Flow Modeling Language) [31] diagram, which describes user interaction, interface structure, and behavior of the software system. The IFML model that describes analytical querying GUI structure and behavior is displayed in Fig. 9.

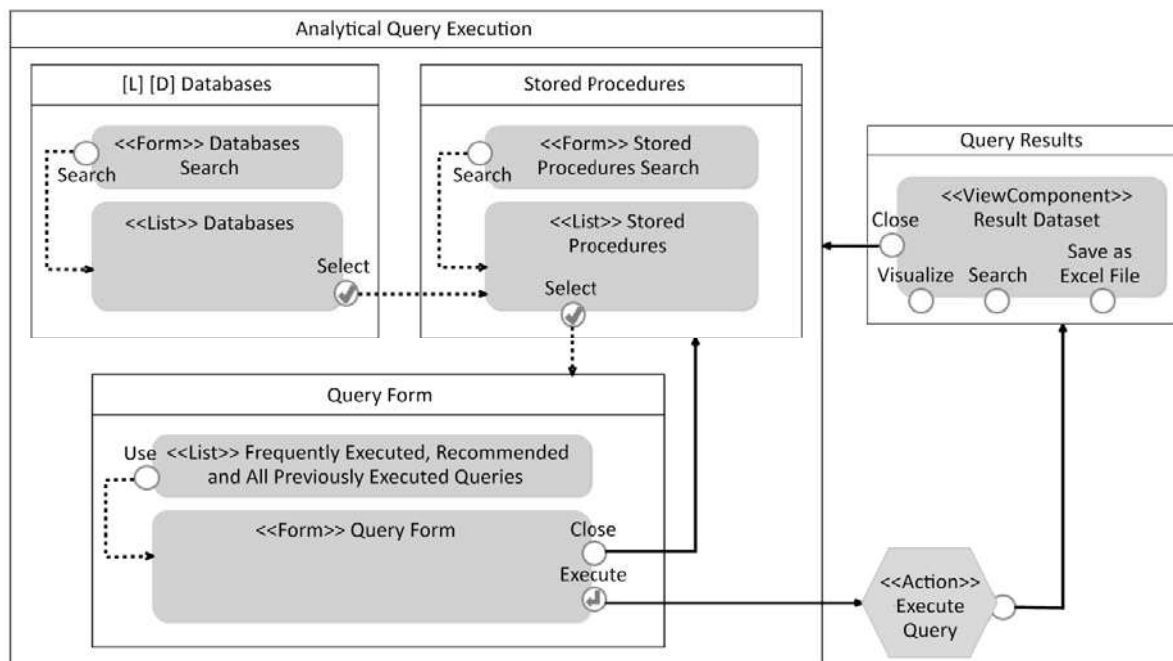


**Figure 9:** Interaction flow model of the interactive adaptive system for data analytics



Obtained result dataset allows search over, export to Microsoft Excel, or visualization using graphs and charts that fit considered data.

## 5. Results and Discussion

### 5.1. Software Architecture

The software prototype was implemented according to design principles of ad-hoc analytical queries support (Fig. 8) and user interaction behavior (Fig. 9) outlined in previous section. The software design is demonstrated in Fig. 10 using components deployment diagram.

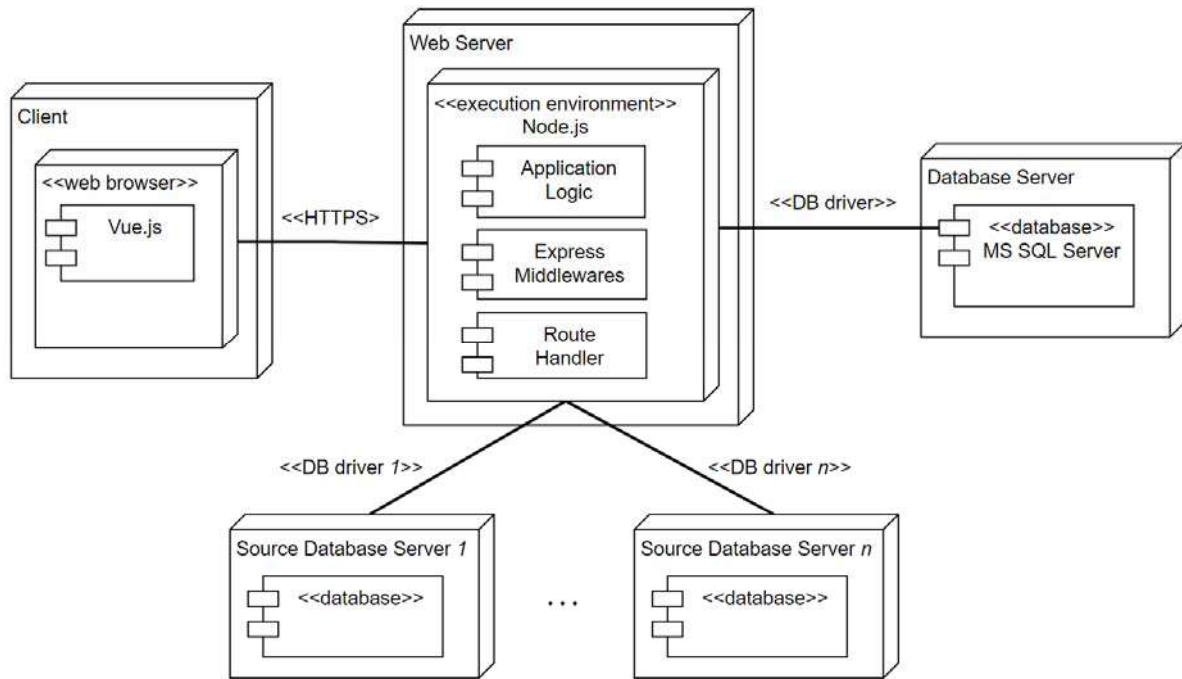


**Figure 10:** Software architecture model of the interactive adaptive system for data analytics

As it is shown in Fig. 10, the software prototype should be implemented using the three-tier client-server architecture consisting of the following nodes and components:

- Client tier containing only the web-browser as a “thin” client that provides user interface built using the Vue.js JavaScript framework.
- Application server tier containing Node.js back-end application that implements business logic and handles HTTP (HyperText Transfer Protocol) requests, and responses to interact with the client side. Application server uses database drivers to interact with database servers.
- Database server containing Microsoft SQL Server database that stores user profiles and granted permissions, data source connection properties, analytical queries, and usage statistics.

At this moment developed software prototype supports connection to Microsoft SQL Server as the source database server, while it is planned to extend its integration capabilities to support other widely used relational DBMS, such as MySQL or PostgreSQL [28].

### 5.2. Software Prototype Demonstration

The software prototype currently is under development, however, its generic functionality is already implemented. Fig. 11 demonstrates querying interface that works with connected database servers and stored procedures hosted on these database servers. The user interface is developed according to ad-hoc data querying principles and information flows outlined in CMMN (Fig. 8) and IFML (Fig. 9) models respectively. Generic GUI areas are:

- Database connection properties.
- Stored procedures displayed for each of attached databases.
- Recommended queries displayed for each of stored procedures.
- Querying form that could be filled by users manually for a certain stored procedure or could be filled automatically based on suggested relevant queries.

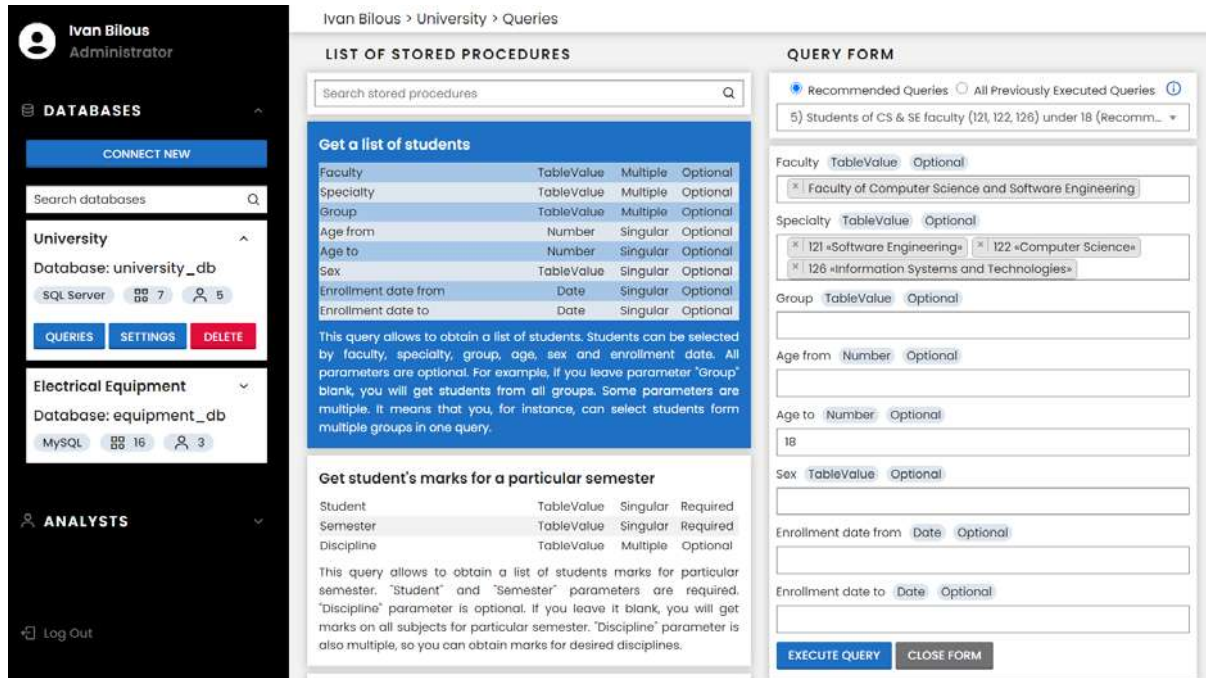


**Figure 11:** Interactive adaptive querying interface of the software prototype

As it is shown in Fig. 11 above, users first of all should select a database to work with. If necessary, new database connection could be created (this feature shall be used only by DBA and will be invisible for users). It is planned to allow using multiple relational DBMS in the same environment (e.g. MySQL, Microsoft SQL Server, Oracle, and PostgreSQL as the most popular and widely used database systems). However, connected database systems should be of client-server architecture, since using of file-server databases, such as Microsoft Access or SQLite, is not planned, since they usually do not support stored procedures and used as embedded or small ledgers. Database connection properties could be configured by DBA or even removed if necessary. Also DBA can manage stored procedures and users assigned to each of database connection profiles.

When a certain database connection selected, the list of stored procedures available for this database is displayed. Besides the meaningful name that hides a system name of the stored procedure, the list of parameters is outlined for each of stored procedures together with its brief textual description. Each of demonstrated parameters is accompanied with the respective data type, multiplicity, and optionality.

For each of stored procedures users can get recommended queries suggested using recommending procedure introduced in section 4.1 (Fig. 4). The querying form is available for manual input, however, when user selects specific item from the list of recommended queries, the form is filled with respective values of filtering parameters. Such feature simplifies significantly repetitive access to certain data sets with minimum efforts required from the user (when no changes are necessary and recommended query could be executed as-is, or when one or two filtering values should be changed for “what-if” analysis or other ad-hoc purposes). The list of recommended queries includes three most popular ones, ordered by their usage frequencies, and then six queries – two most similar for each of three most popular ones, ordered by their similarity values. However, users may expand the list of suggested queries in order to access all of the relevant queries (beyond the threshold of two most similar queries for each of three most frequently used queries) ordered by their similarity values (to

three most popular queries). In order to distinguish obtained recommendations, queries should be identified by their names given by users when these queries are executed for the first time.

## 6. Conclusion and Future Work

In this paper we have proposed an approach to development of interactive adaptive software tool for data analysis based on the content-based filtering approach. Outlined recommending procedure captures usage statistics of database queries and suggests the most relevant ones among previously used queries based on usage frequency and similarity criteria. According to the proposed recommending procedure, the system works with SQL queries used to access attached data sources (we use stored procedures or routines as data sources, since they are representing a good practice of database schema securing and encapsulation for usage by multiple applications and clients). These SQL queries that call data sources are treated as documents, while pairs of parameter names and values are used as terms for content-based filtering. Suggested queries as the most relevant to a particular stored procedure then could be used as given with pre-defined parameter values or partially changed parameter values in order to reuse queries executed before, or run some ad-hoc inquiries respectively. The software tool that should be developed according to the proposed approach currently is implemented as a prototype with limited functionality. However, it allows to connect Microsoft SQL Server databases to access provided data sources, execute SQL queries, and receive recommendations. In future the software should be completed to fulfill all of the elicited requirements, such as user roles and permissions, and connection to multiple DBMS servers. It is also planned to elaborate recommending procedure by considering advanced filtering techniques, similarity measures, and thresholds in order to provide suggestions adjustable by users.

## 7. References

- [1] SHRM Survey Findings, Jobs of the Future: Data Analysis Skills, 2016. URL: <https://www.shrm.org/hr-today/trends-and-forecasting/research-and-surveys/Documents/Data-Analysis-Skills.pdf>.
- [2] SelectHub, Competitive report: Tableau vs. QlikView vs. Power BI, 2019. URL: <https://www.smetrinsicinsights.com/wp-content/uploads/2019/05/Tableau-VS-QlikView-VS-Power-BI-2019-Update.pdf>.
- [3] J. W. Nicklas et al., Supporting distributed, interactive Jupyter and RStudio in a scheduled HPC environment with Spark using Open OnDemand, in: Proceedings of the Practice and Experience on Advanced Research Computing, PEARC'18, 2018, pp. 1–8. doi: 10.1145/3219104.3219149.
- [4] F. Nelli, Python Data Analytics: With Pandas, NumPy, and Matplotlib, Apress, Berkeley, CA, 2018. doi:10.1007/978-1-4842-3913-1.
- [5] L. Tang et al., Dynamic query forms for database queries, IEEE transactions on knowledge and data engineering 9(26) (2013) 2166–2178. doi:10.1109/TKDE.2013.62.
- [6] P. P. Nikam, A Review on Dynamic Query Forms for Database Queries, International Journal of Computer Science and Information Technologies 5(6) (2014) 8079–8081.
- [7] V. Jadhav, A. Priyadarshi, A Survey on Database Queries by using Dynamic Query Forms, International Journal of Science and Research 3(11) (2014) 1157–1159.
- [8] S. S. Baravkar, A. Gupta, A Survey on Dynamic Query Forms for Database Queries, International Journal of Science and Research 3(6) (2017) 1566–1568.
- [9] M. M. Jisha, M. A. Jacob, Dynamic Query Forms for Database Queries, International Journal of Engineering Research & Technology 5(7) (2016) 217–223.
- [10] K. L. Jaiswal, S. S. Joshi, Dynamic Query Form Generation, International Journal of Emerging Technologies in Engineering Research 5(4) (2017) 95–97.
- [11] P. Dagade, M. Bhonsle, Espionage on Search Optimization using Dynamic Query Form, International Journal of Engineering Research and General Science 2(6) (2014) 380–386.
- [12] G. Patle, R. Uikey, E. Meshram, Ranked Based Dynamic Query Forms for Database Queries, International Journal of Scientific Research in Computer Science, Engineering and Information Technology 5(2) (2019) 1209–1212.

- [13] A. V. S. Kumar, O. Devakiran, Dynamic Query Forms for Ad-Hoc Queries on Databases, *International Journal of Scientific Engineering and Technology Research* 3(36) (2014) 7176–7179.
- [14] G. H. K. Reddy, A. Bhattacharjya, S. S. Rawat, S. Reddy, Multiple Methods for Detection of User Priority of Dynamic Query Forms on Databases, *International Journal of Scientific Engineering and Technology Research* 4(31) (2015) 6005-6008.
- [15] R. B. Sangore, P. P. Bhavsar, M. S. Patil, T. M. Chaudhari, An Alternative For Database Queries: Auto Query Forms, *International Journal of Advanced Research in Computer and Communication Engineering* 4(3) (2015) 94–96. doi:10.17148/IJARCCE.2015.4322
- [16] G. Radhakrishnan, S. S. Babu, Creation of Dynamic Query Forms and Ranking of its Components based on User's Preference, *International Journal of Innovative Research in Computer and Communication Engineering* 3(8) (2015) 7326–7332. doi:10.15680/IJIRCCE.2015.0308028
- [17] G. S. G. Pavan Kumar, S. U. Maheswara Rao, Generating Efficiency and Robustness Dynamic Query Forms for Advanced Database Queries, *International Journal of Research in Information Technology* 4(9) (2016) 42–51.
- [18] K. Srinivasarao, G. R. Bharathi, Automated Creation of a Database Queries by using Dynamic Query Forms, *International Journal of Scientific Engineering and Technology Research* 5(4) (2016) 0609–0612.
- [19] K. Ozarkar, R. Rajani, Optimization Technique for Efficient Dynamic Query Forms with NoSQL, *International Journal of Science and Research* 3(11) (2014) 2041–2044.
- [20] D. Pagaro et al., Dynamic Query Forms for Non-Relational Database, *International Journal of Advanced Engineering, Management and Science* 2(6) (2016) 552–555.
- [21] P. S. Revathy, A survey paper on dynamic query forms for database queries, *International Journal of Advance Research in Computer Science and Management Studies* 2(9) (2014) 65–69.
- [22] J. M. Jambukar, M. B. Vaidya, A Review paper on Construction of Query Forms, *International Journal of Computer Science and Information Technologies* 6(3) (2015) 2426–2428.
- [23] F. Ricci, L. Rokach, B. Shapira, B. P. Kantor, *Recommender Systems Handbook*, Springer US, 2011. doi:10.1007/978-0-387-85820-3
- [24] R. Manjula, A. Chilambuchelvan, Content Based Filtering Techniques in Recommendation System using user preferences, *International Journal of Innovations in Engineering and Technology* 7(4) (2016) 149–154.
- [25] J. M. Pazzani, D. Billsus, Content-Based Recommendation Systems, in: Brusilovsky P., Kobsa A., Nejdl W. (Eds.), *The Adaptive Web. Lecture Notes in Computer Science*, Springer, Berlin, Heidelberg, 2007, pp. 325–341. doi:10.1007/978-3-540-72079-9\_10
- [26] S. Gupta, Overview of Text Similarity Metrics in Python, 2018. URL: <https://towardsdatascience.com/overview-of-text-similarity-metrics-3397c4601f50>.
- [27] P. Sitikhu, K. Pahi, P. Thapa, S. Shaya, A Comparison of Semantic Similarity Methods for Maximum Human Interpretability, 2019. URL: <https://arxiv.org/pdf/1910.09129.pdf>.
- [28] DB-Engines Ranking. URL: <https://db-engines.com/en/ranking>.
- [29] SysML FAQ: What is a Requirement Diagram and how is it used? URL: <https://sysml.org/sysml-faq/what-is-requirement-diagram.html>.
- [30] Case Management Model and Notation. URL: <https://www.omg.org/cmmn/>.
- [31] Scope – IFML: The Interaction Flow Modeling Language. URL: <https://www.ifml.org/scope/>.

# Synthesis of Spherical Object Configurations: Models and Information Technologies

Oleksii Kartashov<sup>a</sup>, Kyryl Korobchynskiy<sup>a</sup>, Iryna Yakovleva<sup>b</sup>, Olga Yarovaya<sup>a</sup>  
and Georgiy Yaskov<sup>c</sup>

<sup>a</sup> National Aerospace University "Kharkiv Aviation Institute", 17 Chkalova Street, Kharkiv, 61070, Ukraine

<sup>b</sup> National University of Urban Economy in Kharkiv, 17 Marshal Bazhanov Street, Kharkiv, 61002, Ukraine

<sup>c</sup> A.N. Podgorny Institute for Mechanical Engineering Problems of the National Academy of Sciences of Ukraine

## Abstract

The problem of synthesis of optimal configurations of spherical objects is considered. The basis is a general approach to the formalization of the configuration space of geometric objects. The analysis of models is carried out, their typology and information technologies for structural synthesis of complex spherical objects are offered. Approaches to the development of new information technologies for solving problems based on their generalized information-analytical model are proposed. Data is transformed into an object-oriented structure for visualization of spherical configurations.

## Keywords 1

Spherical object, configuration, analytical model, information technology

## 1. Introduction

The term configuration is understood as appearance, outline and also the relative positioning of objects and their parts. In mathematics, as a rule, the configuration means some placement of points in space. The problem of synthesis of spatial configurations will be considered as the definition of such a property of geometric objects that would satisfy the given relations (constraints, properties) and would deliver an extremum to a certain quality criterion.

In this article, we confine ourselves to a class of geometric objects of spherical form. However, the results obtained can be generalized to broader classes of spatial objects. A significant number of publications testify to the importance of the problems of spatial synthesis of spherical objects. The article [1] presents an overview of the works devoted to this topic, containing more than 60 references. In [2] - [10] a number of different approaches are considered for the search for optimal spherical configurations. Papers [11] - [16] describe practical applications of problems of placing spherical objects with additional constraints. It should be noted that when considering this class of problems, much attention is paid to mathematical modeling of systems with spherical objects and the development of special methods for nonlinear optimization to solve them. Issues related to information support, analysis of the data structure and ways to transform geometric information in the optimization process are not fully covered.

The purpose of this paper is to describe the possibilities of using modern information technologies when considering the investigated class of problems based on existing mathematical models and analysis of optimization methods.

---

CMIS-2021: The Fourth International Workshop on Computer Modeling and Intelligent Systems, April 27, 2021, Zaporizhzhia, Ukraine  
EMAIL: alexeykartashov@gmail.com (O. Kartashov); kirill.korobchinskiy@gmail.com (K. Korobchynskiy); yakovlevairine@gmail.com

(I. Yakovleva); helga.yarovaya@gmail.com (O. Yarovaya); yaskov@ukr.net (G. Yaskov)

ORCID: 0000-0002-6282-553X (O. Kartashov); 0000-0002-3676-6070 (K. Korobchynskiy); 0000-0001-5204-9779 (I. Yakovleva),

0000-0003-4969-1677 (O. Yarovaya), 0000-0002-1476-1818 (G. Yaskov)



© 2021 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

This paper proposes an information technology that connects into a single system modules for solving an optimization problem and solution visualizing. The decision maker is included in the system. Using the results of the decision, especially in its visual representation, the decision-maker can change the parameters of the decision, relying on his experience and intuition. Such transformed solutions are used in the next step as a starting point for finding a local extremum. This approach is because the problems under consideration are multi-extreme and have a large number of local extrema. The suggested method is one of the ways to improve the local solution.

## 2. Geometric information and configuration space of spatial objects

Investigations of configurations as mathematical objects are naturally associated with the notion of configuration space. The configuration space determines the configuration of the system, that is, the set of values of geometric variables called generalized coordinates and sets the location in the space of some system and its parts both relative to one another and with respect to a given fixed point.

In the papers [17] - [19] the concepts of geometric information and geometric objects configuration space are introduced. Geometric information  $\mathbf{g} = (\{s\}, \{\mu\}, \{p\})$  about an object  $S \subset R^3$  includes a spatial form  $\{s\}$  as an equivalence class on a set of point sets; metric parameters of the form  $\{\mu\} = (\mu_1, \dots, \mu_k)$ , specifying the dimensions of the object; placement parameters  $\{p\} = (p_1, \dots, p_l)$ , that determine the position of an object  $S$  in the space  $R^3$ .

In accordance with the general concept of constructing such spaces, we define their structure in the following way. The basis for specifying the components  $\{s\}$  and  $\{\mu\}$  of geometric information  $\mathbf{g}$  is the equation of its boundary

$$f(\xi, \mu) = 0, \quad \xi \in R^3, \quad (1)$$

where the variables  $(\xi, \mu)$  have a domain of admissible values  $D \subseteq R^k$ , and the function  $f(\xi, \mu)$  is defined and continuous everywhere on  $R^3 \times D$ .

Depending on the analytic type of function  $f(\xi, \mu)$  various classes of objects can be formed. At the same time, from a practical point of view it is natural to distinguish classes of models of the most widespread material objects. These objects are called basic.

Different constraints may be imposed on the values of generalized variables in the configuration space  $\Xi(S_B)$ . These restrictions will form the domain of admissible configurations in the configuration space for the class of tasks under consideration. On the domain of admissible values of generalized variables may be imposed other additional restrictions that arise when solving specific classes of tasks, for example, when searching for optimal configurations in a certain sense. We divide the following restrictions into the following groups: restrictions associated with the fixation of some generalized variables, restrictions that arise on the mutual location of geometric objects, and the restrictions imposed by certain physical or mechanical properties of the synthesized system. If to give on a set of geometric objects some relations, binary in particular, we obtain different classes of placement configurations.

If for any pair of objects a non-overlapping binary relation must be performed, that is, these objects do not have common internal points, then the corresponding configuration of the placement is called the packing configuration of geometric objects. Let us note that in most practical packaging tasks an additional object is called a container. All geometric objects that are considered, must belong to the container. Note that the generalized variables in the configuration space may be subject to additional restrictions that generate special classes of configurations [19]. A distinctive feature of such configurations is the presence of restrictions on the minimum and the maximum permissible distances between objects.

In the case when geometric objects are solids with given masses, a balanced system of such bodies gives the balanced packing configuration. Note that the configuration spaces both layout and balanced packing configurations are the same.

We indicate also the covering configurations [20], [21] and partitioning configurations [22], [23]. A special class of configurations are Euclidean combinatorial configurations that are formed in the case when the placement parameters and metric parameters of the objects take discrete values. Such configurations have a number of important properties, the implementation of which allows us to propose new approaches to modeling and solving optimization problems [24], [25].

### 3. Synthesis of spherical object configurations

In this paper, we will examine the class of geometric objects of a spherical shape. Three groups of spherical objects will be considered: two-dimensional (2D), three-dimensional (3D) and multidimensional (nD) objects. Each of the basic objects is given by the equation of its boundary. Then the generalized variables of the configuration space of spherical objects are their radii and the coordinates of the centers.

Consider the spherical object  $S \subset R^n, n=2,3,\dots$  of the radius  $r>0$ . We associate with  $S$  the moving coordinate system by choosing its beginning (the pole of the sphere) in the center of its symmetry. We define the spatial shape  $\{s\}$  of the sphere by the equation of its boundary

$$f(\xi, r) = 0, \xi \in R^n, \quad (2)$$

which, in accordance with the dimension of space, has the form:

$$f(\xi, r) = -x^2 - y^2 + r^2, \xi = (x, y), \quad (3)$$

$$f(\xi, r) = -x^2 - y^2 - z^2 + r^2, \xi = (x, y, z), \quad (4)$$

$$f(\xi, r) = -\sum_{i=1}^n \xi_i^2 + r^2, \xi = (\xi_1, \dots, \xi_n). \quad (5)$$

If  $S \subset R^3$ , then  $\mu = r$  is a metric parameter, and we denote  $p = (v_1, v_2, v_3)$  location parameter. The configuration space of a sphere is given by generalized coordinates  $v_1, v_2, v_3, r$ , and the equation of the general position of the sphere has the form

$$F(x, y, z, v_1, v_2, v_3, r) = r^2 - (x - v_1)^2 - (y - v_2)^2 - (z - v_3)^2 = 0. \quad (6)$$

Suppose  $\Omega = \{S_1, \dots, S_n\}$  is a set of basic spherical objects. We assign the structure of a complex object in the form

$$S_B = B(S_1, \dots, S_n) = \bigcup_{i=1}^n S_i. \quad (7)$$

An object  $S_B$  whose structure is given by the expression (7), we call a composite object.

Based on the general approach and the proposed general classification, we propose the *typology* of spherical objects configuration.

In the configuration spaces of spherical objects  $S_i$  and  $S_j$  with corresponding generalized variables  $g^i = (x_i, y_i, z_i, r_i)$  and  $g^j = (x_j, y_j, z_j, r_j)$  the constraints of non-overlap have the form

$$(r_i + r_j)^2 - (x_i - x_j)^2 - (y_i - y_j)^2 - (z_i - z_j)^2 \geq 0. \quad (8)$$

The condition for the contain of a sphere  $S_i(g^i)$  into a spherical container  $S_0(g^0)$  with generalized variables  $g^0 = (x_0, y_0, z_0, r_0)$  will be written as

$$-(x_0 - x_i)^2 - (y_0 - y_i)^2 - (z_0 - z_i)^2 + (r_0 - r_i)^2 \geq 0, \quad (9)$$

where  $x_0 = y_0 = z_0 = 0$ .

Note that the generalized variables  $g^0, g^1, \dots, g^n$  in the configuration space  $\Xi(S_0) \times \Xi(S_B)$  may be subject to additional constraints that generate special classes of packing configurations. First of all, it is about layout configuration and balanced packing configuration.

A distinctive feature of such configurations is the constraints on the minimum and the maximum admissible distances between objects. In this case, the conditions will be

$$d_{ij}^{min} \leq (x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2 - (r_i + r_j)^2 \leq d_{ij}^{max}, \quad (10)$$

where  $d_{ij}^{min}, d_{ij}^{max}$  are corresponding distances between objects.

In the case when geometric objects  $S_i, i \in J_n$  are solids and have masses  $m_i, i \in J_n$ , a balanced system of such bodies defines the balanced packing configuration. If the poles of the objects  $S_i, i \in J_n$  coincide with the centers of their masses, balanced packing takes place under condition

$$\sum_{i=1}^n x_i m_i = 0, \sum_{i=1}^n y_i m_i = 0, \sum_{i=1}^n z_i m_i = 0. \quad (11)$$

Note that the configuration spaces  $\Xi(S_0) \times \Xi(S_B)$  both layout and balanced packing configuration are the same.

The generalization of the spherical objects placement problem is the layout and packing of composite spheres. For each component of composite sphere (object of the first order) whose structure has the form (1), we fix the distances between the poles of the base spheres  $S_i, S_j, i, j \in J_n, i < j$  (objects of zero order). This will create an additional rigid system of restrictions on the placement parameters  $p^i = (x_i, y_i, z_i), i \in J_n$  of the base spheres

$$(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2 = \rho_{ij}^2. \quad (12)$$

Depending on the possible affine transformations of complex objects (congruence, translation, rotation, etc.) the placement options may be subject to additional restrictions.

To formalize of non-overlap conditions of the components of spherical objects, the decomposition of such conditions is naturally performed on non-overlap of base spheres for various composite objects.

Let a symmetric matrices  $B = [b_{ij}]_{n \times n}$  and  $B^0 = [b_{ij}^0]_{n \times n}$  be given with elements

$$b_{ij} = \begin{cases} 1, & \text{if } S_i * S_j; \\ 0, & \text{otherwise,} \end{cases} \quad b_{ij}^0 = \begin{cases} 1, & \text{if } S_i \in S_j \\ 0, & \text{otherwise,} \end{cases} \quad (13)$$

where  $S_i * S_j$  means non-overlapping  $S_i$  and  $S_j$ .

The configuration of the spheres  $S_i, i \in J_n$  is said to be admissible if the inequalities (8) holds for any  $i, j \in J_n, i < j$  such that  $b_{ij} = 1$ . The configuration of the spheres  $S_i, i \in J_n$  is said to be valid if it is included, if the inequality (9) holds for any  $i, j \in J_n, i < j$  such that  $b_{ij}^0 = 1$ .

The vast majority of modern publications are devoted to the tasks of the layout and packaging of objects, the metric parameters are fixed, with the exception of the container. Introduction of the configuration spaces of geometric objects allows us to investigate fundamentally new, practically important problems from the general positions, one of the important classes of which is the problem of synthesis of optimal configurations of spherical objects.

Suppose there is a certain function on the set of permissible configurations of the spheres, which we call the quality criterion of the configuration. Then the problem of synthesis of optimal configurations of spherical objects is to determine the generalized variables of the configuration space, which has a lot of geometric interpretations and practical applications.



Select a set of numbers  $L' \subset J_n$ . Spheres  $S_i, i \in L'$  are called prohibition zones and fix their position in  $R^3$ , putting

$$x_i = \hat{x}_i, y_i = \hat{y}_i, z_i = \hat{z}_i, i \in L'. \quad (14)$$

We will form a matrix  $B = [b_{ij}]_{n \times n}$  with elements

$$b_{ij} = \begin{cases} 0, \forall i, j \in L'; \\ 1, \text{ otherwise.} \end{cases} \quad (15)$$

A set of admissible configurations will be given by a system of inequalities

$$(r_i + r_j)^2 - (x_i - x_j)^2 - (y_i - y_j)^2 - (z_i - z_j)^2 \geq 0, \quad (16)$$

where  $i \in J_n \setminus L', j \in J_n$ .

Let a quality criterion, which need to be minimum, is

$$\varphi_1(x_1, y_1, z_1, r_1, \dots, x_n, y_n, z_n, r_n) = \max_{i \in J_n \setminus L'} \{x_i + r_i, y_i + r_i, z_i + r_i\}. \quad (17)$$

If  $L' = \emptyset$ , we have a classic problem of placing unequal spheres in a sphere of minimal radius. In this case, we can consider a container  $S_0$  as a sphere of variable radius  $r_0$  with the placement parameters  $x_0 = y_0 = z_0 = 0$ . Then

$$\varphi_1(x_0, y_0, z_0, r_0, x_1, y_1, z_1, r_1, \dots, x_n, y_n, z_n, r_n) = r_0. \quad (18)$$

New classes of problems of synthesis of optimal configurations of spherical objects are considered, generalizing the classical problems of packing and arrangement in the case of variables of metric parameters of objects (radii of spheres).

We fix the generalized variables

$$x_0 = y_0 = z_0 = 0, r_0 = \hat{r}_0, x_i = \hat{x}_i, y_i = \hat{y}_i, z_i = \hat{z}_i, i \in L', \quad (19)$$

in the configuration space  $\Xi(S_B) = \Xi(S_0) \times \Xi(S_1) \times \dots \times \Xi(S_n)$ .

Then, depending on the choice of specific constraints, we describe the appropriate set of admissible configurations. In turn, the variety of quality criteria of the configurations expands. For example, we will consider such a criterion as the maximum of occupied part of a container with variable radii of spheres:

$$\varphi(r_1, \dots, r_n) = \frac{4}{3} \pi \sum_{i \in J_n \setminus L'} r_i^3. \quad (20)$$

In the configuration space  $\Xi(S_B) = \Xi(S_1) \times \dots \times \Xi(S_n)$  we fix the placement parameters  $x_i = \hat{x}_i, y_i = \hat{y}_i, z_i = \hat{z}_i$  of spheres  $S_i, i \in J_n$ .

Let  $\varphi(r_1, r_2, \dots, r_n)$ - arbitrary convex function on  $R_+^n$ . Then we have the problem of convex programming

$$\varphi(r_1, r_2, \dots, r_n) \rightarrow \min \quad (21)$$

with linear constraints

$$r_i + r_j - \rho_{ij} \geq 0, i, j \in J_n, i < j, \quad (22)$$

where

$$\rho_{ij} = \sqrt{(\hat{x}_i - \hat{x}_j)^2 + (\hat{y}_i - \hat{y}_j)^2 + (\hat{z}_i - \hat{z}_j)^2}, i, j \in J_n, i < j. \quad (23)$$

Let a given matrix  $\mathbf{C} = [c_{ij}]_{n \times n}$  whose elements determine the number of connections between the spheres  $S_i$  and  $S_j$ ,  $i, j \in J_n$ ,  $i < j$ . Then the total length of the network connecting the objects is given by the function

$$\varphi(x_1, y_1, z_1, r_1, \dots, x_n, y_n, z_n, r_n) = \sum_{i \in J_n} \sum_{\substack{j \in J_n \\ j > i}} c_{ij} \rho(S_i, S_j), \quad (24)$$

where

$$\rho(S_i, S_j) = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2} - r_i - r_j. \quad (25)$$

Let the spheres have fixed radii and represent a solid density body. We minimize the function that characterizes the deviation of the center of the masses of the combination of bullets from a given point. We have

$$\varphi(x_1, y_1, z_1, r_1, \dots, x_n, y_n, z_n, r_n) = \sum_{i \in J_n} \left[ \left( x_0 - \frac{m_i x_i}{m_\Sigma} \right)^2 + \left( y_0 - \frac{m_i y_i}{m_\Sigma} \right)^2 + \left( z_0 - \frac{m_i z_i}{m_\Sigma} \right)^2 \right], \quad (26)$$

where  $m_i = 4/3 \pi v r_i^3$ ,  $i \in J_n$ ,  $m_\Sigma = \sum_{i \in J_n} m_i$ .

Thus, the formalization of problems of synthesis of optimal configurations of spherical objects allows us to conclude that depending on the quality criteria and the choice of specific constraints that form the set of admissible configurations, such tasks can be attributed to the corresponding class of mathematical programming problems.

#### 4. Informational technology of synthesis of spherical object configurations

The paper proposes a unified approach for solving all types of problems of synthesis of spherical objects under consideration. This approach is based on information technology, which includes the interaction of several interconnected units. Each of these units solves its own specific tasks. In our work, 4 such units can be distinguished:

- optimization;
- visualization;
- consolidated data warehouse;
- modeling and control.

The optimization unit as a basis contains a solver that allows you to effectively solve nonlinear optimization problems to obtain local extrema. In our work, we use the IpOpt software package for this purpose. It uses interior point techniques for high-dimensional NLP problems. Freely distributed, has open source code and does not require obligations when creating commercial software.

The visualization unit is an interactive graphical interface for working with 3D graphics. For this purpose, the work uses software from Autodesk 3ds MAX, which is provided under a free license. To automate the display of data, programs are used - scripts written in the built-in 3ds MAX Scripts language. They are executed in the environment of the internal Listener code interpreter, which renders the objects. Thanks to this unit, you can display the obtained solutions, interactively change their parameters or create a new problem using a graphical interface.

The need for a consolidated data warehouse is caused by the following. When using traditional tools, there is redundancy in data volumes, and when processing significant resources, they are converted. Thus, the data obtained from a variety of sources and different types of information

resources integrated into the system must be consolidated into an adequate information storage model. For this we used ORM (Object Relational Mapping), and to transform ADO.NET Entity Framework Code First into a repository strategy, and to save the results, we used a portable SQLite database.

The modeling and control unit is the core of the entire system. It is implemented in C++ language. In it, using polymorphic classes, information models of both spherical objects of varying degrees of complexity and problems of synthesis of configurations of all considered types are built, stored and transformed. In addition, this unit provides interaction between all units of the system and automatically converts the internal representation of the problem into the form required for an optimization, visualization or data store units.

The model of information technology for synthesis of spatial configurations of spherical objects is shown in Fig. 1.

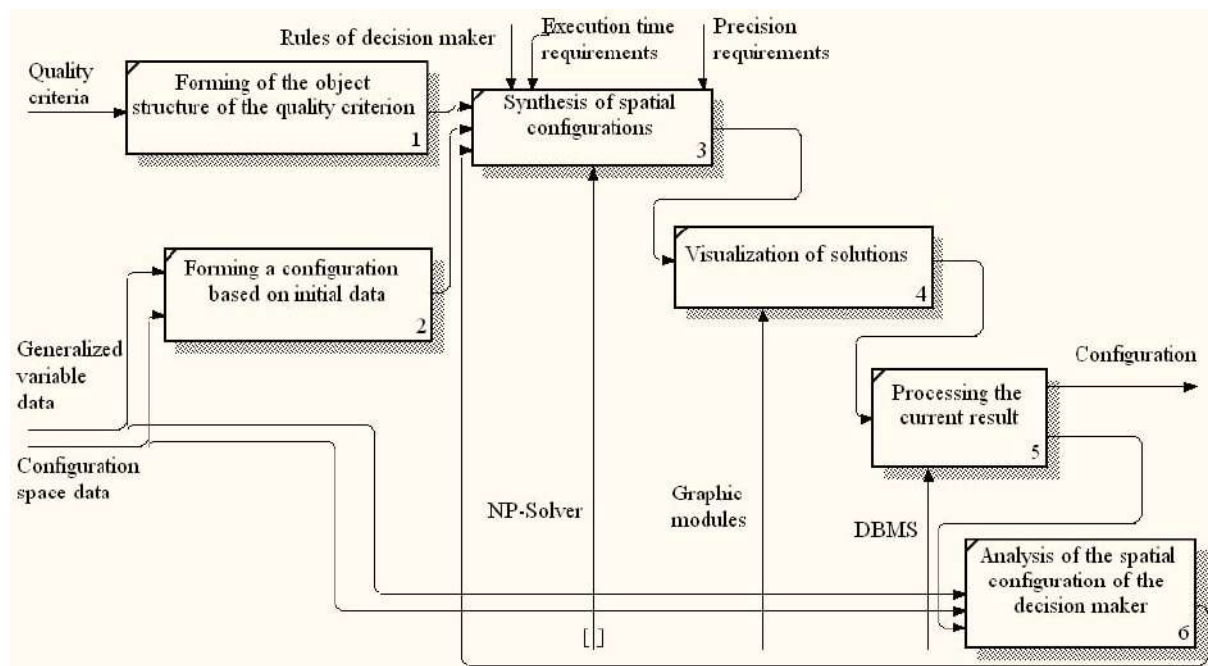


**Figure 1:** The model of information technology for synthesis of spatial configurations

The expediency of this approach is due to the following. All the problems of the considered classes of synthesis of spatial configurations are NP-complex complicated ones, which have a high dimension (in real systems with more than 1000 variables), are multiextreme (with a non-linear number of local extrema). For many of them it is possible to find the local extremum using standard classical methods. One of the ways to improve a local solution, based on the proposed information model, consists in a dialogue between a solver and a decision maker. The dialog is based on a graphical display of the solver result. The decision maker interactively modifies the received locally optimal configuration and sends it as an initial placement to the solver. This dialogue can continue until the decision-maker stops the process. All obtained locally optimal solutions are stored in the database and are the basis for the final selection. This process does not preclude the use of automatic global search methods, for example, an genetic algorithms [26, 27].

Process of synthesis of optimal spherical configurations including the participation of a decision maker is shown on Fig. 2.

The inclusion of a decision maker in the process will also allow solving problems of the types under consideration with additional constraints that are difficult to formalize. In such cases, only the decision maker can choose the final outcome. Consider as an example the spherical objects placement problem. It is necessary to find the minimum spherical shell for a set of spheres of different radii. Starting point is shown on Fig. 3 and a locally optimal solution obtained by the solver given on Fig. 4.

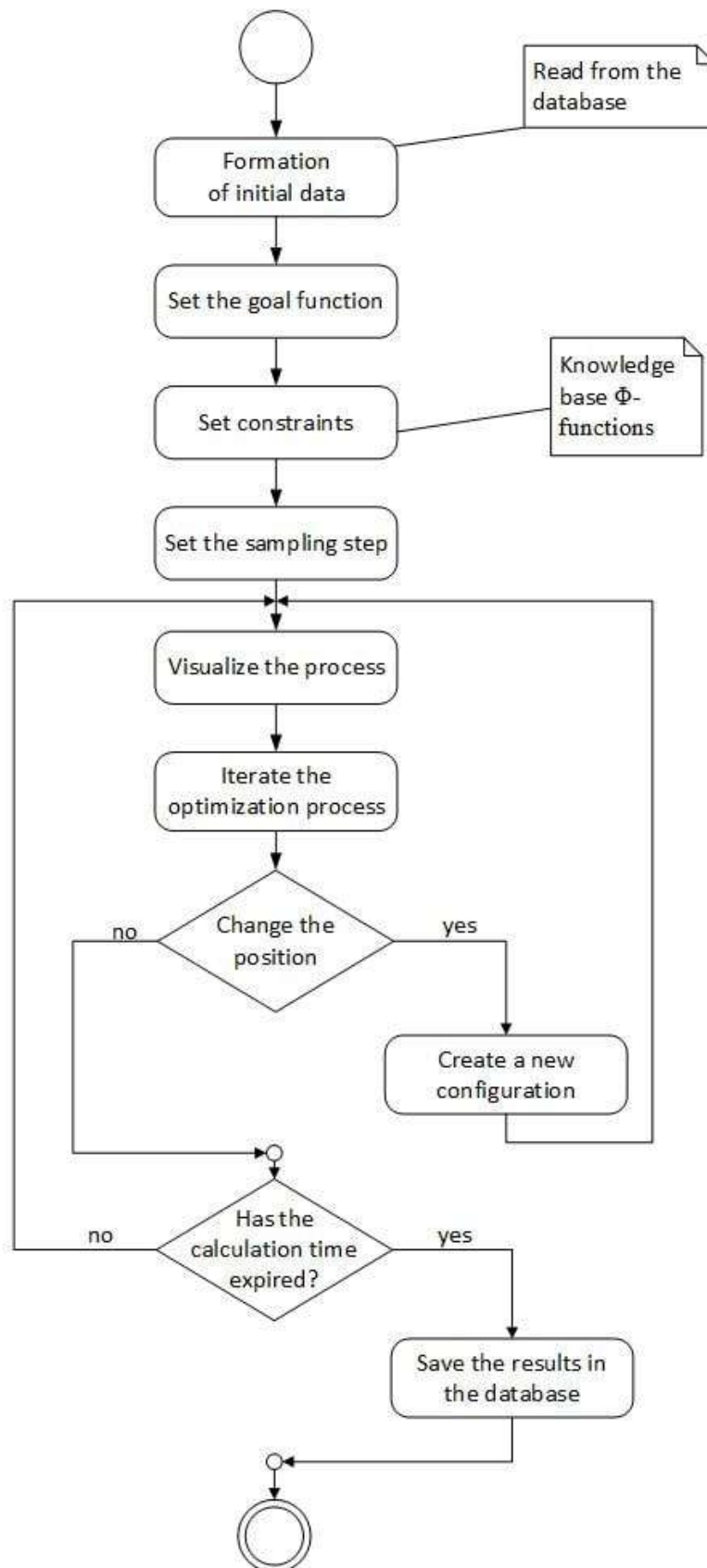


**Figure 2:** UML diagram of a process of synthesis of optimal spherical configurations including a decision maker

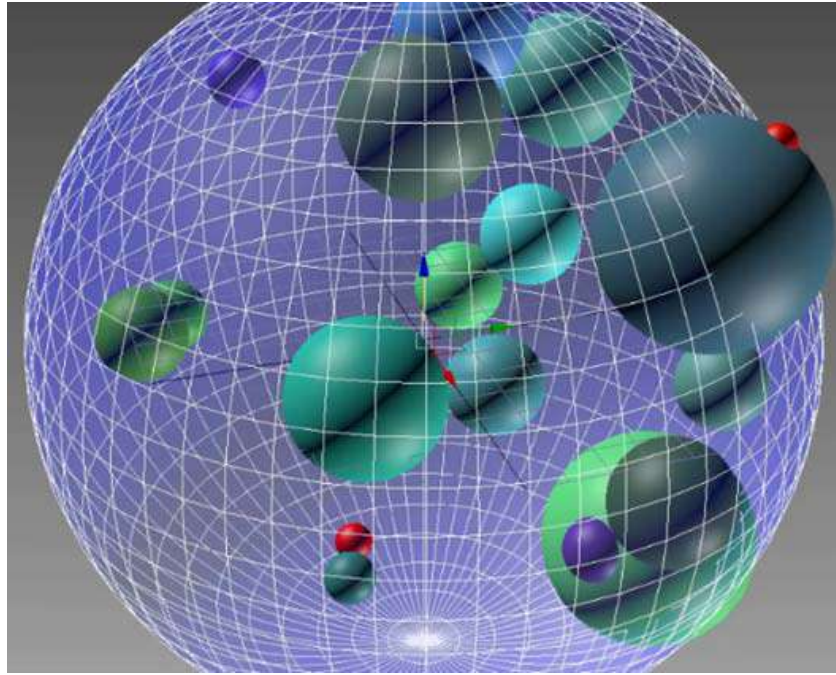


**Figure 3:** Starting point

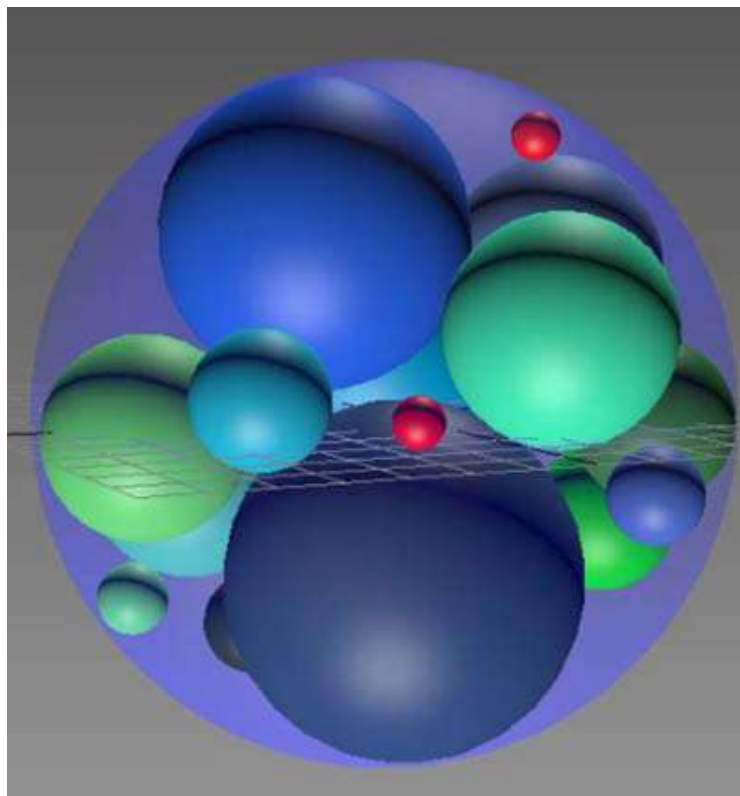


**Figure 4:** Locally optimal solution obtained by the Solver

The decision maker changed the received decision and sent it back to the Solver (Fig. 5). A new locally optimal solution is shown on Fig. 6.

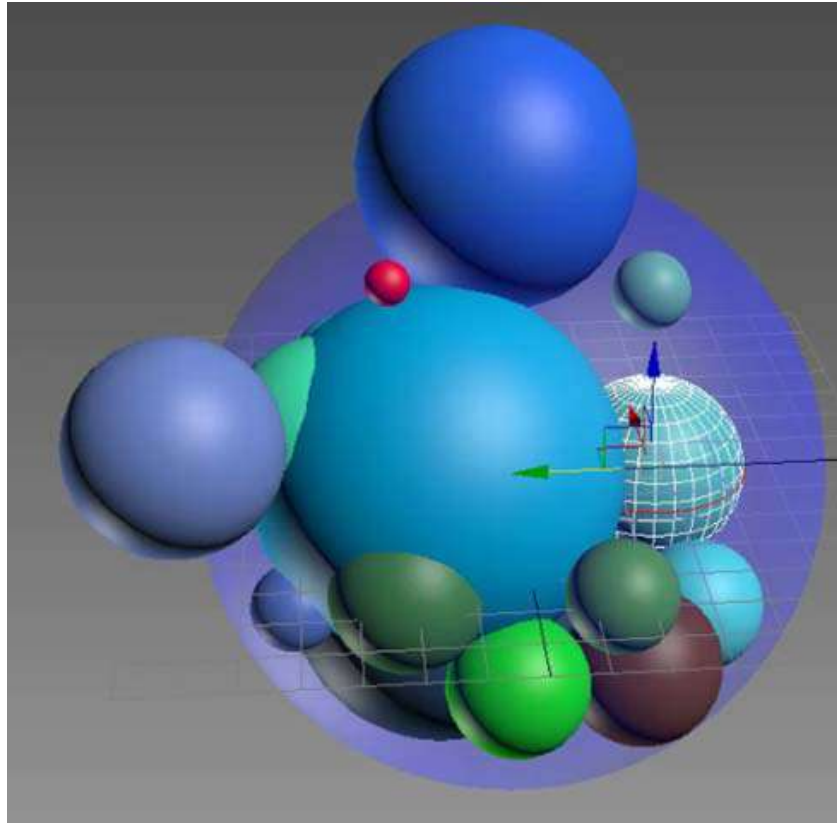


**Figure 5:** New point proposed by the Decision maker

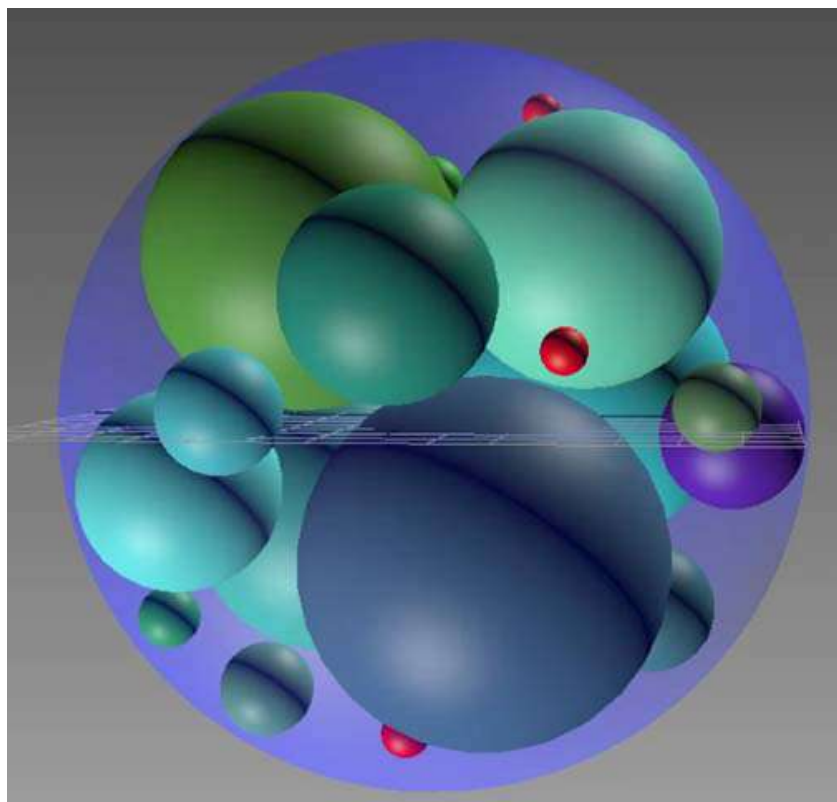


**Figure 6:** Locally optimal solution obtained by the Solver

## 5. Conclusions

The paper provides a general formulation of problems for the synthesis of spatial configurations and the classification of these problems. Analytical models of various types are given for the case of spherical objects. To solve the problems of synthesis of spherical configurations, an information technology has been proposed, including modules for optimization, visualization and the participation of a decision maker. This approach made it possible to use the experience and intuition of a specialist for a global search for a solution, i.e. attempts to improve local extrema. An example of using the proposed technology is given.

## 6. References

- [1] M. Hifi, R. M'Hallah, A literature review on circle and sphere packing problems: models and methodologies, *Advances in Operation Research*, no. 7, pp. 1–22, 2009. doi:10.1155/2009/150624.
- [2] J. Liu, Y. Yao, Yu. Zheng, et al. An effective hybrid algorithm for the circles and spheres packing problems, *Combinatorial Optimization and Applications, Lecture Notes in Computer Science*, vol. 5573, pp. 135–144, 2009. doi:10.1007/978-3-642-02026-1\_12.
- [3] A. Sutou, Y. Day, Global optimization approach to unequal sphere packing problems in 3D, *Journal of Optimization Theory and Applications*, vol. 114, no. 3, pp. 671–694, 2002. doi:10.1023/A:1016083231326.
- [4] E.G. Birgin, F.N.C Sobral, Minimizing the object dimensions in circle and sphere packing problems, *Operations Research*, no. 35, pp. 2357–2375, 2008. doi:10.1016/j.cor.2006.11.002.
- [5] Yu. G. Stoyan, G. Scheithauer, G. N. Yaskov, Packing unequal spheres into various containers, *Cybernetics and Systems Analysis*, vol. 52, no.3, pp. 419–426, 2016. doi:10.1007/s10559-016-9842-1.
- [6] I. Castillo, F.J. Kampas, J.D. Pinter, Solving circle packing problems by global optimization: numerical results and industrial applications, *Eur. J. Oper. Res.*, vol. 191, no. 3, pp. 786–802, 2008. doi:10.1016 / j.ejor.2007.01.054.
- [7] Y. Stoyan, G. Yaskov, T. Romanova et al., Optimized packing multidimensional hyperspheres: a unified approach, *Mathematical Biosciences and Engineering*, vol. 17, no.6, pp. 6601–6630, 2020. doi: 10.3934 / mbe.2020344.
- [8] C.O. López, J.E. Beasley, Packing unequal circles using formulation space search, *Comput. Oper. Res.*, vol. 40, no. 5, pp. 1276–1288, 2013. doi:10.1016/j.cor.2012.11.022.
- [9] S. Yakovlev, O. Kartashov, K. Korobchynskiy, B. Skripka, Numerical results of variable radii method in the unequal circles packing problem, in: *Proceedings of the 15th International Conference on the Experience of Designing and Application of CAD Systems (CADSM 2019)*, pp. 1–4, 2019. doi:10.1109/CADSM.2019.8779288.
- [10] Z.Z. Zeng, W.Q. Huang, R.C. Xu, Z.H. Fu, An algorithm to packing unequal spheres in a larger sphere, *Advanced Materials Research*, vol. 546, pp. 1464–1469, 2012. doi: 10.4028/www.scientific.net/AMR.
- [11] J. Wang, Packing of unequal spheres and automated radiosurgical treatment planning, *Journal of Combinatorial Optimization*, vol. 3, pp. 453–463, 1999. doi:10.1023/A:1022951027498.
- [12] J. Zhou, Y. Zhang, J. K. Chen, Numerical simulation of random packing of spherical particles for powder-based additive manufacturing, *ASME. J. Manuf. Sci. Eng.*, vol. 131, pp.1–8, 2009. doi:10.1115/1.3123324.
- [13] N. Yu. Ilyasova, A. S. Shirokanov, A. V. Kupriyanov, R. A. Paringer, Technology of intellectual feature selection for a system of automatic formation of a coagulate plan on retina, *Computer Optics*, vol. 43, no 2, 2019. doi: 10.18287 / 2412-6179-2019-43-2-304-315.
- [14] S. Bereg, J.M. Diaz-Banez, M.A. Lopez et al., A decentralized geometric approach for the formation keeping in unmanned aircraft navigation, in: *2015 International Conference on Unmanned Aircraft Systems (ICUAS)*, pp. 989–997, 2015.



- [15] O. Blyuss, L. Koriashkina, E. Kiseleva, R. Molchanov, Optimal placement of irradiation sources in the planning of radiotherapy: mathematical models and methods of solving, *Computational and Mathematical Methods in Medicine*, vol. 2015, pp. 1-8, 2015. doi: 10.1155/2015/142987.
- [16] Lin Lu, Y. Guo, R. Fan, et al., Content-aware photo collage using circle packing, *IEEE Transactions on Visualization and Computer Graphics*, vol. 20, no.2, pp. 182–195, 2014. doi:10.1109/TVCG.2013.106.
- [17] Y.G. Stoyan, S.V. Yakovlev, Configuration space of geometric objects, *Cybernetics and Systems Analysis*, vol. 54, no. 5, pp. 716-726, 2018. doi:10.1007/s10559-018-0073-5.
- [18] S.V. Yakovlev, Formalizing spatial configuration optimization problems with the use of a special function class, *Cybernetics and Systems Analysis*, vol. 55, no. 4, pp. 581-589, 2019. doi:10.1007/s10559-019-00167-y.
- [19] S.V. Yakovlev, On some classes of spatial configurations of geometric objects and their formalization, *Journal of Automation and Information Science*, vol. 50, no. 9, pp. 38–50, 2018. doi:10.1615/JAutomatInfScien.v50.i9.30.
- [20] S.V. Yakovlev, On a class of problems on covering of a bounded set, *Acta Mathematica Hungarica*, vol. 53, no. 3, pp. 253-262, 1999. doi:10.1007/BF01953365.
- [21] S.B. Shekhovtsov, S.V. Yakovlev, Formalization and solution of one class of covering problem in design of control and monitoring systems, *Avtomatika i Telemekhanika*, vol. 5, pp. 160–168, 1989. <http://mi.mathnet.ru/eng/at6296>.
- [22] E.M. Kiseleva, The emergence and formation of the theory of optimal set partitioning for sets of the n-dimensional Euclidean space: theory and application, *Journal of Automation and Information Science*, vol. 50, no. 9, pp. 1-24, 2018. doi: 10.1615/JAutomatInfScien.v50.i9.10.
- [23] O.Blyuss, A. Zaikin, V. Cherepanova et al., Development of PancRISK, a urine biomarker-based risk score for stratified screening of pancreatic cancer patients, *British Journal of Cancer*, vol. 122, no. 5, pp. 692–696, 2020. doi:10.1038/s41416-019-0694-0.
- [24] Y.G. Stoyan, S.V. Yakovlev, Theory and methods of Euclidian combinatorial optimization: current status and prospects, *Cybernetics and System Analysis*, vol. 56, no. 3, pp. 366–379, 2020. doi:10.1007/s10559-020-00253-6.
- [25] O. Pichugina, S. Yakovlev, Euclidean combinatorial configurations: typology and applications, in: 2019 IEEE 2nd Ukraine Conference on Electrical and Computer Engineering (UKRCON), pp. 1065–1070, 2019. doi:10.1109/UKRCON.2019.8879912.
- [26] S. Yakovlev, O. Kartashov, O. Pichugina, Optimization on combinatorial configurations using genetic algorithms, in: *CEUR Workshop Proceedings*, vol. 2353, pp. 28–40, 2019. <http://ceur-ws.org/Vol-2353/paper3.pdf>.
- [27] S. Yakovlev, O. Kartashov, O. Yarovaya, On class of genetic algorithms in optimization problems on combinatorial configurations, in: *Proceedings of the 2018 IEEE 13th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT 2018)*, vol. 1, pp. 374–377, 2018. <http://toc.proceedings.com/41726webtoc.pdf>.



# An algorithm for topic modeling of researchers taking into account their interests in Google Scholar profiles

Serhiy Shtovba<sup>a</sup> and Mykola Petrychko<sup>b</sup>

<sup>a</sup> *Vasyl' Stus Donetsk National University, 600-richchia str., 21, Vinnytsia, 21021, Ukraine*

<sup>b</sup> *Vinnytsia National Technical University, Khmelnytske Shose, 95, Vinnytsia, 21021, Ukraine*

## Abstract

An algorithm for topic modeling of researchers based on their interests from Google Scholar's profiles is proposed. As topics for modeling, we took research groups from research classification system ANZSRC – Australian and New Zealand Standard Research Classification. Researchers' distribution to research groups is found based on their interests' usage statistics in categorized publications from Dimensions. Topic modeling is conducted accordingly to principles of statistical support, multi-labeling, noise filtering, ignoring stop-words, solidarities, focusing, compactness and research groups' interactions. We compare topic modeling based on data with low level of information from researchers' profiles in Google Scholar with topic modeling based on a few dozen authored publications categorized by Dimensions. Comparison is made by modified Czekanowski metric that takes into account the interaction between research groups. By comparing the results of topic modeling based on different sources of initial information a good match was found. It allows to use the proposed algorithm as the intellectual core of information technology in regards to scientific staff, in particular, for the selection of candidates as opponents of a dissertation, as reviewers for research projects, for forming a team to collaborate on mutual research projects etc.

## Keywords

topic modeling, Google Scholar, Dimensions, ANZSRC, researcher's profile, research interests, research group, Czekanowski metric, Jaccard index.

## 1. Introduction

Google Scholar aggregates the most volumetric collection of researchers' profiles. The most used information from Google Scholar profiles is citations. It, for example, is used as primary information for university ranking in Webometrics. Several studies, in particular [1, 2], are concerned with comparing concordance of Google Scholar citations with different scientometrics systems such as Scopus, Web of Science, Dimensions and others, that use only meta-information from publishers. A researcher's profile in Google Scholar contains not only publications and their citations, but also other information. In particular, a researcher provides his or her interests. A researcher chooses the interests in a loose manner without any limitations. Google Scholar provides a web interface to search researchers by an interest. However, the results are formed by literal coincidence. That is why the results for *fuzzy set* and *fuzzy sets* are different; the same applies for synonymous interests such as *fuzzy evidence* and *fuzzy inference*. Moreover, Google Scholar does not take into account an interconnection of the interests, that is the search by an interest is done independently and isolated. Given that, the search and analytical services that provide information about many researchers in Google Scholar are relatively straightforward.

The goal of this paper is topic modeling of researchers based on their interests from Google Scholar. Methods that process a researcher's interests from Google Scholar profile are not studied

---

CMIS-2021: The Fourth International Workshop on Computer Modeling and Intelligent Systems, April 27, 2021, Zaporizhzhia, Ukraine

EMAIL: shtovba@donnu.edu.ua (S. Shtovba); mpetrychko@vntu.edu.ua (M. Petrychko)

ORCID: 0000-0003-1302-4899 (S. Shtovba); 0000-0001-6836-7843 (M. Petrychko)



© 2020 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

well. We identified only two relevant publications. One of them is [3], it describes a recommendation system that recommends supervisors based on some information and interests of candidates from Google Scholar profiles as well. Another paper [4] presents an information technology that synthesizes a research profile of institute or research laboratory. It also uses interests of researchers from their profiles at Google Scholar. Articles [3, 4] are based on pairwise comparison using cosine similarity metric between researcher and a set of keywords from a given topic. Such a topic in [3] is an article at Wikipedia. Unlike these methods, we strive to categorize researchers by a given research classification system, that is to assign a research group to each of them.

Automatic researchers' categorization is usually done as a result of generalizing the topics of their publications. One of the methods for this is presented at [5]. The authors present a statistical model "Author-Topic" that is based on topic modeling model Latent Dirichlet Allocation [6]. This model represents a researcher as a distribution over some abstract topics. The topics are clusters of similar words. One of the drawbacks of this model is low interpretation of the topics because they are formed by words frequency in a document. To improve the interpretation another model "Author-Subject-Topic" is proposed in [7]. This model additionally uses a research specialty that is defined by journal in which an analyzed publication is published. In [8] another improvement of "Author-Topic" model is presented – "Author-Persona-Topic" model. In this model rather than representing all researcher's documents as single topic distribution, authors group all documents into different clusters, each with its own topic distribution. These clusters represent "personas" under which an author writes.

Apart from topic modeling methods there are also methods based on word embedding. They generally perform better than topic models because they can incorporate semantic relationships. One of the most popular models of embedding is word2vec [13]. It is used in [9] as a part of similarity metric between researchers using their publications. They assess the similarity between words that comprise publications of different researchers using representation of words defined by word2vec. The authors of [10] use publications' titles as source information for solving the problem of collaboration recommendation. The words from the titles are represented as vectors using word2vec. These vectors are then clustered using k-means to partition researchers into different academic domains. The representation of a researcher is further improved by using his co-authorship and the random walk method to find his influence in different domains. In [11] authors represent a researcher as a set of documents he/she has written. The words of the documents are defined as vectors trained by word2vec model. These representations are then used to solve the problem of expert finding by utilizing a restricted convolutional neural network. In [12] a researcher is represented as a concatenation of all his/hers abstracts. Each word in the concatenation is then represented as a vector from word2vec model to solve the problem of reviewer recommendation.

Analyzed methods assume to have enough number of publications for a given researcher with selected keywords. At the same time, they do not account for the fact that co-author contribution is sometimes relative to a small subset of the paper keywords. A researcher, especially a young one, may have only a few publications that may not be enough for a valid categorization. On the other hand, the researcher can manually specify at the profile a set of keywords that describe his (or her) activities. As the time goes on a researcher may change his research direction, for example, move to another laboratory or another project. Given that there is no change when a researcher is categorized based on his publications, categorization based on his keywords may find these changes. By taking that in mind, we study the topic modeling based on actual interests that a researcher specified by himself (or herself) at the current moment.

## 2. Problem statement

We use the following notations:

$W = (w_1, w_2, \dots, w_n)$  is a set of keywords that are equal to researcher interests in Google Scholar profile;

$T = (t_1, t_2, \dots, t_m)$  is a set of research topics from a research classification system;

$D_1, D_2, \dots, D_m$  is a set of topic-marked collections of texts; each collection contains only publications from topics  $t_1, t_2, \dots, t_m$ , respectively;

$B = D_1 \cup D_2 \cup \dots \cup D_m$  is the general collection of topic-marked texts; each element of  $B$  belongs to one or more topics from the set  $T$ ;

$R(D, T) \subset D \times T$  is a relation that describes membership of a publication to topic-marked collections.

The problem is to find out topics from  $T$  that correspond to the set of interests  $W$ . The results of mapping  $W \rightarrow T$  is a fuzzy set  $\tilde{W}$  defined on the universal set of topics  $T$  as follows:

$$\tilde{W} = \left( \frac{\mu_W(t_1)}{t_1}, \frac{\mu_W(t_2)}{t_2}, \dots, \frac{\mu_W(t_m)}{t_m} \right),$$

where  $\mu_W(t_p) \in [0, 1]$  denotes membership degree of the set of interests  $W$  to topics  $t_p$ ,  $p = \overline{1, m}$ .

We set the following restrictions on  $\tilde{W}$ :

- 1) the cardinality of the fuzzy set support must be small  $1 \leq |\text{sup}(\tilde{W})| \leq T_{\max}$ , for example, with  $T_{\max} \in \{2, 3, 4\}$  a researcher will be assigned only to a few topics;
- 2)  $\sum_{p=\overline{1, m}} \mu_W(t_p) = 1$ , which is equivalent to the topic modeling regularization condition.

### 3. Data acquisition and preprocessing

We use a researcher's profile from Google Scholar to get the keywords. For example, in Figure 1 we have a researcher's profile with three keywords that are marked with blue color. For this researcher:  $w_1 = \text{"Computational Intelligence"}; w_2 = \text{"Fuzzy Logic"}; w_3 = \text{"Artificial Intelligence"}$ . The order of keywords in the set  $W$  is not important, this corresponds to the bag of words model. Interests often complement each other thus making their research topics more focused. To take that into account we synthesize additional keywords defined as pairs of initial interests. Interests in a pair are combined by a logical operation *AND*. For researcher from Figure 1 additional keywords are defined as follows:

$w_4 = \text{"Computational Intelligence" AND "Fuzzy Logic"};$

$w_5 = \text{"Computational Intelligence" AND "Artificial Intelligence"}$

$w_6 = \text{"Fuzzy Logic" AND "Artificial Intelligence"}$

If a researcher's profile has 3 interests, additional 3 keywords are synthesized, if it has 4 interests then 6 additional are synthesized etc.

Ronald R Yager

Professor Iona College

Підтверджена електронна адреса в panix.com

Computational Intelligence Fuzzy Logic Artificial Intelligence

**Figure 1:** An example of a researcher's profile with 3 interests

For researchers' topic modeling, we need to choose a research classification system. There are a lot of them, but when choosing we take into account not only their semantic advantages and disadvantages, but also that there is an information system with this research classification system that has available search services. In addition, we require that the information system must index a large number of categorized publications over all research. The information system that satisfies these requirements is Dimensions.

Dimensions indexes more than 110M of publications. All publications are categorized by the two-level variant of Australian and New Zealand Standard Research Classification (ANZSRC) with 22 research divisions and 154 research groups (Table 1). In this work we use the research groups to model a researcher's interests.

**Table 1.**

Research classification system ANZSRC, that is used in Dimensions

| Research Division                    | Research Group                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
|--------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Mathematical Sciences                | A1 - Pure Mathematics; A2 - Applied Mathematics; A3 - Numerical and Computational Mathematics; A4 – Statistics; A5 - Mathematical Physics                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
| Physical Sciences                    | B1 - Astronomical and Space Sciences; B2 - Atomic, Molecular, Nuclear, Particle and Plasma Physics; B3 - Classical Physics; B4 - Condensed Matter Physics; B5 - Optical Physics; B6 - Quantum Physics; B7 - Other Physical Sciences                                                                                                                                                                                                                                                                                                                                                                                    |
| Chemical Sciences                    | C1 - Analytical Chemistry; C2 - Inorganic Chemistry; C3 - Macromolecular and Materials Chemistry; C4 - Medicinal and Biomolecular Chemistry; C5 - Organic Chemistry; C6 - Physical Chemistry (incl. Structural); C7 - Theoretical and Computational Chemistry; C8 - Other Chemical Sciences                                                                                                                                                                                                                                                                                                                            |
| Earth Sciences                       | D1 - Atmospheric Sciences; D2 - Geochemistry; D3 - Geology; D4 - Geophysics; D5 - Oceanography; D6 - Physical Geography and Environmental Geoscience; D7 - Other Earth Sciences;                                                                                                                                                                                                                                                                                                                                                                                                                                       |
| Environmental Sciences               | E1 - Ecological Applications; E2 - Environmental Science and Management; E3 - Soil Sciences; E4 - Other Environmental Sciences;                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
| Biological Sciences                  | F1 - Biochemistry and Cell Biology; F2 - Ecology; F3 - Evolutionary Biology; F4 - Genetics; F5 - Microbiology; F6 – Physiology; F7 - Plant Biology; F8 - Zoology; F9 - Other Biological Sciences                                                                                                                                                                                                                                                                                                                                                                                                                       |
| Agricultural and Veterinary Sciences | G1 - Agriculture, Land and Farm Management; G2 - Animal Production; G3 - Crop and Pasture Production; G4 - Fisheries Sciences; G5 - Forestry Sciences; G6 - Horticultural Production; G7 - Veterinary Sciences; G8 - Other Agricultural and Veterinary Sciences                                                                                                                                                                                                                                                                                                                                                        |
| Information and Computing Sciences   | H1 - Artificial Intelligence and Image Processing; H2 - Computation Theory and Mathematics; H3 - Computer Software; H4 - Data Format; H5 - Distributed Computing; H6 - Information Systems; H7 - Library and Information Studies; H8 - Other Information and Computing Sciences                                                                                                                                                                                                                                                                                                                                        |
| Engineering                          | I1 - Aerospace Engineering; I2 - Automotive Engineering; I3 - Biomedical Engineering; I4 - Chemical Engineering; I5 - Civil Engineering; I6 - Electrical and Electronic Engineering; I7 - Environmental Engineering; I8 - Food Sciences; I9 - Geomatic Engineering; I10 - Manufacturing Engineering; I11 - Maritime Engineering; I12 - Materials Engineering; I13 - Mechanical Engineering; I14 - Resources Engineering and Extractive Metallurgy; I15 - Interdisciplinary Engineering; I16 - Other Engineering                                                                                                        |
| Technology                           | J1 - Agricultural Biotechnology; J2 - Environmental Biotechnology; J3 - Industrial Biotechnology; J4 - Medical Biotechnology; J5 - Communications Technologies; J6 - Computer Hardware; J7 – Nanotechnology; J8 - Other Technology                                                                                                                                                                                                                                                                                                                                                                                     |
| Medical and Health Sciences          | K1 - Medical Biochemistry and Metabolomics; K2 - Cardiorespiratory Medicine and Haematology; K3 - Clinical Sciences; K4 - Complementary and Alternative Medicine; K5 - Dentistry; K6 - Human Movement and Sports Science; K7 – Immunology; K8 - Medical Microbiology; K9 – Neurosciences; K10 - Nursing; K11 - Nutrition and Dietetics; K12 - Oncology and Carcinogenesis; K13 - Ophthalmology and Optometry; K14 - Paediatrics and Reproductive Medicine; K15 - Pharmacology and Pharmaceutical Sciences; K16 - Medical Physiology; K17 - Public Health and Health Services; K18 - Other Medical and Health Sciences; |

| Research Division                          | Research Group                                                                                                                                                                                                                            |
|--------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Built Environment and Design               | L1 - Architecture; L2 – Building; L3 - Design Practice and Management; L4 - Engineering Design; L5 - Urban and Regional Planning; L6 - Other Built Environment and Design;                                                                |
| Education                                  | M1 - Education Systems; M2 - Curriculum and Pedagogy; M3 - Specialist Studies In Education; M4 - Other Education                                                                                                                          |
| Economics                                  | N1 – Economic Theory; N2 – Applied Economics; N3 – Econometrics; N4 – Other Economics                                                                                                                                                     |
| Commerce, Management, Tourism and Services | O1 - Accounting, Auditing and Accountability; O2 - Banking, Finance and Investment; O3 - Business and Management; O4 - Commercial Services; O5 – Marketing; O6 – Tourism; O7 - Transportation and Freight Services;                       |
| Studies in Human Society                   | P1 - Anthropology; P2 - Criminology; P3 - Demography; P4 - Human Geography; P5 - Policy and Administration; P6 - Political Science; P7 - Social Work; P8 - Sociology; P9 - Other Studies In Human Society                                 |
| Psychology and Cognitive Sciences          | Q1 - Psychology; Q2 - Cognitive Sciences; Q3 - Other Psychology and Cognitive Sciences;                                                                                                                                                   |
| Law and Legal Studies                      | R1 – Law; R2 - Other Law and Legal Studies                                                                                                                                                                                                |
| Studies in Creative Arts and Writing       | S1 - Art Theory and Criticism; S2 - Film, Television and Digital Media; S3 - Journalism and Professional Writing; S4 - Performing Arts and Creative Writing; S5 - Visual Arts and Crafts; S6 - Other Studies In Creative Arts and Writing |
| Language, Communication and Culture        | T1 - Communication and Media Studies; T2 - Cultural Studies; T3 - Language Studies; T4 – Linguistics; T5 - Literary Studies; T6 - Other Language, Communication and Culture                                                               |
| History and Archaeology                    | U1 - Archaeology; U2 - Curatorial and Related Studies; U3 - Historical Studies; U4 - Other History and Archaeology                                                                                                                        |
| Philosophy and Religious Studies           | V1 - Applied Ethics; V2 - History and Philosophy of Specific Fields; V3 - Philosophy; V4 - Religion and Religious Studies; V5 - Other Philosophy and Religious Studies                                                                    |

A query to Dimensions is formed separately by each element of the set  $W$ . If an element is a phrase, then it is surrounded by quotes. As a search scope, we use *Title and Abstract* and we search only the last 5 years – 2016 – 2020. An example of a search result for the query “fuzzy logic” is presented in Figure 2. For each research division and research group there is a number of publications that has the query mentioned in either the title or abstract. The results are sorted by the number of publications descending. We can also find the overall number of publications for each research division and group, that is without any query.

#### 4. Topic modeling algorithm

We perform the topic modeling of researchers based on the following principles:

- *the principle of statistical support* – the more publications from a specific research group a given keyword contains, the more membership degree of this keyword to this research group is;
- *the principle of multi-labeling* – a keyword can belong to a few research groups;
- *the principle of noise filtering* – we ignore research groups with low membership degree to a given keyword;
- *the principle of ignoring stop-words* – we ignore keywords that appear in a very large number of publications;

- *the principle of solidarities* – the more keywords belong to the same research group the larger the chance that the researcher belongs to this research group;
- *the principle of focusing* – if a topic-marked collection of publications contains a few keywords of a researcher at once then the chances to assign this researcher to the respected topic increase;
- *the principle of compactness* – a researcher can only be assigned to a few research groups;
- *the principle of research groups interaction* – when cutting the tail of topic distribution, the contribution of minor research groups is redistributed on leaders by taking into account their similarity.

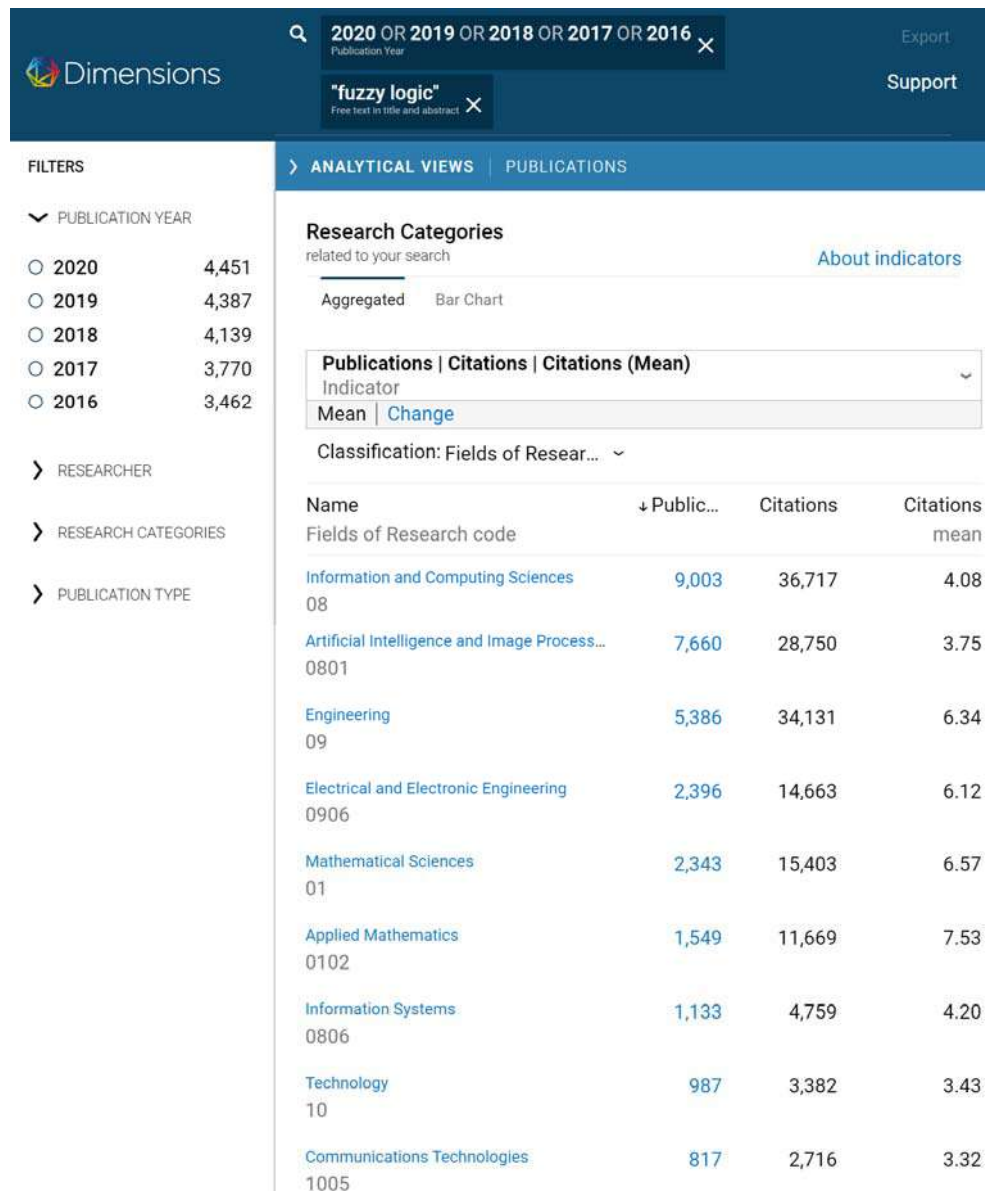


**Figure 2:** The results from Dimensions by search query "fuzzy logic" for the period 2016-2020

We propose an algorithm to implement the proposed principles that consists of 3 stages. On the first stage the set of queries based on keywords and their combination is formed. We use only pairs of keywords because the results using triples of keywords are often empty and increase the processing time. The second stage performs topic modeling by each query separately. Research groups are chosen by the frequency of mentions at a topic-marked collection. Stop-words and noise are filtered by the frequency of mentions in research groups at all topic-marked collections. The minor research groups are left out using cumulative principle, by cutting the tail of the distribution. On the third stage

all membership degrees of queries are averaged, the resulted distribution is cut and research groups with the low membership degree are dropped. To ensure compactness we only allow 1 to 4 research groups.

```
%Topic modeling algorithm
% #1 - creating the set E of search queries from the keywords
E=W
for i=1:length(W)
    for j=i:length(W)
        E={E; ["" W(i) 'AND' W(j) '']} }
    end
end
% #2 - compute membership degrees to research groups by each query
< Find the number of publications at each topic-marked collection
N=[N(1), N(2), ..., N(m)] >
Counter=0 % the counter of successful query responses
for i=1:length(E)
    < Find Q - the number of publications D, that contain E{i} >
    If Q>Threshold_SW continue % stop-words
    end
    If Q<Threshold_noise continue; % noise
    end
    < Find t(1), t(2),..., t(m) - the number of publications in the topic-
        marked collections in each research group for query E{i} >
    % Ignore topics with a low number of publications:
    index=find(t<Threshold_topic)
    t(index)=0
    if max(t)==0 continue
    end
    % Compute the frequency of E{i} at topic-marked collections:
    Gamma=t./N
    < Choose the most popular research groups that have cumulative
        contribution in Gamma not lower than Tail_1. Research groups that have
        cumulative contribution lower than Tail_1 are put in vector Rejected >
    % Ignore research groups with contribution lower than Tail_1:
    Gamma(Rejected)=0
    Gamma=Gamma./sum(Gamma) % norm to be in [0, 1]
    Counter=Counter+1
    Mu(Counter)=Gamma
end
If Counter==0
    return ('Unsuccessful')
end
% #3 - compute membership degrees using all queries
Mu_mean=mean(Mu) % averaging all successful queries
< Form leaders of research groups that have cumulative contribution Mu_mean not
    lower than Tail_2. We restrict the number of leaders to be at most 6 with the
    largest cumulative contribution. If we have more than 6 leaders their numbers
    will be in the vector Rejected >
% Ignore research groups with contribution lower than Tail_2:
Mu_mean(Rejected)=0
Mu_mean= Mu_mean./sum(Mu_mean) % norm to be in [0, 1]
% Current number of research groups:
Current_N_fields=sum(Mu_mean>0)
```

```

% Set the max number of research groups for a researcher:
T_max=min(4, Counter+1)
< Find Mu_worst - the smallest membership degree among the leaders >
while (Current_N_fields>T_max OR Mu_worst<Tail_3)
    < Drop the minor group and redistribute its contribution to others based
    on their similarity >
    Current_N_fields=Current_N_fields-1;
    Mu_mean= Mu_mean./sum(Mu_mean) % norm to be in [0, 1]
    < Find Mu_worst - the smallest membership degree among chosen
    research groups >
end

```

On the last stage of the algorithm when dropping a minor research group its contribution is redistributed to other research groups based on the similarity defined at [14, 15]. For example, let us say that on an intermediate stage a researcher is assigned to research groups in the following way:

$\tilde{W} = \left( \frac{0.5}{H6}, \frac{0.2}{O5}, \frac{0.2}{O6}, \frac{0.1}{O4} \right)$ . Let us drop the minor group *O4*. For this, first using method from [14, 15]

we compute Jaccard indexes between *O4* and other research groups. For the data from 2016 – 2020 they are:

$$J(O4, H6) = 0;$$

$$J(O4, O5) = 0.13;$$

$$J(O4, O6) = 0.22.$$

By taking into account the similarity, the contribution of the minor specialty *O4* is redistributed in the following way:

$$\tilde{W} = \left( \frac{0.5 + 0 \cdot 0.1}{H6}, \frac{0.2 + 0.13 \cdot 0.1}{O5}, \frac{0.2 + 0.22 \cdot 0.1}{O6} \right).$$

As a result, we get:

$$\tilde{W} = \left( \frac{0.5}{H6}, \frac{0.213}{O5}, \frac{0.222}{O6} \right).$$

After norming to be in [0,1] we have:

$$\tilde{W} = \left( \frac{0.535}{H6}, \frac{0.228}{O5}, \frac{0.237}{O6} \right).$$

## 5. Checking example

Let us illustrate how the algorithm works using as an example topic modeling of the researcher from Figure 1. Using three interests, we form six queries. Figure 3 shows frequency of queries at topic-marked collections. Figure 4 shows the results after cutting the first tail of the distribution. Next, we average by all queries (Figure 5) and cut the tail of the distribution (Figure 6). The resulting distribution is overfilled due to the broad usage of interests for the given researcher. To make the results more focused the final stage of the algorithm reduces the number of research groups to 2 (Figure 7). As a result, we get that a researcher with interests at artificial intelligence and neural networks has the largest membership degree in research groups *H1 – Artificial Intelligence and Image Processing* with membership degree 0.441 and *H2 – Computation Theory and Mathematics* with membership degree 0.559. Such a categorization of the researcher does not contradict with the authors' viewpoint. The example shows that even with two initial keywords the proposed algorithm finds a good enough membership relation between the researcher and research groups.



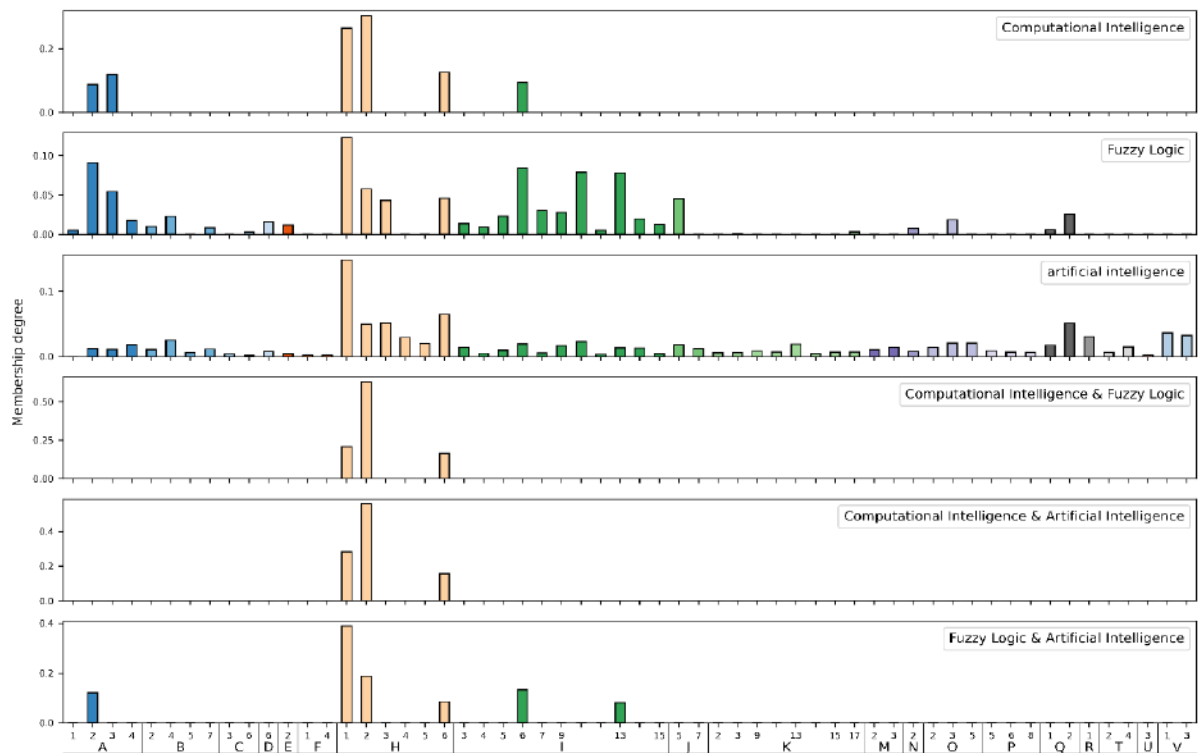


**Figure 3: Initial membership distribution of each interest to research groups**

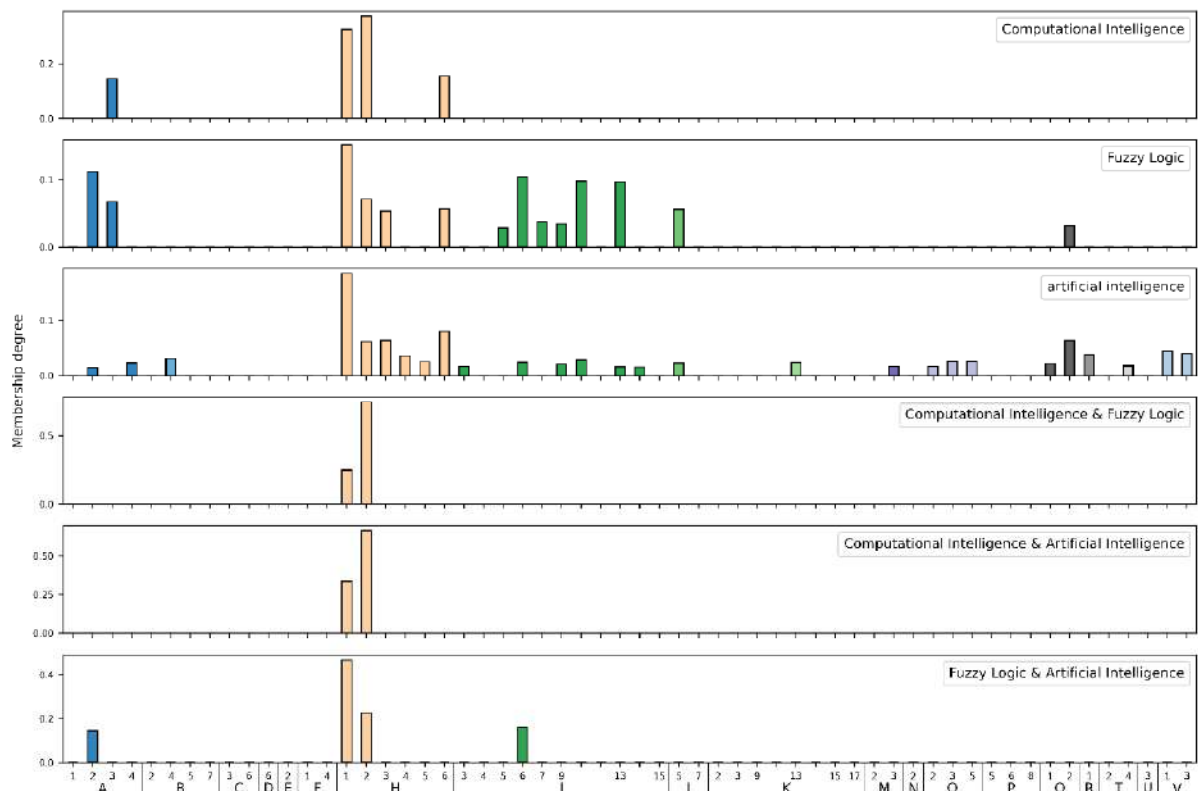


**Figure 4: Distributions after the first noise filtering**

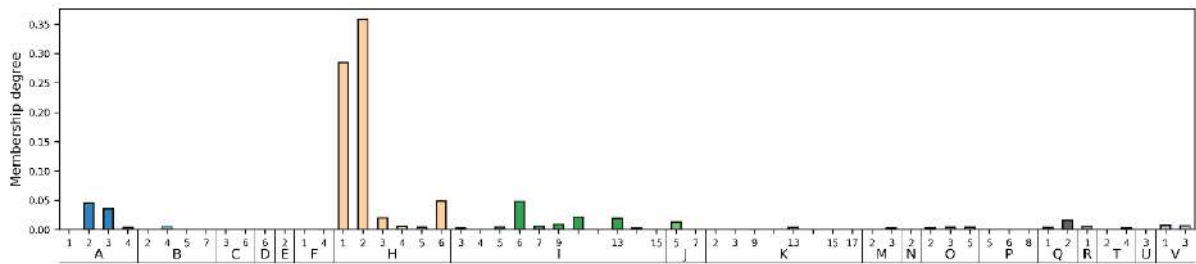


Figure 5: Averaging distributions of all keywords

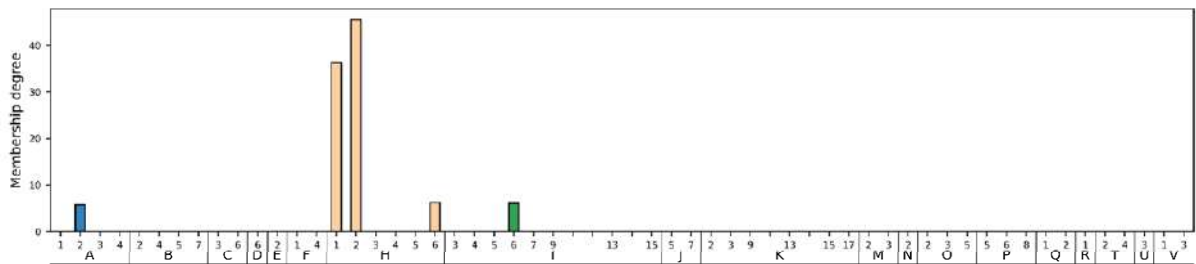


Figure 6: Distributions after the second noise filtering

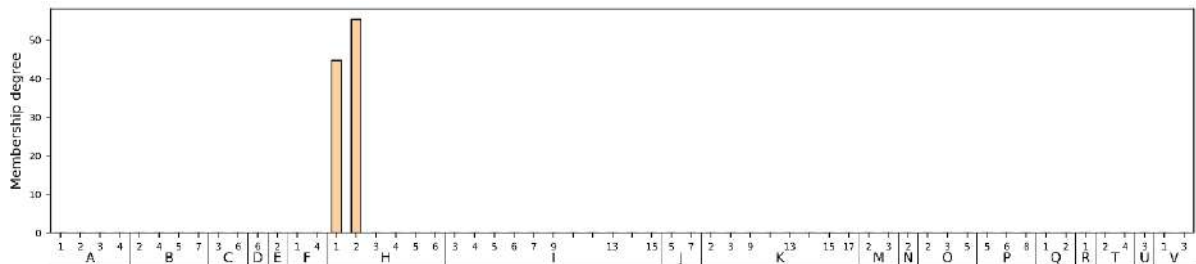


Figure 7: The result of topic modeling for the researcher from Figure 1

## 6. Comparing with categorized papers

Let us compare the topic modeling results based on keywords from researchers' profiles in Google Scholar and based on categorized publications of the researchers in Dimensions. For this we take three researchers:

- Ronald Yager with the interests *Computational Intelligence*, *Fuzzy Logic*, and *Artificial Intelligence*;
- Nataliia Kussul with the interests *Machine Learning*, *Remote Sensing*, *Data Science*, *Disaster Management*, and *Agricultural Monitoring*;
- Yevgeniy Bodyanskiy with the interests *Computational Intelligence*, *Data Mining*, *Data Stream Mining*, and *Big Data*.

The results of topic modeling of researchers with the above-mentioned interests are presented in Table 2.

These researchers have a considerable number of publications in Dimensions for the last 5 years that allows getting statistically significant results.

During the analyzed period, Ronald Yager published 141 papers that are categorized by 20 research groups. The most publications – 63 are assigned to the research group *H1*. Yevgeniy Bodyanskiy published 88 papers. They are categorized by 12 research groups. The most publications – 59 are assigned to the research group *H1*. Nataliia Kussul published 47 papers that are categorized by 14 research groups. The most publications – 21 are assigned to research group *I9*. By using the third stage of the proposed algorithm on papers distributions, we get membership degrees to research groups (Table 2).

**Table 2**

The results of researchers' topic modeling

| Research group | Ronald Yager |                | Nataliia Kussul |                | Yevgeniy Bodyanskiy |                |
|----------------|--------------|----------------|-----------------|----------------|---------------------|----------------|
|                | Dimensions   | Google Scholar | Dimensions      | Google Scholar | Dimensions          | Google Scholar |
| D6             |              |                |                 | 0.283          |                     |                |
| I9             |              |                | 0.675           | 0.447          |                     |                |
| H1             | 0.63         | 0.441          | 0.172           | 0.346          | 0.797               | 0.295          |
| H2             |              | 0.559          |                 |                |                     | 0.199          |
| H6             | 0.37         |                | 0.153           |                | 0.203               | 0.506          |

Comparing the results, we see that topic modeling based on interests from Google Scholar – laconic subjective information, with the proposed algorithm categorizes researchers good enough. For quantitative assessment of the results, we used Czekanowski metric. For the case when membership degrees are in  $[0, 1]$ , Czekanowski metric between two researchers  $W_1$  and  $W_2$  is computed in the following way:

$$Fit(W_1, W_2) = \sum_{p=1, \overline{m}} \min(\mu_{t_p}(W_1), \mu_{t_p}(W_2)). \quad (1)$$

The metric (1) can be interpreted as a sum of membership degrees of the intersection of fuzzy sets  $\tilde{W}_1$  and  $\tilde{W}_2$ , that represent the topic modeling results based on two source of information – interests from Google Scholar and categorized publications in Dimensions.

Based on the data from Table 2 we get the following assessments using metric (1):

$$Fit(Yager) = 0.441;$$

$$Fit(Kussul) = 0.619;$$

$$Fit(Bodyanskiy) = 0.498.$$

Using metric (1) the match is computed with isolate assumption – only in the scope of each individual research group. To take into account the contribution of similar research groups we propose to the value of metric (1) to add the following interactive addend:

$$\Delta Fit(W_1, W_2) = \sum_{v=1, \overline{M}} \sum_{p=1, \overline{M}} J(t_v, t_p) \cdot \min(\varepsilon_{t_v}(W_1), \varepsilon_{t_p}(W_2)) \quad (2)$$

where  $J(t_v, t_p)$  denotes Jaccard index between research groups  $t_v$  and  $t_p$ ;

$\varepsilon_{t_v}(W_1) = \min(0, \mu_{t_v}(W_1) - \mu_{t_v}(W_2))$  denotes residual of membership degree to research group  $t_v$  in  $\tilde{W}_1$  after taking into account the matching between  $\mu_{t_v}(W_1)$  and  $\mu_{t_v}(W_2)$  in (1);

$\varepsilon_{t_p}(W_2) = \min(0, \mu_{t_p}(W_2) - \mu_{t_p}(W_1))$  denotes residual of membership degree to research group  $t_p$  in  $\tilde{W}_2$  after taking into account the matching between  $\mu_{t_p}(W_1)$  and  $\mu_{t_p}(W_2)$  in (1).

To filter information noise, we use the formula (2) only for pairs of research groups with a high level of similarity – with Jaccard index greater than 0.02. For the research groups from the Table 2 we have 2 such pairs. Jaccard indexes for them are the following:

$$J(D6, I9) = 0.083;$$

$$J(H1, H6) = 0.071.$$

Substituting the indexes in (2), we get:

$$\Delta Fit(Yager) = 0;$$

$$\Delta Fit(Kussul) = \min(0.675 - 0.447, 0.283) \cdot 0.083 + \min(0.346 - 0.172, 0.153) \cdot 0.071 = 0.03;$$

$$\Delta Fit(Bodyanskiy) = \min(0.797 - 0.295, 0.506 - 0.203) \cdot 0.071 = 0.022.$$

By taking into account the similarity of research groups the level of matching the topic modeling results takes the following values:

$$Fit_{sim}(Yager) = 0.441 + 0 = 0.441;$$

$$Fit_{sim}(Kussul) = 0.619 + 0.03 = 0.649;$$

$$Fit_{sim}(Bodyanskiy) = 0.498 + 0.022 = 0.52.$$

## 7. Conclusions

We proposed researchers' topic modeling based on their interests in Google Scholar profiles. Interests in profiles researchers specify based on their discretion without using any vocabulary of keywords. In the paper, we propose an approach to categorization of such researchers using the research classification system ANZSRC. A mapping "researcher – research groups" is done using information system Dimensions that contains more than 110 millions of publications categorized according to ANZSRC.

The algorithm of researchers' topic modeling has 3 stages. The first stage forms a set of queries based on keywords and their combination. On the second stage we perform topic modeling using each query separately with filtering stop-words and underused words. On the third stage we average membership degrees of all queries and cut the distribution to a few research groups. When dropping minor research groups their contribution is redistributed to the leaders. As a result, we get membership degrees for a researcher to a few research groups that correspond to the set of his interests the most. Such mapping of interests can be viewed as an analog to the word2vec procedure.

We compared topic modeling based on small amount of information from researchers' profiles at Google Scholar with topic modeling based on a few dozens of authored publications categorized by Dimensions. As a result of comparison, we get a good matching of topic modeling results based on different sources of initial information. It allows using the proposed algorithm as the intellectual core of information technology in regards to scientific staff, in particular, for the selection of candidates as opponents of a dissertation, as reviewers for research projects, for forming a team to collaborate on shared research projects etc.

## 8. References

- [1]. A. Martín-Martín, M. Thelwall, E. Orduna-Malea, E.D. López-Cózar, Google Scholar, Microsoft Academic, Scopus, Dimensions, Web of Science, and OpenCitations' COCI: a multidisciplinary comparison of coverage via citations, *Scientometrics* 126 (2021) 871–906. doi: [10.1007/s11192-020-03690-4](https://doi.org/10.1007/s11192-020-03690-4).
- [2]. A. W. Harzing, S. Alakangas, Google Scholar, Scopus and the Web of Science: A longitudinal and cross-disciplinary comparison, *Scientometrics* 106(2) (2016) 787–804. doi: [10.1007/s11192-015-1798-9](https://doi.org/10.1007/s11192-015-1798-9).
- [3]. B. Rahdari, P. Brusilovsky, D. Babichenko, E. B. Littleton, R. Patel, J. Fawcett, Z. Blum, Grapevine: A profile-based exploratory search and recommendation system for finding research advisors, *Proceedings of the Association for Information Science and Technology* 57(1) (2020). doi: [10.1002/pra2.271](https://doi.org/10.1002/pra2.271).
- [4]. J. Saad-Falcon, O. Shaikh, Z. J. Wang, A. P. Wright, S. Richardson, D. H. Chau, PeopleMap: Visualization Tool for Mapping Out Researchers using Natural Language Processing, *arXiv preprint* (2020), arXiv:2006.06105.
- [5]. M. Rosen-Zvi, T. Griffiths, M. Steyvers, P. Smith, The author-topic model for authors and documents, In *Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence*, AUAI Press (2004) 487-494.
- [6]. D. Blei, A. Ng, M. Jordan, Latent Dirichlet allocation, *Journal of Machine Learning Research* 3 (2003) 993-1022.
- [7]. J. Jian, G. Qian, M. Haikun, C. Chong, Author–Subject–Topic model for Reviewer Recommendation, *JIS-Journal of Information Science* 4 (2019). doi: [10.1177/0165551518806116](https://doi.org/10.1177/0165551518806116).
- [8]. D. Mimno, A. McCallum, Expertise modeling for matching papers with reviewers, in: *Proceedings of the 13th ACM SIGKDD international conference on knowledge discovery and data mining, KDD' 07, ACM, San Jose, CA, 2007*. doi: [10.1145/1281192.1281247](https://doi.org/10.1145/1281192.1281247).

- [9]. C. Sun, T. J. King, P. Henville, R. Marchant, Hierarchical Word Mover Distance for Collaboration Recommender System, Springer 996 (2018) 289-302. doi: [10.1007/978-981-13-6661-1\\_23](https://doi.org/10.1007/978-981-13-6661-1_23).
- [10]. K. Xiangjie, J. Huizhen, Y. Zhuo, A. Tolba, X. Zhenzhen, X. Feng, Exploiting Publication Contents and Collaboration Networks for Collaborator Recommendation, PlosOne 11(2): e0148492 (2016). doi: [10.1371/journal.pone.0148492](https://doi.org/10.1371/journal.pone.0148492)
- [11]. Y. Zhao, J. Tang, Z. Du, EFCNN: A Restricted Convolutional Neural Network for Expert Finding, volume 11440 of Lecture Notes in Computer Science, Springer, Cham, 2019. doi: [10.1007/978-3-030-16145-3\\_8](https://doi.org/10.1007/978-3-030-16145-3_8).
- [12]. A. Omer, G. Hongyu, B. Suma, H. Wen-Mei, X. JinJun, PaRe: A Paper Reviewer Matching Approach Using a Common Topic Space, in: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP/IJCNLP), Association for Computational Linguistics, Hong Kong, 2019. doi: [10.18653/v1/D19-1049](https://doi.org/10.18653/v1/D19-1049).
- [13]. T. Mikolov, I. Sutskever, K. Chen, G. Corrado, J. Dean, Distributed representations of words and phrases and their compositionality, Neural Information Processing Systems 2 (2013) 3111–3119.
- [14]. S. Shtovba, M. Petrychko, Jaccard Index-Based Assessing the Similarity of Research Fields in Dimensions, CEUR Workshop Proceedings 2533 (2019) 117-128.
- [15]. S. Shtovba, M. Petrychko, An Informetric Assessment of Various Research Fields Interactions on Base of Categorized Papers in Dimensions, CEUR Workshop Proceedings 2845 (2021) 159-169.

# Intellectual property assurance method for digital university ecosystem based on blockchain technology

Oleksandr Kapliienko<sup>a</sup>, Galyna Tabunshchik<sup>a</sup>, Tetiana Kapliienko<sup>a</sup> and Carsten Wolff<sup>b</sup>

<sup>a</sup> National University "Zaporizhzhia Polytechnic", 64, Zhukovsky Street, Zaporizhzhya, 69063, Ukraine

<sup>b</sup> Fachhochschule Dortmund, 23, Otto-Hahn-Str., Dortmund, 44227, Germany

## Abstract

In the university digital ecosystem it is very important to have a proper process of collecting, processing and storing the data about students educational results which could confirm the level of the achieved outcomes within their study. In the paper authors provide an analysis of the data which should be stored and existing system which are used for diploma verification. According to the authors' research, the strategy for the digital transformation of the high educational institutes education system should include the modernization of the management of the intellectual property and learning outcomes (competencies and skills). To provide an intellectual property assurance for the students' documents there were chosen blockchain technology. Intellectual property assurance method for digital university ecosystem based on blockchain technology was proposed by the authors.

## Keywords 1

Education, Intellectual property, Learning outcomes, Blockchain, Security, Decentralization, eLearning

## 1. Introduction

The modern high educational institutes (HEIs) should ensure a study process in the relevant professions and specialties according to the demands the labor market. The learning process should be comfortable and based on the principles of academic integrity. The digital university infrastructure should ensure safety and security for staff and students, specially ensure the intellectual property (IP) of the created academic materials, as some if it according to the national law and regulation should be stored from 5 to 10 years [1].

An important feature of the educational system is an implementation of the different learning approaches and standards in different HEI even in one country, which also change year by year. Accordingly, the presence of a diploma in the same specialty from different institutions and / or from different years of graduation does not imply the presence of the same learning outcomes (competencies and skills). It is also necessary to take into account that students could choose alternative disciplines and internships within their study program, or have additional achievements in the learning process.

To support a coherent and understandable system of an individual's lifelong learning there should be stored a sizable amount of data that needs to be saved for the diploma assurance.

There are a lot of requirements which are put forward for this data:

- they must be ordered by the time of creation;
- intellectual property and personal data must be protected from external access;

---

CMIS-2021: The Fourth International Workshop on Computer Modeling and Intelligent Systems, April 27, 2021, Zaporizhzhia, Ukraine  
EMAIL: alexandr.kaplienko@gmail.com (O. Kapliienko); galina.tabunshchik@gmail.com (G. Tabunshchik); bragina.zntu@gmail.com (T. Kapliienko); carsten.wolff@fh-dortmund.de (C. Wolff)  
ORCID: 0000-0003-2080-7875 (O. Kapliienko); 0000-0003-1429-5180 (G. Tabunshchik); 0000-0001-8192-2397 (T. Kapliienko); 0000-0003-3646-5240 (C. Wolff)



© 2021 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

- the data that confirm the acquisition of certain skills and competencies must be located in the public domain;
- the data should be stored decentralized to avoid loss problems;
- the data should be protected against falsifications, for example change protection or adding faking data.

One of the options for meeting these requirements can be the use of blockchain technology, which introduces immutability and trust to the decentralize data. This paper presents an approach for saving education data based on the blockchain for ensuring the safety and reliability of storing information about the entire learning process for each student (from admission stage to graduation with a HEIs diploma).

## 2. State of the art

The primary reason for storing of the individual academic materials is the need to confirm the level of education by a certain student (for example, bachelor's, master's degree). Until 2012 in Ukraine, such confirmation was the physical presence of a state diploma, formed by the authorized agency. The disadvantages of this approach were the possibility of forgery or loss of the diploma itself or its supplementary.

In 2012 year, the Ukrainian government introduced the Unified State Electronic Database on Education (USEDE), which is an automated system for collecting, registering, processing, storing and protecting educational information [2]. Since that the diploma as a physical object is not the main goal, the main achievement is the record in this database which confirms the presence of a degree of education of a certain level in a certain specialty.

There are many similar systems and databases in the world that mainly pursue 2 goals:

- obtaining data for statistics [3, 4, 5, 6];
- confirmation of the authenticity of educational documents [7, 8, 9].

For example, the "UIS.Stat" is a product of the UNESCO Institute for Statistics (UIS), which is the official and trusted source of internationally-comparable data on education, science, culture and communication [3]. This system is a comprehensive browser for viewing and downloading the most popular education data and indicators, such as [3]:

- mean years of schooling;
- number and rates of international mobile students (inbound and outbound);
- number of students and enrolment rates by level of education;
- graduation ratio from tertiary education;
- out-of-school children, adolescents and youth (number);
- percentage of graduates by field of education (tertiary education);
- educational expenditure by nature of spending in public educational institutions;
- government expenditure on education (amount);
- survival rate by grade for primary education;
- population of the official age / school age population;
- official entrance age and theoretical duration by level of education (years) and etc.

The OECD.stat database [4] have the similar information provided by the Organization for Economic Co-operation and Development (OECD). In this database there are also information about graduation and entry rates, profile of graduates and new entrants, distribution of graduates and entrants by field, and other indicators. The World Bank EdStats (Education Statistics) portal [5] and Global Education [6] extensive data and analysis source for key topics and a "meta-entry" in education with the visualizations and research. These data make it possible to determine the state of education in a particular country by many different indicators, the progress of development of a particular area, but described systems do not have the opportunity to confirm the level of education of a particular graduate or student.

"DiplomaVerify" and "DegreeVerify" (National Student Clearinghouse Verification Services) [7] gives possibility to instantly verify diplomas for high school graduates of USA (549 educational institutions for this moment) and provides online verifications of college degrees and attendance.

“Nuffic” is the National Academic Recognition Information Centre for Holland [8], the goal of which is to allow to find an equivalent Dutch level of diploma from another country.

The [9] is a service of the "Institut National des Sciences Appliquées" (INCA) of Toulouse that allows to authenticated all INSA diplomas using ProofTag. ProofTag is a security seal and electronic tag found on INSA's diplomas, which contains a unique code that acts as an electronic birth certificate [10].

The need for the presented systems arose due to the presence of progressive and low-cost scanning and printing technologies, which increase the number of fake academic certificates. These systems represent cooperation within a country, organization or institution.

Since document verification is very important to assure that the diploma is official, some organizations have developed and are presenting their approaches to this systematic process. For example, Documentorum is a global initiative and a new approach to the issuance, registration and verification of certificates and diplomas [11]. It was developed using Hyperledger technology (an open-source collaboration created to advance cross-industry blockchain technologies). It allows students to collect all of their academic data in one place, where it is protected from theft and fraud [11]. Each issued document is automatically verified and approved by the relevant parties and is permanently stored in a tamper-proof ledger on the blockchain, so its authenticity can be always guaranteed.

As we see, most of these systems are oriented on the verification of documents and certificates, which were issued by HEIs and other education organizations, and information about these organizations (formal and statistical). But there is lack of information how HEIs could record information about the learning process that leads to the diploma obtaining. As we mention above, to confirm learning outcomes, competencies and skills of each student, HEI should take into account a lot of information, which must be trustable and with the highest degree of protection against counterfeiting.

For example, storing academic information in learning management systems (LMS) like Moodle does not guarantee that there is no threat of data loss. Conversely, the exclusion of the users who successfully passed the course or training leads to data removal in the system. And even if all information is archived, there is no guarantee of the reliability of the data due to the fact that the system administrator has full access to all data and can change it on the demand.

Since the method of storing data in the blockchain is devoid of these disadvantages, let's look at examples of blockchain technology implementation in the education applications.

There are a number of ways for blockchain implementation in the education field, such as full record of learning trajectory, trusted certification of learning results, decentralized sharing of education resources, etc. The greatest advantage of this network is a decentralized peer-to-peer infrastructure, supporting trust, accountability, transparency, identity management and openness.

In [12] a case study on the decentralization of lifelong learning using blockchain technology was explored. In this case study, the Semantic Blockchain was used as a solution to combine all acquired learning and accreditation experiences to form a holistic picture of a person's lifelong learning. The purpose of this study is to offer transparent and permanent educational accreditation for learners throughout their lives and to support them in their personal and professional progress.

In [13] blockchain technology was used to increase the flexibility of organization the student activity and the results obtained via managing the attendance and the grades obtained by students.

Article [14] describes the problems of online education such as lack of results certification, poor confidentiality and lack of an exchange mechanism. The authors combine blockchain technology with online learning to address these challenges, creating a smart, decentralized and shared online education system.

The article [15] describes research experiments on Smart Pedagogy with a lifelong learning transcript called the passport of knowledge in blockchain architecture. The proposed scenario allows students to publish proof of their academic achievement in blockchain-related formats for instant authentication. The paper describes the experimental network, which consists of three nodes (located in Latvia, in the USA and in Asia). Based on the proposed implementation, this approach validates training evidence much better, eliminates passport fraud, and reduces organizational overhead for stakeholders involved in verifying human knowledge documents.



The authors of the article "Blockchain, IoT and Fog Computing for Smart Education Management" [16] presented a great summary of the literature review of these techniques and, based on this review, applied modern technologies as an infrastructure for the development of educational innovations. Blockchain is used for recording and storing various collaborative data based on consensus in educational institutions, for developing teaching and learning on a digital platform, for storing information about student history, educational history in different grades, registration information, curriculum participation and institutional activities, academic performance and a certificate of education [16]. The Internet of Things and Fog Computing are used to develop infrastructure with the ability to create intelligent educational environments to support and quickly answer questions about the use of all types of equipment for real-time control. They allow to organize and share resources for collaboration work and training to gain access to the organization without restrictions on time or place.

Thus, the usage of blockchain technology in education is widely discussing for the last 5 years in the whole world, and in Ukraine particularly [17]. There is no doubt about the need to develop a strategy for the digital transformation of university education [18] with the modernization of the management of the corporate ecosystem, which should be implemented as a cloud platform. In this article, it is planned to streamline the considered methods and approaches to address the issue of creating an approach for saving intellectual and personal data involved in the training process in a reliable storage protected from the possibility of forgery and unauthorized access.

### 3. An approach for intellectual property assurance in digital university ecosystem

The process of transforming the educational environment in the format of "Smart Environment" [19] and striving for the realization of the idea of "Smart Pedagogy" is a new trend in the global educational field. Smart Pedagogy is a learning process in technology, different forms of information and media-enriched environment. This led to the fact that there is a lot of digital information which should be stored while the learning process in HEI and for a long time after its completion (Figure 1).

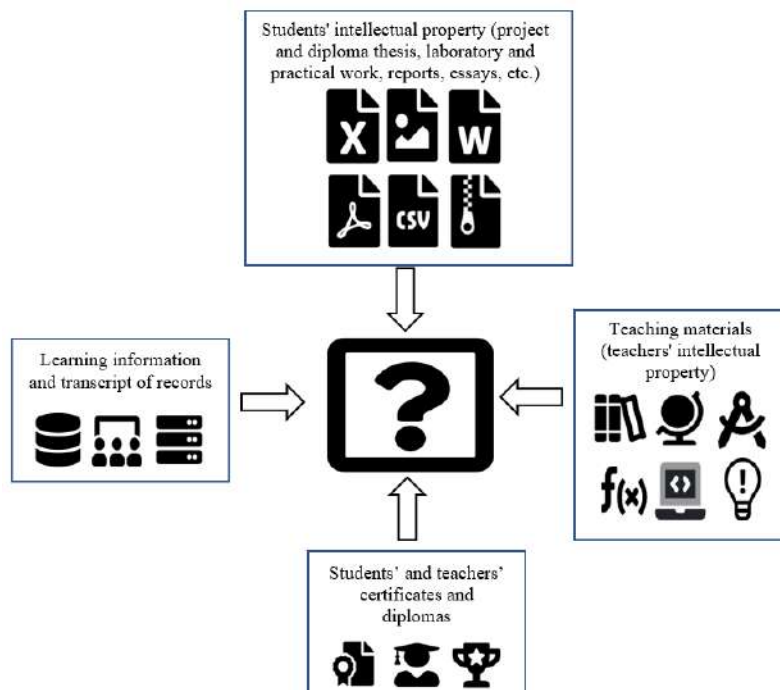


**Figure 1:** Data for integration in the university digital infrastructure

In the university digital ecosystem, it is very important to have a proper process of collecting, processing and storing the data about students' educational results which could confirm the level of the achieved outcomes within their study. The authors of the analyzed publications [12 - 17] agree about the need of using the blockchain to provide a reliable information network.

The blockchain is a database made up of a series of fixed-length blocks ordered over time, which could be integrated with applied information and communication technology in learning and teaching processes [20]. Each block contains one or more confirmed transactions. After a block has been verified and completed, it is added to the blockchain and all blocks are made public. The process of adding a transaction consists of:

1. Adding a transaction to a block. On average, blocks contain approximately 1 MB of information and are time stamped;
2. Cryptographic verification of each transaction;
3. Adding a new block after completion at the end of the block row, no transaction in the block can be deleted or changed (unchangeable) (as shown at the Figure 2, adding a new block fixes the hash of information from the previous block, which excludes the possibility of changing information in the previous block in the future).

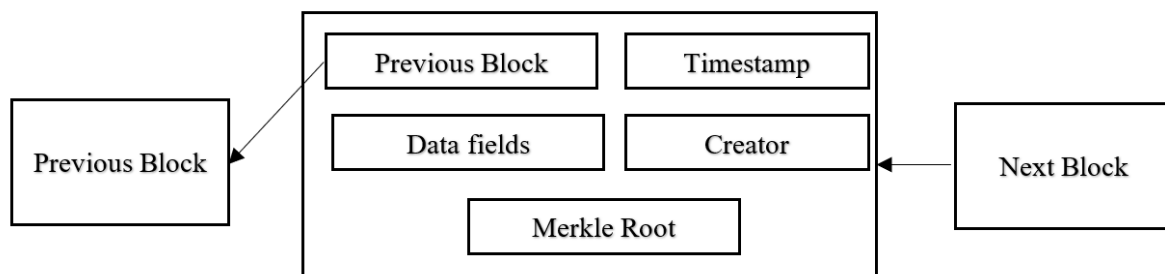


**Figure 2:** Blockchain structure

A Merkle root is a part of the block header and a simple mathematical way to verify the data on a Merkle tree. Merkle trees are an essential part of blockchain technology. A Merkle tree is a structure that allows for efficient and secure verification of content in a large body of data. This structure helps verify the sequence and content of the data. A Merkle tree summarizes all the transactions in a block by producing a digital fingerprint of the entire set of transactions, thus enabling a user to verify whether or not a transaction is included in a block [21].

The Merkle tree will be created based on the decision about what data and in what form should be stored in the blockchain.

Further consider the model for implementation of blockchain technology for intellectual property assurance in the digital university ecosystem.

### **3.1. Intellectual property assurance method for digital university ecosystem based on blockchain technology**

Intellectual property assurance method for digital university ecosystem based on blockchain technology consists of the following steps:

- Step 1: Elicitation of the classes of the storing information. In this phase, all type of data which should be stored in the blockchain should be identified and separated into public and confidential.
- Step 2: Deployment of files storing and implementation of data transferring process into a file server.
- Step 3: Selection of the method and keys for encrypting private information and files.
- Step 4: Creating an interface system for uploading files and information into the blockchain with the LMS integration.
- Step 5: Storing information into the blockchain.

The first step is aimed to determine the information that needs to be stored to confirm the acquired skills and educational level, as well as for the management of the student's learning process itself.

Upon admission, the applicant must provide a document on previously received education and its supplementary, all required certificates, identity documents, a personal photo and other necessary documents. These documents must be stored in an encrypted way during the entire educational process, since contain confidential information.

After students are enrolled, the admission information is created – study degree, specialty/ educational program, form of study, type of financing, list of courses, etc. This information is public and must be available without encryption to confirm the enrollment process. The document that confirms this process is the enrollment order, a link to which can be attached to the enrollment record (the document must be publicly available).

The training process generates the following set of files and data:

- methodological support (should be publicly available with a link to authorship);
- learning outcomes in the form of completed project and diploma thesis (should be publicly available with a link to the owner), laboratory and practical work, reports, essays, etc. (must be in private access with a link to authorship);
- transcript of records – marks for all types of control and types of assignments (must be in private access with a link to the student and teacher);
- received learning outcomes (competencies and skills) (must be publicly available with a link to the student);
- student's and teacher's certificates and diplomas (should be publicly available with a link to the owner).

In the second step, it is proposed to store the files on a backup server, where each file will be assigned a unique number. Thus, a link to the file uploaded to the server will be saved in the blockchain. Depending on the file type, it will be stored in its original or encrypted form.

To encrypt private files, in the third step a unique key should be chosen, ideally which will uniquely identify the person who has access to this information. In each particular HEI, this key can be selected based on their own preferences. As an example, in Ukrainian HEI it is proposed to use identifiers from USEDE, i.e., person code in USEDE was chosen as the key for encrypting personal data and documents and education code in USEDE was chosen as the key for encrypting learning data and documents.

This choice depends on following factors:

- until the student is enrolled, there is no education code, but there is already a need to preserve the applicant's documents;
- a person can study in several programs (at the same time or not) and thus the use of an education code will allow separating this information, but at the same time personal information will not be duplicated (it is tied to the person code in USEDE);
- access to the person code and the training code has a limited circle of persons in the learning process (university administration);
- the graduate gets access to the education code (it is registered to the diploma supplement) for further demonstration of the training outcomes to employers.

In step 4 should be realized the integration with the used training platform to avoid data duplication and integrate the process of saving files and data into the blockchain without the need for additional actions from the participants of the educational process.

In the fifth step information which was elicitation at the first step should be saved into blockchain.

### **3.2. Implementation of the blockchain for intellectual property assurance method**

A framework of the blockchain consist of the next layers, key ones of which will be described below:

- Data layer;

- Network layer;
- Consensus layer;
- Incentive layer;
- Contract layer;
- Application layer.

### 3.2.1. Data layer

At the data layer storing information about data block, chain structure, timestamp, hash function, Merkle tree and the encryption. The information shown in Figure 3 is entered into the blockchain transaction. The proposed blockchain architecture consists of the transaction head and the transaction data. The transaction data contains the nine columns where the data field carries information about the education process, such as the Creator, Leaner (person code), Education id (can be null), BLOB data (Smart Pedagogy Data), Transaction Type, File unique ID (from the file server), File hash (from the file server via SHA-256 cryptographic hash function), Outcomes, Mark. And the transaction head contains the four columns, such as the previous transaction key, the current transaction key, Merkle root (hash) and Date of creation (timestamp).

Transaction Type can be one of:

- Personal document (id = 1);
- Education document (id = 2);
- Intellectual property (learning outcome) (id = 3);
- Mark for learning activity (competencies and skills) (id = 4).

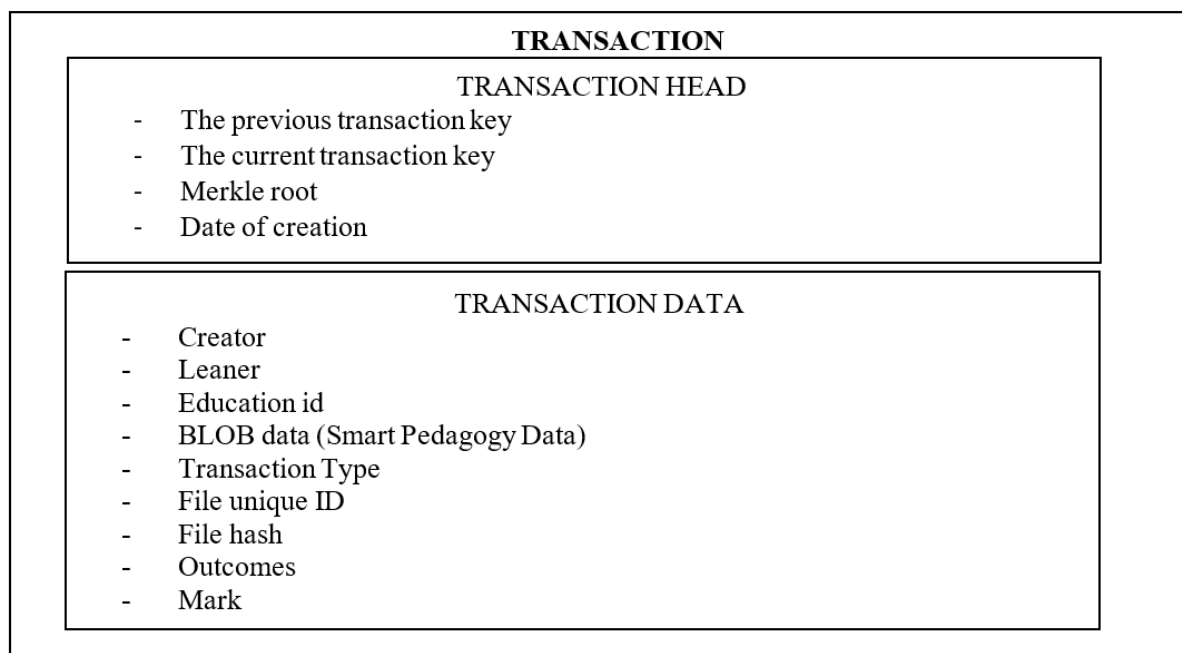


**Figure 3:** Transaction structure

For producing a digital fingerprint of the transaction propose to use the hash of the next data (Figure 4) using SHA-256 cryptographic hash function [22].

In the situation when the learner completed any activity for the second time, some documents should be updated on the server or any information should be changed, and results which were published in the blockchain before, this updated data will be uploaded and added into the blockchain again. Then all other nodes would have both the previous and actual assessment values for the same learner but with different timestamps (so the task of choosing the actual information will be the easiest).

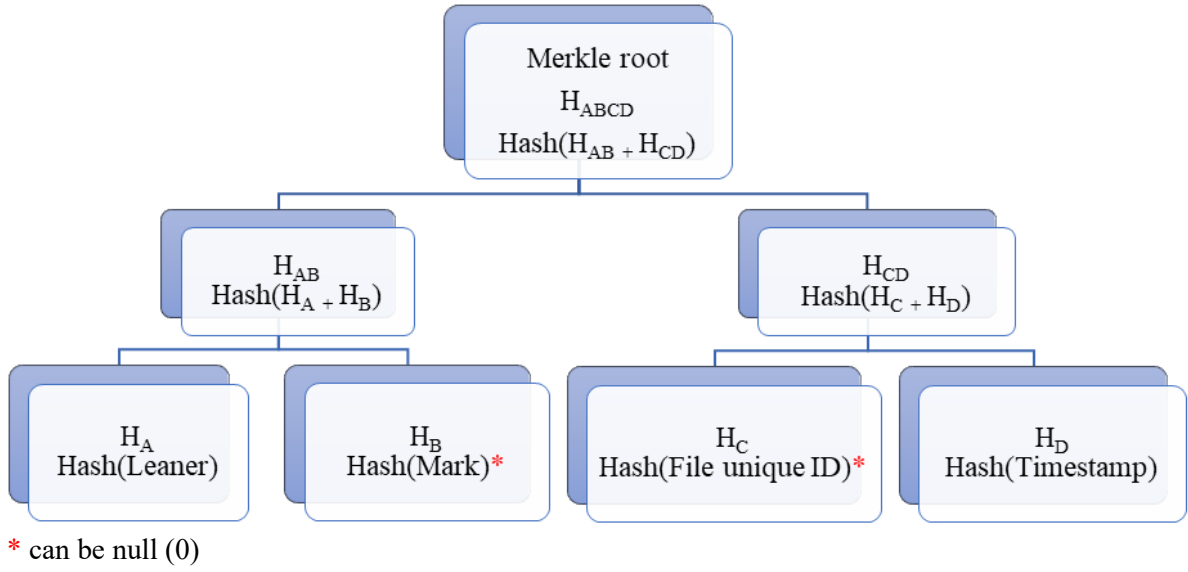


**Figure 4:** Merkle tree for learning information transaction

### 3.2.2. Consensus layer

The security of blockchain technology is achieved by the usage of the consensus algorithm. There are different types of consensus algorithms that might be executed before inserting a transaction in a block, comparison of which is presented in Table 1.

Based on the analysis of the advantage and disadvantages described in the Table I, the Delegated Proof-of-Stake (DPoS) consensus algorithm was chosen. According to this algorithm, selected delegates make the validation of the blocks by themselves. A relatively small number of delegates (about a couple dozen) can organize themselves efficiently and create designed time slots to publish blocks. With a collaborative effort and a partially centralized process, DPoS has been able to run orders of magnitude which are faster than any other consensus algorithms [23]. Also disadvantage of a partially centralized process in case of HEI's blockchain organization become the advantage on practice. Designing trusted delegates owned by the HEI as elected delegates will prevent a 51% attack when an attacker or a group of attackers gain control of 51% or more of the computing power or hash rate [24].

**Table 1**

## Consensus algorithms comparison

| Consensus algorithms name                         | Basis                                                                                                                                                     | Disadvantages                                                                                                                                   | Advantages                                                                                                                                              |
|---------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------|
| Proof-of-Work (PoW)                               | Complex mathematical calculations                                                                                                                         | High CPU computations, is energy-intensive and costly                                                                                           | Highly scalable                                                                                                                                         |
| Proof-of-Stake (PoS)                              | The PoS randomly selects validators, the higher chances of getting have the most coin holders                                                             | Complex implementations, vulnerable, only the richest can have control of the consensus                                                         | Higher speed, lesser energy consumption and hardware requirements                                                                                       |
| Delegated Proof-of-Stake (DPoS)                   | Selected delegates make the validation                                                                                                                    | A partially centralized process                                                                                                                 | Offers all the benefits of the PoS, faster than any other consensus algorithms, secure real-time voting                                                 |
| Byzantine Fault Tolerance (BFT)                   | Nodes regularly vote in order to identify the true transaction                                                                                            | There is no central authority that can step in to correct it, low scalability, semi-centralized system                                          | High throughput, cost-efficiency, smart-contract support                                                                                                |
| Practical Byzantine Fault Tolerance (PBFT), SIEVE | This computation process asks the individual general about the opinion on the message                                                                     | Susceptible to Sybil attacks, does not scale well                                                                                               | Forces a low overhead on the performance of the replicated service, energy efficiency, transaction finality, low reward variance                        |
| Proof-of-Weight (PoW)                             | Mechanism gives users a 'weight' based on how much cryptocurrency they are holding                                                                        | Incentivization can be hard                                                                                                                     | Energy efficient, highly customizable and scalable                                                                                                      |
| Proof-of-Burn (PoB)                               | By committing coins to an address where they are irretrievable, user earns a lifetime privilege to mine on the system based on a random selection process | The protocol wastes resources needlessly, mining power goes to those who are willing to burn more money                                         | An interesting alternative to Proof-of-Work                                                                                                             |
| Proof of Authority (PoA)                          | An optimized Proof-of-Stake model that leverages identity as a form of stake rather than actually staking tokens                                          | Strongly lack decentralization, the risk of damaging the reputation does not necessarily keeps a person from participating in malicious actions | Low requirements for computational power, no requirement for communication between nodes in order to reach consensus, and the continuity of the network |

### 3.2.3. Application layer

At the application level, the principle of interaction between applications and the blockchain is described. The process of saving student project thesis with IP data in the blockchain is shown in Figure 5.

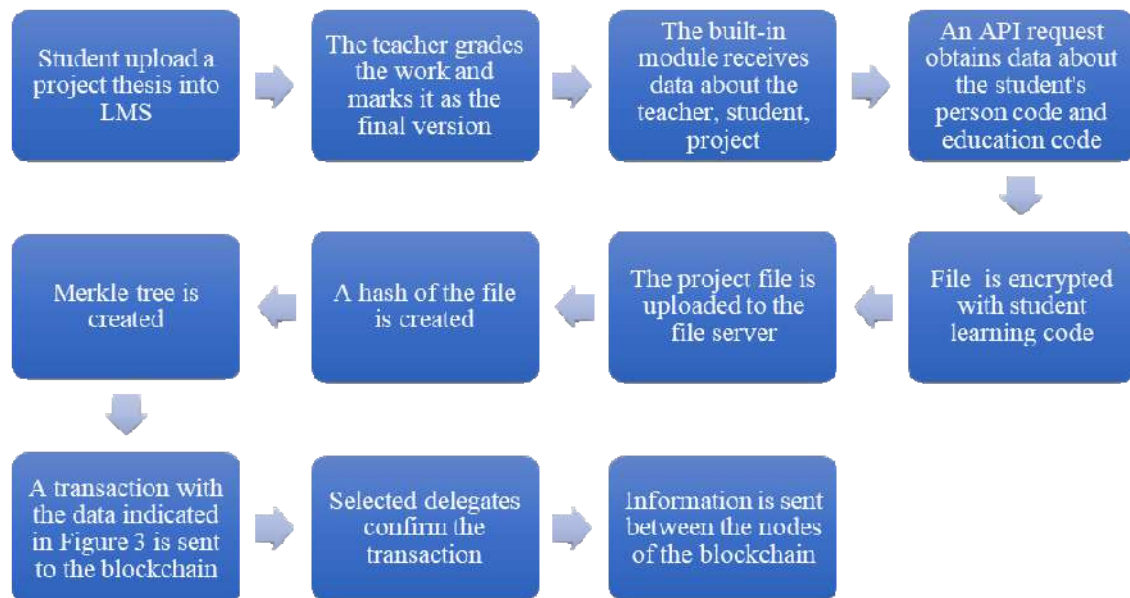


**Figure 5:** An example of loading a project thesis into the blockchain of intellectual property of HEI's participants

For application layer realization an interface system for uploading files and information into the blockchain should be developed. One of the features of this system is the LMS integration and API integration with the database with information about students and teachers. The LMS integration should be used for a simple, protected and user-friendly process of data and files transfer. API integration with the database is requested to obtain data about the student's person code and education code. The specification for the development of this system will depend on the LMS used in HEI and the access to the database with students' and teachers' information.

Taking into account the systems used in National University "Zaporizhzhia Polytechnic", the following application integration is planned (Figure 6). It should integrate information from the university Moodle system, USEDE and data from the ECS eLearning Community Server, which is supporting international projects ViMaCs: Virtual Master Cooperation Data Science(ID: 57513461) [25] and Cross-domain competences for healthy and safe work in the 21st century WOR4CE [26].

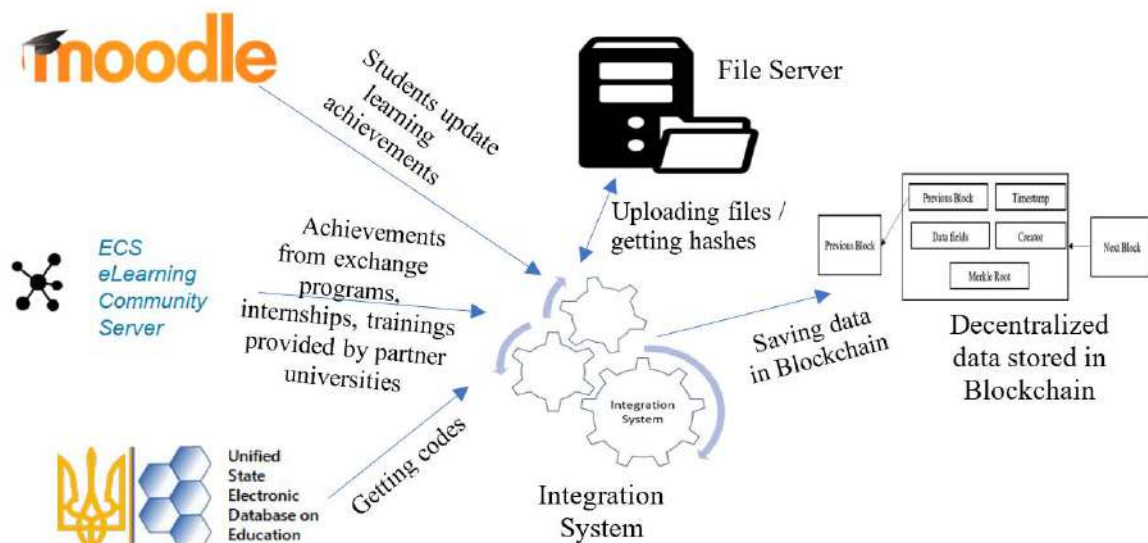


**Figure 6:** An example of the Application layer

## 4. Discussion

The proposed method can be used as a base for integration the data about students' educational achieved outcomes within their study and IP. In line with the hypothesis that digital university ecosystems are going to grow and expand the area of penetration, the application of this method can be used for different purposes, affording transparency and secure data archiving into infrastructure. The further implementation required integration of data from different sources so the cloud infrastructure of the ecosystem is also of great importance.

## 5. Acknowledgements

This research was partly done within the framework of international project “Cross-domain competences for healthy and safe work in the 21st century” (Ref. no. 619034-EPP-1-2020-1-UA-EPPKA2-CBHE-JP).

## 6. Conclusions

Intellectual property assurance method for digital university ecosystem based on blockchain technology was proposed in the article. The implementation of the proposed approach in different HEIs could become the basis for creation of the global cloud platform for intellectual property. It aimed to support the principles of academic integrity, afford transparency and secure archiving for intellectual property into university digital infrastructure. Based on the data stored in the blockchain it will be easy to confirm the presence of a diploma, learning competencies and skills, gained in the educational process, particularly the results achieved within alternative disciplines, training, internship as well as additional individual educational achievements.

## 7. References

- [1] Order of the Ministry of Justice of Ukraine dated 12.04.2012 No. 587/5, 2012. URL: <https://zakon.rada.gov.ua/laws/show/z0571-12#Text>.
- [2] About USEDE, 2020. URL: <https://info.edbo.gov.ua/about>.
- [3] Welcome to UIS.Stat, 2021. URL: <http://data.uis.unesco.org>.



- [4] Indicators databases, 2021. URL: <https://www.oecd.org/education/database.htm>.
- [5] Education Statistics (EdStats), 2021. URL: <https://datatopics.worldbank.org/education/indicators>.
- [6] Global Education, 2021. URL: <https://ourworldindata.org/global-education>.
- [7] National Student Clearinghouse Verification Services, 2021. URL: <https://nscverifications.org>.
- [8] Nuffic. The Dutch organisation for internationalisation in education, 2021. URL: <https://www.nuffic.nl/en>.
- [9] How to verify the authenticity of a diploma, 2021. URL: <http://www.insa-toulouse.fr/en/industrial/how-to-verify-the-authenticity-of-a-diploma.html>.
- [10] ProofTag, 2021. URL: <http://www.prooftag.net>.
- [11] Documentorum, 2021. URL: <https://exelior.net/documentorum>.
- [12] A. Mikroyannidis, A. Third, J. Domingue, A Case Study on the Decentralisation of Lifelong Learning Using Blockchain Technology, *Journal of Interactive Media in Education* 1 (2020) 23–32. doi: 10.5334/jime.591.
- [13] P. L. Ogrutan, An Analysis on the Opportunity of Introducing Blockchain Technology in Education - A Case Study, *TEM Journal* (2020) 971-976. doi:10.18421/TEM93-19.
- [14] H. Sun, X. Wang, X. Wang, Application of Blockchain Technology in Online Education, *International Journal of Emerging Technologies in Learning (iJET)* 13 (2018) 252-259. doi:10.3991/ijet.v13i10.9455.
- [15] A. Arins, Blockchain architecture in Smart pedagogy, in: *Innovations, Technologies and Research in Education*, 2019, pp. 207-217. doi: 10.22364/atee.2019.itre.13.
- [16] S. Chaiyarak, A. Koednet, P. Nilsook, Blockchain, IoT and Fog Computing for Smart Education Management, *International journal of education and information technologies* 14 (2020) 52-61. doi: 10.46300/9109.2020.14.7.
- [17] O. Kaminskyi, Y. Yereshko, S. Kyrychenko, Digital transformation of university education in ukraine: trajectories of development in the conditions of new technological and economic order, *Information Technologies and Learning Tools* 64 (2018) 128-137. doi: 10.33407/itlt.v64i2.2083.
- [18] G. Tabunshchyk, C. Wolff, P. Arras, Development of Cross-Domain Competences for Work 4.0, Springer Nature Switzerland AG 2021, M. E. Auer and T. Rüttmann (Eds.): *ICL 2020, AISC 1328*, 2021, pp. 205–211. doi:10.1007/978-3-030-68198-2\_18.
- [19] O. Kapliienko, S. Tabunshchyk, G. Tabunshchyk, T. Kapliienko, S. Sylenko, Virtual Reality Implementation for Design of Warehouse Lighting, in: *The 10th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications*, 18-21 September, 2019, Metz, France, pp. 969-973. doi: 10.1109/IDAACS.2019.8924272.
- [20] O. Petrova, G. Tabunshchyk, T. Kapliienko, O. Kapliienko, Fuzzy Verification Method for Indoor-Navigation Systems, in: *Proceedings of 14th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering (TCSET)*, Lviv-Slavske, Ukraine, February, 20 – 24, 2018, pp. 65 – 68. doi: 10.1109/TCSET.2018.8336157.
- [21] Merkle Trees, 2017. URL: <https://medium.com/hackernoon/merkle-trees-181cb4bc30b4>
- [22] H. Handschuh, SHA Family (Secure Hash Algorithm), in: van Tilborg H.C.A. (eds) *Encyclopedia of Cryptography and Security*. Springer, Boston, MA, 2005. doi: 10.1007/0-387-23483-7\_388
- [23] What is Consensus Algorithm In Blockchain & Different Types Of Consensus Models, 2018. URL: <https://medium.com/@BangBitTech/what-is-consensus-algorithm-in-blockchain-different-types-of-consensus-models-12cce443fc77>.
- [24] Understanding 51% Attacks On Blockchains, 2020. URL: <https://www.jumpstartmag.com/understanding-51-attacks-on-blockchains/>
- [25] ViMaCs: Virtual Master Cooperation Data Science URL: <https://go-study-europe.de/vimacs/>
- [26] WORK4CE: Cross-domain competences for healthy and safe work in the 21st century. URL: <http://work4ce.eu/>

# Steganographic Resources of FPGA-based Systems for Approximate Data Processing

Kostiantyn Zashcholkin<sup>a</sup>, Oleksandr Drozd<sup>a</sup>, Svetlana Antoshchuk<sup>a</sup>, Olena Ivanova<sup>a</sup> and Oleg Sachenko<sup>b</sup>

<sup>a</sup> Odessa National Polytechnic University 1, Ave. Shevchenko, Odessa, 65044, Ukraine

<sup>b</sup> West Ukrainian National University, 11, Lvivska Str., Ternopil, 46009, Ukraine

## Abstract

The program code controls the operation of FPGA (Field Programmable Gate Arrays) chips. Through the use of program code it is possible to interfere with the operation of these chips. Operational monitoring (integrity, authenticity, distribution paths) of program code is one of the most effective means of preventing illegal interference with FPGA-based systems. Traditional methods of monitoring program code have the problem of storing control data. A promising approach is to embed the monitoring data into the FPGA program code in a steganographic manner. In this case, the fact of the availability of monitoring data and the fact of monitoring is not obvious to an external observer. There are methods for steganographic embedding of additional data into the FPGA program code. Methods of this kind are based on equivalent transformations of the program code. This paper proposes and experimentally substantiates the presence in the FPGA program code of new resources for steganographic data embedding. These resources arise in the programming code of FPGA systems that perform approximate data processing. In the paper, inessential LUT (Look Up Table) FPGA units are classified as such resources. These units are involved in the calculation of the least significant (discarded during rounding) bits of the result and do not participate in the calculation of the most significant (retained) bits. Also inessential bits of the program code of LUTs are referred to new resources. These bits are bits that are not accessed during the calculation of the result of the operation. The experiments performed have shown the possibility of organizing stego containers in the FPGA program code using the identified additional resources. The volume of such stego containers has been estimated. It is shown that this volume is sufficient for storing monitoring data, which are used for operational monitoring of the FPGA program code.

## Keywords 1

Steganographic approach, FPGA-Based Systems, program code of FPGA, integrity monitoring, LUT-oriented architecture, stego container, floating-point format, approximate data processing, complete arithmetic operation

## 1. Introduction

Programmable components are now very widely used in computer systems. The ability to programmatically control the operation of programmable chips provides significant advantages over specialized chips. The main advantage is that the operation of programmable components can be modified at any time. This makes it possible:

- eliminate errors found in the program code after putting the system into operation;
- improve the functioning of the component at any stage of system lifecycle;
- adapt the functioning of the device to changes in its operating conditions.

CMIS-2021: The Fourth International Workshop on Computer Modeling and Intelligent Systems, April 27, 2021, Zaporizhzhia, Ukraine  
EMAIL: const-z@te.net.ua (K. Zashcholkin); drozd@ukr.net (O. Drozd); asgonpu@gmail.com (S. Antoshchuk); en.ivanova.ua@gmail.com (O. Ivanova); olsachenko231@gmail.com (O. Sachenko)  
ORCID: 0000-0003-0427-9005 (K. Zashcholkin); 0000-0003-2191-6758 (O. Drozd); 0000-0002-9346-145X (S. Antoshchuk); 0000-0002-4743-6931 (O. Ivanova); 0000-0001-9337-8341 (O. Sachenko)



© 2020 Copyright for this paper by its authors.  
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).  
CEUR Workshop Proceedings (CEUR-WS.org)

All these modifications can be done by replacing the program code or parts of it. However, this ease of modifying the functioning of a programmable component also has a negative side. It consists in the possibility of malicious interference in the functioning of a computer system through the program code.

One of the most effective ways to counteract unauthorized actions against program code is operational monitoring [1]. There are considerable developments (methods, means, tools) which provide different variants of program code monitoring: integrity monitoring [2]; authenticity monitoring [3]; distribution monitoring [4]; monitoring the legitimacy of the use [5] of the program code. All of these types of monitoring are based on the well-researched mathematical apparatus of hash functions [6, 7] and message authentication codes [8, 9]. Based on the methods built on this mathematical apparatus, monitoring data are created with the help of which monitoring is performed.

However, all existing types of monitoring have a problem. This problem is the way monitoring data is stored and accessed. There are three most common approaches to storing monitoring data:

1. attaching monitoring data to program code [10];
2. storage of monitoring data together with the program code in memory or in file system [11];
3. storage of monitoring data in a remote database (monitoring data are requested from the database at the time of monitoring) [12].

All three of the above approaches to storing monitoring data make open the fact of the availability of monitoring data and the fact that monitoring is performed [13, 14]. The protection of monitoring data (including by means of encryption) is leveled by the openness of the fact of their presence. Knowledge of this fact creates the possibility of manipulating monitoring data to falsify monitoring.

There is an approach to storing monitoring data that differs from these common approaches. This approach is based on the use of the theory of digital steganography [15]. Steganography aims to protect data by secretly embedding (hiding) data of one type in data of another type. Protected data are embedded in a stego container [16], which is a specially prepared information object. When using the steganographic approach, the very fact that the data is protected becomes hidden. The main requirement for this approach is the absence of any changes (functional, visual, structural) in the stego container after the embedding of additional data into it.

The use of digital steganography methods for storing monitoring data is as follows. The monitoring data is calculated in the traditional way. After that, they are embedded in the program code in a steganographic manner. At the same time, the embedded monitoring data must not affect the functioning of the device which is programmed by this program code. Also, the embedding should not lead to structural or visual changes in the program code. Thus, the program code acts as a stego container, and monitoring data acts as additional data embedded into this container. With this approach, the very existence of monitoring data is hidden from the outside observer. The fact that monitoring is carried out in relation to the program code is also hidden. At the time of the active monitoring stage, the monitoring data embedded in the program code is extracted. Extraction is possible if there is a stego key that describes the rules for localizing additional data in the program code space.

The approach to steganographic storage of monitoring data appeared later than other approaches to solving this problem. Because of this, traditional approaches, while having disadvantages, are used much more frequently. The steganographic approach has the prospect of taking a significant place along with the traditional approaches. Therefore, the research and development of stenographic methods and means of storing monitoring data is an important and urgent task.

## **2. Literature Review and Goal of the Paper**

Among the programmable components of computer systems a significant place belongs to FPGA chips [17, 18]. These chips have a programming principle that differs from microprocessors and microcontrollers. Microprocessors change their behavior as a result of changing the program. However, in this case, the structure of the chip of the microprocessor remains unchanged. FPGA chips have a different programming principle. An FPGA chip consists of a large number of programmable elementary units. Modification the FPGA program code leads to the reconfiguration of the elementary units and changes in the connections between them. That is, the program code is able to change the structure of the FPGA chip [19].

In addition, the difference between microprocessors and FPGAs lies in the centralization of computations. In microprocessors, the computing process is concentrated in one or more central computing nodes. In FPGA, the computation process is distributed over a large number (hundreds of thousands, millions, tens of millions, depending on the generation of the FPGA chip) of elementary computing units. These features determine the higher performance of FPGA in comparison with microprocessors [20].

High computing performance and the possibility of reprogramming have led to the fact that FPGAs are often used as components of safety-critical systems [21, 22]. Control systems for power engineering, high-speed transport, and aerospace objects often contain FPGA components. The ability to change the program code for the components of such systems has two sides. On the one hand, thanks to this feature, the functioning of the component can be improved at any time and errors found can be eliminated. On the other hand, this possibility creates the danger of destructive influences on the functioning of the system. Under these conditions, operational monitoring of the program code (its integrity, authenticity, ways of its distribution) of FPGA chips is a very urgent task.

As already noted, the problematic place of methods for program code monitoring is the way of storing control data and the way of accessing them. One of the most effective approaches to storing monitoring data is the steganographic approach. This approach has received the greatest development for multimedia stego containers such as bitmaps, digital video and digital audio [23]. This is due to the fact that these containers are of analog origin. Digital representations of images, video and sound are obtained by registration analog phenomena. The results of this registration have been digitized and presented in the appropriate multimedia formats. As a result of such transformations, the data of elementary units (pixels, samples) of multimedia containers have an approximate representation. This means that minor distortions of this data do not lead to degradation of the container.

The low-order bits of the elementary units of the container in the spatial or transformation domain have such a small contribution to the basic function of the container that their distortion cannot be detected. Therefore, with regard to multimedia containers, steganographic embedding of additional data is efficiently implemented through non-equivalent transformations. Namely, the low-order bits of the elementary units of the spatial or frequency domains of the container are used as storage for steganographic data [24, 25].

The program code (and in particular the program code of FPGA chips) is a container with exact data. Distortions of the program code bits lead to distortions in the functioning of the device which is programmed by this code. For steganographic embedding of additional data into the FPGA program code, methods of equivalent transformation are used [26, 27]. These methods do not change the functioning of the chip and the size of program code, but allow to embedding additional data into this code.

However, when implementing approximate data processing on FPGAs, this approximation is indirectly transferred to the exact represented program code. Some of the FPGA elementary units are involved in the calculation of only least significant bits of the results [28]. The program code of such units can be changed non-equivalently for the embedding of additional data into it. Moreover, such an unequal embedding will not affect the correctness of the function of device.

Thus, in the case of approximate data processing, the FPGA program code has the potential to apply non-equivalent transformations to it. These transformations are similar to those used for multimedia containers. However, such distorting transformations do not affect the main function of the code [29].

Based on this, the *goal of this paper* is to identify and estimate the volume of resources of the FPGA program code, which allow performing non-equivalent embedding of additional data. The identification of these resources in this paper is performed for FPGA chips, which carry out approximate data processing. Resource estimation is done for the task of steganographic storage of monitoring data necessary to monitor the FPGA program code.

### **3. The proposed approach to the use of FPGA program code resources for non-equivalent steganographic embedding**

Many operations that are used in approximate data processing have the same operand and result format. For example, specifications for floating-point arithmetic operations in most cases dictate the

same format and size for operands and results of operations. However, a situation often arises in which the result of an operation is larger than the size of the operands. In this case, the size of the result is brought to the size of the operands by discarding a certain number of low-order bits, followed by rounding [30, 31].

So, for example, one of the main operations – the operation of multiplying floating-point numbers, in accordance with most specifications, forms the result mantissa, the size of which is equal to the size of the operand mantissa. Multiplying the mantissa of operands that are  $n$ -bits in size naturally produces a result which contains  $2n$ -bits. However, the specification of the multiplication operation requires the mantissa to have the same size as the mantissa of the operands, that is,  $n$ -bits. To bring the mantissa of the result to the required format, rounding is performed, which consists in discarding the  $n$  least significant bits [32, 33].

The main elementary units of the structure of FPGA microcircuits are LUT (Look Up Table) units. In modern FPGA series, the number of such units is in the millions and tens of millions. The LUT unit has  $m$  inputs ( $m$  for different FPGA families ranges from 4 to 8) and performs the calculation of one logic function from  $m$  variables. LUTs are programmable units. Setting up a unit to implement a specific logical function is performed by a  $2^m$  bit binary program code. Programming the FPGA microcircuit leads to tuning each of the LUT units to implement the required logic function and creating the required connections of such units with each other.

### 3.1. The first kind of resources for non-equivalent data embedding into the FPGA program code

The structure of the implementation on FPGA of arithmetic operations performed in approximate data processing is focused on calculating the full result with its subsequent reduction to the required format. In floating-point arithmetic, converting the result to the format of the operands is performed in two stages: first, the complete result is calculated, and then its incorrect bits are discarded during the rounding process.

Let  $A = \langle a_n, a_{n-1}, \dots, a_1 \rangle$  and  $B = \langle b_n, b_{n-1}, \dots, b_1 \rangle$  these are the mantissa of the operands. The result of the multiplication is the complete  $2n$ -bit product  $C = \langle c_{2n}, c_{2n-1}, \dots, c_{n+1}, c_n, c_{n-1}, \dots, c_1 \rangle$ . Rounding is performed to bring the result to the format of the operands. As a result is formed  $C_{Result} = \langle c_{2n}, c_{2n-2}, \dots, c_{n+1} \rangle$ , which consists of the most significant  $n$  bits of the complete result. In this case, the multiplier circuit calculates both the most significant bits of the result and the least significant bits (discarded) bits of the result.

Based on this, it can be expected that the structure of the FPGA system can contain a subset of LUT units that perform the calculation of only the discarded least significant bits of the result and do not participate in the calculation of the retained most significant bits. If such LUT units are present in the FPGA system, then their program code can be corrupted, and this corruption will not lead to an incorrect result. These LUT units will hereinafter be referred to as inessential LUTs. The rest of the LUT units have a direct impact on the formation of the rounded result and therefore belong to the essential ones.

Under these conditions, the program code of inessential LUT units (in whole or its individual bits) can be used as a storage of secret data, embedded in the FPGA program code in a steganographic way.

Thus, inessential LUTs represent *the first kind of FPGA code resources* that can be used to unequally embed additional data. The experimental part of this paper aims to confirm the availability of this resource and estimate its volume.

### 3.2. The second kind of resources

When performing approximate arithmetic operations, not only LUT units can be inessential, but also individual bits of the program code of LUTs. We will consider as inessential bits of the program code of the LUTs the bits that are not accessed during the calculation of the most significant bits of the result. It is useful to highlight such bits only in the program code of the essential LUT units. This

is due to the fact that the program code of inessential units has been identified by us as an independent type of steganographic resource.

If the program code of the LUT units contains bits that are not accessed during the calculation of the most significant bits of the result, then these inessential bits can be corrupted. Such a distortion will not affect the correctness of the calculations when forming the rounded result of the operation.

Based on this, we define the inessential bits of the program code of the essential LUTs as the *second kind of resources* that can be used for unequal embedding of additional data. The experimental part of this paper aims to confirm the availability of this resource and estimate its volume.

### 3.3. The main provisions of the proposed approach

The proposed approach to non-equivalent steganographic embedding of data in FPGA code is summarized in the following principles.

*Principle 1:* Schema of the FPGA system for approximate data processing contains inessential LUT units. These units are involved in the calculation of the least significant bits discarded during rounding of the result and do not participate in the calculation of the retained most significant bits.

*Principle 2:* if the program code of irrelevant LUTs is distorted, it will not distort the final result of the arithmetic operation.

*Principle 3:* the program code of inessential LUT units can be used as storage for steganographic data;

*Principle 4:* the program code of essential LUT units may contain inessential bits, which are not accessed during the calculation of the most significant bits in result of the operation.

*Principle 5:* Corruption of inessential bits of the LUT program code does not affect the correctness of the rounded result of an arithmetic operation.

*Principle 6:* Program code of inessential bits can be used as storage for steganographic data.

It is necessary to prove the availability of these resources and estimate their volume. For this, an experimental study is carried out. The experimental material is FPGA systems that perform the operation of multiplying floating-point numbers of various bits.

## 4. Experimental research and estimation of resources for non-equivalent data embedding into the FPGA program code

The *purpose of the experiment:* to prove the existence of the two types of resources described in this paper, suitable for non-equivalent steganographic embedding of additional data into the FPGA program code space. Also it is necessary to estimate the volume of these resources and justify its sufficiency for application in the tasks of monitoring the FPGA program code.

To carry out the experiment in the CAD system Intel Quartus Prime 20.1 Lite Edition [34], five FPGA systems were designed. Each of them is a floating-point mantissa multiplier. The size of the operands used in these systems is 4, 6, 8, 10 and 12 bits, respectively. The multiplier circuits are derived from the Intel Quartus Prime CAD library components. The implemented multipliers allow obtaining a  $2n$ -bit result for  $n$ -bit operands. The result of calculations before rounding and discarding inessential bits was available at the output of experimental multiplication circuits. The synthesis of FPGA systems was performed for target FPGA chips Intel Cyclone IV EP4CE15F23A7 [35].

To carry out the experiment, three software applications were developed. They will hereinafter be referred to as App1, App2 and App3, respectively.

The App1 application was implemented in the TCL language. It runs in the Intel Quartus Prime CAD environment. The purpose of App1 is to retrieve detailed information about the FPGA design structure from the Intel Quartus internal CAD database. This information contains a list of FPGA LUT units, the values of the program codes for these units and a list of links that connect the LUT units to each other. The result of the work of the App1 application was presented in a form suitable for submitting the App2 and App3 applications to the input.

App2 is developed in Delphi 10 Seattle demo version [36]. App2 receives information about the structure of the FPGA system from App1 and determines which LUT units in the system are

inessential. The functioning of the App2 application is as follows. The application sequentially iterates over the values of the operands at the input of the multiplier and stores the correct  $2n$ -bit multiplication result and the correct values at the outputs of each of the LUT units in the system. Thereafter, the output values from the LUT units were sequentially corrupted. In doing so, the application analyzed which bits of the result (discarded or retained) are affected by each of the distortions. LUT units, whose inverted outputs only distorted the discarded result bits, were categorized as inessential units.

Fig. 1 shows the results of an experiment to identify inessential LUT units. These results show the percentage of inessential LUT units in experimental FPGA systems, which ranges from 39.3% to 42.5%.

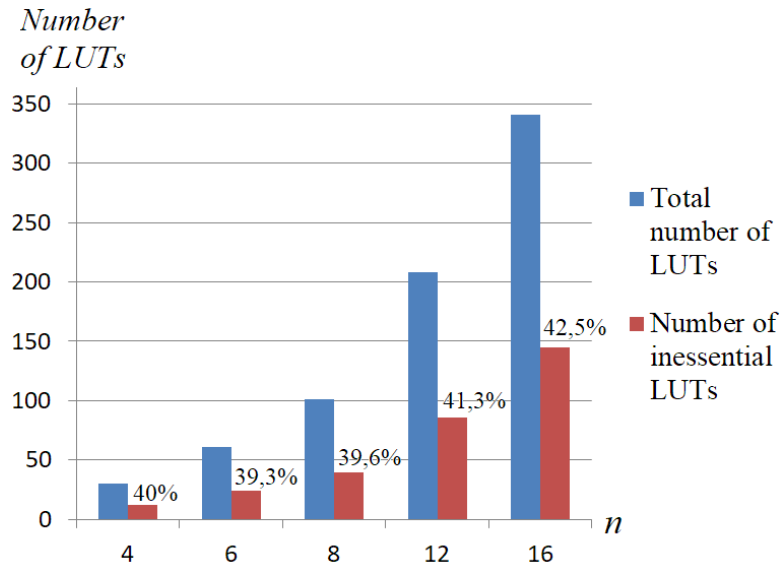


**Figure 1:** Number of inessential LUT units

The next part of the experiment was to determine the number of inessential bits in the program codes of the LUT units. This research was done using App3, also developed in the Delphi 10 Seattle demo version. The main window of the App3 application for 8-bit multiplier FPGA systems is shown in Fig.2.

App3 also iterates over the multiplier operands. At the same time, the application determines the bits for each of the LUT units of the FPGA-system to which the addressing is carried out. At the end of this study, each of the LUT units 1 ... 101 is evaluated by the following indicators (Fig. 2): N is the number of used inputs of the LUT unit; B1 is the number of bits that could be addressable in the LUT based on the number of its inputs used; B2 and B3 are the numbers of bits addressed in the LUT unit when multiplying the binary codes of numbers and binary codes of the normalized mantissa. Binary codes of numbers take all possible values. The binary codes of the normalized mantissa are defined at half of these values. Table 1 and Fig. 3 show the generalized count of the number of inessential bits for each of the multiplier widths.

The results of the experiments made it possible to show that inessential LUTs are present in FPGA-based floating point multipliers. The portion of inessential LUT units ranged from 39.3% to 42.5% of the total number of units in the circuit. Since the distortion of the program code of such LUT units does not lead to distortion of the results, these units constitute a resource for steganographic data embedding.

An LUT with  $m$  inputs is programmed by  $2^m$  bit program code. Based on this, each of the inessential LUTs can provide steganographic storage of up to  $2^m$  bits of additional data. The Intel FPGA Cyclone IV EP4CE15F23A7 chip, for which the synthesis was carried out, contains LUT units with the number of inputs  $m = 4$ . The program code of each LUT of this chip has a size of 16 bits. In the experimental FPGA projects, the number of inessential LUTs ranged from 12 to 145. Thus, for these cases, the volume of the stego container provided by the resource of inessential LUTs was from 192 to 2320 bits. These volumes are sufficient to store hash-sums obtained using commercially used hash functions. With an increase in the size of the multiplier operands, the volume of the stego container also becomes sufficient to store additional monitoring service information.

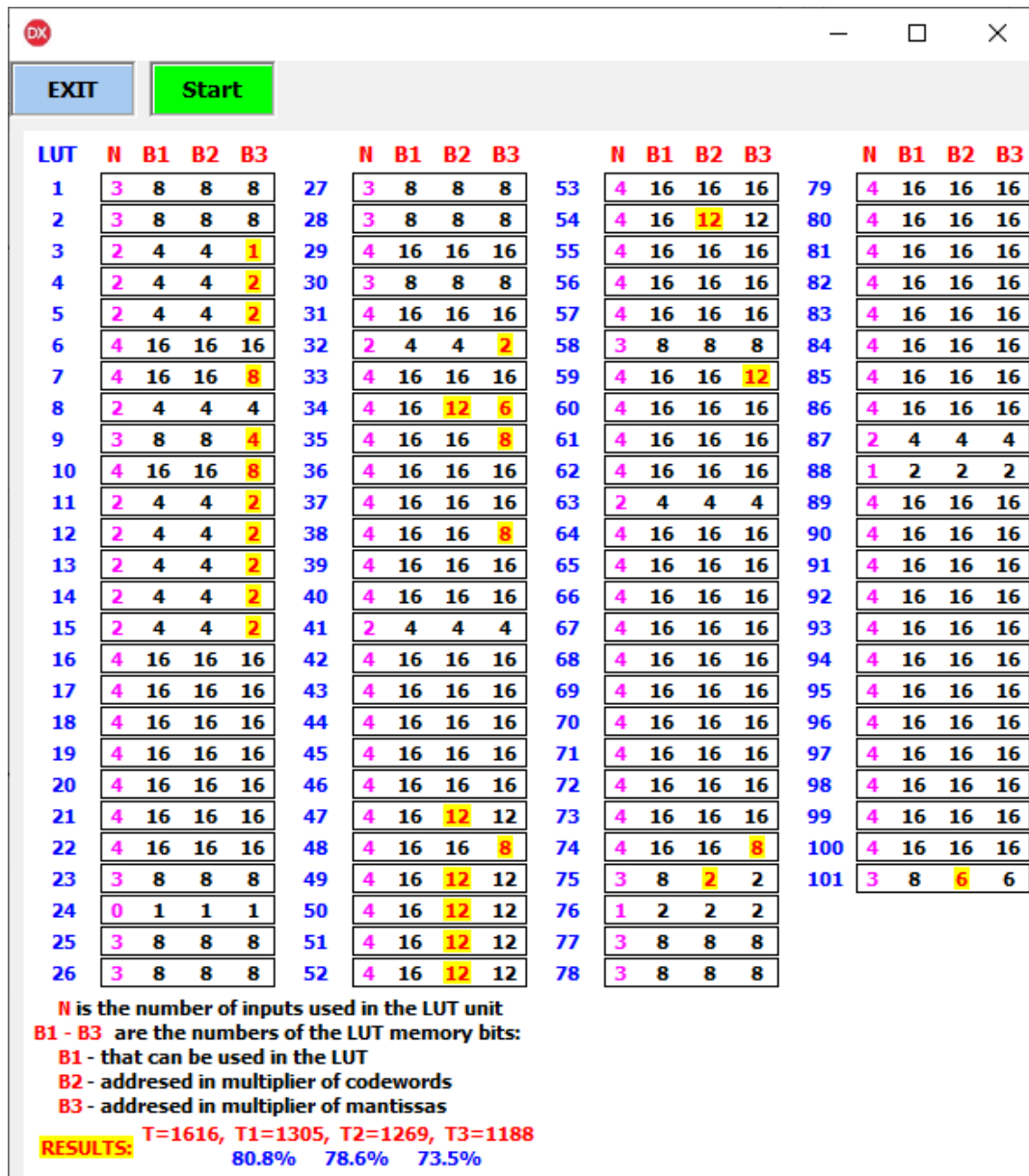


**Figure 2:** Window of the results for the estimation in the number of inessential bits for an 8-bit multiplier

**Table 1**  
Experiment Results

| Size of operands and result          |                     | 4       | 6       | 8       | 10      | 12      |
|--------------------------------------|---------------------|---------|---------|---------|---------|---------|
| Size of the result before rounding   |                     | 8       | 12      | 16      | 20      | 24      |
| Total memory volume of the LUT units |                     | 480     | 976     | 1616    | 2384    | 3328    |
| LUT units memory volume used         |                     | 396     | 747     | 1305    | 1991    | 2811    |
|                                      |                     | (82.5%) | (76.6%) | (80.8%) | (83.5%) | (84.5%) |
| Essential bits                       | Binary code         | 391     | 719     | 1269    | 1983    | 2752    |
|                                      | multiplication      | (81.5%) | (73.7%) | (78.6%) | (83.2%) | (82.7%) |
|                                      | Normalized mantissa | 287     | 658     | 1168    | 1882    | 2638    |
| Inessential bits                     | multiplication      | (59.8%) | (67.4%) | (73.5%) | (79.0%) | (79.3%) |
|                                      | Binary code         | 89      | 257     | 347     | 401     | 576     |
|                                      | multiplication      | (18.5%) | (26.3%) | (21.4%) | (16.8%) | (17.3%) |
| Inessential bits                     | Normalized mantissa | 193     | 318     | 448     | 502     | 690     |
|                                      | multiplication      | (40.2%) | (32.6%) | (26.5%) | (21.0%) | (20.7%) |



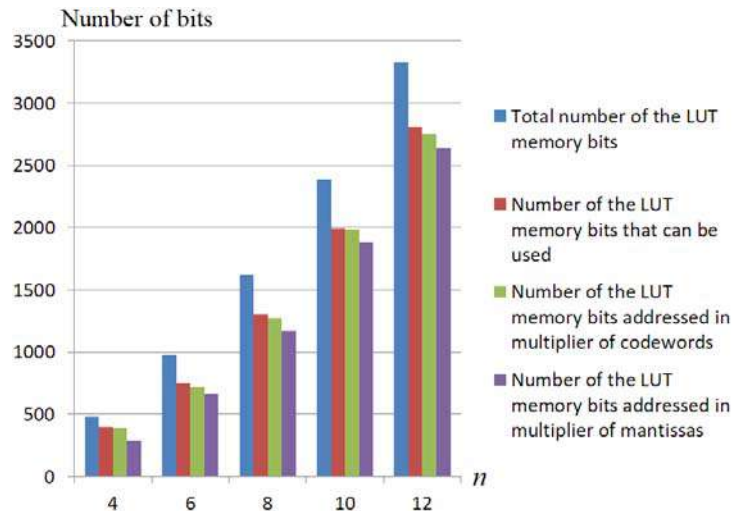


**Figure 3:** Number of addressable container bits

Also the results of the experiments showed the presence in the studied FPGA-systems of the second resource for non-equivalent steganographic data embedding. This resource is the inessential bits of the program code of the LUTs. These bits are not addressed during the calculation of the result. Therefore, the distortion of such bits of the program code does not lead to distortion of the calculation result. Based on this, these bits form a resource of the second kind for non-equivalent steganographic data embedding into the FPGA program code.

For experimental FPGA systems, the number of inessential bits ranged from 89 to 576. The portion of inessential bits in this case ranges from 16.8% to 26.3% of the total number of bits in program codes of LUTs. When multiplying the normalized mantissa, the number of inessential bits of the program code ranged from 193 to 690. In this case, the portion of inessential bits from the total number of bits of the program codes is from 21% to 40.2%. The resulting volumes are sufficient for storing checksums. With an increase in the size of the operands, these volumes of the stego container are sufficient to store the additional monitoring service data as well.

Thus, the experiment confirmed the presence of two kinds of FPGA program code resources that can be used for non-equivalent steganographic embedding of additional data. The experiment also made it possible to estimate the volume of stego containers, which are based on the use of these resources. It is shown that these volumes are sufficient for storing monitoring data.

## 5. Conclusions and Potential Directions of the Further Research

An approach to steganographic data embedding into the program code of FPGA chips is proposed. It differs from existing approaches in that it uses non-equivalent transformations of the FPGA program code. It is shown that software code of FPGA systems that perform approximate data processing, there are new resources. These resources can be used for non-equivalent steganographic embedding. The first of these resources are inessential LUTs. These units are involved in the calculation of the least significant (discarded during rounding) bits of the result and do not participate in the calculation of the most significant (retained) bits. Distortion of the program code of inessential LUT units does not lead to distortion of the calculation result. The second resource is the inessential bits of the LUTs code. These are the bits that are not addressed when calculating the most significant bits of the operation result. Distorting these bits does not distort the result. It is shown that these two kinds of resources make it possible to non-equivalently embed additional data into the FPGA program code. This does not change the operation of the device and the size of the program code.

The experiment performed made it possible to confirm the presence of these two kinds of resources. Also, the experiment made it possible to estimate its volume of stego containers resulting from the use of these resources. The volume of the stego container provided by the resource of inessential LUTs was from 192 to 2320 bits. The portion of inessential bits in this case ranges from 16.8% to 26.3% of the total number of bits in program codes of LUTs. When multiplying the

normalized mantissa, the number of inessential bits of the program code ranged from 193 to 690. In this case, the portion of inessential bits from the total number of bits of the program codes is from 21% to 40.2%. These volumes are sufficient to store hash-sums obtained using commercially used hash functions.

The proposed approach can be used both independently and in conjunction with an equivalent approach to embedding additional data into the FPGA program code. Given the prevalence of approximate data processing among the functions of modern computer systems, embedding based on the features of operations with such data is very promising.

The direction for further research is to determine how to jointly use the equivalent and non-equivalent approaches to steganographic data embedding. Of interest is the study of a hybrid stego system in which both approaches play the same or different roles. For example, the roles of embedding, code obfuscation, or false embedding aimed at countering stego analysis.

## 6. References

- [1] W. Stallings, *Cryptography and Network Security: Principles and Practice*, 7th ed., Pearson Education Limited, United Kingdom, Harlow, 2017.
- [2] A. Awad, M. Fairhurst, *Information Security. Foundations, technologies and applications*, The Institution of Engineering and Technology, UK, London, 2018.
- [3] J. Katz, *Introduction to Modern Cryptography*, 2nd Edition, CRC Press, Boca Raton, 2018.
- [4] M. Miyanaga, H. Irie and S. Sakai, "Accelerating Integrity Verification on Secure Processors by Promissory Hash," 2017 IEEE 22nd Pacific Rim International Symposium on Dependable Computing (PRDC), Christchurch, 2017, pp. 20-29, doi: 10.1109/PRDC.2017.13.
- [5] M. Bishop, *Computer Security*, 2nd edn., Addison-Wesley, USA, Boston, 2018.
- [6] Y. Yang, F. Chen, X. Zhang, J. Yu, P. Zhang, Research on the Hash Function Structures and its Application, in *Proceedings of International Conference Wireless Personal Communications*, 2016.
- [7] J. Vacca, *Computer and information security handbook*, 3rd ed., Morgan Kaufmann, Waltham, Mass, 2017.
- [8] X. Wu, Z. Yang, C. Ling, X. Xia, Artificial-Noise-Aided Message Authentication Codes With Information-Theoretic Security, *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 6 (2016) 1278-1290. doi: 10.1109/TIFS.2016.2524514.
- [9] S. Hirose, J. Shikata, Aggregate Message Authentication Code Capable of Non-Adaptive Group-Testing, *IEEE Access*, vol. 8 (2020) 216116-216126. doi: 10.1109/ACCESS.2020.3041638.
- [10] T. Hovorushchenko, O. Pomorova. Evaluation of Mutual Influences of Software Quality Characteristics Based ISO 25010:2011, in: *Proceedings of 2016 IEEE International Scientific and Technical Conference "Computer Science and Information Technologies, CSIT-2016*, Lviv, Ukraine, 2016, pp. 80-83.
- [11] M. Zaman, T. Shen and M. Min, Hash Vine: A New Hash Structure for Scalable Generation of Hierarchical Hash Codes, 2019 IEEE International Systems Conference, Orlando, FL, USA, 2019, pp. 1-6. doi: 10.1109/SYSCON.2019.8836921.
- [12] T. Hovorushchenko, O. Pavlova. Evaluating the Software Requirements Specifications Using Ontology-Based Intelligent Agent, in: *Proceedings of 2018 IEEE International Scientific and Technical Conference "Computer Science and Information Technologies CSIT-2018*, Lviv, Ukraine, 2018, pp.215-218.
- [13] S. Wang, C. Li, H. Shen, Distributed Discrete Hashing by Equivalent Continuous Formulation, *IEEE Transactions on Signal and Information Processing over Networks*, vol. 6 (2020) 196-210. doi: 10.1109/TSIPN.2020.2975356.
- [14] S. Lysenko, O. Savenko, K. Bobrovnikova, A. Kryshchuk. Self-adaptive system for the corporate area network resilience in the presence of botnet cyberattacks, *Communications in Computer and Information Science*, 860 (2018) 385-401.
- [15] F. Shih, *Digital Watermarking and Steganography: Fundamentals and Techniques*. 2nd ed., CRC Press, USA, Boca Raton. 2017.

- [16] K. Zashcholkin, O. Drozd, R. Shaporin, O. Ivanova, Y. Sulima, Increasing the effective volume of digital watermark used in monitoring the program code integrity of FPGA-based systems, in Proceedings of 2019 IEEE East-West Design and Test Symposium, EWDTS 2019.
- [17] Z. Zhang, L. Njilla, C. A. Kamhoua, Q. Yu, Thwarting Security Threats From Malicious FPGA Tools With Novel FPGA-Oriented Moving Target Defense, IEEE Transactions on Very Large Scale Integration Systems, vol. 27, no. 3 (2019) 665-678. doi: 10.1109/TVLSI.2018.2879878.
- [18] H. Amano, Principles and Structures of FPGAs, Springer, 2018.
- [19] J. Andina, FPGAs: Fundamentals, Advanced Features, and Applications in Industrial Electronics, CRC Press, USA, Boca Raton, 2017.
- [20] A. Melnyk, A. Salo, Automatic generation of ASICs, in: Proceedings of the Second NASA/ESA Conference on Adaptive Hardware and Systems, Edinburgh, UK, 2007, pp. 311-317, doi: 10.1109/AHS.2007.36.
- [21] O. Drozd, K. Zashcholkin, R. Shaporin, J. Drozd, Y. Sulima, Development of ICT Models in Area of Safety Education, in: Proceedings of the 18th IEEE East-West Design & Test Symposium, Varna, Bulgaria, 2020, pp. 115–119. doi:10.1109/EWDTS50664.2020.9224861.
- [22] O. Drozd, K. Zashcholkin, O. Martynyuk, O. Ivanova, J. Drozd, Development of Checkability in FPGA Components of Safety-Related Systems, CEUR Workshop Proceedings, vol. 2762 (2020) 30-42.
- [23] O. Y. Abdulhammed, Improving Encryption Digital Watermark by Using Blue Monkey Algorithm. International Journal of Computing, vol. 20, no. 1 (2021) 129-136. doi: 10.47839/ijc.20.1.2101
- [24] Y. Ke, J. Liu, M. Zhang, T. Su, X. Yang, Steganography Security: Principle and Practice, IEEE Access, vol. 6 (2018) 73009-73022. doi: 10.1109/ACCESS.2018.2881680.
- [25] T. Korkishko, A. Melnyk, Cryptographic processor architectures for DES algorithm, in: Proceedings of the 5th IEEE Africon Conference (Cat. No.99CH36342), Cape Town, South Africa, 1999, vol. 1, pp. 175-180. doi: 10.1109/AFRCON.1999.820788.
- [26] K. Zashcholkin, O. Drozd, O. Ivanova, P. Bykovyy, Formation of the interval stego key for the digital watermark used in integrity monitoring of FPGA-based systems. CEUR Workshop Proceedings (CEUR-WS) 2623 (2020), 267–276.
- [27] A. Drozd, J. Drozd, S. Antoshchuk et. al., Objects and Methods of On-Line Testing: Main Requirements and Perspectives of Development, in: Proceedings of the IEEE East-West Design & Test Symposium, Yerevan, Armenia, 2016, pp. 72–76. doi:10.1109/EWDTS.2016.7807750.
- [28] J. Drozd, A. Drozd, M. Al-dhabi, A resource approach to on-line testing of computing circuits, in: Proceedings of the IEEE EWDTS Symposium, Batumi, Georgia, 2015, pp. 276–281. doi:10.1109/EWDTS.2015.7493122.
- [29] T. Chen, K. Hou, W. Beh and A. Wu, Low-Complexity Compressed-Sensing-Based Watermark Cryptosystem and Circuits Implementation for Wireless Sensor Networks, IEEE Transactions on Very Large Scale Integration Systems, vol. 27, no. 11 (2019) 2485-2497. doi: 10.1109/TVLSI.2019.2933722.
- [30] IEEE Standard for Floating-Point Arithmetic, in IEEE Std 754-2019 (Revision of IEEE 754-2008), pp.1-84, 22 July 2019. doi: 10.1109/IEEESTD.2019.8766229.
- [31] A. Anderson, S. Muralidharan, D. Gregg, Efficient Multibyte Floating Point Data Formats Using Vectorization, IEEE Transactions on Computers, vol. 66, no. 12 (2017) 2081-2096. doi: 10.1109/TC.2017.2716355.
- [32] O. Drozd, I. Perebeinos, O. Martynyuk et. al., Hidden fault analysis of FPGA projects for critical applications, in: Proceedings of the IEEE International Conference TCSET, paper 142, Lviv–Slavsko, Ukraine, 2020. doi:10.1109/TCSET49122.2020.235591.
- [33] O. Pomorova, O. Savenko, S. Lysenko, A. Kryshchuk, K. Bobrovnikova, Anti-evasion Technique for the Botnets Detection Based on the Passive DNS Monitoring and Active DNS Probing, Communications in Computer and Information Science, 608 (2016) 83-95.
- [34] Intel Quartus Prime Standard Edition User Guide, 2020. URL: [https://www.intel.com/content/dam/alterawww/global/en\\_US/pdfs/literature/ug/ug-qps-getting-started.pdf](https://www.intel.com/content/dam/alterawww/global/en_US/pdfs/literature/ug/ug-qps-getting-started.pdf).
- [35] Cyclone IV Device Handbook. URL: <https://www.intel.com/content/dam/www/programmable/us/en/pdfs/literature/hb/cyclone-iv/cyclone4-handbook.pdf>.
- [36] Delphi 10 Seattle: Embarcadero. URL: <https://www.embarcadero.com/docs/datasheet.pdf>.

# A reasonable smart home technology on the Arduino

Dmitry Lubianov<sup>a</sup>, Kostiantyn Kasian<sup>a</sup>, Mykola Kasian<sup>a</sup>

<sup>a</sup> National university "Zaporizhzhia polytechnic", Zhukovskogo street 64, Zaporizhzhia, 69063, Ukraine

## Abstract.

The article demonstrates the results of the development and implementation of low cost smart home technology programmed on the FPGA Arduino. The purpose of the work is to program remote control gadgets, with which the user has the ability to control all devices that support infrared signal technology. The scientific novelty of the work results is the developed set of programmable gadgets can replace a significant number of different remotes and switches by using only one gadget as a control device, which is an alternative to a smart home. The practical value of the work results is that the developed set of programmable gadgets will ease the trivial life of the handicapped and the elderly.

## Keywords 1

Android, Arduino, FPGA, IR-signal, Remote control, Smart home

## 1. Introduction

Nowadays, concepts like a smart vacuum cleaner, a smart refrigerator, and in general, a smart home is not something fantastic. Smart devices are approaching to the 21st century people. However, there are certain restrictions that inhibit "smart development" in human life. The first reason is people's skepticism about this kind of thing. The second is the high cost of such devices. Solving the first problem is much more difficult, so we need to focus on the second, which can affect the previous one.

The simplest option is to upgrade existing appliances in compare with replace them with new ones. In this way, a person gets a full-fledged smart device, or even a smart home.

A smart home is a wide concept that includes a system of home devices that capable of performing actions and solving certain everyday tasks without human activity. Electrical appliances of the building, which are controlled centrally (from the remote control). The devices are connected to a computer network, which allows us to control them with a PC and provide remote access via the Internet [1].

However, there is an even simpler and cheaper option. The only disadvantage is all appliances go into remote control mode (the previous method included automatic mode).

The obvious and effective solution to this problem is to replace all remote controls with a universal remote control or smartphone. Although the last option has been practiced in everyday life for some time, but not all devices are equipped with infrared (IR) technology.

Furthermore, the smart home technology will be really handy the handicapped and the elderly life who find using a large amount of different remotes very tough or even impossible. The system settles all those problems.

This article describes methods and tools for the development and deployment of a ready-made system of the smart house, which will ease the trivial life for all people, especially the elderly and the

handicapped including devices analysis, hardware and software system review and development and testing of the system.

## **2. Devices analysis**

To develop a low cost “smart home” technology we need to purchase a device for encoding/decoding signals. There are no specialized devices for such purposes, but there are FPGAs, the functionality of which fully covers these requirements. Among all the options, the Arduino FPGA family famous of its ease of use and low price (compared to BroadLink RM Pro, which is several times more expensive than the chosen integrated board) [2].

FPGA is an electronic component used to create digital integrated circuits. Unlike conventional digital chips, the logic of FPGA operation is not determined during manufacture, but get set by programming. For this purpose, there are programmers and debugging environments are used, which allow to set the desiring structure of the digital device using one of the special languages of hardware description (Verilog, VHDL, AHDL and others) [3].

The most popular and one of the budget options is the purchase of FPGA family Arduino. To build a smart home, the Arduino UNO is commonly used as a motherboard, and the Arduino Mini / Nano as an additional option to control certain devices.

The essence of smart home technology is not difficult to deploy, we just need to connect the main board to a computer, laptop or smartphone that has Internet access and organize communication with other modules and sensors via Wi-Fi or Bluetooth. Thus, a properly configured system will be able to remotely read the values of the installed sensors (temperature, fire, moving, lighting etc.) and send them to the user's device through the Wi-Fi or Bluetooth modules [4]. In turn, the user can program the delayed start of devices, adjust their work mode or control electrical appliances remotely (vacuum cleaner, air conditioner, light switch, etc.).

Bluetooth's big advantage for smart home products is low energy use. With Bluetooth LE, you can have a standalone device that does not need charging for weeks. That may come in handy when we get to the point where all our clothes, all our furniture, all our possessions are smart-home enabled.

Generally, there are a lot of appliances that can be controlled remotely. For instance:

- TV
- audio devices
- slide projector
- conditioner
- purifier
- led-string
- smart-things and so on

However, the simplest option is to buy some of the cheapest FPGAs Arduino Mini / Nano, Bluetooth or WiFi modules and signal transmitters. This is the case considered in this work.

### **2.1. Hardware system review and development**

The reasonable smart home technology development includes two components: hardware and software.

The hardware includes selecting and connecting the required components. For one-time and more convenient FPGA programming, the Arduino UNO board is selected (after that, all devices are connected to the Arduino Mini / Nano).

To get started the system, there are three additional modules are needed for Arduino including:

- IR-receiver VS1838B
- infrared LED
- bluetooth module HC-05

IR receivers, sometimes called IR sensors or IR detection diodes, usually come in two different form factors. You can either buy the diodes separately or mounted on a small breakout board.

They work exactly the same, so it does not matter which one you use. The only difference is that the breakout board often contains a small LED that blinks every time the receiver detects a signal which can be handy for debugging.

First, we need to connect the power supply to the board. Except for providing power, we need to connect an IR receiver.

In our case, the IR receiver (TSOP 1838) is connected to 11 pins and as well as to 5V and GND power supply. FPGA power supply, in our case, is provided by two options: from the 12V power supply or from the USB Type B connector [5].

The VS1838B infrared receiver is a receiver of IR signals that operates at a frequency of 38 KHz. The IR receiver can receive and process an infrared signal, which is an IR pulse of fixed frequency and a certain duration (a bundle of pulses) [6].

In IR signal modulation, an encoder on the IR remote converts a binary signal into a modulated electrical signal. This electrical signal is sent to the transmitting LED. The transmitting LED converts the modulated electrical signal into a modulated IR light signal. The IR receiver then demodulates the IR light signal and converts it back to binary before passing on the information to a Minicontroller.

The scheme of the button-code receiving through IR-receiver and the follow displaying them on the screen is shown in Figure 1.

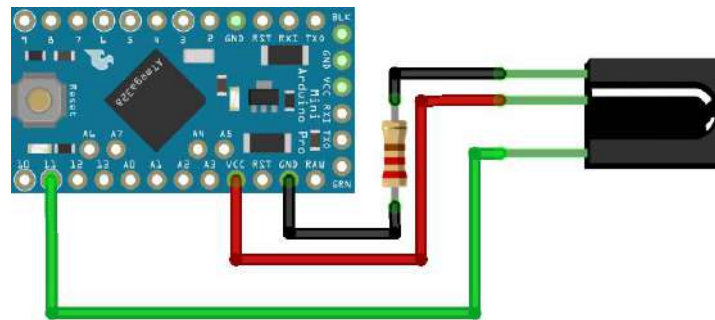


**Figure 1:** Arduino Mini connection scheme with IR-receiver VS1838B [7]

Features of IR receiver:

- frequency 38 KHz
- supply voltage 2.7-5.5 V
- current 0.4-1.5 mA
- signal reception angle: 90 °
- signal reception range up to 20 m

Infrared LED allows us to create a wireless optical method of data transmission, because infrared rays do not distract a person due to their invisibility [8].

From a physical point of view, an infrared LED is a diode or a simple semiconductor.

To operate other slave modules, it is necessary to connect an IR transmitter and, in our case, a Bluetooth module. The IR transmitter is connected to second pin and GND through a resistor.

The Bluetooth module HC-05 is connected to TXD, RXD, 5V and GND through a resistor. By the way, this module has one more pin called Key. If it is input low level or connect to the air, the module is at paired or communication mode. If it is input high level, the module will enter to AT mode.

Bluetooth module HC-05 is one of the best solutions for two-way Bluetooth communication of our Arduino device with a smartphone, laptop or other Bluetooth devices.

To compare with the HC-06 module, which can only be installed as a Slave, the HC-05 can also be installed as a Master (searching for Bluetooth devices and initiating a connection), which allows communication between two separate Arduino boards [9].

Technical features of the HC05 module:

- bluetooth chip: HC-05 (BC417143)
- radio frequency range: 2.4-2.48 GHz
- speed: Asynchronous: 2.1Mbps(Max) / 160 kbps, Synchronous: 1Mbps/1Mbps

- power supply: 3.3-5 V
- current: 50 mA
- range: up to 10 meters
- profiles: serial port
- modes: master, slave
- working temperature: -25... 75 ° C
- dimension: 27 x 13 x 2.2 mm

The scheme of the button-code receiving device on Bluetooth technology and its decoding in a signal is shown in Figure 2.

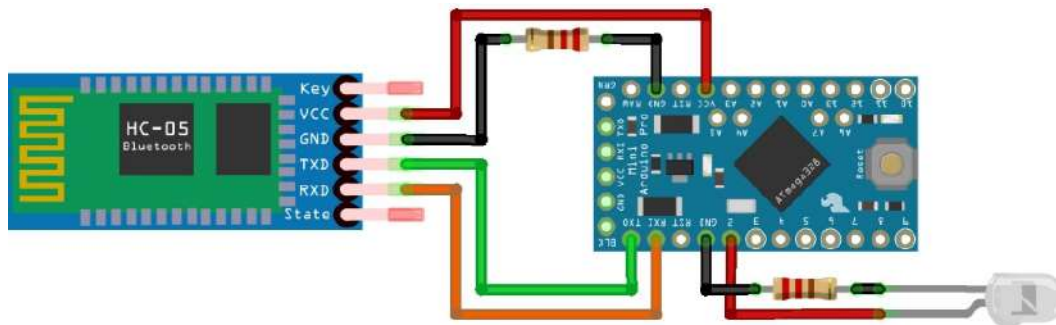


**Figure 2:** Arduino Mini connection scheme with IR-LED and HC-05 module [7]

By the way, we can combine two schemes, which shown in Fig.1 and Fig. 2, in only one. All we need is to connect IR-receiver to the last scheme and change listing a little. So that, we can adapt each board not only for transmitting signal but for receiving it through IR-technology.

To summarize, we need at least four the most significant modules to build the system (Arduino board, IR-receiver, IR-LED, Bluetooth module) and, of course, a few resistors to prevent a module failure because of the high amount of current, and some wires to combine all the components.

## 2.2. Software system review and development

The software part includes three components: writing a program for reading codes of different signals from remote controls; writing a program for remote signal transmission using Bluetooth technology; writing a program to receive, decode and transmit a signal to a household appliance.

Arduino programs are written in the C or C++ programming language. Users need to define only two functions in order to create a program that works on a cyclical basis [10]:

- `setup()`: the function is executed only once at program start and allows to set initial parameters
- `loop()`: the function is performed periodically until the board is turned off

An IR receiver was connected to the board to receive data from different remotes. It should be said that this operation can be performed without a computer. All we have to do is pre-load the required sketch to the Arduino board and all the codes are displayed on the FPGA monitor itself. However, this option is not so convenient due to the need for manual data recording [11].

An IR transmitter was connected to the Arduino UNO for preliminary testing. The program for decoding and encoding the remote control signal was written and the first results were obtained (in the 16th format) [12]. The principle of operation is to receive a signal from the parent remote control to the IR receiver of the board. After that, a sketch is performed and the recoded code is sent to the IR transmitter, which sends a signal to the device. An example of the obtained codes is shown in Table 1.

The first column is a short description of each command. The second is HEX code that was received from TV-box appliance and the third one is HEX code that was received from TV. The last column is testing buttons to send chosen command through the IR-LED.

To control household appliances with a smartphone, we need to write an application that transmits the received signals using Bluetooth technology to the HC-05 module. However, we can use a ready-made application, many of which are in the Internet and are very similar to each other.



First, we need to connect a Bluetooth module and an IR transmitter to our FPGA Arduino Mini / Nano.

**Table 1.** Codes from the two tested devices

| Commands   | SAT      | TV     | Code |
|------------|----------|--------|------|
| power      | 9966DA25 | FE50AF | 0    |
| mute       | 99667A85 | FED02F | 1    |
| menu       | 99669A65 | FE2AD5 | 2    |
| exit       | 9966B24D | FEEA15 | 3    |
| left       | 9966926D | FEDA25 | 4    |
| right      | 996650AF | FE1AE5 | 5    |
| up         | 9966D827 | FE7A85 | 6    |
| down       | 9966F00F | FE6A95 | 7    |
| epg/source | 9966708F | FECA35 | 8    |
| ok         | 9966D02F | FE5AA5 | 9    |

A feature of the HEX encoding format is the almost 100% absence of signal code repetition on different devices [13]. Therefore, programming of all installed Arduino boards is made according to one template.

The algorithm of the program includes several key stages of obtaining, recognizing and decoding commands (Fig. 3). First, a signal transmits via Bluetooth technology to the HC-05 Arduino Mini / Nano module. Then the installed sketch works and after that, through the IR transmitter, the signal transmits the appropriate code to the IR receiver of the device.

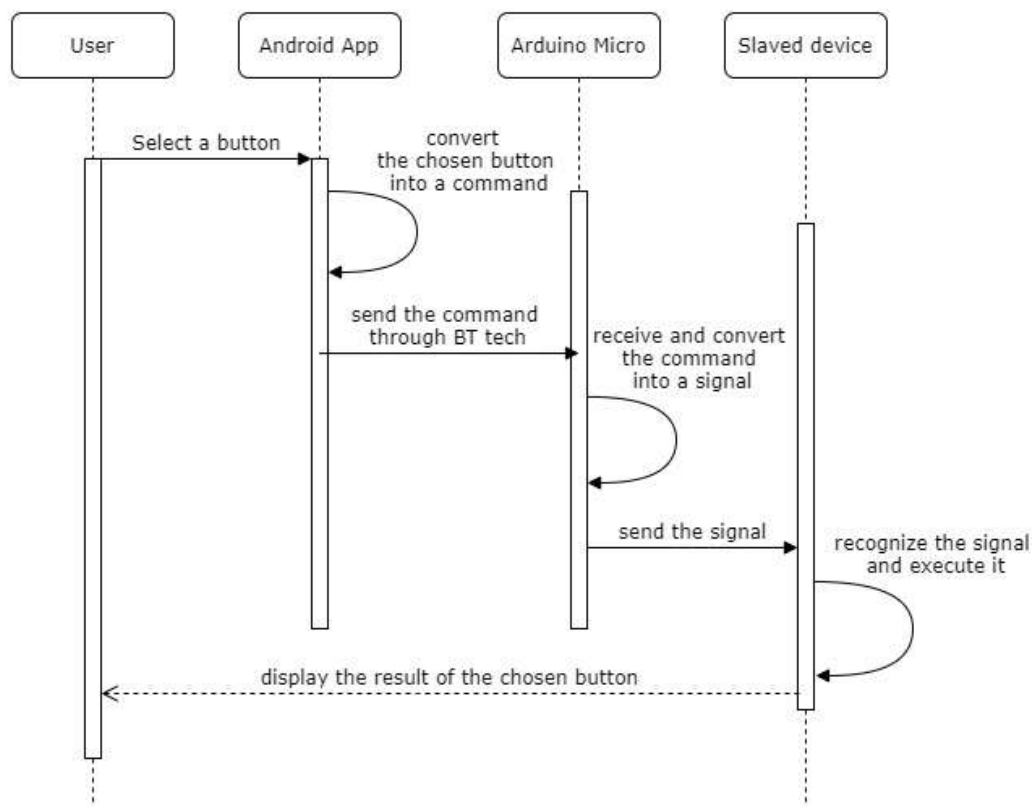


**Figure 3:** The main algorithm of transferring user's command

The main screen of the application has all the necessary functionality and an intuitive interface (Fig. 4), on which we can control the connected devices or perform other manipulations. The blue



Bluetooth icon means that the device is connected properly and is ready to use. The red Bluetooth icon, in contrast, means that there are problems with the connection [14].

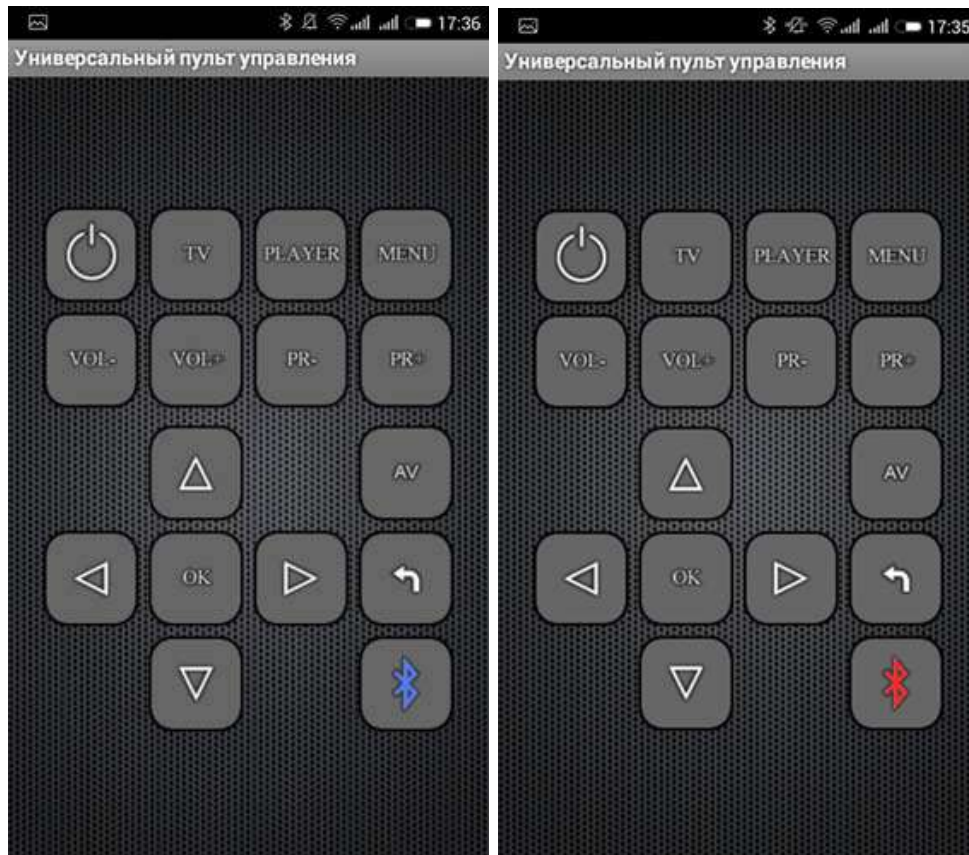


**Figure 4:** The main menu of the Android application [15]

As a result, we configured all the necessary components including hardware and software parts. We combined two schemes and programmed Arduino boards to execute two types of code: receiving and display code of signal and receiving and transmitting received signal.

### 3. Testing of the system

At the system testing stage, we need to make sure that the sent signal on the programmable remote control responds correctly and the device correctly receives the command.

First we need to establish connection with all the home appliances. We turn on the Bluetooth on a smartphone and open the application. Then we start searching Bluetooth devices around, connect to them and after that, devices are ready to use.

For testing features of the system, six devices were programmed: two different TVs, a set-top box, a CD player, a DVD player and an LED strip. We can connect up to 10 different devices (that amount was set at the programming stage), but these amount enough for an ordinary user.

The smart home structure is shown in Figure 5.

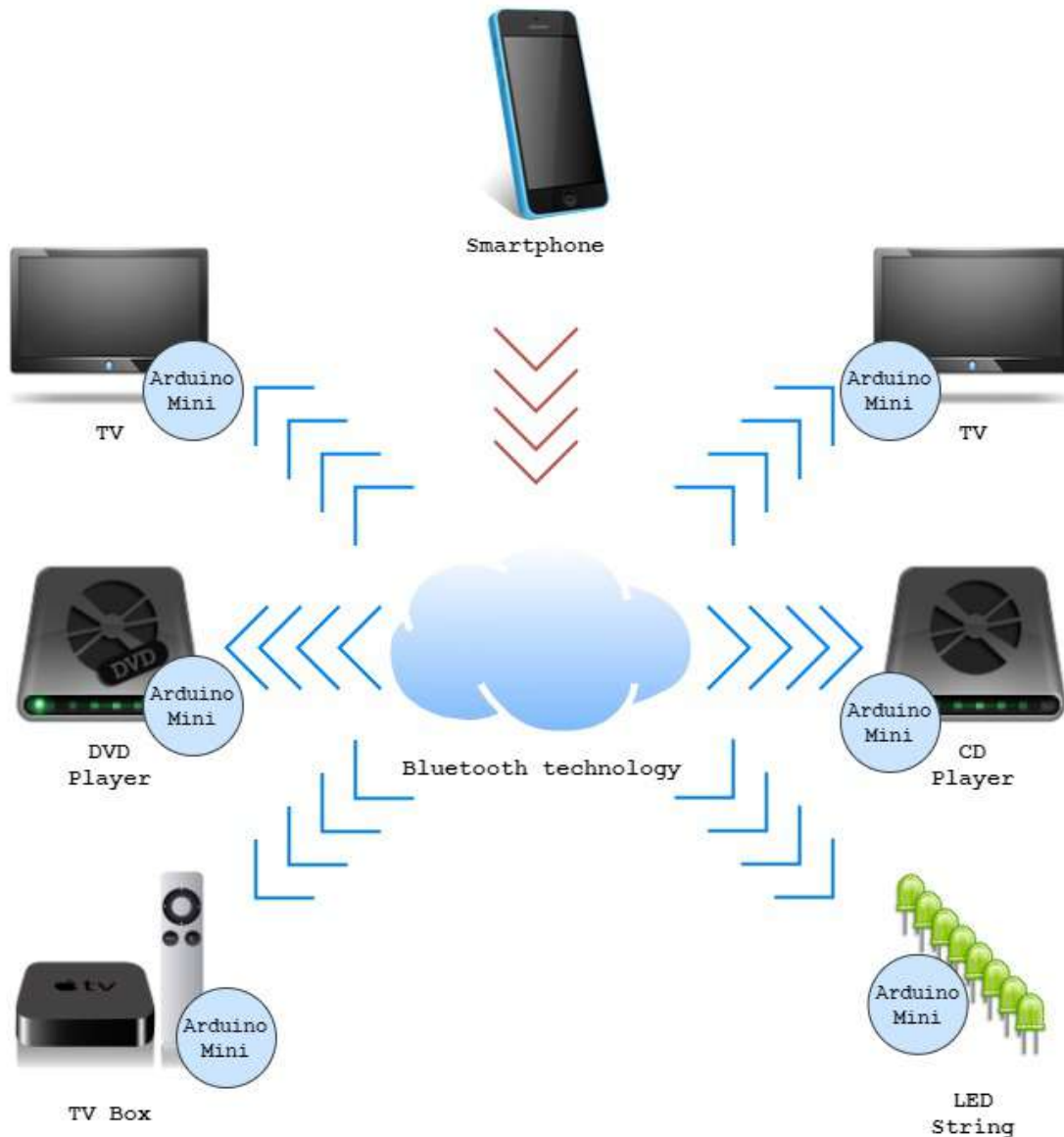


**Figure 5:** The smart home structure

In order to demonstrate correction of work we join the Arduino Mini system with HC-05 Bluetooth module to laptop and open “Port monitor” tab to see sent signals.

On Uno, Nano, Mini, and Mega, pins 0 and 1 are used for communication with the computer. Connecting anything to these pins can interfere with that communication, including causing failed uploads to the board.

We can use the Arduino environment’s built-in serial monitor to communicate with an Arduino board. So that way, we can check which information transmitted through the Arduino board to the terminal device.

The result of work is showed in Figure 6.



access. Our option is simpler and reasonable, even without great abilities in programming a person is able to build and deploy the smart home technology.

## 4. Conclusions

The result of research work is a universal complex for controlling devices with an infrared sensor, which opens up a wide range of opportunities for its users. Devices are controlled by a smartphone, even if it does not have an infrared port. This allows the user to set up the technology of a smart home, which is very popular and relevant now, in any room.

It is perfect, fully functional, low cost and reliable replacement to regular smart home technology, because not every person can afford to deploy true smart home technology in their house.

All components to build the complex system, including several Arduino Mini/Nano boards, cost even less than only one branded smart device. Once you need to write a necessity code and then just upload it to all the boards. It takes a few minutes, but the result is terrific.

Besides, the developed set of programmable gadgets will help the handicapped and the elderly, who find it very difficult to operate multiple remotes or master a new one. The only thing they should get is getting used to the android application.

As a result, the reasonable smart home technology for controlling household appliances has been developed, programmed and tested.

## 5. Acknowledgement

The system was implemented in the work process at the LLC "STYLE MAX" enterprise and used as a device for testing different appliances at certain stages of production which controlled by an infrared signal.

## 6. References

- [1] Dmitry Lubyantsev, Max Tverdokhlib, Tatiana Dyachuk. "Complex of programmable gadgets for household appliances management". The modern science potential (part IV): materials of the IV International scientific-practical conference (2019): 49-50.
- [2] Simon Monk. "Programming Arduino. Professional work with sketches", 2016.
- [3] Jeremy Blum. Exploring Arduino: Tools and Techniques for Engineering Wizardry 1st Edition, (2016): 90-93.
- [4] Simon Monk. Arduino + Android Projects for the Evil Genius: Control Arduino with Your Smartphone or Tablet, 2017.
- [5] Becky Stewart. "Adventures in Arduino", 2015.
- [6] Simon Monk. Programming Arduino: Getting Started with Sketches (Tab) Second Edition (2011): 140-154.
- [7] Fritzing. URL: <https://fritzing.org/>.
- [8] Alasdair Allan, Don Coleman, Sandeep MistryMake. "Bluetooth LE Projects with Arduino, Raspberry Pi, and Smartphones", 2018.
- [9] John Boxall. Arduino Workshop: A Hands-On Introduction with 65 Projects, (2013): 35-43.
- [10] Jody Culkin, Eric Hagan. "Learn Electronics with Arduino: An Illustrated Beginner's Guide to Physical Computing (Make: Technology on Your Time) 1st Edition", 2014.
- [11] Richard Blum. Arduino Programming in 24 Hours, Sams Teach Yourself 1st Edition, (2014): 111-119.
- [12] Turner Ryan. Arduino Programming: The Ultimate Beginner's Guide to Learn Arduino Programming Step by Step Paperback, 2019.
- [13] Mark Geddes. Arduino Project Handbook: 25 Practical Projects to Get You Started 1st edition, 2016.
- [14] Jack Purdum. Beginning C for Arduino, Second Edition: Learn C Programming for the Arduino, (2015): 167-170.

- [15] A universal remote on the Arduino. URL: <https://habr.com/ru/post/254761/>.
- [16] Massimo Banzi, Michael Shiloh. Getting Started with Arduino: The Open Source Electronics Prototyping Platform (Make) 3rd Edition, 2014.
- [17] Mr. Lawrence Harte. Introduction to Bluetooth: Technology, Operation, Profiles, and Services 2nd Edition, 2019.
- [18] Marco Schwartz. Internet of Things with Arduino Cookbook, 2018.

# Conceptual model of IoT-based Laboratory for study the Electrical Engineering and Electronics

Oleksandr Osolinskyi<sup>a</sup>, Liubomyr. Kolodiichuk<sup>b</sup>, Hrystyna. Lipyana-Goncharenko<sup>a</sup>, Anatoliy. Sachenko<sup>c,a</sup>, Lukasz Kopania<sup>c</sup>, Volodymyr Kochan<sup>a</sup>, Diana Zahorodnia<sup>a</sup>

<sup>a</sup> West Ukrainian National University Department for Information Computer Systems and Control, 3 Peremoga Square, Ternopil, 46020, Ukraine

<sup>b</sup> SE NULES of Ukraine "Berezhany Agrotechnical Institute", 20 Akademichna st., Berezhany, Ternopil region, 47501, Ukraine

<sup>c</sup> Kazimierz Pulaski Technology and Humanitarian University, Malczewskiego St 29, 26-600, Radom, Poland

## Abstract

Abstract - The popularity of IoT has been growing rapidly in recent years. This is due to the reduction in the cost of devices that take place in IoT solutions, the creation of user-friendly software development systems and the development of cloud services. This, in turn, has led to the transition of the educational process to a completely different level, where higher education students can obtain most of the knowledge (both theoretical and practical) remotely without the need to physically visit the classroom. In this regard, a conceptual model of IoT hybrid laboratory is proposed, in which students have the opportunity to conduct their research both remotely and in patiently. The structure of the control system and general operating scenarios are described. The issue of optimizing the use of electricity in the proposed laboratories, which is associated with an increase in electrical load several times relative to classical laboratories.

## Keywords 1

Conceptual model, Arduino; Raspberry; IoT, IoT-based Laboratory

## 1. Introduction

The Internet of Things is a concept of building an environment in which static physical objects (mechanical, digital, people, animals and other objects) are connected to the World Wide Web and can communicate with each other and exchange information to solve everyday problems [1-3]. The main idea of IoT [4,5] is that these objects can interact with each other, perceive and collect data from the environment without the need for human intervention or the need to communicate with her. According to researchers, IoT infrastructure could reach 1.5 billion devices in 2022 [6-9]. Given that the prices of microcontrollers have fallen to the prices of everyday food [10], it is possible to automate and monitor all processes in life. This trend has not bypassed the learning process either [11]. In particular, there is a growing tendency to conduct remote experiments without spending time on physical visits to laboratories, which does not reduce the workload of laboratories and staff, but also allows not to reduce the quality of training, because students still work with real equipment. However, this approach makes it difficult to monitor all devices in such laboratories. Requires control of equipment that is not currently used, but consumes electricity [12]. In addition, given that the

---

Proceedings Name, Month XX-XX, YYYY, City, Country

EMAIL: osolinskyi.oleksandr@gmail.com (O. Osolinskyi); kollub@ukr.net (L. Kolodiichuk); xrustya.com@gmail.com (H. Lipyana-Goncharenko); as@wunu.edu.ua (A. Sachenko); l.kopania@uthrad.pl (L. Kopania); volodymyr.kochan@gmail.com (V. Kochan); dza@wunu.edu.ua (D. Zahorodnia)

ORCID: 0000-0002-0136-395X(O. Osolinskyi); 0000-0003-1172-8972 (L. Kolodiichuk); 0000-0002-2441-6292 (H. Lipyana-Goncharenko); 0000-0002-0907-3682 (A. Sachenko); 0000-0002-7318-4803 (L. Kopania); 0000-0001-8376-4660 (V. Kochan); 0000-0002-9764-3672 (D. Zahorodnia)



© 2020 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)



laboratories conduct not only remote experiments, but also the actual presence of students, it is also necessary to control the level of lighting and air conditioning in the room.

For such purposes, the concept of IoT is best suited, which will allow for mixed experiments in laboratories, with some students working with equipment and others conducting experiments remotely in parallel. The system fully controls the energy consumption and the turn of the experiments. This will save electricity in general in the laboratory and increase the number of students involved without increasing the physical presence of people.

## 2. Related work

In [19], a comprehensive approach to the implementation of a hardware laboratory based on FPGA and ignited as a service in a hybrid cloud. This laboratory can also be scaled in the public cloud with the possibility of free connection of IoT devices, which are implemented using FPGA [15]. The presented laboratory uses FPGA to work with peripherals and cloud services.

The implemented controller on the FPGA simultaneously works with the server and receives commands, as well as, on the other hand, receives data from sensors, which are sent to the cloud using the Internet protocol [16]. After that, cloud services use FPGA to store, store data, analyze and visualize data. In order for such an environment to work at the IoT level, the student first prepares IoT peripheral designs in Verilog-HDL and server-side software. In addition, Docker [17] and Docker Swarm containers were used for the operation of the system. This part is implemented in Go in the form of templates and each student can easily customize them for their own needs. Using Docker tools also reduces application development and human resource usage time. Thanks to the microservice architecture, the laboratory can dynamically change its configuration according to the workload of students. There are several work strategies in this system: students can conduct experiments on their own or join group experiments. But first, the student must develop their own Verilog-HDL for each FPGA, as well as management software to process data and work with the cloud using Docker tools that will implement the IoT infrastructure.

Another implementation of the remote laboratory [11] is based on the most popular systems for development and prototyping, namely Arduino and Raspberry PI. The main module of this system is the Arduino debug board, which serves as a means of data collection and control for other devices. In this implementation, the board is connected to the LED cube, mechanical crane, sensor, etc. Execution of all functionality is performed on the Arduino board, and the assembly and download of executable code is implemented using Raspberry PI. Hardware communication between the two devices is via a USB port. The number of devices that can be connected to the Raspberry PI is limited by their technological capabilities. At the same time, the laboratory administrator once a week sets up stands for research according to user requests, indicating the port - the device file that will be used by the Arduino board, the device name - the device ID that will be displayed on the website where the Arduino processor type is specified, the programmer is the type of programmer according to the processor to be used, Baud is the connection speed and the termination program. This program configures the device with its default configurations. Thus, the laboratory allows students to create a full-fledged system of measurement or control. In addition, the system uses webcams so that the student can visually check the behavior of the test (device) that he performed (created).

The system [12] is a smart laboratory, where all devices are connected to a set of intelligent IoT equipment, and the whole set is placed on a single printed circuit board. All physical parameters in the laboratory are measured in real time and can be viewed on the information panel. Data transfer between all devices is performed using the MQTT protocol. All devices are connected to the Wi-Fi module ESP 8266 and act as MQTT clients, the tool for visual programming of data streams Node-RED (Raspberry Pi 3) acts as an MQTT broker (server). The user can use the smartphone to turn on / off any device in the laboratory via the MQTT server. Also, the condition of the equipment and information about the consumed electricity can be seen on the control panel. The ThingSpeak IoT service was used for data processing and transmission. This whole complex allows you to significantly reduce electricity consumption in the laboratory.

Another interesting implementation is the learning management system using IoT described in [13]. This system implements software and hardware, as well as IoT devices, which are used for

interesting experiments in the field of chemistry and robotics. This system is also essentially a module and integrated into the Moodle LMS system, which uses IoT devices and services that improve the learning process for both teachers and students. With this application, students can use Internet of Things sensors and robotic tools to remotely perform an experiment via LMS (Learning Management Systems). At the same time, students gain full control over the equipment and devices for the experiment. In addition, during the experiment, the student always has access to real-time information about the characteristics of the environment and the parameters of the experiment using sensors of the Internet of Things. All these indicators are transmitted to the LMS through the IoT module. The proposed application is integrated into LearnSmart LMS, which is based on Moodle.

The authors of [14] proposed a remote laboratory for the study of photovoltaic systems for remote experiments. The system consists of 3 subsystems: power electronics, sensors and communications. The remote lab is controlled by a dsPIC30F microcontroller, which acts as an intermediary between the web server and the power system. The communication subsystem consists of two parts: communication with the application on the PC via EM203 and communication with the power system and sensors via dsPIC. The graphical interface was implemented in Visual Basic Programming for date storage, user identification, development of measurement strategies, power system management. The maximum duration of one session is one hour. But in reality, the experiment can be performed in 30 minutes, in addition, for each lesson is offered double the time in case students have problems, or there is a need to conduct more tests. Another interesting feature of this system is the feedback of students and teachers through the system or service feedback on the experiment.

From the literature it is clear that there are two main areas of development of smart laboratories:

- the first is the so-called virtual laboratories, where all work with the equipment is performed remotely and students do not have direct access to the equipment, which in some cases is necessary, for example, when monitoring power equipment;
- the second - smart laboratories that provide comfort in a real laboratory and optimization of the workflow and electricity consumption. But they in turn impose restrictions on the number of students who can experiment with real equipment.

For example, if several research stands are connected to a specialized PC and the student uses only one, then why is it not possible for another student to be able to use the free stand remotely. Therefore, it is proposed to combine these two concepts and develop a smart laboratory where students could work both permanently and remotely. In addition, the system should optimize the energy consumption of such a complex. It is also advisable to use an artificial intelligence system for the so-called deferred experiments, which is described below, as well as introduce into the system a submodule of equipment accounting [20] based on NFC technology.

### 3. System structure

Given the previous work, it is proposed to develop the concept of a smart laboratory, which would combine the capabilities of a remote laboratory and a regular classroom for practical training. In addition, it would be advisable to integrate a system of control and management of electricity consumption by devices located in the laboratory. Due to the fact that such a concept involves complex management of facilities, it is proposed to use as a management manager the ANN apparatus, which would work in conjunction with the IoT cloud service. Figure 1 shows a simplified structure of such a laboratory. As can be seen from Figure 1, the system is divided into two subsystems - control of electrical appliances to ensure comfort in the room and control of training stands. These two subsystems are independent of data transmission and control channels.

The control subsystem of electrical equipment includes all the equipment that provides the educational process, namely the projector, air conditioners and laboratory heating, lighting and other equipment needed to ensure a comfortable learning environment in the laboratory.

In the current implementation, the control subsystem includes the following equipment:

- Smart projector for presentation of lectures or other teaching materials
- Smart Condition for air conditioning control and temperature control in the laboratory room.
- Smart Light - intelligent lighting control system in the laboratory



Additional control systems can be included in the current subsystem based on the wishes of lecturers and students or faculty management.

Each such device, such as a projector, according to Figure 2, is equipped with additional modules - a smart meter for electricity consumption and a specialized relay to control the operation of the projector.

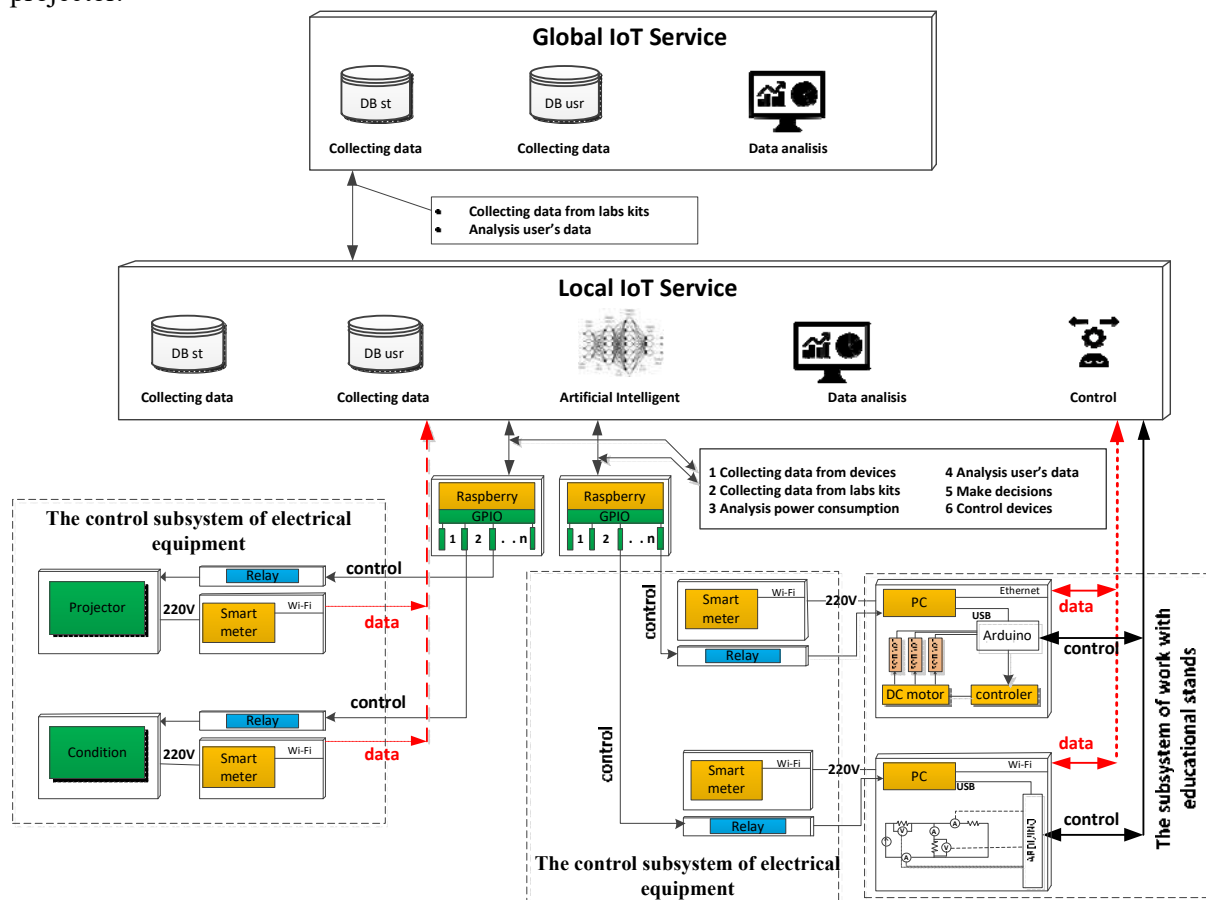


**Figure 1:** Simplified structure of IoT-based Laboratory

Data collection from the SmartMeter and relay control is performed via a network control controller, such as a Raspberry Pi or BeagleBone board. Such a device serves as a gateway between the IoT service and the equipment. Its main functions include collecting data on the state of the equipment (data is read directly from smartmeters or from the cloud, or simultaneously) and control the inclusion of devices (commands from the local IoT service).

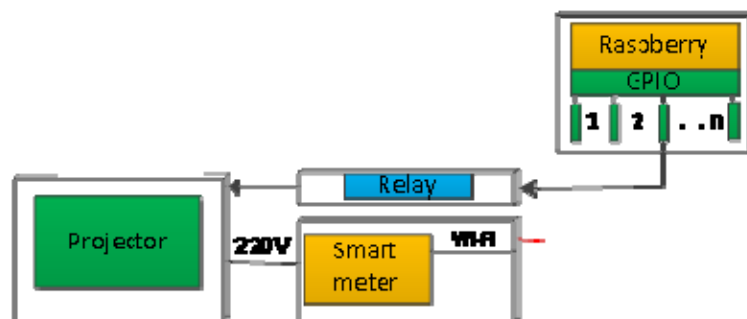


**Figure 2:** Smart Projector System Structure

The subsystem of work with educational stands works on a similar principle. The stand consists of a PC, where the necessary software to remote access and communication with a debug board (in this case Arduino) or other controller is installed in accordance with the tasks of the stand, sensors and

controls. For example, we will consider a stand for Investigation of the simplest DC circuits, the structure of which is shown in Figure 3. The stand includes a PC with software installed, board for getting electric values from investigated DC circuit, a set of sensors for measuring the current and the voltage, etc.

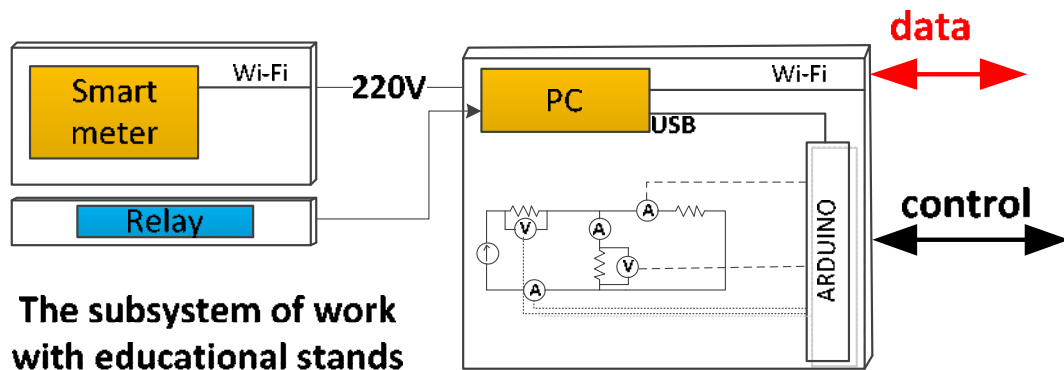


Figure 3: "Smart" stand for Investigation of the simplest DC circuits

To measure the electricity consumed by the stand, a smart meter has been introduced, which sends data on the power consumption of the PC and the controller or the computer power control relay together with the stand. All data is collected and sent to a local cloud service where the student can see the measurement results. Communication with the local cloud service and the PC is provided by a secure local area network (Ethernet).



Figure 4: Local IoT Service

The local IoT service, presented in Figure 4, consists of the following subsystems: 1) user database, which contains information about users and their access rights; 2) databases on measuring equipment and experimental data for each user and stand separately; 3) apparatus of artificial neural networks to control all elements of the laboratory in automatic mode [18]; 4) a set of software for analyzing user data; 5) specialized software for equipment management in remote control mode and in remote experiment mode.

The system also provides a global IoT service for working with remote users, the structure of which is shown in Figure 5. This subsystem has limited functionality. The system has user and device databases, as well as a data analysis system. That is, users can see the results of their experiments, perform data analysis and order a so-called remote experiment, when measurements will be performed only when the stand is vacated. In this implementation, the user will not have direct access to the equipment, but only pre-order the experiment according to the configuration specified by him.



Figure 5: Global IoT Service

This layout enables dividing the work process in the laboratory and ensuring the security of the whole complex [21], because remote users will not have a direct access to the PC and the stand.

In addition, learning should take place where students spend the most time. Today it is social networks. The introduction of social networks [26-28] in a smart laboratory will allow first of all the placement of a variety of materials:

- Announcements - a function when it is necessary to inform as many students as possible in a short time about changes in the educational process (change of class schedule), competitions, Olympiads that will take place in the near future;
- Survey - allows you to take into account the wishes and suggestions through various surveys;
- Photo album - a function that allows you to place educational material in drawings, illustrations, photographs. This is necessary in order to speed up and facilitate the assimilation of educational material;
- Video - this feature allows you to demonstrate the experimental basis for the learning process;
- Materials for learning - diagrams, tables, text material, interesting facts, etc. (everything that reflects the subject of study).

The social network allows the teacher to better remember students (correlate names and faces in the audience), understand their interests and allows them to jointly create and improve the course. Instead of simply consuming information, students become developers and experts in a virtual learning group environment (creating messages, discussions, resources, and more).

Thus, the use of social networks can give the learning process more interactivity, positively affect the results of cognitive activity of students, become an effective means of increasing motivation and quality of learning, organizing teamwork, joint project activities, individualize the student's virtual learning space.

#### **4. Description the stand for Investigation of the simplest DC circuits and experiment for it**

The stand (see Fig. 3) includes a PC with the following software: LabVIEW [22] and Arduino\_cc [23]. LabVIEW is a system engineering software for applications that require testing, measurement, and control with fast access to hardware and data analysis. The process of developing a SCADA system in LabVIEW is simpler than in "traditional" development tools. It is also a fundamentally different programming language called "G" and is a functional language that is similar in concept to C ++. Arduino\_cc is an integrated cross-platform Java development environment that includes a code editor, compiler and firmware transfer module. The environment is based on the programming language Processing. It is similar to the Wiring language. In general, it is C ++, supplemented by special libraries. This combination is chosen because different students have different levels of programming skills and it is desirable for beginners to use the Arduino environment, and senior students can develop projects of a more complex level, where it is advisable to use LabVIEW. The Arduino Expansion Shield for Raspberry board was chosen as the hardware data acquisition and control module [24] because it has more functionality than standard boards and the ability to integrate with the Raspberry Pi.

A classic voltage divider circuit is proposed as a research electrical circuit (Figure 6)

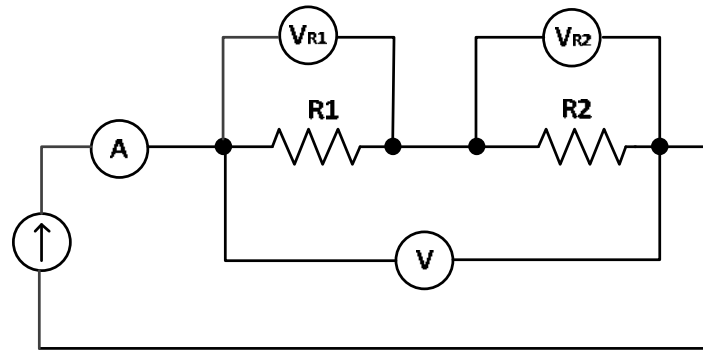


**Figure 6:** Voltage divider

The stand is based on the previous scheme and includes the following components:

1. controller of data collection from sensors and supply voltage control. In this particular implementation (Figure 7), an Arduino board is proposed, but you can also use another module based on different microcontrollers;
2. analog current sensor ACS712 on the Hall effect;
3. analog voltage sensor (V1 ... V3);
4. variable resistors R1 and R2 based on digital potentiometer X9C103S.

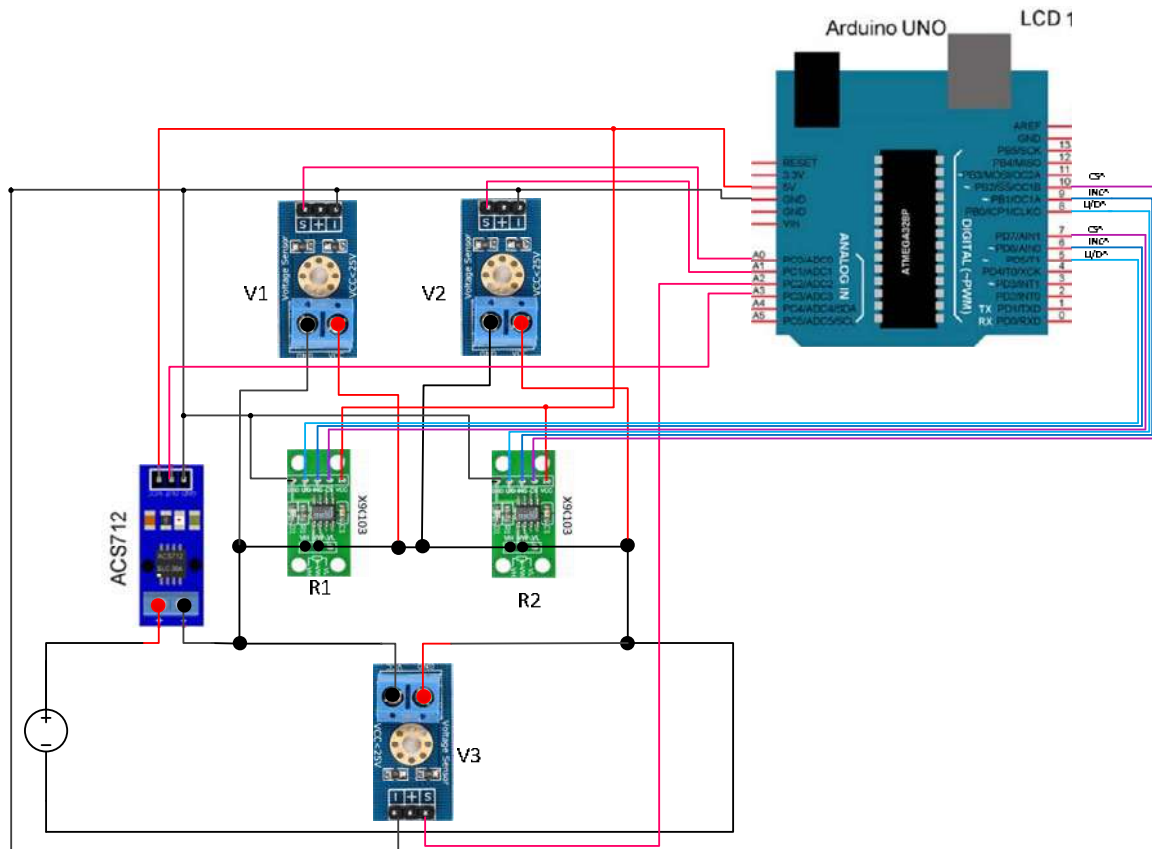


**Figure 7:** Measurement part of stand

#### 4.1. The experiment

The essence of the experiment is to experimentally check the basic relationships of electrical quantities for DC circuits with series connection of resistors. Student Increasing the value of one of the resistors from  $100 \cdot (\text{Number of students})$  Ohm to 200 Ohm in 20 Ohm steps, fills in table 1.

**Table 1**

Basic parameters

| R1 Ohm | R2 Ohm | I, Amps | V, Volts | VR1, Volts | VR2, Volts |
|--------|--------|---------|----------|------------|------------|
| 700    | 200    | 0.0133  | 12       | 9.31       | 2.66       |
| 700    | 220    | 0.013   | 12       | 9.1        | 2.86       |
| 700    | 240    |         |          |            |            |

## 4.2. Software for experiment and remote access

Figure 8 shows the concept of a graphical user interface for conducting an experiment. Despite the simplicity of the interface, it fully provides the necessary functionality and is a network client, and provides control of the queue of experiments and access rights of users (students) to the measuring stand, collection of measurements and control of the change in resistance on digital potentiometers for a specific student card.

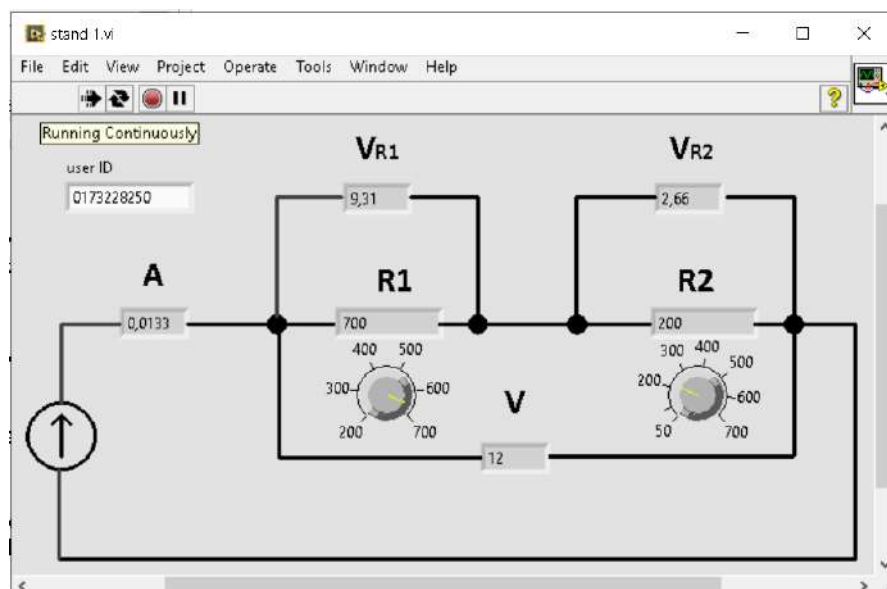


**Figure 8:** User interface for experiment

In addition to this, the program transmits data to a foggy service for remote access and control. There are two options for working:

1. When a student conducts an experiment locally, the control interface on the fog service will become inactive and will be asked to stand in the queue for the experiment, although the process itself will be visible to other students.
2. If the control and experiment are removed, the main interface on the PC will also become inactive and the student will be asked to enroll in the queue or select another free stand in the laboratory.

All parameters will be collected by the Arduino board and transferred to the PC, from where, after pre-processing (if necessary), they are transferred to the local IoT service.

The control system of the PC itself includes two devices:

1. DELOCK 25242 power supply control controller: ATX Power Supply Controller or its analogue, which will be connected to the Raspberry Pi (see Figure 2);
2. smart current-consumption meter smart-MAC [25], which will independently transmit data to the local IoT service on the electricity consumed by a specific PC.

This information is needed to analyze the total power consumption of the laboratory stands and to further intelligently control the PCs and stands described above, as well as to automatically control the on / off and start of the PCs themselves.

## 5. Basic functions and algorithms of IoT laboratory

From the above structure of the system (see Fig. 1), the laboratory must have many scenarios for working with users and operating scenarios. In addition, the structure and algorithms of work in this system should be divided into two main classes - algorithms of the subsystem of control of educational equipment in the laboratory and algorithms of software operation of test benches, where research will be conducted (Fig. 9).

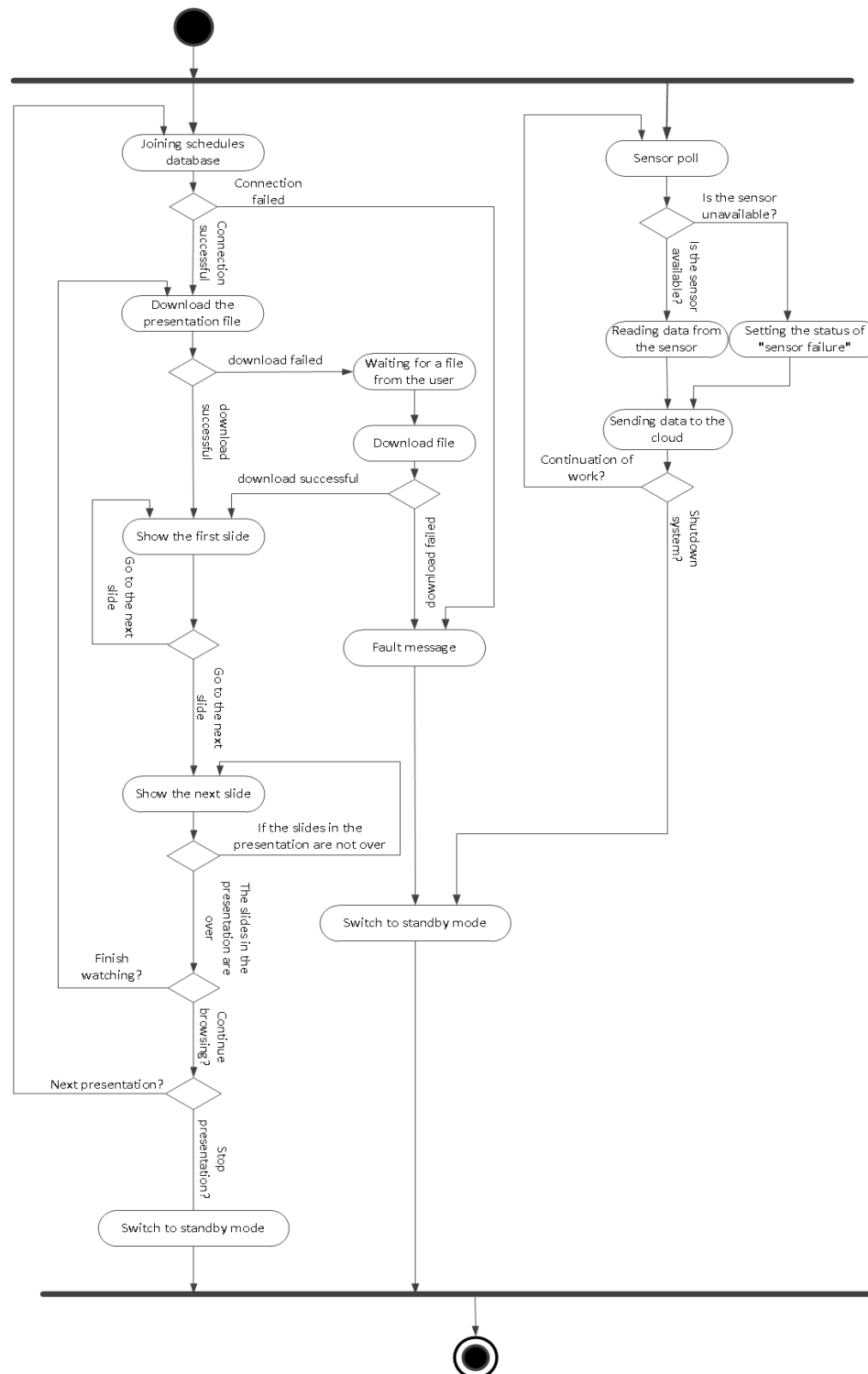


Figure 9: Activity diagram of smart projector system

The main functions of the control system of educational equipment include: control of equipment operation, equipment management, collection of data on the use of equipment, collection of energy consumption statistics of classroom equipment for each device separately and in total.

Below is a scenario of the projector, which is located in the laboratory. The main controller of the projector control is a script that operates on a local IoT service and has access to a database of schedules in a particular laboratory, where it loads the time of laboratory, also downloads presentations of laboratory classes, if already downloaded by the teacher. If (before a certain lesson the teacher) did not download the presentation files, the system simply turns on the projector in standby mode relative to the schedule, and the teacher can use the Web interface to download the presentation or directly connect to the projector and give a presentation. In parallel with the projector works a smart electricity meter, which sends data on energy consumption by the projector. These data are collected on the local cloud service for further analysis by the laboratory administrator (if necessary), as well as for reporting to the accounting department.

## 6. Conclusions

In this paper, the authors proposed a conceptual model of a training laboratory based on IoT, which combines the functions of a remote laboratory and a regular classroom for practical classes. The general structure of such system and the basic modules of functioning are presented.

Despite the fact that only one test bench is described, it is planned to implement about 10 test benches for each PC (see Figure 1) for different types of tests.

The subsystem of control and management of electricity consumption by the equipment working in the laboratory and the scenario of operation of measuring stands is described. Scenarios for managing objects in this system and working with users are revealed.

In further works, the authors plan to implement the main modules of the system and the use of Artificial Neural Networks for intelligent control and protection of the system as a whole.

## 7. Acknowledgements

The authors would like to thank the Ukrainian-German Education and Research Center of West Ukrainian National University for the opportunity to use their training laboratories and support from the Internet of Things project: Emerging Curriculum for Industry and Human Applications ALIOT (573818-EPP-1-2016-1-UK -EPPKA2-CBHE-JP) for the provided equipment for measuring stands.

## 8. References

- [1] D. Singh, G. Tripathi, and A. J. Jara, "A survey of internet-of-things: Future vision, architecture, challenges and services." in 2014 IEEE worldforum on Internet of Things (WF-IoT). IEEE, (2014): 287–292. DOI: 10.1109/WF-IoT.2014.6803174
- [2] Srinivasan, C. R., Rajesh, B., Saikalyan, P., Premasagar, K., & Yadav, E. S. "A review on the different types of internet of things (IoT)." Journal of Advanced Research in Dynamical and Control Systems, 11(1), (2019). 154-158.
- [3] Jeretta Horn Nord, Alex Koohang, Joanna Paliszkievicz, "The Internet of Things: Review and theoretical framework." Expert Systems with Applications, Volume 133, (2019). 97-108, ISSN 0957-4174, DOI: 10.1016/j.eswa.2019.05.014.
- [4] P.P. Ray, "A survey on Internet of Things architectures." Journal of King Saud University - Computer and Information Sciences, Volume 30, Issue 3, (2018). 291-319. DOI: 10.4108/eai.1-12-2016.151714.
- [5] O. Kanoun, T. Keutel, C. Viehweger, X. Zhao, S. Bradai, S. Naifar, C. Trigona, B. Kallel, I. Chaour, G. Bouattour et al., "Next generation wireless energy aware sensors for internet of things: A review." 15th International Multi-Conference on Systems, Signals & Devices (SSD). IEEE, (2018).1–6. DOI: 10.1109/SSD.2018.8570695

- [6] Ericsson Mobility Report, November 2020. URL: <https://www.ericsson.com/4adc87/assets/local/mobility-report/documents/2020/november-2020-ericsson-mobility-report.pdf>
- [7] IoT evolution towards a super-connected world, URL: <https://arxiv.org/abs/1907.02589>
- [8] Saravanan K., Julie E.G., Robinson Y.H. "Smart Cities & IoT: Evolution of Applications, Architectures & Technologies, Present Scenarios & Future Dream." Internet of Things and Big Data Analytics for Smart Generation (2018).135-151. DOI: 10.1007/978-3-030-04203-5\_7.
- [9] Ruth Ande, Bamidele Adebisi, Mohammad Hammoudeh, Jibran Saleem, "Internet of Things: Evolution and technologies from a security perspective." Sustainable Cities and Society, Volume 54, (2020) 101728. DOI: 10.1016/j.scs.2019.101728.
- [10] The shape of the MCU market URL: <https://www.embedded.com/the-shape-of-the-mcu-market/>
- [11] A. Fernández-Pacheco, S. Martin and M. Castro, "Implementation of an Arduino Remote Laboratory with Raspberry Pi." 2019 IEEE Global Engineering Education Conference (EDUCON), Dubai, United Arab Emirates, 2019, pp. 1415-1418, DOI: 10.1109/EDUCON.2019.8725030.
- [12] M. Poongothai, P. M. Subramanian and A. Rajeswari, "Design and implementation of IoT based smart laboratory." 2018 5th International Conference on Industrial Engineering and Applications (ICIEA), Singapore, 2018, pp. 169-173, DOI: 10.1109/IEA.2018.8387090.
- [13] K. Merashad and A. Hamieh, "Using Internet of Things to Enhance Remote Experiments in Learning Management Systems." 2019 IEEE International Smart Cities Conference (ISC2), Casablanca, Morocco, 2019, pp. 458-464, DOI: 10.1109/ISC246665.2019.9071786.
- [14] A. D. Martin, J. M. Cano, J. R. Vazquez and D. A. López-García, "A Low-Cost Remote Laboratory for Photovoltaic Systems to Explore the Acceptance of the Students." 2020 IEEE Global Engineering Education Conference (EDUCON), Porto, Portugal, 2020, pp. 1333-1337, DOI: 10.1109/EDUCON45650.2020.9125211.
- [15] T. Gomes, S. Pinto, T. Gomes, A. Tavares and J. Cabral, "Towards an FPGA-based edge device for the Internet of Things," 2015 IEEE 20th Conference on Emerging Technologies & Factory Automation (ETFA), Luxembourg, 2015, pp. 1-4, DOI: 10.1109/ETFA.2015.7301601.
- [16] P. G. Lopez , A. Montresor, D. Epema, A. Datta, T. Higashino, A. Iamnitchi, M. Barcellos, P. Felber and E. Riviere, "Edge-centric Computing: Vision and Challenges", in ACM SIGCOMM Computer Communication Review, Volume 45, Number 5, October 2015, pp.37-42. DOI: 10.1145/2831347.2831354
- [17] Docker URL: <https://docs.docker.com/>
- [18] W. Raad, M. Bueno-Delgado, M. Deriche and W. Suliman, "An IoT Based Inventory System for High Value Laboratory Equipment." 2019 Sixth International Conference on Internet of Things: Systems, Management and Security (IOTSMS), Granada, Spain, 2019, pp. 314-319. DOI: 10.1109/IOTSMS48152.2019.8939259.
- [19] N. Fujii and N. Koike, "IoT Remote Group Experiments in the Cyber Laboratory: A FPGA-based Remote Laboratory in the Hybrid Cloud." 2017 International Conference on Cyberworlds (CW), Chester, 2017, pp. 162-165, DOI: 10.1109/CW.2017.29.
- [20] I. Turchenko et al., "Approach to neural-based identification of multisensor conversion characteristic." 2009 IEEE International Workshop on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, Rende, Italy, 2009, pp. 27-31, DOI: 10.1109/IDAACS.2009.5343030.
- [21] M. Komar et al., "High performance adaptive system for cyber attacks detection." 2017 9th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), Bucharest, 2017, pp. 853-858, DOI: 10.1109/IDAACS.2017.8095208.
- [22] LabVIEW Community Edition, URL: <https://www.ni.com/ru-ru/shop/labview/select-edition/labview-community-edition.html>
- [23] Arduino IDE URL: <https://www.arduino.cc/>
- [24] Arduino Expansion Shield for Raspberry Pi model B, URL: <https://www.dfrobot.com/product-1148.html>
- [25] Smart meters for any kind of consumptions! URL: <https://smart-mac.com/en/>



- [26] G. Marques and R. Pitarma, "Using IoT and Social Networks for Enhanced Healthy Practices in Buildings," *Smart Innovation, Systems and Technologies*, (2018) 424–432. DOI: 10.1007/978-3-030-03577-8\_47
- [27] V. Tyagi and A. Kumar, "Internet of Things and social networks: A survey." 2017 International Conference on Computing, Communication and Automation (ICCCA), Greater Noida, India, (2017) 1268-1270, DOI: 10.1109/CCAA.2017.8230013.
- [28] G. A. Stelea, V. Popescu, F. Sandu, L. Jalal, M. Farina and M. Murrone, "From Things to Services: A Social IoT Approach for Tourist Service Management." in *IEEE Access*, vol. 8, pp. 153578-153588, 2020, DOI: 10.1109/ACCESS.2020.3018331.

# Self-Improvable Computer System Model and Architecture Based on Reconfigurable Hardware, Automatic Design and Synthesis Tools and Artificial Intelligence Technologies

Anatoliy Melnyk<sup>a,b</sup> and Viktor Melnyk<sup>a,b</sup>

<sup>a</sup> Lviv Polytechnic National University, Lviv, Ukraine,

<sup>b</sup> The John Paul II Catholic University of Lublin, Lublin, Poland

## Abstract

In the article the idea on self-improvable computer systems creation is formulated and directions of their self-improvement and the proposed approach to self-improvement are described. This approach is based on methods and tools for specialized processors high-level automatic design and synthesis, new computational models, including the one that uses principle of self-configuration, new architecture of self-configurable FPGA-based computer systems and computational intelligence technologies. An FPGA was chosen as the hardware basis of self-improvable computer systems. The directions of the FPGA-based reconfigurable computer systems development over the next years are considered. The self-improvable computer system operation model and architecture are proposed, and their components are described. Among them there are method of self-configuration of a computer system with reconfigurable logic, Artificial Intelligence technologies, and software basis of self-improvable computer systems. The core of this software are specialized processors high-level design tools, which are used as a basic mechanism for computer system self-improvement. Expected benefits of self-improvable computer systems creation are estimated at the end of the article.

## Keywords

Self-improvable computer systems, reconfigurable computing, automatic design and synthesis tools, self-configurable computer systems, FPGA, Artificial Intelligence.

## 1. Introduction

At the junction of the development of reconfigurable computing and Artificial Intelligence a new scientific direction has emerged in recent years, and it was entitled with the term “Evolvable Hardware”. Under the term “Evolvable Hardware” is understood a hardware, that is able to permanently solely change its architecture and behavior, interacting with the environment it is placed in, possessing necessary for this embedded means for self-learning, self-optimization, self-debugging, self-improvement etc. with the aim of more effective on those or other criteria execution of the tasks assigned to them [1]. An ability to evolve in the evolvable hardware is ensured by the deployment of the evolutionary algorithms – the direction in Artificial Intelligence, which is used for evolutionary modelling. These algorithms simulate processes that reflect the basic provisions of the theory of biological evolution - the processes of selection, recombination, mutation and reproduction [2], [3].

The base of the evolvable hardware is reconfigurable logic devices, structure and performed functions of which can be changed in accordance with the task requirements. Today reconfigurable hardware means are mainly implemented in the form of semiconductor integrated circuits entitled FPGAs – Field Programmable Gate Arrays [4]. A direction of research and engineering activities on the creation and application of reconfigurable logic devices in computer technology is called Reconfigurable computing [5], [6], while computer systems based on these devices are called reconfigurable computer systems (RCCS) [7], [8].

CMIS-2021: The Fourth International Workshop on Computer Modeling and Intelligent Systems, April 27, 2021, Zaporizhzhia, Ukraine

EMAIL: anatoliy.melnyk@kul.pl (Anatoliy Melnyk); viktor.melnyk@kul.pl (Viktor Melnyk)

ORCID: 0000-0002-8981-0530 (Anatoliy Melnyk); 0000-0002-5046-8002 (Viktor Melnyk)



© 2021 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

RCCSs are able to perform high-performance computations owing to high performance characteristics of their processing components – FPGAs, and advances in design methodology of specialized processors, which are synthesized in FPGA.

The direction of the reconfigurable computing is based on the idea of computer hardware adjustment to perform the specific computational tasks. Feasibility of the reconfigurable computing approach in the real computer systems became possible with the advent of modern FPGA with high integration degree. FPGAs are such integrated circuits that can store and, when it is needed, can change the configuration. After loading the configuration into the FPGA chip, it turns into the specialized computing module, for example, into the hardware-oriented specialized processor.

Combining in the computer system general-purpose processors with specialized processors implemented in reconfigurable logic allows one to increase its overall performance by 2-3 orders of magnitude [9]. The functional commitment of RCCS can be changed with preserving its high performance on another class of problems by implementing in reconfigurable logic of a new specialized processor with a corresponding structure and functionality.

However, along with high performance, that is the advantage of RCCSs, there are also challenges associated with their use. These challenges are the significant timing expenses for load balancing between the processing units and often unavailability of previously designed specialized processor IP cores, required for implementation in the reconfigurable logic. It forces to develop them from scratch. Also, there are high additional requirements to the qualification of RCCS operators, as they, besides modeling and programming, have to perform system analysis, design of the specialized processors architecture, their synthesis and implementation in reconfigurable logic.

In the self-configurable computer systems (SCCS), these labor-intensive and time-consuming operations are fully automated and replaced from an operator to a computer system. The method of self-configuring allows one to exploit all the potential provided by the property of reconfiguration and thus ensures one of the leading places for SCCS in high-performance computing means.

## 2. Problem Statement

The beginnings of Evolvable Hardware research date back to the mid-90s, when the first attempts were made [10], [11] to create evolvable hardware based on Xilinx FPGA. This was preceded by the publication of several theoretical works, in particular [12], [13], which described the conceptual basis of Evolvable Hardware. Subsequently, a number of scientific papers were published, including articles and monographs describing theoretical research, and some practical works in the direction of Evolvable Hardware [14-21], a number of scientific conferences and seminars, including a series of such conferences: IEEE International Conference on Evolvable Systems, IEEE International Conference on Self-Adaptive and Self-Organizing Systems, Evolvable Systems: From Biology to Hardware; NASA/DoD Conference on Evolvable Hardware.

Along with significant progress in the developments in the reconfigurable computing, today there is some "slack" in the direction of Evolvable Hardware. The reason is, most of all, in the absence of significant, fundamental steps in the development of computer systems and artificial intelligence systems, such that would change approaches to their design, information processing methods they use, methods of tasks execution, their architecture. In fact, if to go a little away from the subject of Evolvable Hardware and reconfigurable computing, and analyze actual developments in computer technology in general, we can conclude, that despite the permanent technological progress (improvement of integrated technologies), methodological advances (multiple cores in a processor, multiple threads in a process), computer architecture in general, remains unchanged and continues to be based on the computing model proposed by John von Neumann more than 70 years ago.

So, it is needed to search for a new theoretical basis, which will become a platform for future developments in the Evolvable Hardware direction, on which ideas and statements of this direction will obtain an effective embodiment and will give an opportunity to make significant progress in the development of computer technology.

In this article, the authors estimate the possibilities of creating an architecture of computer systems of a fundamentally new type - self-improvable CS, which are able to independently improve their technical characteristics and functionality.

### **3. The Idea on Self-Improvable Computer Systems and Directions of Their Self-Improvement**

The conception of self-improvement has so far applied only to a human as a carrier of intelligence. Moreover, self-improvement is multifaceted and covers almost all areas of human activity. In particular, people train their bodies in order to achieve certain physical results, such as running speed, swimming, jumping range or throwing various shells, and so on. In this sense, activities in the field of physical culture and sports are aimed at self-improvement. It is common knowledge that trained people in their activities significantly outnumber the untrained ones, which is the result of their self-improvement.

Naturally, the highest level of self-improvement of a person is his education in its broadest sense: at home, at school and other educational institutions, in clubs, interest groups, independently, and so on. Education allows to realize to the greatest extent in the person its potential intelligence and is a basis of all-human progress.

Applying the above considerations and the conception of self-improvement to computer technology, we can formulate a definition of self-improvable computer system (SICS) as a computer system capable of solely (independently) improving its technical characteristics and functional capabilities.

It should be noted that SICS is a CS that has already been created and is in operation. SICS as technical systems should be able to improve their technical characteristics and functional capabilities just as a human is able to improve her physical results.

When we talk about the self-improvement of the CS, we mean first and foremost, the ability of the CS to increase the efficiency of use of equipment available in it during its operation, because its amount in the already created CS is constant. Therefore, the higher is the performance, the more efficiently this equipment is operated. Moreover, to increase efficiency of use of the equipment it is possible only by increase of performance of CS on certain classes of tasks. Taking into account the complexity of modern CS and their components, we can expect a significant (in times, if not dozens or hundreds of times) self-improvement of computer systems on task classes. After all, it is common knowledge that in some classes of tasks, due to the lack of harmonization of computer architecture to the features of these tasks, only units of the percentage of peak performance of modern supercomputers are used [22].

Another important technical characteristic that the CS could improve on its own is the value of its power consumption. After all, not all tasks solved in the CS require the operation of all components of the CS at the maximum frequency, when the power consumption is greatest. Therefore, by controlling the frequency of operation, the CS can reduce power consumption. In addition, significant energy can be saved by disabling the components of the CS that are not involved in certain tasks. That is, the prospects for self-improvement in this direction are also promising.

Is the list of those that the CS can improve by itself limited to these technical characteristics of the CS? Of course not. In particular, SICS can independently, if necessary, increase the accuracy of computations and the dynamic range of processed data, using pre-created virtual bit grids.

### **4. The Main Hypothesis of the Study and the Proposed Approach to Self-Improvement of the Computer System**

In order to solve arbitrary problems, computer systems must be universal and based on a universal computational model. On the other hand, the use of specialized processors or specialized parts of heterogeneous chips speeds up the operation of universal computers and increases their efficiency because their hardware can be customized to the specifics of the computations [23, 24]. This explains the widespread use of specialized processors in the latest high-performance computer systems [25]. However, not always at the design stage of computer systems are known all the features of the problems that will be solved by them in the future, which does not allow to take into account these features in their architecture in order to increase efficiency. It should also be noted that the extremely high complexity of modern computer systems does not allow to provide the highest values of their technical characteristics at the design stage.

The main hypothesis of the study is that the introduction of self-improvement mechanisms into the architecture of computer systems should provide an opportunity at the stage of their operation to improve the efficiency of equipment usage by increasing performance on classes of algorithms up to 2-3 orders of magnitude, or to decrease the power consumption the same level down.

The question of the implementation of this hypothesis and the creation of self-improvable computer systems was made possible by the creation of methods and tools for high-level automatic design of specialized processors [26], new computational models, including the one that uses principle of self-configuration [27], new architecture of self-configurable FPGA-based computer systems [28, 29] and computational intelligence technologies [30, 31].

## **5. Hardware Basis of Self-Improvable Computer Systems**

As mentioned above, the hardware basis of self-improvable CS are reconfigurable components – FPGAs, the structure and functions of which can be changed according to the requirements of the problem to be solved. FPGAs have begun to be used as elements for the implementation of computer devices quite long ago. However, they have become used in a high-performance computing relatively recently. Thus, the creation of reconfigurable high-performance computer systems is relatively new scientific and technical direction. There are many reconfigurable high-performance computer systems today, some of which can be classified as supercomputers [32-34]. We believe that the trend of reconfigurable technologies implementation in the high-performance computing will continue in the coming years. Along with this, we should expect further increase of level of the application-specific computing devices design and implementation in FPGAs, and simplification of the design process.

In our view, the development of reconfigurable computer systems over the next years will take place in several directions:

1. Through the developments in integrated circuit manufacturing technologies and FPGA architecture, that will improve their technical parameters (rise the performance, increase the amount of on-chip resources, decrease power consumption ratio).
2. Design technology and tools for computer devices implementation in FPGA will further develop, and will reach a level at which rapid automatic synthesis of highly optimized specialized processors for execution of algorithms based on their high-level representation, becomes a reality, and will largely replace the use of hardware description languages in computer devices design by the high-level programming.
3. The software will be created, that run on the level of operating system, that will enable computing load automatic balancing between fixed and reconfigurable components of a computer system, automatically manage the FPGA operation, in particular, carrying out their reconfiguring during the computing processes, and use FPGAs in multitask mode.
4. Application of the reconfigurable logic in computer systems will become common, including placing the reconfigurable logic resources into the same chip with the general-purpose processors, e.g., as we can already see in Intel Acquisition of Altera.
5. Efforts will be made in order to introduce self-monitoring, self-optimization and self-improvement technics into the reconfigurable computer systems (one such approach is shown in [35]).

The processing core of self-improvable computer system are general-purpose processor and FPGA-based reconfigurable environment (RCE) connected between themselves as shown in Figure 1. Here, the universal processor is aimed, among other things, to execute programs for the initial creation of hardware configurations that are implemented in the reconfigurable environment, and their improvement in the process of CS operation.

## **6. Methodological Basis of Self-Improvable Computer Systems**

As already mentioned, the question of creation of self-improvable computer systems has become possible due to the creation in recent years of new methods of high-level automatic design of specialized processors, new computational models, including the one that uses principle of self-configuration, as a basis for self-improvement, and a new architecture of self-configurable FPGA-based computer systems, as well as the creation and development of modern computational intelligence technologies, including self-learning technologies, neural networks, genetic algorithms, collective intelligence, which can be used as a basis for self-improvement.

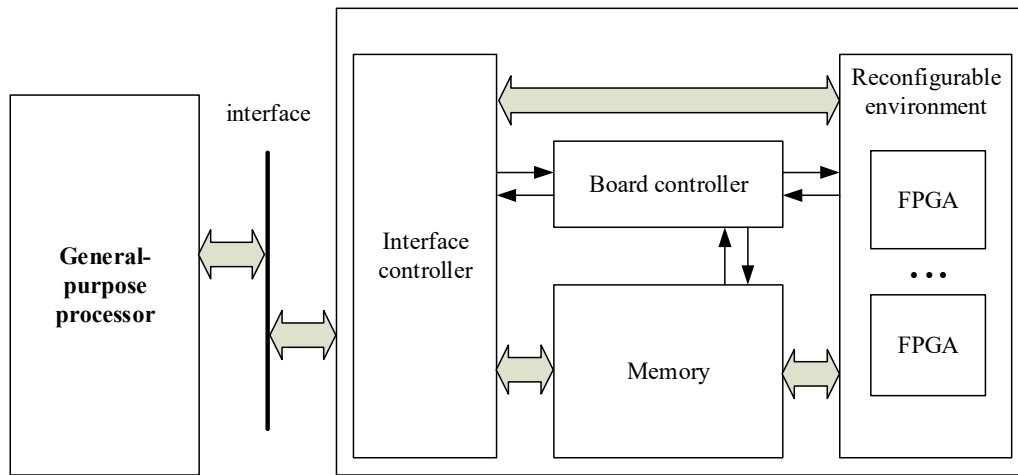


**Figure 1:** The processing core of self-improvable computer system

A description of the proposed mechanism of self-improvement of the CS together with the above-mentioned innovations used in this mechanism, are given below.

## 6.1. Model and Architecture of Self-Improvable Computer System

Given the iterative nature of self-improvement processes in general, we can determine the necessary conditions for self-improvement of the CS, namely:

- Knowledge of technical characteristics that need to be improved;
- Knowledge of ways to improve technical characteristics;
- Ability to measure the values of current technical characteristics and compare them with previous values;
- Availability of hardware, the structure and functions of which can be changed;
- Availability of methods and means of self-adjustment of the structure and functions of hardware by the computer system for self-improvement.

To meet these conditions in the CS at the design stage it is necessary to provide the ability to measure, store and compare its technical characteristics, which can be the subject for improvement, i.e., the efficiency of hardware use and power consumption, as well as to implement the means for the self-improvement to be performed.

Let's define the basic components of the CS, the availability of which will ensure the fulfillment of the above conditions and, accordingly, the possibility of its self-improvement.

1. the information base of technical characteristics that need to be improved, to which these characteristics are submitted and accumulated during the operation of the CS;
2. the information base of knowledge, in which algorithms and corresponding software to improve the technical characteristics of the CS should be stored;
3. the information base of specifications, in which the specifications of tasks and the corresponding descriptions of hardware configurations of CS should be stored;
4. hardware, which, in addition to a fixed part in the form of one or more universal programmable processors (UPP), contain a reconfigurable environment, the presence of which will provide the ability to specify the initial structure and functions of the CS and their subsequent changes to improve technical characteristics;
5. software tools for CS self-configuring, the goal of which is to automatically create a configuration of the reconfigurable hardware to perform user tasks;
6. means of measuring technical characteristics of CS during the execution of user tasks. These means can be hardware or software;
7. artificial intelligence software, the goal of which is to introduce changes to previously generated hardware configurations based on data from the knowledge information base, the information base of technical characteristics, and data obtained from the means of measurement of the technical characteristics, in order to improve the technical characteristics of the CS.

These means must operate in a CS in parallel with the execution of user's tasks.

High-level structure of self-improvable CS, which shows the principles of interaction of its basic components and the functions they perform, is shown in Figure 2.

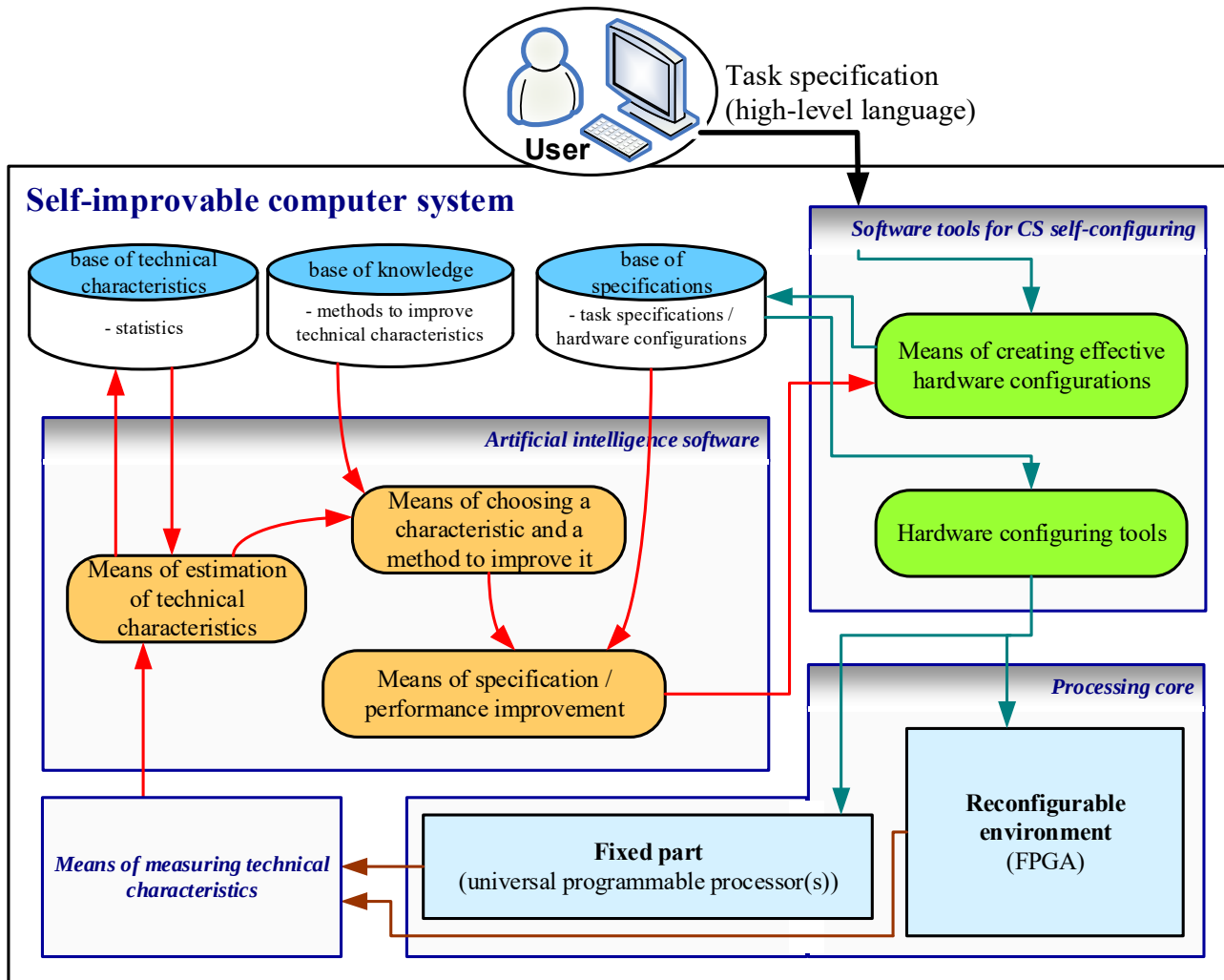


**Figure 2:** High-level structure of self-improvable computer system

Self-Improvable computer system operation model consists in following. The user submits a task specification in the form of computer program to the SICS. This program, using software tools for CS self-configuration, is automatically compiled first into subprograms for UPP and RCE, then the second one into a software model of a specialized processor to execute RCE subprogram, and then both into executable file. This file consists of object codes for execution by the universal processor(s) and RCE configuration codes, and specifies the dedicated hardware configuration corresponding to submitted input program. The created executable file together with the corresponding input program, subprograms for UPP and RCE and the software model of the specialized processor to execute RCE subprogram, are stored in the information base of specifications.

After program initialization, from the information base of specifications the corresponding hardware configuration (executable file) is being loaded. Then, during the program execution, the technical characteristics of the CS are measured by appropriate means and submitted to the artificial intelligence software. After that, the process of self-improvement of the CS begins, which is iterative. In the first iteration, the obtained values of the characteristics are stored in the appropriate information base and simultaneously submitted for processing to the means of choosing a characteristic and a method to improve it. These means select some characteristics (according to predefined criteria) and a method to improve it. The corresponding specification is being read from the information base of specifications and a trial of its improvement is performed. The improvement can be performed at various levels: at the level of the input

program, at the level of obtained subprograms of UPP and RCE, at the level of software model of specialized processor, and at the level of FPGA configuration. The obtained improved specification is submitted to software tools for CS self-configuration, which automatically create a new configuration and store it to the appropriate database. This configuration can be loaded either with the next initialization of this program, or during its current execution (which is more difficult to implement).

In each subsequent iteration, the obtained values of the technical characteristics are compared with their previous ones, read from the information base, and depending on the results, the means of artificial intelligence decide to further improve this characteristic, choose another characteristic, or return back to the previous configuration. As a result, the CS improves itself.

Next, we consider the method of self-configuration of CS with reconfigurable logic, which is realized by software tools for CS self-configuration. An overview of these software tools will be given in the next section.

## **6.2. Method of Self-Configuration of Computer System with Reconfigurable Logic**

The principle of self-configuration, which is a basis of self-configurable computer systems design concept, provides operation of a computer system with reconfigurable logic by independent and automatic determining and creating of hardware configuration effective for the submitted task – its architecture and behavior of its hardware (in the form of a data structure - a file specifying the configuration) and automatic and immediate acquisition of this configuration when the task is executed.

To implement the principle of self-configuration in a computer system with reconfigurable logic, a method of self-configuring was developed, as well as a method of information processing based on it, which are described in detail in [27-29].

The method of self-configuring of a computer system with reconfigurable logic, which contains general-purpose processors and reconfigurable environment built with FPGA or other types of reconfigurable integrated circuits, is that a software is introduced into computer system, that during compilation of user programs solely and independently extract from it such fragments, the execution of which in the reconfigurable environment accelerates the computer system, and distribute this program to the subprograms for GPP and for RCE formed from the selected fragments. Next, the GPP subprogram is compiled into object code, and the RCE subprogram into its configuration file so that first from this subprogram a software model of a specialized processor for its execution is generated, then its logic synthesis is performed and thus an executable file from these two components is created.

At the stage of program loading after its initialization the executable code of the GPP subprogram is being loaded into a main memory and, at the same time, the configuration code is being loaded into a RCE and thus creates an ASP in there. Then program execution is being performed.

This ensures automatic and transparent to the user and efficient (in terms of performance and hardware use) use of reconfigurable logic in the computer system. The method of self-configuring given above is one of the basic mechanisms for the CS self-improvement.

## **6.3. Artificial Intelligence Technologies**

CS self-improvement tools will also be based on the use of artificial intelligence technologies. The first in this context is the use of self-learning technologies, which allow to significantly improve the services of CS through the accumulation and application of historically acquired knowledge. Secondly, it is the use of mechanisms of expert systems that allow for in-depth analysis of the possibility of improving the service programs of CS. Third, it is the use of computational intelligence technologies: neural networks, genetic algorithms, fuzzy logic, etc., for the fast execution of optimization procedures [31]. And fourth, it is the use of technologies of collective intelligence: multi-agent systems, ant intelligence, swarm intelligence, etc., to quickly find the best solutions, in particular, described in [36].

The integration of the above technologies in order to obtain a synergistic effect from their joint use is a factor that will ensure the development of the architecture of self-improvable computer systems.



## 7. Software Basis of Self-Improvable Computer Systems

Below we will consider the basic software tools of self-improvable computer systems. First of all, we consider the tools for generating software models of specialized processors from description of their operation in a high-level programming language. Next, we provide a brief overview of other software tools used to implement self-configuration of the CS.

### 7.1. Specialized Processors High-Level Design Tools as a Basis of the Mechanism of Computer Systems Self-Improvement

The basis of the mechanism of computer systems self-improvement, which provides the ability to solely and independently improve the efficiency of hardware use, are specialized processors high-level design tools. These tools, in contrast to the conventional design on the register transfer level, suppose a description of the algorithm executed by the specialized processor in high-level language, a separate description of the specialized processor interface and its technical characteristics, and, based on this information, further automatic generation of the spectrum of possible VHDL-models of specialized processor, their synthesis, estimation of characteristics and selection of the most effective one according to the set criteria.

That is, the use of high-level design tools allows one to design specialized processors and choose among the designed the best in terms of technical characteristics. As an example, Figure 3 depicts the dependency of the operation frequency to the ALU number with the preserved level of performance of FFT processor, VHDL models of which have been generated automatically with use of the Chameleon C2HDL design tool developed by Intron Ltd [26], [37], [38], [39], to the development and implementation of which the authors have made significant efforts.

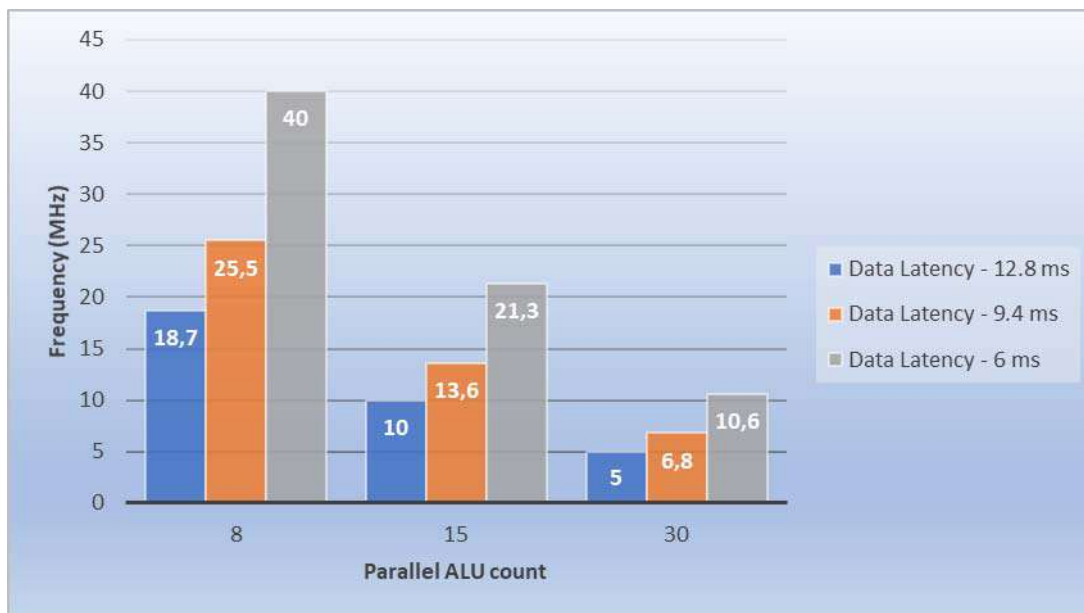


**Figure 3:** Dependency of the FFT processor operation frequency to the ALU number with the preserved level of performance

The Chameleon C2HDL automatic design tool is also the basis of self-configurable computer systems and the method of self-configuration. Note that besides to the Chameleon, there are a number of other high-level design software tools that are designed to generate HDL code of computer devices from the description of the algorithm of their operation in high-level programming languages. These, for example, are: C-to-Verilog [40], System-C [41] and Catapult-C [42] from Calypto Design Systems, SPARK [43] from University of California. These tools can be used with various efficiency for the implementation of self-improvable CS.

VHDL-model of specialized processor can be further implemented in FPGA using CAD tools provided by their manufacturers (Intel Quartus, Xilinx ISE, etc.).

## 7.2. Other Software Tools of SICS, with which Self-Configuration is Realized

To the other software tools of SICS, with the use of which self-configuration is implemented, belong computational load balancing system, compiler for GPP subprograms compilation, and specialized processor logic synthesis tools.

The software means are used for computational load balancing between GPP and RCE. This software has to automatically extract fragments from the input (user) program, execution of which in RCE reduces this program execution time, and divide this program into the GPP subprogram, and the RCE subprogram, formed with the extracted fragments. An example of such system implementation is shown in [44]. The RCE subprogram here is created in the x86 assembly language, and therefore it must be supplemented by some software for the assembly language code translation into a high-level programming language to be used in SICS. The mentioned software is also available on the market, e.g., Relogix Assembler-to-C Translator [45] from MicroAPL.

A compiler is used for compilation of the GPP subprogram from the language that it is represented in into the object codes that can be directly executed by the general-purpose processors.

Logic synthesis tools and FPGA configuring tools are used for the specialized processor HDL-models logic synthesis during the program compilation stage and FPGA configuring during the program loading stage. These tools are available from the FPGA vendors, for example, Vivado Design Suite and ISE from Xilinx; Quartus II from Intel.

## 8. Expected benefits brought by creation of SICS

Computers have become a mandatory component of our life. They are used for solving of science and engineering problems including the design of automobiles, buildings, electronic devices, aircrafts, medications, and robots. Their use makes automobiles safer, more aerodynamic, more comfortable, and more energy-efficient. The more powerful computer is used the more considerable effect of it is gained. For example, exascale computing is used to understand economics, national security, and setting public policy. Billions of processor hours are devoted to understanding and predicting climate change. With faster computer systems, scientists and engineers could simulate critical details - such as clouds in a climate model or mechanics, chemistry, and fluid dynamics in the human body.

Development of the new computer systems architecture has gained significant attention mainly because of its core influence on the computer system characteristics improvement, especially during the past decade when using microelectronic technology improvements for it became very low.

Further practical implementation of self-improvable computer systems based on new computing models will be a fundamental innovation which has no analogues in the field of computer technology. According to preliminary estimates, self-improvable computer systems will be able to increase their technical characteristics by 2-3 orders of magnitude. This assessment is based on the fact that self-improvable CS will be able to adapt their architecture to the executed tasks by its specialization, and according to many years of experience in the production and operation of computer systems, confirmed by data thoroughly covered in [9], the value of improving computer systems performance due to their specialization reaches 2-3 orders of magnitude. A similar indicator is observed in the reduction of power consumption of computer systems. This will dramatically increase the efficiency of their operation.

The following example shows what benefits can be obtained as a result. In 2018, the global electricity consumption was 23 809 TWh out of which 2 476 TWh or 10% were consumed by information and communication technology. The trend suggests that in 2030, the global consumption will be around 40 000 TWh out of which around 8 000 TWh or 20% will be information and communication technology [46], [47]. It is obvious that the reduction of energy consumption by at least 1 order, which is quite possible to achieve by self-improvement of computers, in 2030 will bring, based on this estimate, savings of 7200 TWh! This figure is impressive!

Thus, the creation of such systems will have a positive impact on the further development of society and civilization in general, taking into account the use of computer systems in all spheres of today's human activity.

## 9. Conclusions

1. SICS is a new type of computer systems, and today there is already available all methodological and technological base necessary for their practical implementation, namely - high-performance hardware of universal processors and reconfigurable logic devices, technologies and software of self-configuration of computer systems with reconfigurable logic and artificial intelligence technologies. Therefore, the creation of SICS is likely to be one of the significant stages of further development of the computer technology in the near future.
2. Self-improvement can be carried out at different levels of the CS specification. At the highest - the level of the input program that specify the algorithm of the CS operation, and at the lower - at the level of UPP and RCE subprograms, at the level of software model of a specialized processor, and at the level of FPGA configuration.
3. An important software basis for the mechanism of self-improvement of computer systems are specialized processors high-level automatic design tools, which allow one to automatically obtain a number of software models of a specialized processor from description of the algorithm of its operation in high-level language, and then choose the model with best characteristics.
4. Supplementing the architecture of computer systems with the self-improvement mechanisms should provide an opportunity at the stage of their operation to solely and independently increase the efficiency of hardware use by increasing performance in the classes of algorithms up to 2-3 orders of magnitude, or decreasing power consumption by the same level. Based on the fact that the share of information and communication technology in total global energy consumption today is more than 10%, in the global dimension, as the result of such innovations the energy savings can have a huge impact!
5. SICS are multifaceted, which requires a wide range of research.

## 10. References

- [1] Yao X., Higuchi T., Promises and challenges of Evolvable hardware. In: Higuchi T., Iwata M., Liu W. (eds) *Evolvable Systems: From Biology to Hardware*, ICES 1996, volume 1259 of *Lecture Notes in Computer Science*, Springer, Berlin, Heidelberg, 1997. doi: 10.1007/3-540-63173-9\_38.
- [2] E. Cantu-Paz, A survey of parallel genetic algorithms, *Calculateurs Paralleles, Reseaux et Systems Repartis*, 10(2), 141–171, 1998.
- [3] D. Goldberg, *Genetic Algorithms in search, optimization, and machine learning*, Addison– Wesley, 1989. ISBN:978-0-201-15767-3.
- [4] S. Trimberger, *Field-programmable gate array technology*, Kluwer Academic Publishers, Boston, 1994. doi: 10.1007/978-1-4615-2742-8.
- [5] S. Hauck, A. DeHon, *Reconfigurable Computing: The Theory and Practice of FPGA-Based Computation*, Morgan Kaufmann, 2007. ISBN: 9780123705228.
- [6] C. Bobda, *Introduction to Reconfigurable Computing: Architectures, Algorithms, and Applications*, Springer, 2007. doi: 10.1007/978-1-4020-6100-4.
- [7] T. Todman, G. Constantinides, S. Wilton, O. Mencer, W. Luk and P. Cheung, *Reconfigurable Computing: Architectures, Design Methods, and Applications*, in: *IEE Proceedings on Computers and Digital Techniques*, 152 (2), 2005, pp.193-207. doi: 10.1049/ip-cdt:20045086.
- [8] V. Melnyk. "Principles of Design and Operation of Reconfigurable Computer Systems." *Scientific journal "Bulletin of Khmelnytskyi National University"*, series "Technical sciences", №6, (2012): 212 - 217.
- [9] Shao, Y.S., Reagen, B., Wei, G.Y., and Brooks, D. Aladdin, A pre-RTL, power-performance accelerator simulator enabling large design space exploration of customized architectures, in: *Proceedings of the Int. Symp. Comput. Archit.*, 2014, pp. 97–108. doi:10.1109/ISCA.2014.6853196.
- [10] Upegui A., Sanchez E., *Evolving Hardware by Dynamically Reconfiguring Xilinx FPGAs*, in: Moreno J.M., Madrenas J., Cosp J. (eds), *Evolvable Systems: From Biology to Hardware*, ICES 2005, volume 3637 of *Lecture Notes in Computer Science*, Springer, Berlin, Heidelberg, 2005. doi:10.1007/11549703\_6.
- [11] A. Thompson, An evolved circuit, intrinsic in silicon, entwined with physics, in: *First International Conference on Evolvable Systems (ICES-96)*, *Lecture Notes in Computer Science*, Springer-Verlag, 1996, pp. 390- 405. doi: 10.1007/3-540-63173-9\_61.

- [12] T. Higuchi et al., Evolvable hardware: A first step towards building a Darwin machine, in Proceedings of the 2nd International Conference on Simulated Behaviour, MIT Press, 1993, pp. 417–424.
- [13] J.R. Koza et al., The importance of reuse and development in evolvable hardware, in: J. Lohn et al. (Eds.), Proceedings of the 2003 NASA/DoD Conference on Evolvable Hardware, Chicago, IL, USA, 2003, pp. 33–42. doi: 10.1109/EH.2003.1217640.
- [14] M. Sipper, E. Sanchez, D. Mange, M. Tomassini, A. Perez-Urbe, and A. Stauffer. "A Phylogenetic, Ontogenetic, and Epigenetic View of Bio-Inspired Hardware Systems." IEEE Transactions on Evolutionary Computation, vol. 1, 1997: 83-97.
- [15] J. F. Miller, Digital filter design at gate-level using evolutionary algorithms, in: W. Banzhaf et al. (Eds.), Proceedings of the Genetic and Evolutionary Computation Conference (GECCO'99), Morgan Kaufmann, 1999, pp. 1127–1134.
- [16] M. Murakawa et al. "The grd chip: Genetic reconfiguration of dsps for neural network processing." IEEE Transactions on Computers, 48(6), June 1999: 628–638.
- [17] X. Yao, T. Higuchi. "Promises and challenges of evolvable hardware." IEEE Transactions on Systems, Man, and Cybernetics, Part C, vol. 29, February 1999: 87 - 97.
- [18] M. Murakawa, S. Yoshizawa, I. Kajitani, T. Furuya, M. Iwata, and T. Higuchi, Hardware evolution at function level, in: Proceedings of Parallel Problem Solving from Nature IV (PPSN IV), volume 1141 of Lecture Notes in Computer Science, Springer-Verlag, 1996, pp. 62–71.
- [19] Sakanashi et al., Evolvable hardware chip for high precision printer image compression, in Proceedings of 15th National Conference on Artificial Intelligence (AAAI-98), 1998.
- [20] R. Porter, K. McCabe and N. Bergmann, An applications approach to evolvable hardware, in: Proceedings of the First NASA/DoD Workshop on Evolvable Hardware, Pasadena, CA, USA, 1999, pp. 170-174. doi: 10.1109/EH.1999.785449.
- [21] L. Sekanina, Evolvable Components: From theory to Hardware Implementations, SpringerVerlag, 2004. doi: 10.1007/978-3-642-18609-7.
- [22] C. Hsu, W. Feng and J. S. Archuleta, Towards efficient supercomputing: a quest for the right metric, in: 19th IEEE International Parallel and Distributed Processing Symposium, Denver, CO, USA, 2005, p. 8. doi: 10.1109/IPDPS.2005.440.
- [23] Małgorzata Płaza, Stanisław Deniziak, Mirosław Płaza, Radosław Belka, Paweł Pięta, Analysis of parallel computational models for clustering, in: Proceedings SPIE 10808, Photonics Applications in Astronomy, Communications, Industry, and High-Energy Physics Experiments 2018, 108081O (1 October 2018). doi: 10.1117/12.2500795.
- [24] Leiserson CE, Thompson NC, Emer JS, et al., There's plenty of room at the Top: What will drive computer performance after Moore's law?, Science, New York, N.Y., 2020 Jun, 368(6495). doi: 10.1126/science.aam9744.
- [25] Neil C. Thompson, Svenja Spanuth, The Decline of Computers as a General Purpose Technology, Communications of the ACM, Vol. 64 No. 3, (2021) 64-72. doi: 10.1145/3430936.
- [26] Melnyk, A., Salo, A., Automatic generation of ASICs, in: Proceedings of 2007 NASA/ESA Conference on Adaptive Hardware and Systems, AHS-2007, 2007, pp. 311–317.
- [27] Melnyk, A., Melnyk, V., Self-Configurable FPGA-Based Computer Systems, Advances in Electrical and Computer Engineering, vol. 13, no. 2 (2013) 33-38. doi: 10.4316/AECE.2013.02005.
- [28] V. Melnyk. "Conceptual bases of the self-configurable computer system components design." Lviv Polytechnic National University Journal "Computer Systems and Networks" №773, 2013: 73-81.
- [29] Anatoliy Melnyk, Viktor Melnyk, Heterogeneous computing: from the ASIC-based hardware accelerators to the reconfigurable and self-configurable computer systems, in: Proceedings of the 6-th International Conference on Advanced Computer Systems and Networks, ACSN-2013, Lviv, 2013, pp. 9-12.
- [30] R. Kruse, C. Borgelt, F. Klawonn, C. Moewes, M. Steinbrecher, P. Held, Computational Intelligence: A Methodological Introduction, Springer-Verlag London, 2013. doi: 10.1007/978-1-4471-5013-8.
- [31] L. Rutkowski, Computational Intelligence: Methods and Techniques, Springer, 2008. doi: 10.1007/978-3-540-76288-1.
- [32] J. Leijten, G. Burns, J. Huiskens, E. Waterlander, A. Van Wel, AVISPA: a massively parallel reconfigurable accelerator, in: Proceedings of International Symposium on System-on-Chip (IEEE Cat. No.03EX748), Tampere, Finland, 2003, pp. 165-168. doi: 10.1109/ISSOC.2003.1267747.

- [33] T. El-Ghazawi, Reconfigurable supercomputing, in: Proceedings of the IEEE/ACS International Conference on Pervasive Services, ICPS-2004, Beirut, Lebanon, 2004, p. 163. doi: 10.1109/PERSER.2004.1356790.
- [34] Sandeep Kumar, Christof Paar, Jan Pelzl, Gerd Pfeiffer, Manfred Schimmmler, COPACOBANA: A Cost-Optimized Special-Purpose Hardware for Code-Breaking, in: Proceedings of the 14th Annual IEEE Symposium on Field-Programmable Custom Computing Machines, FCCM'06, 2006, pp. 311-312. <https://doi.org/10.1109/FCCM.2006.34>.
- [35] O. Drozd, K. Zashcholkina, O. Martynyuk, O. Ivanova, J. Drozd, Development of Checkability in FPGA Components of Safety-Related Systems, in: CEUR Workshop Proceedings, vol. 2762, pp. 30-42 (2020). URL: <http://ceur-ws.org/Vol-2762/paper1.pdf>.
- [36] Melnyk A., Golemba V., Bochkaryov A., Multiagent approach to the distributed autonomous explorations, in: Proceedings of NASA/ESA Conference on Adaptive Hardware and Systems, AHS-2007, Edinburgh, UK, 2007, pp. 606-610.
- [37] A. Melnyk, V. Melnyk., Personal Supercomputers: Architecture, Design, Application, Lviv Politechnic National University Publishing, 2013.
- [38] Chameleon – the System-Level Design Solution, 2008. URL: [http://intron-innovations.com/?p=sld\\_chame](http://intron-innovations.com/?p=sld_chame).
- [39] Melnyk, A., Salo, A., Klymenko, V., Tsyhylyk, L. “Chameleon – system for specialized processors high-level synthesis.” Scientific-technical magazine of National Aerospace University "KhAI", Kharkiv, N5 2009: 189-195.
- [40] Giang Nguyen Thi Huong, Seon Wook Kim, GCC2Verilog Compiler Toolset for Complete Translation of C Programming Language into Verilog HDL, ETRI Journal, 2011, 33(5), pp. 731-740. <https://doi.org/10.4218/etrij.11.0110.0654>.
- [41] IEEE Standard for Standard SystemC Language Reference Manual, IEEE Std 1666-201, 9 January 2012, 638 p.
- [42] Catapult High-Level Synthesis and Verification, Siemens EDA Software, 2021. URL: <https://eda.sw.siemens.com/en-US/ic/catapult-high-level-synthesis/>.
- [43] S. Gupta, N.D. Dutt, R.K. Gupta and A. Nicolau, SPARK: a high-level synthesis framework for applying parallelizing compiler transformations, in: Proceedings of the 16th International Conference on VLSI Design, New Delhi, India, 2003, pp. 461-466, doi: 10.1109/ICVD.2003.1183177.
- [44] V. Melnyk, V. Stepanov, Z. Sarairesh. “Computational Load Balancing System Between the Host-Computer and Self-Configurable Accelerator.” Scientific Bulletin of Chernivtsi National University “Computer systems and components”, Chernivtsi: Yuriy Fedkovych Chernivtsi National University, Vol.3, Issue 1, 2012: 6-16.
- [45] Relogix Assembler-to-C translator, 2017. URL: <http://www.microapl.co.uk/asm2c/>.
- [46] Anders S. G. Andrae, Tomas Edler, On Global Electricity Usage of Communication Technology: Trends to 2030, Challenges 2015, 6(1), 117-157. doi:10.3390/challe6010117.
- [47] Between 10 and 20% of electricity consumption from the ICT sector in 2030? 9 Aug 2018. URL: <https://www.enerdata.net/publications/executive-briefing/between-10-and-20-electricity-consumption-ict-sector-2030.html>.

# The contour of causality in control automata of systems

Mykhailo Poliakov<sup>a</sup>, Sergey Subbotin<sup>a</sup> and Oleksii Poliakov<sup>b</sup>

<sup>a</sup> Zaporizhzhia Polytechnic National University, vul. Zhukovskoho, 64, Zaporizhzhia, 69063, Ukraine

<sup>b</sup> National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", prosp. Peremohy, 37, Kyiv, 03056, Ukraine

## Abstract

Known models of control automata of systems describe only the necessary cause-and-effect relationships between sets of an automaton, except for the relationship between sets of outputs and inputs. This complicates the analysis of non-standard situations associated with the diagnosis and development of machines. A model of an automaton in the form of activity and control loops is proposed, which reflects the cause-and-effect relationships in the control automata of the system, including the relations between the elements of the set of outputs and the set of inputs of the automaton. These circuits interact through the machine states.

The classification and models of the machine states are proposed. The structure of the knowledge base along the automaton contours and the predicates types describing causal relationships in the automaton are proposed. An example of constructing the contours of activity, control and Prolog program for diagnosing the technical state of the machine using the knowledge base is given.

## Keywords 1

Control system, control automata, causal relationships, contour of activity and control, knowledge base.

## 1. Introduction and Analysis of Publications

In systems for technical purposes, it is customary to distinguish an object and a control unit [1]. The system behavior is determined by the structure of the functional converter [2], acting as control automata, as well as by the nomenclature and characteristics of the operating automata of the control unit.

The control machine is described by a logical and semantic model. The logical model of control automata is a FSM, which is described by a tuple [1]

$$A = \langle X, Y, S, s_0, \delta, \psi \rangle, \quad (1)$$

where  $X$  is a set of inputs;  $Y$  is a set of outputs;  $S$  is a set of states;  $s_0$  is an initial state;  $\delta, \psi$  is a function of outputs and transitions of the finite automaton, respectively.

The logical model of the control automaton arose as a result of the formalization of the verbal description of the system's behavior in the form of a set of statements [3]. The use of this model in the construction of control systems has shown its limitations, which lies in the fact that it takes into account only the predetermined behavior of the control object and does not take into account the ring causal relationships in the contours "control object – control unit" in case of unplanned changes in the control object, the external environment and the technical condition of the elements of the control unit.

In [4], the influence of the outputs of the control automaton on its inputs is modeled by constructing a control loop, which includes the control automaton and the finite automaton representing the control object. This circuit is used to synthesize a control automaton, which specifies

a certain behavior of the system object. At the same time, the work [4] does not detail the structure of the control loop, does not consider objects with continuous behavior, as well as cause-and-effect relationships that arise in the event of a change in the technical state of system elements.

Further development of logical programming methods and the growth of requirements for adaptation and development of system behavior led to the use of semantic information to describe the behavior of the control automata.

The sematic model of control automata element is a system of tuples [5]

$$SMS = \langle S_a, S_c \rangle; S_i = \langle N_i, CD_i, KB_i \rangle; N_{ij} = \langle N_p, N_{pr}, N_f \rangle; \quad (2)$$

where  $S_a, S_c$  is a set of activity and control contours, respectively;  $S_i$  is a set included in the sets  $S_a$ , or  $S_c$ ;  $N_{ij}$  is a complex of names of the  $j$ -th element in the  $i$ -th contour;  $CD_i$  is direction of causality in the  $i$ -th contour;  $KB_i$  is knowledge base of the  $i$ -th contour;  $N_p, N_{pr}, N_f$  are the set of names of the  $j$ -th element in the  $i$ -th contour in the style of the prehistory, the current choice and the expected events, respectively. The knowledge base consists of rules of the form  $KB_{ij}$ : IF  $CD_i$  THEN  $N_{ik} \leq N_{im}$ , where  $KB_{ij}$  is an element of the knowledge base of the  $j$ -th element in the  $i$ -th contour. The expression  $N_{ik} \leq N_{im}$  indicates the fact that  $N_{im}$  is a consequence of  $N_{ik}$ , where  $N_{ik}, N_{im}$  are knowledge in the form of names for elements  $k$  and  $m$  in the  $i$ -th contour. At the same time, work [5] does not detail the role of states of an automaton in the interaction of activity and control circuits, does not describe a method for constructing cause-and-effect relationships in hierarchical states.

## 2. Research Objectives

The purpose of this article is to provide research on aspects of the behavior of systems associated with a change in the technical state of their elements and parameters of the control object by expanding the knowledge base of the system, detailing the interaction of the activity contours and control of the system through the states of the automaton.

## 3. Research Content

As you know, the "classic automaton" has three sets, which are connected by two functional relationships, as shown in Figure 1a.

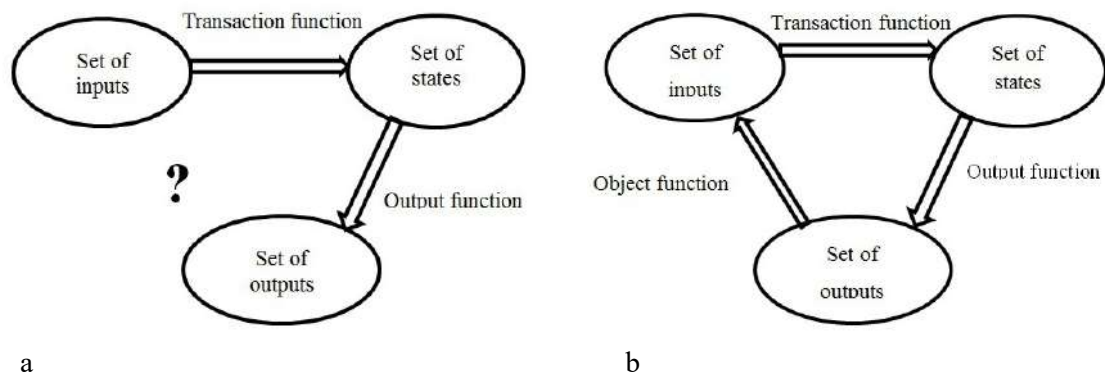


**Figure 1:** Functional relations in the "classic" (a) and the proposed (b) automaton

In this version of the description of the automaton, and its semantic version, there are causal relationships between states and outputs (actions), as well as between current states, inputs and subsequent states. But in this description there are no causal links between outputs (actions) and inputs (events). In addition to the control machine, the control object and the operating machines of the control device take part in the formation of these connections.

The proposed function of the object (Figure 1b) establishes the functional causal dependence of the inputs of the control machine on its outputs. In a more detailed description of the function of the object, the functions of influences and reactions should be taken into account, as shown in Figure 2.

The elements of the "control automaton – object of control" system form the contours of the activity and control of the finite automaton, which interact in the states of the automaton. Let us clarify that the activity contour describes a chain of cause-and-effect relationships ensuring the continuation of the impact on the system object in a given active state. A control loop describes the cause-and-effect relationship that controls the transition of the automaton to another state. Accordingly, the number of control loops passing through this state is equal to the number of transitions from this state to others.

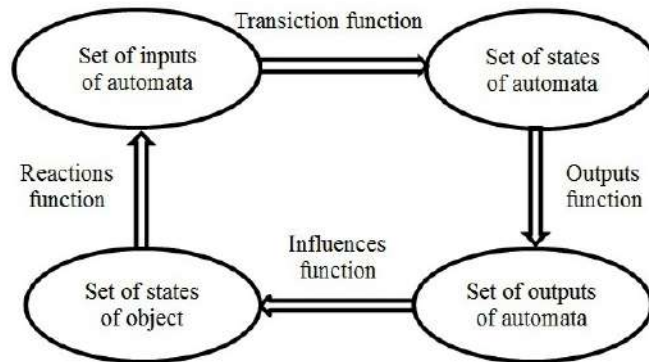


**Figure 2:** Functional relationships in the control machine detailing the functions of the object

To describe the relationship of the state with the rest of the machine, we introduce the following elements: event control inputs (ECI) – inputs and event control outputs (ECO) of control loop events, activity outputs (AO) – and reaction of object inputs (RI) for these actions in the contours of activity [5]. Between these elements of the external interface of the state in the framework of Laplace causality and the direction of causality from entry to exit of the state, the following types of causal relations are possible:  $ECI \leq AO$ , (activity is a consequence of the state transaction);  $RI \leq ECO$  (transaction is a consequence of the reaction).

If in some state of the machine intersect certain circuits of activity and control, then the above causal links provide the interaction of these circuits as shown in Figure 3.

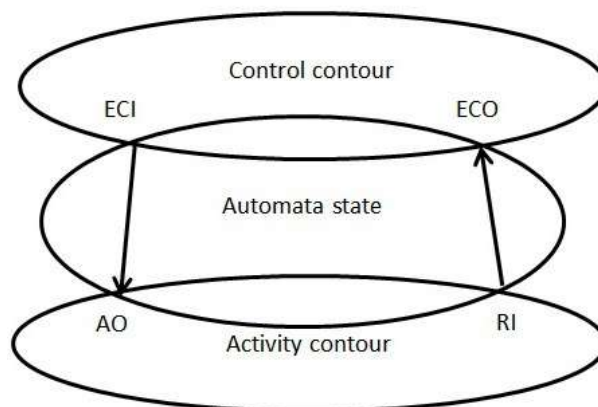


**Figure 3:** The interaction of the contour of activity and control in a certain state of the machine

The mechanism of interaction is as follows:

1. The change in the circuit of activity comes to the input of the reactions of the RI state.
2. Due to internal causation in the state, it is transmitted to the ECO output and is the cause of changes in the control loop.
3. Changes in the control loop lead to activation of the ECI input.



4. Due to the internal causal links in the state, Due to the internal causal links in the state, this activation is manifested at the output AO and initiates changes in the contour of activity.

A special case of the control circuit is the arc of the graph of the automaton, which exits state A and closes in the same state.

With the help of causation and the external arc of the graph of the automaton, which comes out of state A and closes in the same state, you can perform a transition forward (conditional causality) or backward (looping of the site) in the control or activity contour.

To enter the cause-and-effect relationship  $ECI1 \leq ECO2$  (transaction2 is a consequence of transaction1 or the original transaction event is a consequence of the input transaction event) in state A we use the external arc of the automaton graph, which exits AO of state A and closes to RI input in the same state and the internal connection  $RI \leq ECO$ . Similarly, using an external arc between the ECO output and the ECI input, you can create a ratio of  $RI \leq AO$  (activity is a consequence of the reaction).

The proposed classification of states of the control automaton depending on the logic of processing input events is given in Table 1.

**Table 1**

Classification of states of the control machine

| Type of condition | Classification feature    |                           |
|-------------------|---------------------------|---------------------------|
|                   | Incoming event processing | Memory of previous events |
| Switching         | is absent                 | is absent                 |
| Combinational     | is present                | is absent                 |
| Automatic         | does not affect           | is present                |

In the "classic" finite state machine, the output (activity) does not depend on how the activation of the state is performed. That is, if such a state has more than one input event, then it is a combinational state with the processing of input events by logic OR in relation to the activity output of the state. And the automatic state in such automatic machines represents the control subautomata unit of the lower level. This machine is executed if the state in which it is placed is currently active.

The state of the  $S_i$  machine can be described by a tuple

$$S_i = \langle ECI, RI, ECO, AO, \mu_s, t_s \rangle, \quad (3)$$

where  $ECI$  is a set of inputs of control events;  $RI$  is a set of inputs of reactions to activities;  $ECO$  is a set of outputs of control events;  $AO$  is a set of outputs of activity events;  $\mu_s, t_s$  is a function of activity and state transactions, respectively.

The  $\mu_s$  function describes the mapping of the set of  $ECI$  inputs to the set of  $AO$  outputs  $\mu_s: ECI \times AO$ , and the  $t_s$  function describes the mapping of the set of  $RI$  inputs to the set of  $ECO$  outputs  $t_s: RI \times ECO$ .

The type of functions  $\mu_s, t_s$  depends on the type of state. For the switching state it is enough to list the relations  $ECI_i \leq AO_j$  and  $RI_i \leq ECO_j$ , where  $i, j$  are the numbers of elements in the corresponding sets.

To describe the combinational state, you need to specify the output functions of the inputs for all  $i$

$$AO_i = F(ECI_1, \dots, ECI_n); \quad (4)$$

$$ECO_i = F(RI_1, \dots, RI_m), \quad (5)$$

From expressions (4) and (5) it is seen that the switching state is a special case of combinational state, when the function in the right part of these equations takes a constant value equal to the value of a certain input event.

In the automatic state, the functions  $\mu_s, t_s$  describe two automata – activities and transactions, each of which is a tuple similar to the tuple (1)

$$\mu_s = \langle \mathbf{ECI}, \mathbf{AO}, \mathbf{S}_\mu, s_{\mu 0}, \mu_\mu, t_\mu \rangle; \quad (6)$$

$$t_s = \langle \mathbf{RI}, \mathbf{ECO}, \mathbf{S}_t, s_{t 0}, \mu_t, t_t \rangle, \quad (7)$$

where  $\mathbf{ECI}, \mathbf{AO}, \mathbf{RI}, \mathbf{ECO}$  are the sets described in expression (3);  $\mathbf{S}_\mu, \mathbf{S}_t$  is the set of states of automatic machines of activity and transactions, respectively;  $s_{\mu 0}, s_{t 0}$  is an initial state of automatic machines state of activity and transactions, respectively;  $\mu_\mu, \mu_t$  is a function of outputs of automatic machines state of activity and transactions, respectively;  $t_\mu, t_t$  is a function of transactions of automatic machines state of activity and transactions, respectively. In this way we can describe the hierarchy in the semantic and logical model of automata/

We write the functions  $\mu_s, t_s$  in the form of information arrays of size  $m \times n$  ( $MS$ ) and  $l \times k$  ( $TS$ ), where  $m, n, l, k$  are the power of the sets  $\mathbf{ECI}, \mathbf{AO}, \mathbf{RI}, \mathbf{ECO}$ , respectively. Herewith

$$MS[i, j] = \begin{cases} 1, & \text{if } ECI_i \Leftarrow AO_j, \\ 0, & \text{else} \end{cases}, \quad (8)$$

where  $i, j$  are the numbers of the elements in the array,  $i \leq m, j \leq n$ .

$$TS[i, j] = \begin{cases} 1, & \text{if } RI_i \Leftarrow ECO_j, \\ 0, & \text{else} \end{cases}, \quad (9)$$

where  $i, j$  is the row and column number of the element in the array,  $i \leq l, j \leq k$ .

An example of a graphical interpretation of the causal relationship of the simple state of the control machine is shown in Figure 4.

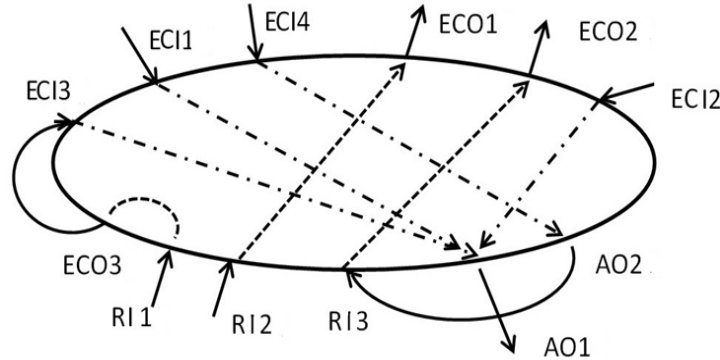


**Figure 4:** Graphical interpretation of the causal relationship of the simple state of the control machine

The set of  $\mathbf{ECI}$  in Figure 4 is represented by inputs  $ECI1, ECI2, ECI3, ECI4$  ( $\mathbf{ECI} = \{ECI1, ECI2, ECI3, ECI4\}$ ). Similarly, we describe the sets  $\mathbf{RI} = \{RI1, RI2, RI3\}$ ,  $\mathbf{ECO} = \{ECO1, ECO2, ECO3\}$  and  $\mathbf{AO} = \{AO1, AO2\}$ . The function of the state activity includes the following causal relationships:  $\mu_s = \{ECI1 \Leftarrow AO1, ECI2 \Leftarrow AO1, ECI3 \Leftarrow AO1, ECI4 \Leftarrow AO2\}$  (event-driven activity). According to (8), this function is also described by an  $MS$  data array of size  $3 \times 3$

$$MS = \begin{vmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{vmatrix}.$$

And the function of state transactions is represented by the following relations  $t_s = \{RI1 \Leftarrow ECO3, RI1 \Leftarrow ECO2, RI3 \Leftarrow ECO2\}$  (event controlled by the sensor) or in the form of a data array  $TS$  of size  $4 \times 2$  which according to (9) has the form

$$TS = \begin{vmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{vmatrix}.$$

In Figure 4, the relationship of activities and transactions is represented by dashed and dotted lines, respectively. Note also that in the presence of an external connection  $ECO3 - ECI3$ , you can implement the ratio  $RI1 \Leftarrow ECO3 \Leftarrow ECI3 \Leftarrow AO1$  (the sensor controls the activity), using external connection  $AO2 - RI3$  ratio  $ECI4 \Leftarrow AO2 \Leftarrow RI3 \Leftarrow ECO2$  can be implemented (event driven event). Figure 4 shows a general case of the state structure. In some cases, the individual elements of the tuple (3) may be empty. For example, in the initial state there may be no  $ECI$ , and in the final –  $ECO$ . The empty set  $AO$  states are possible as well.

An example of a graphical interpretation of the causal relationship of the automatic state of the control machine is shown in Figure 5.

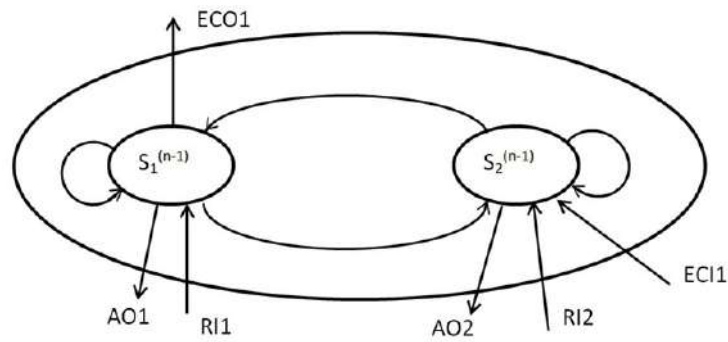


**Figure 5:** Graphical interpretation of the causal relationship of state of the automata

This state has a set of external connections  $ECO1, ECI1, AO1, AO2, RI1, RI2$ . Its internal structure is an automaton of the  $(n-1)$ -th level with two states  $S_1^{(n-1)}$  and  $S_2^{(n-1)}$ . In the general case, the automatic state may represent more than one parallel operating control machines.

The process of building a model of causal relations in the system will be considered on the example of the temperature control system of the object [5, 6]. The control automata of the control unit in such a system performs the functions of a relay controller with thresholds of temperatures  $\theta_{on}$  and  $\theta_{off}$  ( $\theta_{on} < \theta_{off}$ ). The control unit has two states: "HEAT"  $S_1$  and "COOLING"  $S_2$ . State  $S_2$  is initial ( $s_0 = S_2$ ). The operating machines of the system generate the following events:  $E1$  – if the current temperature  $\theta > \theta_{on}$ ;  $\neg E1$  – if  $\theta \leq \theta_{on}$ ;  $E2$  – if  $\theta > \theta_{off}$ ;  $\neg E2$  – if  $\theta \leq \theta_{off}$ .

Figure 6 shows a modified graph of the machine with additional arcs coming from the states. These are the arcs "state - action name - output operating machine - input of the control object" and "output of the control object input operating machine - reaction name - state input". Arcs, transitions from one state to another, are not directly related to the activities and values of the outputs of the control object.

The main functions of the machine shown in Figure 6 are as follows:

- The output function is represented with signals on, off;
- The transition function is represented with events  $E1, E2, \neg E1, \neg E2$ ;
- The function of influences is represented by a controlled heat flow;
- The reaction function is represented by the temperature at the input of the TS sensor.

The states  $S_1$  and  $S_2$  of the automaton in Figure 6 are described by the tuple (3)

$$S_1 = \langle 1ECI, 1RI, 1ECO, 1AO, 1\mu_s, 1t_s \rangle, \quad (10)$$

$$S_2 = \langle 2ECI, 2RI, 2ECO, 2AO, 2\mu_s, 2t_s \rangle, \quad (11)$$

where  $1ECI, 1RI, 1ECO, 1AO$  are a set of interface elements of state  $S_1$ ;  $2ECI, 2RI, 2ECO, 2AO$  are a set of interface elements of state  $S_2$ ;  $l\mu_s, lt_s$  is a function of activity and state transactions of state  $S_1$ , respectively;  $2\mu_s, 2t_s$  is a function of activity and state transactions of state  $S_2$ , respectively.

These sets consist of the following elements:  $1ECI = \langle 1ECI1, 1ECI2 \rangle$ ,  $1RI = \langle 1RI1, 1RI2 \rangle$ ,  $1ECO = \langle 1ECO1, 1ECO2 \rangle$ ,  $1AO = \langle 1AO1 \rangle$ ,  $2ECI = \langle 2ECI1, 2ECI2 \rangle$ ,  $2RI = \langle 2RI1, 2RI2 \rangle$ ,  $2ECO = \langle 2ECO1, 2ECO2 \rangle$ ,  $2AO = \langle 2AO1 \rangle$ .

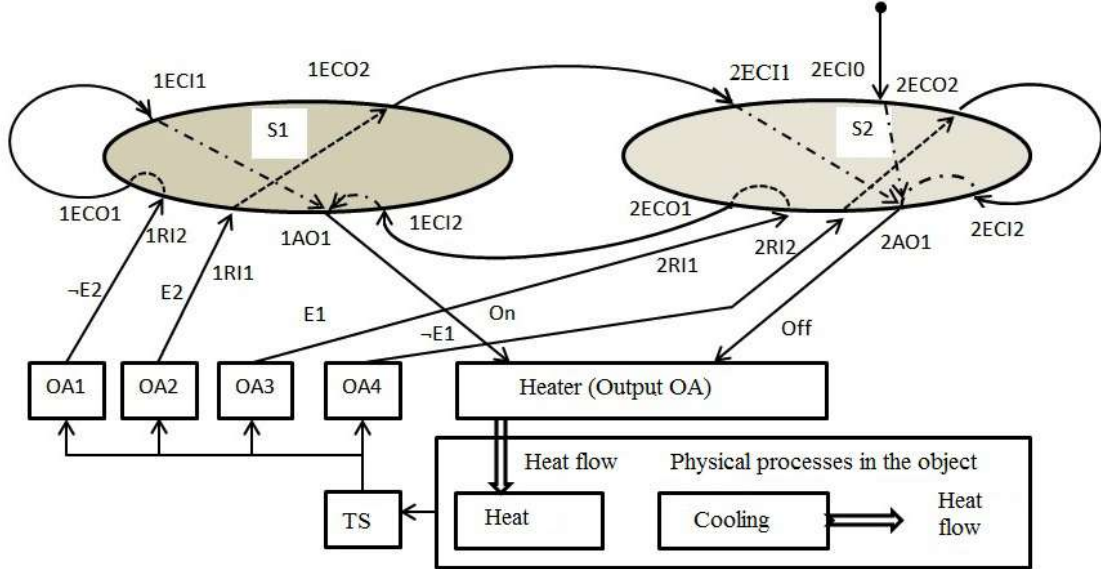


**Figure 6:** Graph of the control machine with contours of control and activity

In terms of activity, states  $S_1$  and  $S_2$  are combinational. We detail expression (4) using the logical OR function

$$1AO1 = 1ECI \vee 1EC2; \quad (12)$$

$$2AO1 = 2ECI1 \vee 2ECI2 \vee 1EC2. \quad (13)$$

In terms of transactions, states  $S_1$  and  $S_2$  are switching. We detail expression (5) using the following connections

$$1RI1 \Leftarrow 1ECO2; 1RI2 \Leftarrow 1ECO1; \quad (14)$$

$$2RI1 \Leftarrow 1ECO1; 2RI2 \Leftarrow 1ECO2. \quad (15)$$

The proposed causal relations of the states of the automaton allow to determine the causal relations in the contours of the automaton. For example, the following relationship results from diagram Figure 6:

1. The output  $1AO1$  is a consequence of any entry into state  $S_1$ .
2. If the active output  $1AO1$ , the input of the heater receives the command "on".
3. If the heater input receives the command "on", the heater generates heat flow to the control object.
4. If there is heat flow to the control object, its temperature rises.
5. If the temperature of the control object rises, sooner or later there will come a time when the output of the operating machine  $OA2$  will form an event  $E2$  which is transmitted to the input  $1RI1$ .
6. The consequence of the event at the input  $1RI1$  is the event of transition to the state  $S_2$ , which is formed at the output  $1ECO2$ .
7. At any input to state  $S_2$  according to (13) the command "Off" is formed at the input of the heater.

In cases where you want to focus on the management structure, it is advisable to show the outline of the generalization (see Figure 7).

The heating continuation circuit in this case contains:  $1AO1$  – Heating circuit –  $1RI2$  –  $1ECO1$  –  $1ECI1$  –  $1AO1$ , and the cooling switch control circuit:  $1AO1$  – Heating circuit –  $1RI1$  –  $1ECO2$  –  $2ECI1$  –  $2AO1$ .

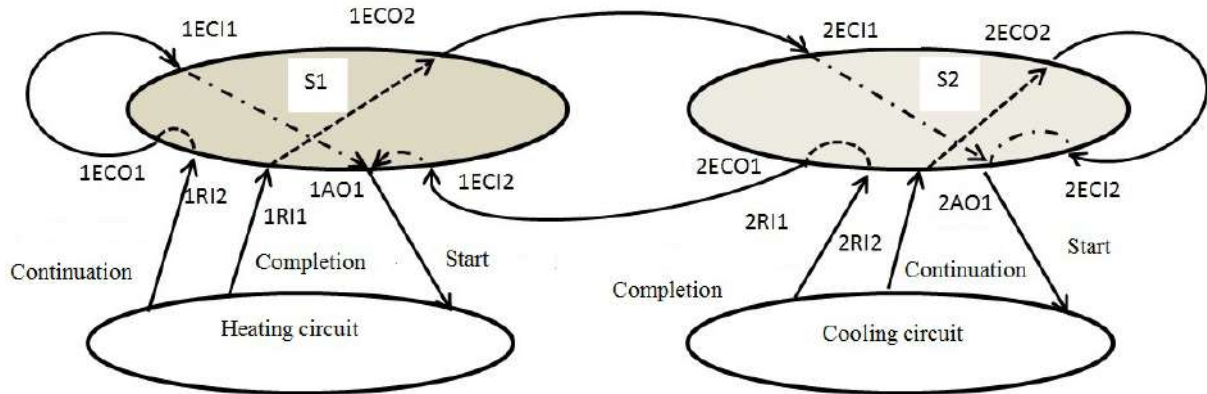


**Figure 7:** Graph of the control machine with generalized contours of activity

The above causal relationships in the contours of activity and control describe the behavior of control of the object of the system. Modern systems, for example, cognitive systems [7], have other behaviors, such as target, scenario, emergency, diagnostic and other behaviors that are associated with the behavior of controlling the object of the system.

As an example, let us consider the use of cause-and-effect relationships for diagnosing the technical state of the elements of the control device and changing the parameters of the object of the temperature control system. An enlarged block diagram of this system is shown in Figure 8.

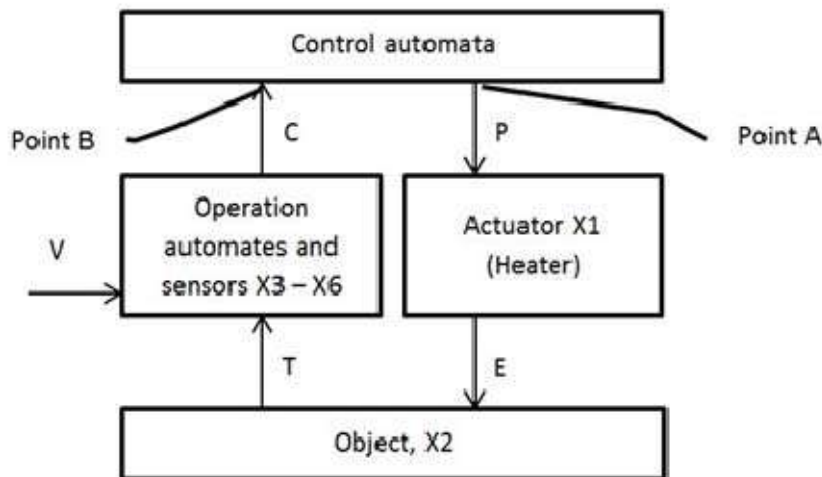


**Figure 8:** Enlarged block diagram of a temperature control system

Figure 8 uses the following notation:

P is a control command (on, off);

E is an energy (h (heat flux), nh (no energy));

T is a temperature change (inc (increasing), const (constant), dec (decreasing));

V is a time from begin of operation (little (), middle (), big ());

C is a heating / cooling conditions (C1 (cond1) – continue heating; C2 (ncond1) – finish heating; C3 (ncond2) – finish cooling; C4 (cond2) – continue cooling;

X is a serviceability (true) / malfunction (false) of the heater (X1), object (X2), operating machines (X3 – X6).

Based on the diagram in Figure 8, we describe the predicates of the input / output ratios with the assessment of the health of the node: heater (P, E, X1); object (E, T, X2); sensor1 (T, V, C1, X3), sensor2 (T, V, C2, X4), sensor3 (T, V, C3, X5), sensor4 (T, V, C4, X6). An example of the knowledge base of the system Figure 8 (facts and rules) is shown in Figure 9 as a fragment of the Prolog program of the SWI version.

The facts of this program describe the functioning of the system with the following simplifications:

- The heater has one level of output power;
- The rate of heating and cooling of the object is constant;
- The time markers of the experiment are chosen in such a way that the “big” time is sufficient to heat the object from  $\theta_{on}$  to  $\theta_{off}$  and its cooling from  $\theta_{off}$  to  $\theta_{on}$ , but the time “little” and “middle” is not enough;
- Events C1 – C4 are considered mutually exclusive.

```
heater(on,h,true).
heater(on,nh,false).
heater(off,nh,true).
heater(off,h,false).
object(h,inc,true).
object(h,const,false).
object(h,dec,false).
object(nh,inc,false).
object(nh,const,true).
object(nh,dec,true).
sensor1(inc,little,cond1,true).
sensor1(inc,middle,cond1,true).
sensor1(inc,big,cond1,false).
sensor1(const,little,ncond1,false).
sensor1(const,middle,ncond1,false).
sensor1(const,big,ncond1,false).
sensor2(inc,little,ncond1,false).
sensor2(inc,middle,ncond1,false).
sensor2(inc,big,ncond1,true).
sensor3(dec,big,ncond2,true).
sensor3(dec,little,ncond2,false).
sensor3(dec,middle,ncond2,false).
sensor4(dec,little,cond2,true).
sensor4(dec,middle,cond2,true).
sensor4(dec,big,cond2,false).
sensor4(const,little,cond2,true).
sensor4(const,middle,cond2,true).
sensor4(const,big,cond2,true).
temperature(P,E,T,V,C1,C2,C3,C4,X1,X2,X3,X4,X5,X6):-
    heater(P,E,X1),
    object(E,T,X2),
    (sensor1(T,V,C1,X3);
    sensor2(T,V,C2,X4);
    sensor3(T,V,C3,X5);
    sensor4(T,V,C4,X6)).
```

**Figure 9:** An example of a system knowledge base in the form of a fragment Prolog program of the SWI version

In the experiments, the head of the rule with different variants of the predicate temperature is selected as the target of the Prolog program. The experimental results are shown in Figure 10.

As you can see from Figure 10, the results of the responses differ, which serves as a diagnostic sign of a malfunction. Opposite answers are marked with arrows, the answer typical for a faulty sensor 1 is underlined.

The program (figure 9) describes a chain of cause-and-effect relationships in the contour from point A to B (Figure 8), namely,  $B \Leftarrow A$ . For example, from the command “to turn on the heater” (point A) it follows that the temperature of the object will rise, but for some time the event “continue heating” (point B) will remain true. On the other hand, from the event “continue heating”, the command “to turn on the heater” follows. Accordingly,  $A \Leftarrow B$ .



|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>In what cases is "all right"?</p> <p>?-<br/>temperature(P,E,T,V,C1,C2,C3,C4,true,true,true,tru<br/>e,true,true).</p> <p>P = on, E = h, T = inc, V = little, C1 = cond1<br/>P = on, E = h, T = inc, V = middle, C1 = cond1<br/>P = on, E = h, T = inc, V = big, C2 = ncond1<br/>P = off, E = nh, T = const, V = little, C4 = cond2<br/>P = off, E = nh, T = const, V = middle, C4 = cond2<br/>P = off, E = nh, T = const, V = big, C4 = cond2<br/>P = off, E = nh, T = dec, V = big, C3 = ncond2<br/>P = off, E = nh, T = dec, V = little, C4 = cond2<br/>P = off, E = nh, T = dec, V = middle, C4 = cond2</p> | <p>In what cases is "Sensor1" faulty?</p> <p>?-<br/>temperature(P,E,T,V,C1,C2,C3,C4,true,true,fals<br/>e,true,true,true).</p> <p><u>P = on, E = h, T = inc, V = big, C1 = cond1</u><br/>P = on, E = h, T = inc, V = big, C2 = ncond1<br/>P = off, E = nh, T = const, V = little, C1 = ncond1<br/>P = off, E = nh, T = const, V = middle, C1 = ncond1<br/>P = off, E = nh, T = const, V = big, C1 = ncond1<br/>P = off, E = nh, T = const, V = little, C4 = cond2<br/>P = off, E = nh, T = const, V = middle, C4 = cond2<br/>P = off, E = nh, T = const, V = big, C4 = cond2<br/>P = off, E = nh, T = dec, V = big, C3 = ncond2<br/>P = off, E = nh, T = dec, V = little, C4 = cond2<br/>P = off, E = nh, T = dec, V = middle, C4 = cond2</p> |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

**Figure 10:** Results of Experiments with the System Knowledge Base

The structure of the causal relationships of the automaton that controls the temperature of the object from its input events through its states and arcs to the outputs is shown in Figure 11a.

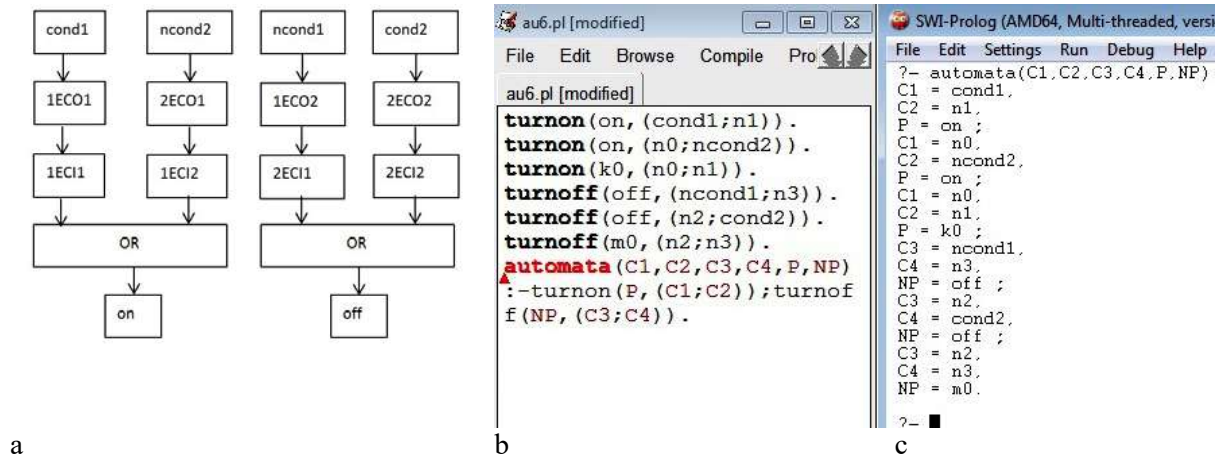


**Figure 11:** The structure of the causal relationships (a) Prolog program (b) and modeling results (c) of the control automata

The knowledge base of this automaton in the form of a Prolog program in the SWI version is shown in Figure 11b. This program contains predicates **turnon**, **turnoff** and **automata** that associate input conditions (cond1, ncond1, cond2, ncond2) or their absence (n0, n1, n2, n3) with the automaton outputs (on, off) or their absence (k0, m0). Figure 11c shows the goal and simulation results, which confirm the dependence of the outputs of the machine on their inputs.

## 4. Conclusion

Traditional design technologies use formal models of control automata obtained as a result of the analysis of the verbal description of the control problem. The use of control automata narrows the knowledge base that is used by the system for control and reduces the quality of control under conditions of unpredictable changes in the system object, the environment and the technical state of the control device elements. The desire to take into account unused knowledge leads to the need to return to the verbal description of the problem, the identification and use of cause-and-effect relationships in the system, which are combined into the contours of activity and control through the object of the system.

The loops of activity in each active state support the conditions for the continuation of activity in this state, and the loops of control contain causal relationships for the completion of activities and transition to other states. The states of automata interact with contours through their interface

elements: inputs and outputs of control events, outputs of activity and inputs of object reactions. A classification of states of an automaton is proposed, based on the logic of transformation of input elements of their interface into outputs, as well as a model of switching, combinational and automatic states.

The introduction of contours into control automata allows building chains of cause-and-effect relationships in the models of the system based on the principles of circular causality. The elements of these chains are described by the predicates "cause", "should" and others with arguments in the form of elements of the structure of the automaton or signals at its points. The arity of predicates can be expanded if we relate knowledge about the connections in the structure of the system with knowledge in the aspect of the research goal: usefulness, serviceability, testability, adaptability, reliability, and others. The knowledge base based on facts and rules with these predicates is explored by means of logical programming.

The proposed models have been tested on an object temperature control system. For this purpose, a knowledge base on the contours of systems has been developed in the form of facts and program rules in the Prolog language of the SWI version, describing the structure of cause-and-effect relationships in the system with the account the technical state of its elements. The program responses to the set goals contain diagnostic signs of malfunctions of system elements, such as the system object and its operating automata. In order to check the completeness of the behavior of the control automaton of the system, a knowledge base has been developed on the structure of causal-logical relationships in this automaton, which is also implemented as a program in the Prolog language of the SWI version. The simulation results confirmed the completeness of the behavior of the control automaton of the temperature control system.

The considered models of the contours of causality of links in the system are supposed to be used when assessing the usefulness of changes in the structures of hierarchical control automata in the processes of adaptation of the system.

## 5. References

- [1] V. M. Glushkov, Synthesis of digital automata (Sintez tsifrovyykh avtomatov), Fizmatizdat, Moscow, 1962.
- [2] S. P. Eugene Xavier, Theory of automata, formal language and computation, New age international Pvt Ltd, New Delhi, 2008.
- [3] V. N. Zakharov, D. A. Pospelov, V. Ye. Khazatskiy, Control systems. The task. Design. Implementation (Sistemy upravleniya. Zadaniye. Proyektirovaniye. Realizatsiya), Ed. 2nd, rev. and add., "Energy", Moscow, 1977.
- [4] V.V. Rudnev, Finite automaton as a control object ( Konechnyy avtomat kak ob"yekt upravleniya), Automation and Telemekhanics, 9, 1978: 126–135.
- [5] M. Poliakov, S. Subbotin, I. Andrias, Control System Control Unit FSM Semantic Models, System technologies. Regional interuniversity compendium of scientific works, Dnipro, 5 (2019): 43–53.
- [6] M. Polyakov, A. Polyakov, State Models for FSM Semantic Models (Modeli sostoyaniy konechnogo avtomata dlya yego semanticheskoy modeli), in S.V. Morshchavka (Ed.), Abstracts reports of the X International scientific-practical conference of Modern problems and achievements in the field of Radio Engineering, Telecommunications and Information Technology, 07–09 October 2020, Zaporizhzhia [Electronic resource] Electron. data. – Zaporizhzhia, NU "Zaporizhzhia Polytechnic", 2020: 86–87.
- [7] D. Vernon Artificial Cognitive Systems: A Primer, Cambridge, Massachusetts: The MIT Press, 2014.



# Forms of Additions to Physical Models of Objects of Study in Remote Laboratories

Karsten Henke<sup>a</sup>, Mykhailo Poliakov<sup>b</sup>, Heinz-Dietrich Wuttke<sup>a</sup>, Johannes Nau<sup>a</sup> and Oleksii Poliakov<sup>c</sup>

<sup>a</sup> Ilmenau University of Technology, Ilmenau, D-98684, Germany

<sup>b</sup> Zaporizhzhia Polytechnic National University, vul. Zhukovskoho, 64, Zaporizhzhia, 69063, Ukraine

<sup>c</sup> National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", prosp. Peremohy, 37, Kyiv, 03056, Ukraine

## Abstract

The trend of expanding the use of remote laboratories in engineering education is noted. This requires an increase in the variety of experiments, including the use of physical models of objects under study as part of a remote laboratory. In addition, an increase in the complexity and cognition of objects of study and the appearance of properties in them that are difficult to simulate on physical models are noted.

It is proposed to supplement the physical models of technical systems with virtual add-ons. The forms of these additions and the interfaces of the remote laboratory system in which they appear are described.

The basic forms of reality and functionality of physical models, as well as forms of augmented reality, functionality and cognition of these models have been determined. Examples of additions to the physical models of the GOLDi remote laboratory are given.

It is proposed to change the interface of video surveillance of the physical model during the experiment. It is proposed to replace the video of the behavior of a physical model in real time with a demonstration of typical fragments that were filmed earlier, including on a real object. Methods of synchronizing the time of demonstration of these fragments are described. Such a replacement allows you to preserve and even increase the feeling of the reality of the experiment even in the absence of physical models of the object of study.

## Keywords 1

Remote laboratories, cognitive systems, augmented reality, augmented functionality, augmented cognition

## 1. Introduction

A relevant argument in favor of distance learning is the continued isolation of university students in the context of the coronavirus pandemic in 2020 and 2021. The relevance of the transition from training in the design of technical systems to the design of cognitive systems (CS) is due to forecasts of the deployment of a cognitive revolution in the world [1]. The hierarchical nesting of physical models improving of cognitive objects of study in remote laboratories (RL) is illustrated in Figure 1.

A cognitive system is defined as a system that uses knowledge to achieve the goals of its functioning, has cognitive abilities, such as perception, planning, training, reasoning [2]. The cognitive system includes the subsystems of cognition, activity, cognitiveness and the knowledge base

---

CMIS-2021: The Fourth International Workshop on Computer Modeling and Intelligent Systems, April 27, 2021, Zaporizhzhia, Ukraine  
EMAIL: karsten.henke@tu-ilmenau.de (K. Henke); dieter.wuttke@tu-ilmenau.de (H.-D. Wuttke); johannes.nau@tu-ilmenau.de (J. Nau);  
polyakov@zntu.edu.ua (M. Poliakov); poliakovalyosha@gmail.com (O. Poliakov);  
ORCID: 0000-0001-5424-9053 (K. Henke); 0000-0002-7772-3122 (M. Poliakov); 0000-0001-8030-647X (H.-D. Wuttke), 0000-0001-7538-2283 (J. Nau); 0000-0002-9355-7056 (O. Poliakov)



© 2020 Copyright for this paper by its authors.  
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).  
CEUR Workshop Proceedings (CEUR-WS.org)

in various forms - from data to wisdom [3]. This system has a hierarchical tree structure with levels of target, conceptual, behavioral, and computational and other controls [4].

Important elements of distance learning environments for learning system design are remote laboratories [5], which provide a number of services to a remote student. First, this is a service for creating a project of behavior of the selected physical model object of study. Secondly, it is a service for remote monitoring of an experiment with an object of study, during which verification and validation of a student's project is performed. In the considered case, a video and an animated image of the object of study in the user interface are used. Therefore, the RL “Grid of Online Lab Devices Ilmenau (GOLDi)” [6] has a set of physical models of an elevator, a crane, an automatic line, a warehouse (manufacturer - Staudinger GmbH Automatisierungstechnik [7]) and others. Their interface with the control board has an electromechanical level. From the standpoint of remote laboratory computing nodes, these models are represented by a set of sequences of sensor and actuator values. That is, these models are not cognitive systems.

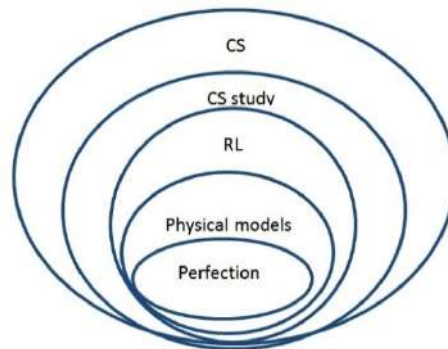


**Figure 1:** Nesting issues of improving physical models of cognitive objects of study

The pedagogical problem when using physical models in the process of teaching systems design lies in a very limited number of technological operations linking the state of the model with its sensors and actuators. For example, for the remote laboratory GOLDi warehouse model, the typical operations are “put a box in a cell with coordinates Y, Z” and “take a box from a cell with coordinates Y, Z”. Many operations performed by the warehouse during the experiment differ only in the coordinates of the cells, which do not provide the required number of design tasks for students.

Another problem of remote laboratories is the small number of simultaneous experiments with physical models of research objects. Virtual experiments solve this problem, but, with such an experiment, the quality of the animated image of the object is significantly reduced.

The authors see the way out of this situation in the use of augmented methods.

## 2. Research Objectives

The purpose of this article is to expand the functionality of physical models of objects of study, increase the variety of experiments with them, increase the level of systems under study to cognitive ones by using various augments, starting with augmented reality and improve interface in experiments with virtual models.

## 3. Analysis of Publications

The obvious advantage of studying objects using their physical models is the reduction in equipment costs. You have to pay for it with reduced functionality. The desire to maintain both low cost and high functionality led to the creation of cyber physical systems (CPS) [8]. In CPS, an extension of functionality is performed by the software of a node with a computing resource that is part of the system. For example, it is possible to expanding the capabilities of a physical model by improving the human perception interface of simulation results. This method, which was called augmented reality (AR), was devoted to approximately 20% of the reports at the REV2020 conference, for example [9]. Varieties of this method differ in the ratio of reality and virtuality of

objects introduced into the field of human perception [10]. At the same time, the essence of AR lies in the fact that under the influence of virtual facts in the field of perception, the flow of influences on the object of study by a person changes. So, the driver of the car, perceiving augmented reality, begins to go around virtual obstacles.

A variation of the AR method can be considered to virtually introduce an interface with the sensors of the physical model. In [11], the sensors of a physical model are affected by the virtual flow of elevator calls from floors from the computer side of the system. Thus, the AR method does not expand the functionality of the physical model, but introduces comfortable options for using the model in experiments.

In [12], a method of additional functionality (AF) was proposed, according to which, using software, the functions of the object of study are realized that are absent in its physical model and even in the object of study. An example is the added functionality of a traffic light in terms of continuous monitoring of the performance of its lamps / indicators. The implementation of this AF is associated with the introduction of additional sensors, events and states of the state machine of traffic light behavior.

The AF functions include the following functions:

- Account the influence of non-deterministic changes in the technical state of its elements [13] - the elevator cable burst, the position sensor does not work, there is no contact in the drive circuit, the power supply is malfunctioning
- Interaction of the manipulation object with the system - the storage box tells the system its number, type, weight or coordinate of the cell in which it must be placed
- Interaction of the manipulation object with each other and with the system - the exchange of information about the intentions of the movement of cars at the intersection and the actions of the traffic light can change these intentions
- RL interaction with a student project - online monitoring, diagnostics, correction / blocking and analysis of student project quality [13].

These changes in functionality are characteristic of behavior that is based on additional information, but they are not related to the intellectual properties of the system.

## 4. Research Content

Additions of a cognitive system with a physical model are complex. They include both well-known types of augmented (AR, AF) and specific augmented, which we will call cognitive (AC) entities. The term “augmented” entity (reality, functionality, cognitivity) implies the definition of a basic entity with respect to which additions are made. The contents of the basic and added entities will be examined using physical models of the GOLDi laboratory warehouse and passenger elevator, which are shown in Figure 2.

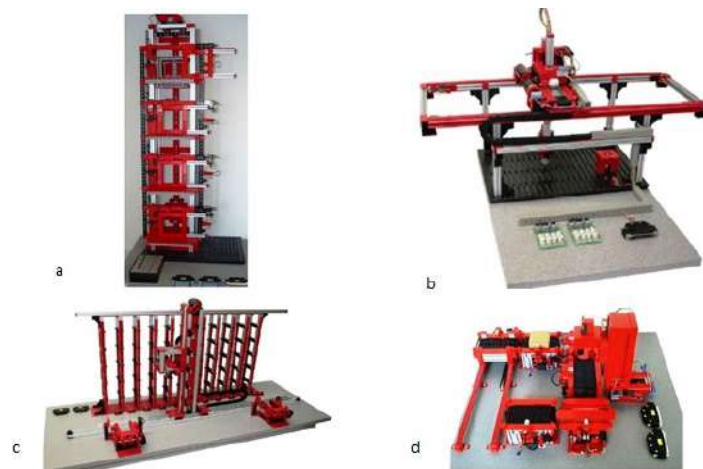


**Figure 2:** Appearance of physical models of the Elevator (a), 3-Axis-Portal (b), Warehouse (c) and Production Cell (d) of the GOLDi laboratory [6]

The basic and augmented entities of these models are manifested in the structure of the model interfaces, which includes the following and interfaces:

- Parallel interface of electromechanical sensors and actuators with a control board (SAPI)
- WEB-camera interface (video stream and model view control) (WCI)
- Serial interface for data transfer between control boards, laboratory server (SASI)
- The user interface of the object of study (UI)
- User perception interface (UPI).

The basic reality forms of the physical models of the GOLDi laboratory storage and passenger elevator shows Table 1.

**Table 1**

Forms of the basic reality of physical models

| Interface type | Forms of the basic reality of Warehouse, Elevator                                    |
|----------------|--------------------------------------------------------------------------------------|
| SAPI           | Parallel binary code of current values of sensors and actuators                      |
| WCI            | Video/ livestream of a physical model during an experiment                           |
| SASI           | Transfer protocol for the current values of sensors and actuators                    |
| MAI            | Visual and virtual model of static and moving visible elements of the physical model |
| USI            | Binary status code for sensors controlled by a remote user                           |

By basic functionality, we mean the minimum set of functions performed by a normal physical model in normal mode with basic reality. The warehouse model has basic functionality, which includes the following basic operations:

- Load a box (cargo) from a warehouse cell or input drive into a movable module (cargo manipulator)
- Move the movable module to a given warehouse cell
- Unload the box (cargo) from the moving module to the warehouse cell or to the output drive
- Waiting for a box operation.

The basic functionality of the elevator model includes the following basic operations:

- Waiting to elevator call from the floors
- Close the door, move to the call floor and open the door
- Close the door, move to the destination floor and open the door
- Slow down the speed of movement when approaching the destination floor
- Indication the current position of the elevator.

The augmented realities of the warehouse model include the following:

1. AR flow of incoming and outgoing boxes, which is entered in the USI inter-face and, as a result, manifests itself in the WCI and MAI interfaces. The essence of this AR is to generate a stochastic sequence of incoming boxes in the input drive with a random set of parameters, such as number, weight, color, etc. As well, as generate random times and requirements for the composition of the batch of boxes that need to be loaded in the output buffer.
2. Failure of the elements of the electro mechanics of the warehouse AR, which is introduced into the SAPI interface. The essence of this AR is to generate the stochastic flow of sensor and actuator failures, which should be taken into account in the student's project.

Consider the possible augmented realities of the elevator model:

1. Virtually elevator calls, which is introduced in the USI interface and, as a result, manifests itself in the WCI and MAI interfaces. The essence of this AR is to generate a stochastic sequence of elevator calls from different floors at the same time, as well as generate destination floors for these calls. At the same time, in the user screens that are controlled by the WCI and MAI interfaces, the queue of users waiting for the arrival of the elevator on call is visualized.
2. AR failures of the elements of the electro mechanics of the elevator, which is introduced into the SAPI interface. The essence of this AR is to generate a stochastic flow of sensor and actuator failures, which should be taken into account in the student's project.

The presence of AR allows the study of more complex systems with physical models of control objects as subsystems that have augmented functionality. Therefore, based on the elevator model and AR, AF, you can build an elevator dispatch system. In addition, the sorting and optimal cargo placement systems can be built based on the warehouse models with AR and AF. Consider the AFs that are needed in an elevator dispatch system:

1. The analysis of the load of the elevator and the optimization of replanting passengers along the route of the elevator.
2. Calculation of the parameters for the elevator latency and user query time
3. Analysis of the parameters of the flow of requests for calling and moving the elevator.

Systems for sorting and optimal placement of goods based on the warehouse can have the following AF:

1. Obtaining information about filling the cells of the warehouse with boxes and information about the numbers and parameters of the boxes stored in the store, the timing of sending batches of boxes and the composition of the boxes in the lot.
2. Sorting boxes in the warehouse according to a specified criterion, for example, by numbers, arrival time, parameters of boxes and their storage areas – rows and floors of the warehouse.
3. Sorting boxes in the warehouse if there are restrictions on the performed rearrangements of the boxes, for example, as in the famous game “fifteen” [14].
4. Analysis of the effectiveness of search algorithms for sorting boxes.
5. Optimization of the placement of boxes according to various criteria, for example, the minimum length of movements of the movable module, the timing of sending boxes, increasing the speed of receiving boxes through the input drive.

At the same time, we note that the tasks of developing tools for analyzing flows (boxes, calls) and choosing effective algorithms are research tasks for students. Moreover, RL turns from a design laboratory into a research tool. In [15], an example of an experiment of sorting boxes in the warehouse according to the principle of the game “fifteen” is given. The initial placement of the boxes is randomly performed by the warehouse animation program, which analyzes the possible existence of a solution. A greedy algorithm on graph search is used to solve the problem. Animation screens of the initial and final placement of boxes in the warehouse are shown in Figure 3.

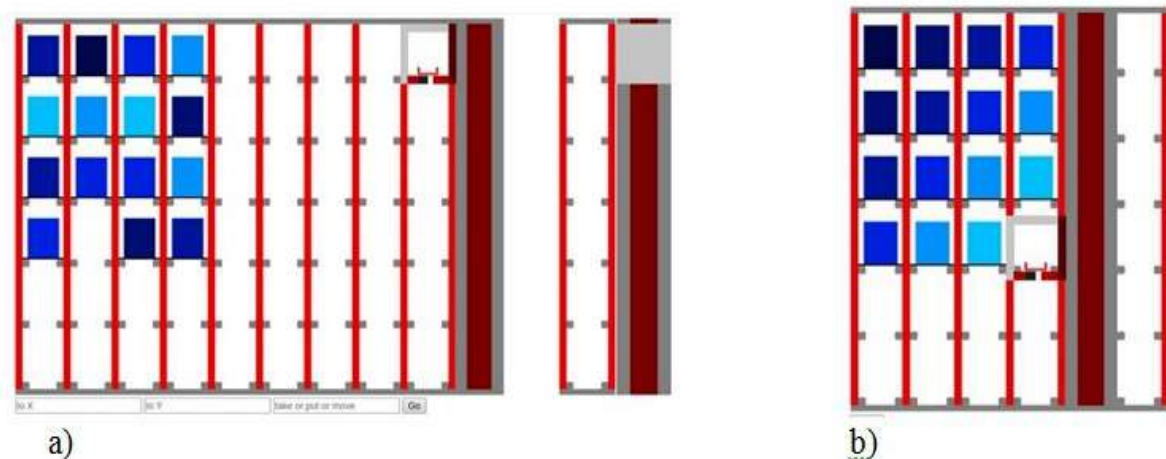


**Figure 3:** Screens of initial (a) and final (b) placement of boxes in the Warehouse

The basic and augmented realities of perception and functionality of behavior discussed above and the related information base are the basis for constructing cognitive systems. When constructing these systems, additional means are used to convert higher forms of knowledge, implement target and scenario management, cognitive perception, planning, training and reasoning. The complex of these tools will be called augmented cognition. Consider this complex on systems that include the physical model of the warehouse and elevator described above.

Cognitive perception of the results of studying an object is possible, for example, because of the transformation of secondary information about added reality. For example, let the signal “Input drive is busy” be received from the receiving device of the warehouse, which is interpreted by the system as

a signal to place a new box in an empty storage cell. The system registers the time of receipt of the box and determines its number. Then it determines the position of the box in the batch (first, intermediate, last), the number and the time interval for receipt of boxes and other parameters are calculated. The statistics of the flow of boxes at the entrance to the storage train the neural network, which forms a model of the flow of boxes. This model is used to predict future receipts of boxes and plan their placement. Options for predicted situations and placements are analyzed by a block of reasoning, taking into account the possible goals of the system:

1. Minimize the movement of the movable module during the operations of receiving, sorting and sending a batch of boxes ("Save resource").
2. Minimize the time for receiving a batch of boxes ("Quick reception").
3. Minimize the time it takes to send a batch of boxes ("Quick shipment").
4. Sort batches of boxes ("Sort").

The graphic form of reasoning, which justifies the change in the goals of the functioning of the system in relation to the situation with filling the warehouse and flows of incoming, sent and sorted boxes, can be represented as FSM graph.

A graph of a state machine of scenarios is a graphical interpretation of reasoning when choosing a scenario change in the course of implementing the chosen goal. Examples of goal and scenario machines graphs are shown in Figure 4.

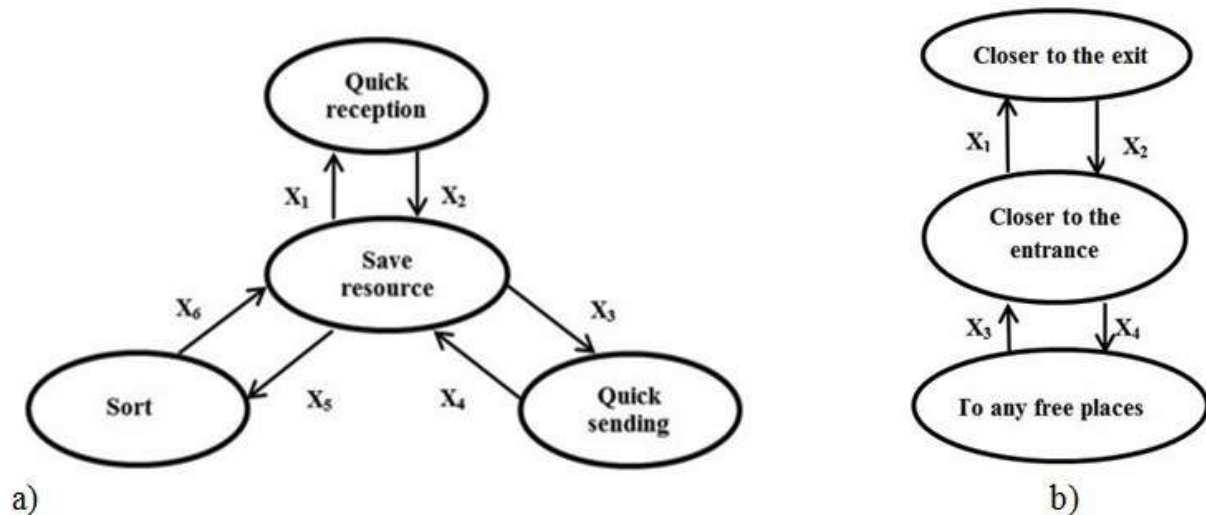


**Figure 4:** Examples of graphs of goal (a) machines and scenarios (b)

The purpose of the "Save resource" is basic: to place the boxes so that the sum of their movements in the mobile unit during reception and sending is minimal. The transition to other goals is due to special conditions:  $X_1$  - the receipt of boxes in the warehouse with a minor interval;  $X_3$  - requirement for quick shipment of a batch of boxes;  $X_5$  - the priority of storing boxes in sorted form. Under condition  $X_1$ , the boxes are placed in the closest available storage cells closest to the input drive. To fulfill condition  $X_3$ , the sent boxes move closer to the output drive. Under  $X_5$  priority conditions, the system strives to sort the boxes according to a given criterion. Conditions  $X_2$ ,  $X_4$ ,  $X_6$  return to the goal "Save resource" are opposite in meaning to the conditions  $X_1$ ,  $X_3$ ,  $X_5$ . As an external constraint, conditions  $X_1$ ,  $X_3$ , and  $X_5$  can only occur exclusively, otherwise, there would be a contradiction.

The selection of any goal activates the corresponding output of the target automaton and the scenario selection automaton of its achievement. Figure 4b shows possible scenarios for achieving the "Quick reception" goal. The priority scenario is "Closer to the exit", in which the system tries to place incoming mailboxes closer to the output drive. However, if the rate of receipt of the boxes is high (event  $X_2$ ), the system begins to place the boxes near the input drive (the "Closer to the entrance" scenario), with the hope of rearranging them after the completion of the fast reception of a batch of boxes. Finally, when free cells end (event  $X_4$ ), it proceeds to the "To any free places" scenario.



The examples of finite state machines of goals and scenarios considered in Figure 4 require detailing for practical use. For example, taking into account the priorities of events X1 – X6 (Figure 4a) when they are simultaneously received at the input of the machine.

We estimate the technical efficiency of the proposed methods by increasing the number of experiments on the existing laboratory equipment. As an indicator of technical efficiency, we will choose the relative increase in the number of experiments

$$K = N/N_b,$$

where  $N$ ,  $N_b$  are the achieved and baseline number of experiments, respectively.

Theoretically, the  $K$  index can take values in the range from 0 to  $\infty$ . The efficiency of the method increases with increasing  $K$ . The criterion of technical efficiency:  $K > 1$ . The economic efficiency of the proposed methods consists in reducing the cost of purchasing and operating a remote laboratory per one type of experiment. These are determined by the formulas

$$\begin{aligned} C_{1b} &= C_b/N_b, \\ C_1 &= C/N, \end{aligned}$$

where  $C_{1b}$ ,  $C_1$  are the cost per one type of experiment for the basic and proposed options, respectively;  $C_b$ ,  $C$  are the costs of purchasing and operating a remote laboratory according to the basic and proposed options, respectively.

As an indicator of economic efficiency, we will choose the relative decrease in the cost of one experiment

$$\Delta C = C_1/C_{1b} = CN_b/C_bN = (C/C_b)/K.$$

If we neglect the additional costs for the software implementation of additions to physical models (that is,  $C = C_b$ ), then  $\Delta C = 1/K$

Theoretically, the  $\Delta C$  indicator can take values in the range from 0 to  $\infty$ . The economic efficiency of the method increases with decreasing  $\Delta C$ . The criterion of the technical efficiency of the proposed methods:  $\Delta C < 1$ .

When implementing the proposed method of additions in the GOLDi laboratory, it is expected that the variety of experiments will increase by at least two times. Accordingly, the efficiency of experiments on the proposed indicators will increase at least twice.

## 5. Additions to the WCI interface

The WCI interface is the most important element of the remote student's perception to get a realistic impression of his remote experiment. Previously it is discussed inserting additional AR-elements into this interface, for example, images of the queue of passengers waiting for the arrival of the elevator on the floor, text comments to the image and the results of calculations. These images should be synchronized with the video stream, taking into account the time delay in the transmission of a video stream over the Internet if we use the GOLDi-lab in its real experimentation mode. Alternatively to this approach we want to propose an additional virtual experimentation mode to the variants, published in [16]. In that virtual mode we use pre-recorded video fragments of typical actions / movements / operations of a physical model instead of the livestream from a web-cam. The display of these fragments should be synchronized with the operation of the executive mechanisms of the experiment object and the triggering of its sensors.

Let us consider the essence of the method using the example of experiments with the model of the elevator of the 4th floor of the GOLDi-lab. In the work of this model, the following typical fragments of movement can be distinguished, which are given in Table 2. These are only the video fragments for a correct control algorithm. If the algorithm, programmed during the experiment is wrong, many alternative video frames are needed.

The duration of each fragment, as a rule, is no more than 10 seconds. Therefore, for a one-time download and storage of video of all fragments (in total, this will be 2-3 minutes of video) on the computer of a remote student, high network exchange speeds and large amounts of memory are not required.

**Table 2**

Typical fragments of movement of the physical model 4-floor elevator

| Designation     | Description                                                                                                                      |
|-----------------|----------------------------------------------------------------------------------------------------------------------------------|
| Fr1             | Cab doors are open                                                                                                               |
| Fr2/Fr3/Fr4/Fr5 | Cab doors close on the first / second / third / fourth floor                                                                     |
| Fr6/Fr7/Fr8/Fr9 | Cab doors open on the first / second / third / fourth floor                                                                      |
| Fr10/Fr11/Fr12  | The lift moves up from the first floor to the second / from the second floor to the third / from the third floor to the fourth   |
| Fr13/Fr14/Fr15  | The lift moves down from the fourth floor to the third / from the third floor to the second / from the second floor to the first |

In the process of synchronizing the video display and the operation of the physical model of the elevator of the 4th floor, the sensors and actuators shown in Table 3 [6] are used.

**Table 3**

Sensors and actuators of the physical model of the 4-floor-elevator to synchronize the video display

| Variable       | Name                      |
|----------------|---------------------------|
| x0/x1/x2/x26   | Elevator on floor 1/2/3/4 |
| x7/x9/x11/x29  | Door open Floor 1/2/3/4   |
| x8/x10/x12/x30 | Door closed Floor 1/2/3/4 |
| y0/y1          | Drive upwards/downwards   |
| y3/y5/y7/v24   | Door floor 1/2/3/4 open   |
| y4/y6/y8/y25   | Door floor 1/2/3/4 close  |

Examples of conditions for displaying video fragments are given in Table 4.

The experiment of moving the elevator from the first floor to the third is accompanied by showing a chain of video fragments Fr1, Fr2, Fr10, Fr11, Fr8 and Fr1.

The fact that the time for creating video clips is practically unlimited makes it possible to use a number of techniques that are used in cinematography to improve the quality of production. This is a preliminary development of the script, improved video cameras and stage lighting, duplication of video fragments, changing shooting angles, alternating shots, zooming in / out of the shooting point, various frame modes, soundtrack, titles and much more.

**Table 4**

Examples of conditions for displaying video fragments

| Fragment<br>video | Demonstration conditions (sensors and actuators state) |            |                |
|-------------------|--------------------------------------------------------|------------|----------------|
|                   | Start                                                  | During     | Stop           |
| Fr1               | Before the experiment                                  |            | Run experiment |
| Fr2               | Run experiment & x0 & y4                               | y4         | x8             |
| Fr10              | x8 & x0 & y0                                           | y0         | x1             |
| Fr11              | x10 & x1 & y0                                          | y0         | x2             |
| Fr8               | x12 & x2 & !y0                                         | y7         | x11            |
| Fr1               | x11                                                    | No new run | Run experiment |

Further development of this method is the use of video fragments of the work of a real object during its operation instead of a physical model or their combination. In this case, the goals of increasing the realism of the video are pursued, showing the details of the object that are absent in its model, demonstrating the video in the absence of a physical model, i.e. experiment in virtual mode and others.

The composition of video fragments of the movement of a physical model should visually reflect the variety of its behaviors during the experiment. The variety of behaviors in the production cell model (Figure 2d) includes:



1. The main option for moving a part along a circular route using three continuously moving belt conveyors, two rotary tables with conveyor chains, a rail carriage with a conveyor belt and a vertical milling machine.
2. A variation of the main variant of movement, when the belt conveyor stops after the part passes through it and turns on again before the part approaches on the next cycle.

In this case, a saving of the resource of conveyors is achieved, which gives the task a practical meaning and increases the variety of educational experiments with this model.

This variety is also achieved by changing the direction of movement of parts along conveyors. By recording video clips depicting the movement of a part along one or two conjugate nodes of a production cell (e.g., conveyor belt rotary table), video support of experiments with virtual models of objects of study showing elements of real physical models is possible. Figure 5 shows a structural diagram of a combined model "warehouse-production cell".

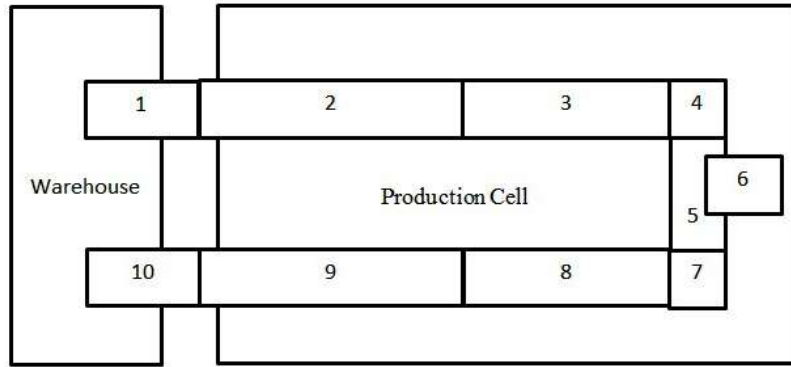


**Figure 5:** Structural diagram of the integrated model: 1, 10 are in-/ output buffers; 2, 9 are rail carriages; 3, 5, 8 are conveyor belts; 4, 7 are rotary tables; 6 is a vertical milling machine

In this model, the circular path of the production cell is closed through the bins of the warehouse model (Figure 2c). Video fragments show the work of the warehouse to unload the next box into the output hopper. Then the next video fragment shows the movement of the box using a rail carriage, then a fragment of the transition of the box to the belt conveyor, and so on.

For some objects, the set of video fragments that demonstrate all its possible movements can be quite large. The number of options for moving the box using the mobile module of the physical model of the warehouse of the GOLDi laboratory is equal to the number of 2-permutations of  $K$  elements, where  $K$  is the number of cells in the warehouse. For the physical model warehouse of the laboratory GOLDi  $K = 50$  and the number of permutations, respectively, of video fragments is  $K(K - 1) = 2450$ . Here we must add the same number of fragments of moving the mobile module without a box. The total duration of these video fragments with an average duration of one fragment of 5 seconds will be about 7 hours. Of course, the creation of such a number of video clips requires a significant amount of time, and during storage, they take up gigabytes of memory.

To assess the effectiveness of the proposed method for improving the WCI interface, we will assume that the quality of perception of the experiment is proportional to the functionality of the experiment object displayed on the computer screen of the remote student. Let's take the  $Q_{RO}$  functionality of a real object as a unit of quality. Then the video image quality of the experiment with the physical  $Q_{PM}$  model will be less than one, and the image quality of the  $Q_{AM}$  animation model will be inferior to the  $Q_{PM}$  quality ( $Q_{AM} < Q_{PM}$ ). It is expected that as a result of inserting previously recorded video fragments of the behavior of an object or its physical model into the interface of a virtual experiment, the quality of perception will increase to the level of  $Q_{PM}$ .

Moreover, this increase in quality was achieved without the cost of purchasing and operating additional physical models. The cost savings for the proposed method for improving the WCI interface depends on the required number of concurrent experiments. If it is required to simultaneously conduct  $G$  experiments with  $Q_{PM}$  quality, then the cost of physical models according to the basic method will be  $C_{PM}G$ , where  $C_{PM}$  is the cost of purchasing and operating  $G$  physical

models. These costs can be avoided by applying the proposed method for improving the WCI interface.

It should be noted that not for all objects it is possible to obtain a set of video fragments that comprehensively describe its visible movements. This applies to physical models, the position of which is described in continuous coordinates: robots, cars, quadcopters, and portal cranes. Accordingly, the number of options for possible movements can be practically infinite. But even in these cases, video clips embedded in the animation screen can improve the quality of the end-user interface.

## 6. Conclusion

The method of additions allows you to expand the range of creative / design tasks for students without additional costs for the acquisition of physical models of the studied systems. The principles of the formation of additions to systems with physical models can be useful in the formation of educational and pedagogical problems for students.

The add-ons presented in the remote lab are complex and include augmented reality, functionality, and cognition that interact with the underlying properties of the object.

Augmented reality is implemented both in the user perception interface and in the interface of hybrid models of the research object in order to visualize the flow of virtual events associated with the research or control object.

Extended functionality concerns the behavior and activity of the system under study, taking into account changes in the technical state of its elements, behavior in emergency situations, quality control of a student project, building higher-level systems in which the original physical model is one of the elements. For example, a dispatching system for two elevators, a system for sorting goods in a warehouse.

Enhanced cognition improves or creates an object's cognitive abilities in terms of perception, learning, planning, and reasoning in a targeted management process.

The paper presents an example of a virtual experiment with a model of a storage warehouse of the GOLDi laboratory, in which the added functionality implements an algorithm for sorting boxes according to the principle of the game "fifteen".

The expected efficiency of experiments using additions is at least twice the baseline in terms of the proposed indicators.

A method is proposed for obtaining a video image of an object in the process of a virtual experiment, based on the use of typical fragments of the object's motion.

It is expected that as a result of inserting previously recorded video fragments of the behavior of an object or its physical model into the interface of a virtual experiment, the quality of perception of the experiment by a distant student will increase to the level of an experiment with a physical model. At the same time, additional costs are not required for the purchase and operation of physical models, and the number of simultaneously conducted experiments does not depend on the number of physical models.

## 7. References

- [1] G. A. Miller, The cognitive revolution. Trends in Cognitive Sciences, 7, 2003.
- [2] D. Vernon, Artificial Cognitive Systems. Carnegie Mellon University Africa 2014.
- [3] M. Poliakov, "Cognitive Control Systems: Structures and Models." Electrical and Computer Systems 25 (2017) 387-393. doi:10.15276/eltecs.25.101.2017.46
- [4] M. Poliakov, K. Henke, H.-D. Wuttke, Prospects for Constructing Remote Laboratories to Study Cognitive IoT Systems in. V. Kharchenko, A. L. Kor, A. Rucinski (Eds), Dependable IoT for Human and Industry. Modeling, Architecting, Implementation, RIVER PUBLISHERS. 2018: 503-513.
- [5] A.K.M. Azad, M.E. Auer, and V.J. Harward (Eds.) (2012) Internet Accessible Remote Laboratories: Scalable E-Learning Tools for Engineering and Science Disciplines, Engineering Science Reference,

- [6] The Grid of Online Lab Devices Ilmenau (GOLDi). <https://goldi-labs.net>
- [7] Staudinger GmbH Automatisierungstechnik, site. [www.staudinger-est.de](http://www.staudinger-est.de)
- [8] H. Giese, B. Rumpe, B. Schätz, and J. Sztipanovits, "Science and Engineering of Cyber-Physical Systems (Dagstuhl Seminar 11441)," Dagstuhl Reports , vol. 1, no. 11, 2012: 1–22.
- [9] N. Schiffeler, V. Varney, E. Borowski, I. Isenhardt, Basic Requirements to Designing Collaborative Augmented Reality Status Quo and First Insights to a User-Centered Didactic Concept, in: Proceedings of the 17<sup>th</sup> International Conference on Remote Engineering and Virtual Instrumentation (REV2020). 26–28 February 2020, University of Georgia, Athens, GA, USA, pp. 216-230.
- [10] M. Poliakov, K. Henke, H.-D. Wuttke. The augmented functionality of the physical models of objects of study for remote laboratories *M.E. Auer, D.G. Zutin (eds.) Online Engineering & Internet of Things*, in: Proceedings of the 14th. International Conference on Remote Engineering and Virtual Instrumentation REV'2017, held 15-17 March 2017, Columbia University, New York, USA Series: Lecture Notes in Networks and Systems, Vol. 22, Springer, Cham. 2018. doi: 10.1007/978-3-319-64352-6.
- [11] M. Poliakov, T. Larionova, H.-D. Wuttke, K. Henke, Automated testing of physical models in remote laboratories by control event streams, in: Proceeding 2016 International Conference on Interactive Mobile Communication, Technologies and Learning IMCL'2016, 17–19 October 2016, San Diego, CA, USA. 94 p., pp. 10–13.
- [12] M. Poliakov, T. Larionova, G. Tabunshchyk, A. Parkhomenko, K. Henke. "Hybrid models of studied objects using remote laboratories for teaching design of control systems". *Int. J. Online Eng. (iJOE)* 9, 7–13 (2016). doi:10.3991/ijoe.v12i09.6128.
- [13] H.-D. Wuttke, M. Hamann and K. Henke. "Learning analytics in online remote labs." 2015. 3rd Experiment International Conference (exp.at'15) (2015): 255-260: Online Experimentation 2016 doi:10.1109/EXPAT.2015.7463275.
- [14] George F. Luger. Artificial Intelligence: Structures and Strategies for Complex Problem Solving, 4th. ed., Williams Publishing House, 2003.
- [15] Example of sorting boxes in the Warehouse,  
URL: <https://drive.google.com/drive/folders/1jzUECoI2ZU0nhrHCPQwvYFFpnxHHKhKu>
- [16] H.-D. Wuttke, K. Henke, R. Hutschenreuter, Digital Twins in Remote Labs. In: Auer M., Ram B. K. (eds.) *Cyber-physical Systems and Digital Twins. REV'2019. Vol. 80 of Lecture Notes in Networks and Systems*. Springer. Cham. 2020. doi:10.1007/978-3-030-23162-0\_26.

# SVD-Aided EKF with Process Noise Covariance Adaptation Applied to Satellite Attitude Dynamics

Chingiz Hajiyeve and Demet Cilden-Guler

*Faculty of Aeronautics and Astronautics, Istanbul Technical University, 34469, Maslak, Istanbul, Turkey*

## Abstract

An attitude estimation scheme is designed with the process noise adaptation rule in the extended Kalman filter (EKF) algorithm for a spacecraft in low Earth orbit. The adaptations are designed for compensating the external disturbances by updating the process noise covariance. Satellite attitude estimation algorithm based on the Singular Value Decomposition (SVD) as a preprocessing step in EKF is proposed. The process noise covariance (Q) adaptation rule is incorporated into the previous filter design. The proposed filter has the capability to be robust against initialization errors and the dynamics modeling errors of the satellite. Numerical simulations based on several scenarios are employed to investigate the robustness of the filter.

## Keywords 1

Singular value decomposition, extended Kalman filter, attitude estimation, adaptive filtering, rate gyro, small satellite.

## 1. Introduction

Attitude and angular rate of a satellite can be estimated using a conventional Kalman filter based on nonlinear measurements of reference directions (e.g. Earth magnetic field direction, Sun direction, etc.) [1–3]. On the other hand, linear measurements-based approach using vector observations in a preprocessing method can also be used for attitude estimation purposes. This approach is called nontraditional Kalman filter [3–7]. Here, a single-frame method determining the attitude angles as linear measurement inputs for the attitude filter such as Extended Kalman filter (EKF) or Unscented Kalman filter (UKF). The measurement model becomes linear as the state vector is composed of attitude angles that are directly observed by the preprocessing step. As an example, in [4], authors present an integrated satellite attitude determination system based on EKF and a single-frame estimator to estimate the attitude angles and angular velocities. The single-frame attitude estimator uses the TRIAD (Three-Axis Attitude Determination) method.

As the measurement model is linear in the nontraditional attitude filter, the overall computational load for the attitude estimation algorithm can be decreased at a great extent. In addition, if the filter is correctly tuned the estimation accuracy may increase compared to the estimates of a traditional filter. Because of these benefits the nontraditional approach for attitude filtering is a promising method for attitude estimation of small satellites including the nanosatellites [3, 5, 8, 9].

One of the difficulties for the non-traditional attitude filters that were met in previous investigations is the tuning of the filter in terms of the process noise covariance. Variations in the disturbance torques such as the residual magnetic torque, modeling errors in the satellite's dynamics and any fault in the actuators, they all cause changes in the process noise covariance of the filter. Despite being robust to any kind of variation in the measurements due to its nature, the nontraditional attitude filter is not capable of tuning itself against this kind of process noise covariance variations. The traditional Kalman filtering algorithms including the EKF, also have the same drawback when using any of them for attitude estimation: they are not robust against the environment changes. In case

CMIS-2021: The Fourth International Workshop on Computer Modeling and Intelligent Systems, April 27, 2021, Zaporizhzhia, Ukraine

EMAIL: cingiz@itu.edu.tr (C. Hajiyeve); cilden@itu.edu.tr (D. Cilden-Guler)

ORCID: 0000-0003-4115-341X (C. Hajiyeve); 0000-0002-3924-5422 (D. Cilden-Guler);



© 2021 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

of such changes, the estimation accuracy of the filter deteriorates and if the faults are long lasting, the filter may diverge. For any satellite, harsh operation conditions and anomalies in the geomagnetic field and sun are very likely in space environment. This situation leads us to a search for an Adaptive EKF (AEKF) algorithm that can be implemented in the nontraditional attitude filter.

The paper [10] proposes a robust adaptive UKF (RAUKF) to improve the accuracy and robustness of state estimation with having uncertain noise covariance. An online fault-detection mechanism is adopted to decide whether there is a need to update the current noise covariance or not. If necessary, innovation-based method and residual-based method are used to calculate the estimations of current noise covariance of process and measurement, respectively. The filter combines the last noise covariance matrices with the estimations for the new noise covariance matrices by using a weighting factor.

Like space anomalies the marine environment is also very complex; therefore, it is difficult to accurately obtain the system noise. The acoustic observations usually have gross errors, which affects the positioning accuracy of classical UKF. An RAUKF algorithm is proposed in [11] based on the Sage-Husa filter. The RAUKF can adaptively adjust the system noise to solve the divergence problem, and control the effects of gross measurement errors on dynamic state estimates.

In [12], the authors proposed an integrated algorithm which adapts both the measurement and process noise covariance matrices of the UKF for satellite attitude estimation. However, in this paper, the traditional approach is used for the attitude estimation.

In this study, we extend and add on to our previous researches about non-traditional filtering by proposing an adaptation method for tuning the Q matrix of the filter. The SVD method runs using the magnetometer and Sun sensor measurements as the first stage of the algorithm, and estimates the attitude of the satellite, giving one estimate at a single-frame. Then, these estimated attitude terms are used as inputs to the Q-adaptive EKF (AEKF). In different from the previous works of authors [28, 29], where SVD-Aided UKF adaptation is performed using single scale factors, in this study adapting the attitude filter is fulfilled using multiple scale factors.

The paper presents the mathematical models for a nanosatellite in Section 2. The SVD-aided EKF algorithm is introduced in Section 3 in detail for each sub-algorithm. The proposed filter adaptation method is also presented in this section. In Section 4, the simulation results of the proposed SVD-aided AEKF algorithm for a hypothetical nanosatellite are given. In Section 5, a brief summary and conclusions are given.

## 2. Singular Value Decomposition Method

Singular value decomposition (SVD) method is one of the single-frame methods (SFMs) used for minimization of Wahba's loss function [13]. At least two vectors need to be measured in order to determine the attitude using the optimal attitude matrix. The loss function is defined as,

$$L(\mathbf{A}) = \frac{1}{2} \sum_i a_i \|\mathbf{b}_i - \mathbf{A}\mathbf{r}_i\|^2 \quad (1)$$

$$\mathbf{B}_{\text{svd}} = \sum_i a_i \mathbf{b}_i \mathbf{r}_i^T \quad (2)$$

$$L(\mathbf{A}) = \lambda_0 - \text{tr}(\mathbf{A}\mathbf{B}^T) \quad (3)$$

where  $\mathbf{b}_i$  is the measurement vector in body coordinate system and  $\mathbf{r}_i$  is the vector in reference coordinate system,  $a_i$  is the non-negative weight with their sum of  $\lambda_0$ .  $\mathbf{B}_{\text{svd}}$  can be seen in Equation (2) has singular value decomposition:

$$\mathbf{B}_{\text{svd}} = \mathbf{U}\mathbf{S}\mathbf{V}^T = \mathbf{U}\text{diag}[S_{11} \ S_{22} \ S_{33}]\mathbf{V}^T \quad (4)$$

$$\mathbf{A}_{\text{opt}} = \mathbf{U}\text{diag}[1 \ 1 \ \det(\mathbf{U})\det(\mathbf{V})]\mathbf{V}^T \quad (5)$$

where the left and right matrices are  $\mathbf{U}$  and  $\mathbf{V}$ ,  $\{S_{11}, S_{22}, S_{33}\}$  are the primary singular values and  $\mathbf{A}_{\text{opt}}$  is the optimal attitude transformation matrix. To examine the rotation angle error, covariance matrix can be defined as,

$$\mathbf{P}_{\text{svd}} = \mathbf{U}\text{diag}[(s_2 + s_3)^{-1} \ (s_3 + s_1)^{-1} \ (s_1 + s_2)^{-1}]\mathbf{U}^T \quad (6)$$

where  $s_1 = S_{11}$ ,  $s_2 = S_{22}$ ,  $s_3 = \det(\mathbf{U})\det(\mathbf{V}) S_{33}$ . In applications where only two vectors are available and one of them has faulty measurements, the SVD method cannot determine the attitude as at least two vector observations are required for the method to work properly.

## 2.1. Modeling of Earth's Magnetic Field Vector

International Geomagnetic Reference Field (IGRF) is used as the Earth's magnetic field model in this study [14]. Model needs two inputs, position of the near-Earth spacecraft and the time, for calculating the geomagnetic field vector as,

$$\mathbf{B}_o(k) = -\nabla \left\{ a \sum_{n=1}^N \sum_{m=0}^n \left( \frac{a}{r_{sc}} \right)^{n+1} \left[ g_n^m(t) \cos(m\phi_{sc}) + h_n^m(t) \sin(m\phi_{sc}) \right] \times P_n^m(\cos\theta_{sc}) \right\} \quad (7)$$

where  $r_{sc}$  is the spacecraft's distance from Earth's center in km,  $a = 6371.2$  km (magnetic reference spherical radius),  $\theta_{sc}$  is colatitude in deg,  $\phi_{sc}$  longitude in deg of the spacecraft at specified time,  $t$ . In the equation, the global variables  $\{g, h\}$  are Gauss coefficients while  $P$  denotes the Legendre function of degree  $n$  and order  $m$ . The magnetometer measurements can be defined using the transformed magnetic field model using the attitude transformation matrix  $\mathbf{A}$  as,

$$\mathbf{B}_m(k) = \mathbf{A}(k)\mathbf{B}_o(k) + \mathbf{v}_B(k) \quad (8)$$

where  $\mathbf{B}_m$  is the magnetometer measurement vector and  $\mathbf{u}_B$  is the zero-mean Gaussian noise of the magnetometer measurements.

## 2.2. Modeling of Sun Direction Vector

The reference sun direction can be calculated in inertial frame at a specified time [15]. The sun sensor measurements can be defined using the transformed sun direction model using the attitude transformation matrix  $\mathbf{A}$  as,

$$\mathbf{S}_m(k) = \mathbf{A}(k)\mathbf{S}_o(k) + \mathbf{u}_S(k) \quad (9)$$

where  $\mathbf{S}_m$  is the sun sensor measurement vector,  $\mathbf{S}_o$  is the reference sun direction vector and  $\mathbf{u}_S$  is the zero-mean Gaussian noise of the sun sensor measurements.

## 3. Mathematical Models

### 3.1. Satellite's Rotational Motion

The kinematics equation of motion of the satellite using Euler angle representation (roll ( $\phi$ ), pitch ( $\theta$ ), and yaw ( $\psi$ )) is given as [16],

$$\begin{bmatrix} \dot{\phi} \\ \dot{\theta} \\ \dot{\psi} \end{bmatrix} = \begin{bmatrix} 1 & \sin(\phi)\tan(\theta) & \cos(\phi)\tan(\theta) \\ 0 & \cos(\phi) & -\sin(\phi) \\ 0 & \sin(\phi)/\cos(\theta) & \cos(\phi)/\cos(\theta) \end{bmatrix} \begin{bmatrix} p \\ q \\ r \end{bmatrix} \quad (10)$$

where the components of the vector  $\boldsymbol{\omega}_{BR}$  in body frame with respect to the reference frame are given as  $p, q, r$ . The body angular rate with respect to the inertial frame is expressed as,

$$\boldsymbol{\omega}_{BI} = [\omega_x \quad \omega_y \quad \omega_z]^T \quad (11)$$

with the relationship

$$\boldsymbol{\omega}_{BR} = \boldsymbol{\omega}_{BI} - \mathbf{A}[0 \quad -\omega_o \quad 0]^T \quad (12)$$

where  $\omega_o$  defines the orbital angular velocity with respect to the inertial frame,  $\mathbf{A}$  is the attitude transformation matrix composed of Euler angles.

For estimating the attitude and attitude rates together, rate gyro driven kinematics or both the dynamics and the kinematics of the satellite are utilized. We use dynamic equations in this study. The satellite's rotational motion dynamics are derived based on the Euler's equations as,

$$J_x \frac{d\omega_x}{dt} = N_x + (J_y - J_z) \omega_y \omega_z, \quad (13a)$$

$$J_y \frac{d\omega_y}{dt} = N_y + (J_z - J_x) \omega_z \omega_x, \quad (13b)$$

$$J_z \frac{d\omega_z}{dt} = N_z + (J_x - J_y) \omega_x \omega_y, \quad (13c)$$

where  $J_x$ ,  $J_y$  and  $J_z$  are the principal moments of inertia and  $N_x$ ,  $N_y$  and  $N_z$  are the terms of the external disturbance torque.

### 3.2. Measurement Models

Single-frame method aided filters use the SFM estimates and the noise covariance calculations. Since the SFM determines the coarse attitude angles, the three elements of the measurements for the filter are directly composed of these angles as,

$$\begin{aligned} z_\phi(k) &= \phi_{svd}(k) + v_\phi(k) \\ z_\theta(k) &= \theta_{svd}(k) + v_\theta(k) \\ z_\psi(k) &= \psi_{svd}(k) + v_\psi(k) \end{aligned} \quad (14)$$

where  $v_{(\cdot)}(k)$  is the measurement noise and  $\mathbf{z}_\Phi = [z_\phi \ z_\theta \ z_\psi]^T$ . The rate gyros are used as the measurement input to the filter as well,

$$\mathbf{z}_\omega(k) = \boldsymbol{\omega}_{Bl}(k) + \mathbf{u}_\omega(k) \quad (15)$$

where  $\mathbf{u}_\omega$  is the zero-mean Gaussian noise of the rate gyro measurements and  $\mathbf{z}_\omega = [z_{\omega_x} \ z_{\omega_y} \ z_{\omega_z}]^T$ . The variances and the expectations of the measurement noises are,

$$\begin{aligned} E[v_{(\cdot)}(k)] &= 0, \ E[v_\phi^2(k)] = \text{Var}(v_\phi(k)), \ E[v_\theta^2(k)] = \text{Var}(v_\theta(k)), \ E[v_\psi^2(k)] = \text{Var}(v_\psi(k)), \\ E[v_{\omega_x}^2(k)] &= \text{Var}(v_{\omega_x}(k)), \ E[v_{\omega_y}^2(k)] = \text{Var}(v_{\omega_y}(k)), \ E[v_{\omega_z}^2(k)] = \text{Var}(v_{\omega_z}(k)). \end{aligned} \quad (16)$$

## 4. SVD-Aided EKF for Attitude and Rate Estimation

For the full-state estimation, the state vector ( $\mathbf{x}$ ) is composed of attitude angles and angular rates as,

$$\mathbf{x} = [\phi \ \theta \ \psi \ \omega_x \ \omega_y \ \omega_z]^T \quad (17)$$

The process and measurement models can be expressed in the discrete-time as,

$$\hat{\mathbf{x}}(k+1) = \mathbf{f}[\hat{\mathbf{x}}(k), k] + \mathbf{w}(k) \quad (18)$$

$$\mathbf{z}(k) = \mathbf{H} \hat{\mathbf{x}}(k) + \mathbf{u}(k) \quad (19)$$

where  $\mathbf{f}[\cdot]$  is the state transition matrix,  $\mathbf{w}$  is Gaussian process noise,  $\mathbf{u}$  is gaussian measurement noise,  $\mathbf{z}$  is the measurement vector,  $\mathbf{H}$  is the measurement matrix.

The process and measurement noises  $\mathbf{u}(k) = [v_\phi(k) \ v_\theta(k) \ v_\psi(k) \ v_{\omega_x}(k) \ v_{\omega_y}(k) \ v_{\omega_z}(k)]^T$  and  $\mathbf{w}(k)$  are assumed linearly additive Gaussian, temporally uncorrelated with zero mean. Their corresponding covariances are,

$$E[\mathbf{w}(i)\mathbf{w}^T(j)] = \mathbf{Q}(i)\delta(ij), \ E[\mathbf{u}(i)\mathbf{u}^T(j)] = \mathbf{R}(i)\delta(ij) \quad (20)$$

Here,  $\delta(ij)$  is the Kronecker symbol. The process and measurement noises are assumed to be uncorrelated as,

$$E[\mathbf{w}(i)\mathbf{u}^T(j)] = 0, \forall i, j \quad (21)$$

The state vector estimate is,

$$\hat{\mathbf{x}}(k+1) = \hat{\mathbf{x}}(k+1/k) + \mathbf{K}(k+1) \times \{\mathbf{z}(k+1) - \mathbf{H} \hat{\mathbf{x}}(k+1/k)\} \quad (22)$$

where the measurement vector is presented as  $\mathbf{z}(k+1) = [\mathbf{z}_\phi(k+1) \quad \mathbf{z}_\omega(k+1)]^T$  with the measurement matrix  $\mathbf{H}$ . For our case,  $\mathbf{H}$  is a 6x6 unit matrix because of the linearity. The extrapolation value is,

$$\hat{\mathbf{x}}(k+1/k) = \mathbf{f}[\hat{\mathbf{x}}(k), k] \quad (23)$$

The gain of the filter is,

$$\mathbf{K}(k+1) = \mathbf{P}(k+1/k) \mathbf{H}^T \times [\mathbf{H} \mathbf{P}(k+1/k) \mathbf{H}^T + \mathbf{R}(k)]^{-1} \quad (24)$$

The extrapolation error covariance matrix is,

$$\mathbf{P}(k+1/k) = \frac{\partial \mathbf{f}[\hat{\mathbf{x}}(k), k]}{\partial \hat{\mathbf{x}}(k)} \mathbf{P}(k/k) \frac{\partial \mathbf{f}^T[\hat{\mathbf{x}}(k), k]}{\partial \hat{\mathbf{x}}(k)} + \mathbf{Q}(k) \quad (25)$$

The estimation error covariance matrix is,

$$\mathbf{P}(k+1/k+1) = [\mathbf{I} - \mathbf{K}(k+1) \mathbf{H}] \mathbf{P}(k+1/k) \quad (26)$$

The filter is represented by the Equations (22) – (26).

One of the difficulties for the nontraditional attitude filters is the tuning of the filter's process noise covariance matrix,  $\mathbf{Q}$  [17]. Especially, the process noise covariance needs to be tuned when the environment changes [18]. Here, any condition change is referred as a change that may affect the process model of the filter. These might include uncertainties in the inertia parameters, disturbance torques (as it may happen when the satellite goes in/out of the eclipse) as well as the changes in the controller parameters (such as a malfunction) although a controller is not considered in this paper.

## 5. SVD-Aided Q-Adaptive EKF

The SVD method runs as the first stage of the algorithm and estimates the attitude of the satellite giving one estimate at a single-frame. Then, these estimated attitude terms are given as inputs to the EKF. This forms the SVD-Aided EKF scheme. An adaptation using process noise covariance matrix is also added to this filter called SVD-Aided AEKF. The algorithms in SVD-Aided AEKF is summarized in the block diagram in Figure 1.

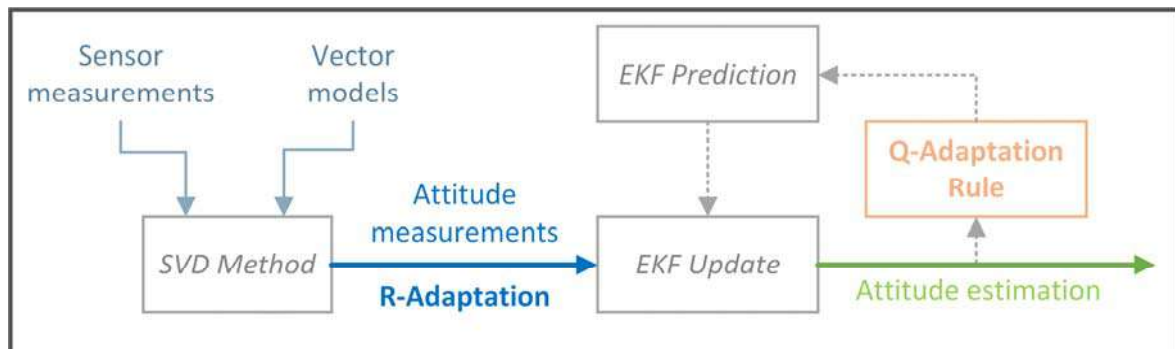


**Figure 1:** Relationships between the algorithms used for SVD-Aided AEKF

Here, there are two measurements used in SVD for attitude determination: magnetic field and sun direction. The determined attitude and corresponding covariance ( $\mathbf{R}$ ) are the filter's inputs from the SVD. In addition to SVD's observations, rate gyro measurements are used as well in the full-state estimation process. Also, one more adaptation rule (Q-adaptation) is defined for making the filter adaptive to the environmental changes. This way, the filter is expected to be robust against process noises.



In case of process noise covariance change, the real error of the innovation covariance will exceed the theoretical one. Thus, the basic of the Q-adaptation is to obtain an appropriate multiplier matrix for the  $\mathbf{Q}$  matrix such that the real and theoretical values of the innovation covariance could match.

The innovation of the EKF in Equations (22) - (26) is,

$$\tilde{\mathbf{e}}(k) = \mathbf{z}(k+1) - \mathbf{H} \hat{\mathbf{x}}(k+1/k) = \mathbf{z}(k+1) - \mathbf{H} \mathbf{f}[\hat{\mathbf{x}}(k), k] \quad (27)$$

In order to adapt the covariance, an adaptive scaling factor is put into the procedure. Hence, if a scaling matrix ( $\mathbf{\Lambda}$ ) built of multiple scaling factors, is added in to the algorithm as,

$$\mathbf{P}(k+1/k) = \frac{\partial \mathbf{f}[\hat{\mathbf{x}}(k), k]}{\partial \hat{\mathbf{x}}(k)} \mathbf{P}(k/k) \frac{\partial \mathbf{f}^T[\hat{\mathbf{x}}(k), k]}{\partial \hat{\mathbf{x}}(k)} + \mathbf{\Lambda}(k) \mathbf{Q}(k) \quad (28)$$

The scaling matrix  $\mathbf{\Lambda}$  can be determined from equality,

$$\frac{1}{\mu} \sum_{j=k-\mu+1}^k \tilde{\mathbf{e}}(j) \tilde{\mathbf{e}}^T(j) = \mathbf{H} \mathbf{P}(k+1/k) \mathbf{H}^T + \mathbf{R}(k) \quad (29)$$

The left side of equality (29) represents the real, and the right side represents the theoretical value of the covariance of the innovation sequence.

The innovation covariance can be expressed in the form below by substituting (28) into (29),

$$\frac{1}{\mu} \sum_{j=k-\mu+1}^k \tilde{\mathbf{e}}(j) \tilde{\mathbf{e}}^T(j) = \mathbf{H} \left[ \frac{\partial \mathbf{f}[\hat{\mathbf{x}}(k), k]}{\partial \hat{\mathbf{x}}(k)} \mathbf{P}(k/k) \frac{\partial \mathbf{f}^T[\hat{\mathbf{x}}(k), k]}{\partial \hat{\mathbf{x}}(k)} + \mathbf{\Lambda}(k) \mathbf{Q}(k) \right] \mathbf{H}^T + \mathbf{R}(k) \quad (30)$$

Here,  $\mu$  is the width of the moving window.

From (30), it can be written,

$$\mathbf{H} \mathbf{\Lambda}(k) \mathbf{Q}(k) \mathbf{H}^T = \frac{1}{\mu} \sum_{j=k-\mu+1}^k \tilde{\mathbf{e}}(j) \tilde{\mathbf{e}}^T(j) - \mathbf{H} \left[ \frac{\partial \mathbf{f}[\hat{\mathbf{x}}(k), k]}{\partial \hat{\mathbf{x}}(k)} \mathbf{P}(k/k) \frac{\partial \mathbf{f}^T[\hat{\mathbf{x}}(k), k]}{\partial \hat{\mathbf{x}}(k)} \right] \mathbf{H}^T - \mathbf{R}(k) \quad (31)$$

By multiplying (31) with  $\mathbf{H}^T$  on the left and with  $\mathbf{H}$  on the right, the scale matrix  $\mathbf{\Lambda}$  can be determined as,

$$\mathbf{\Lambda}(k) = [\mathbf{H}^T \mathbf{H}]^{-1} \mathbf{H}^T \times \left\{ \frac{1}{\mu} \sum_{j=k-\mu+1}^k \tilde{\mathbf{e}}(j) \tilde{\mathbf{e}}^T(j) - \mathbf{H} \left[ \frac{\partial \mathbf{f}[\hat{\mathbf{x}}(k), k]}{\partial \hat{\mathbf{x}}(k)} \mathbf{P}(k/k) \frac{\partial \mathbf{f}^T[\hat{\mathbf{x}}(k), k]}{\partial \hat{\mathbf{x}}(k)} \right] \mathbf{H}^T - \mathbf{R}(k) \right\} \times \mathbf{H} [\mathbf{Q}(k) \mathbf{H}^T \mathbf{H}]^{-1} \quad (32)$$

For a specific case where the all states are measured ( $\mathbf{H}=\mathbf{I}$ ) as in case, (32) reduces to

$$\mathbf{\Lambda}(k) = \left\{ \frac{1}{\mu} \sum_{j=k-\mu+1}^k \tilde{\mathbf{e}}(j) \tilde{\mathbf{e}}^T(j) - \left[ \frac{\partial \mathbf{f}[\hat{\mathbf{x}}(k), k]}{\partial \hat{\mathbf{x}}(k)} \mathbf{P}(k/k) \frac{\partial \mathbf{f}^T[\hat{\mathbf{x}}(k), k]}{\partial \hat{\mathbf{x}}(k)} \right] - \mathbf{R}(k) \right\} \mathbf{Q}^{-1}(k) \quad (33)$$

If the process noise does not change, the scale matrix will be a unit matrix as  $\mathbf{\Lambda}=\mathbf{I}$ , since the innovation values already match. Here,  $\mathbf{I}$  represents the unit matrix.

The obtained scaling matrix should be diagonalized since the  $\mathbf{Q}$  matrix must be a diagonal, positive definite matrix. Remark that, as  $\mu$  is a limited number because of the number of the measurements and the computations performed with the computer implies errors such as the approximation errors and the round off errors;  $\mathbf{\Lambda}$  matrix, found by using Equations (32) or (33), may not be diagonal and may have diagonal elements which are “negative” or less than “one” (although the computer simulations may bring out such results that is physically impossible in reality). Therefore, in order to avoid such situation, composing the scale factor by the following rule is suggested by [19] as,

$$\mathbf{\Lambda}^* = \text{diag}(\lambda_1^*, \lambda_2^*, \dots, \lambda_n^*), \quad (34)$$

where,

$$\lambda_i^* = \max\{1, \Lambda_{ii}\} \quad i=1, \dots, n. \quad (35)$$

Here,  $\Lambda_{ii}$  represents the  $i^{th}$  diagonal element of the matrix  $\mathbf{\Lambda}$ . Apart from that point, if there is an actuator fault in the system,  $\Lambda_k^*$  must be included in the estimation process as,

$$\mathbf{P}(k+1/k) = \frac{\partial \mathbf{f}(\hat{\mathbf{x}}(k), k)}{\partial \hat{\mathbf{x}}(k)} \mathbf{P}(k/k) \frac{\partial \mathbf{f}^T(\hat{\mathbf{x}}(k), k)}{\partial \hat{\mathbf{x}}(k)} + \mathbf{\Lambda}^*(k) \mathbf{Q}(k) \quad (36)$$

Therefore, when process noise change, the related element of the scale matrix, which corresponds to the changed component of the state vector, increases and that brings out a smaller Kalman gain, which reduces the effect of the innovation on the state update process. As a result, the robustness of the filter against the process noise change is ensured and deterioration of the estimation procedure caused by the fault is prevented.

## 6. Analysis and Results

A low-Earth orbiting spacecraft with the principal moment of inertia  $\mathbf{J} = \text{diag}(2.1 \times 10^{-3} \quad 2.0 \times 10^{-3} \quad 1.9 \times 10^{-3}) \text{ kg} \cdot \text{m}^2$  is considered in the simulations. The spacecraft is tumbling during the simulations with an initial state of  $\mathbf{x}_0 = [0.03 \text{ rad} \quad 0.02 \text{ rad} \quad 0.01 \text{ rad} \quad 0.001 \text{ rad/s} \quad 0.0015 \text{ rad/s} \quad 0.002 \text{ rad/s}]$ . The sun sensors are processed at 1 Hz and corrupted by Gaussian zero-mean noise with a standard deviation of 2% whereas the magnetometers are corrupted by 8%. The process noise covariance is selected as  $\mathbf{Q} = 10^{-4} \mathbf{I}_{6 \times 6}$ . There is an eclipse period along the orbit between 1500 and 2500 seconds, which causes the sun sensors to give zero-output.

This section analyzes the application of the process noise covariance adaptation on the filter. The Q-adaptive filter is called SVD-Aided AEKF in this study. The adaptation against process noise is investigated by applying noise increment on the system with a standard deviation of  $\sigma_{1,2,3}^{\text{process}} = 0.01 \text{ rad}$  and  $\sigma_{4,5,6}^{\text{process}} = 0.01 \text{ rad/s}$  between 3000 and 4000 seconds. The inherently R-adaptive filter called SVD-Aided EKF does not cope well with the process noise increment and deteriorates as seen from Figure 2 whereas simultaneous R- and Q-adaptive filter called SVD-Aided AEKF estimates more accurate attitude angles. On the other hand, angular velocity estimations deteriorate for both filters as seen in Figure 3 and are not improved as much as the attitude angles in Q-adaptation.

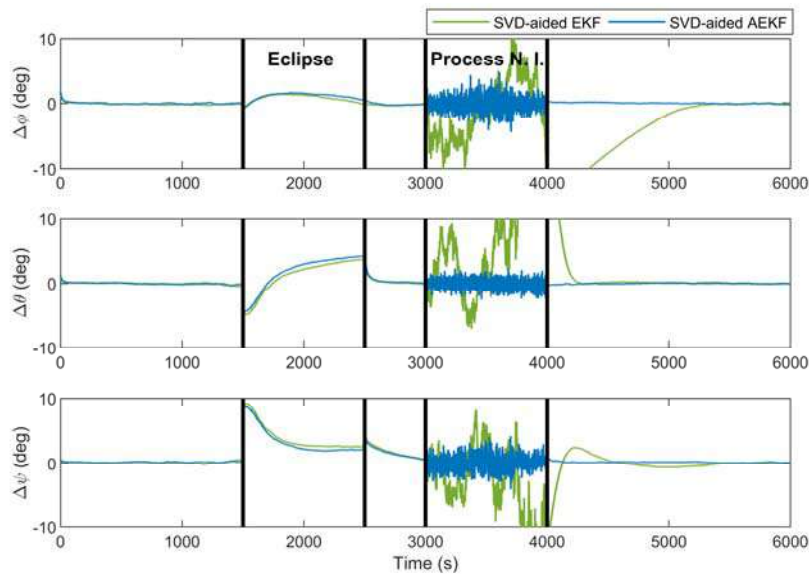


**Figure 2:** Attitude estimation errors for process noise increment (N.I.) case

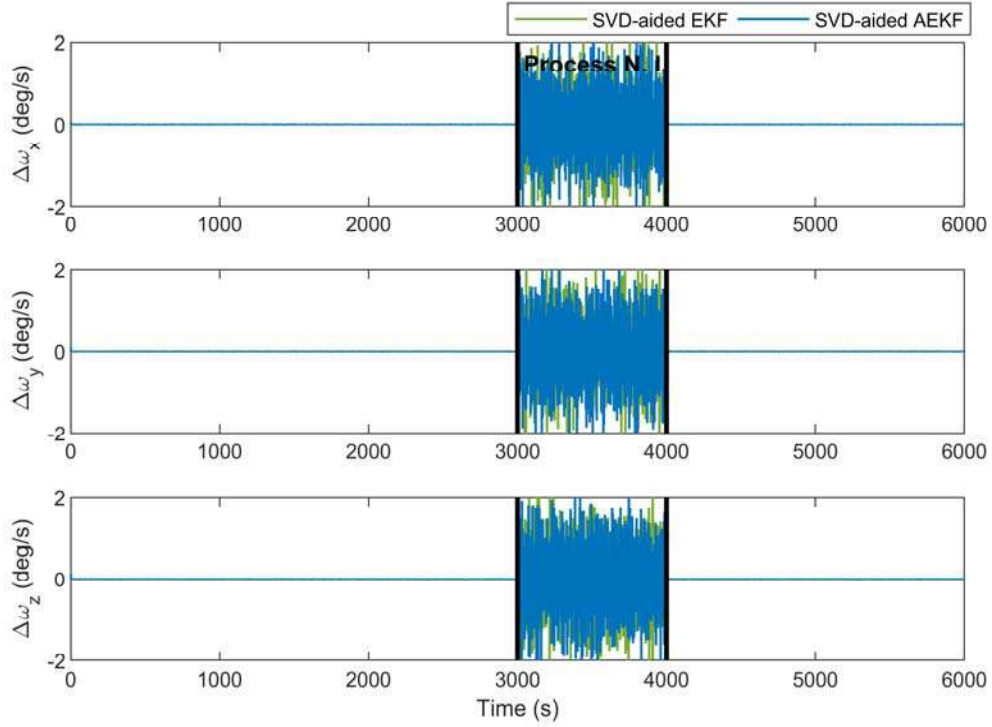


**Figure 3:** Angular velocity estimation errors for process noise increment (N.I.) case

**Table 1**

RMS errors of SVD-Aided EKF and AEKF during nominal mode and process noise increment (N.I.)

| RMS Error          | SVD-aided EKF |              | SVD-aided AEKF |              |
|--------------------|---------------|--------------|----------------|--------------|
|                    | Nominal       | Process N.I. | Nominal        | Process N.I. |
| $\phi$ (deg)       | 0.9479        | 4.8026       | 0.0971         | 1.2094       |
| $\theta$ (deg)     | 0.0568        | 11.7243      | 0.0993         | 0.8061       |
| $\psi$ (deg)       | 0.4194        | 6.7010       | 0.3377         | 1.1692       |
| $\omega_x$ (deg/s) | 0.1545        | 0.8462       | 0.1546         | 0.7900       |
| $\omega_y$ (deg/s) | 0.0845        | 0.7890       | 0.0850         | 0.8197       |
| $\omega_z$ (deg/s) | 0.0055        | 0.7954       | 0.0057         | 0.7915       |

Table 1 is composed of root mean square (RMS) error values of the estimations of the filters considered in this study. The column ‘Nominal’ takes the period outside of the eclipse and outside of the process noise increment into account. The results do not differentiate much between method-to-method as there is no fault considered in this period. On the other hand, SVD-Aided AEKF improves the attitude results with respect to SVD-Aided EKF around 4 to 10 times for the process noise increment interval. Figure 4 demonstrates the adaptive factor used in the Q-adaptation process, which seem to be successful in compensating the noise on the process for estimating the attitude angles.

The adaptive algorithm performs the correction only when the real values of the process noise covariance does not match with the model used in the synthesis of the filter. Otherwise, the filter works in a regular manner.

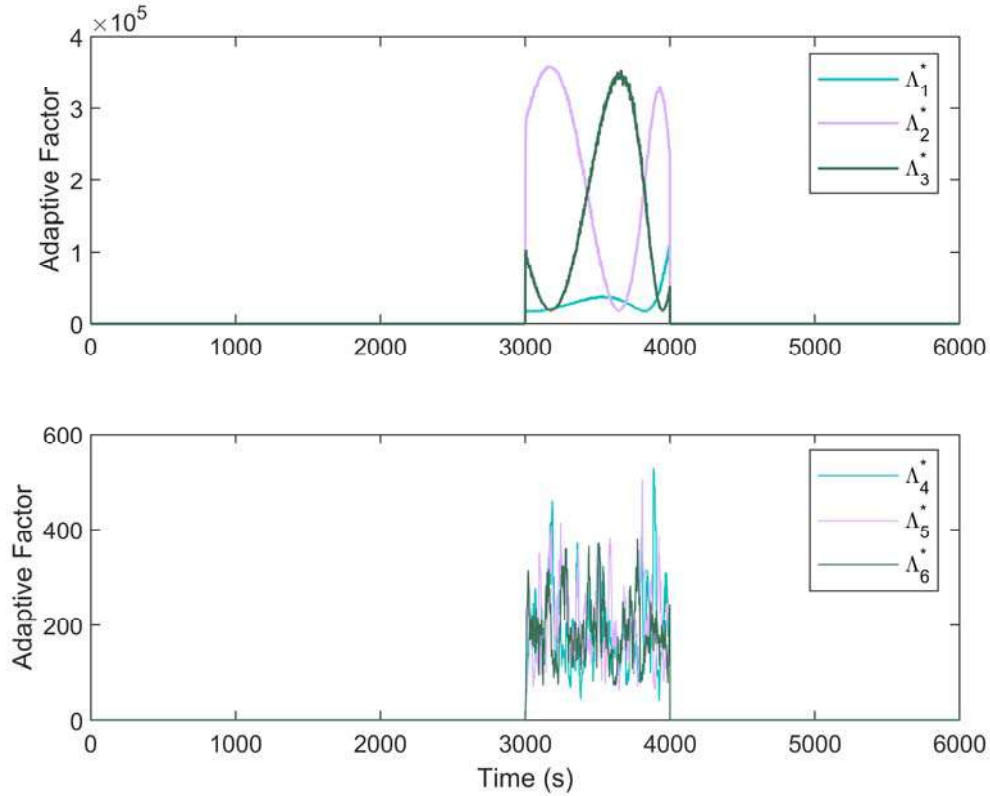


**Figure 4:** Adaptive factor of the SVD-aided AEKF

## 7. Conclusions

This paper presents the design and numerical analyses of Q- adaptive extended Kalman filtering algorithm. An attitude estimation algorithm is designed by integrating SVD and EKF methods to estimate the attitude angles and angular velocities. Processing of attitude sensors (magnetometer and sun sensor measurements in this case) within the SVD method computes the Euler angle measurements and their variances, and provides them to the filter as inputs. This filter also eliminates the possibility of poorly-chosen initial attitude as it initializes the attitude using SVD method.

Specifically, the Q adaptation method is proposed in this study. In case of process noise increment, which may be caused by the changes in the environment or satellite dynamics, performance of the Q- adaptive EKF is investigated. To adapt the EKF to the changing conditions its Q matrix is tuned by an adaptive method. Since the measurement model of the SVD-aided EKF algorithm is linear, the adaptation rule is similar to linear KF and easy to apply.

For investigating the system modelling errors, one more case is implemented by implementing a noise increment on the system. Here, the Q-adaptation rule is introduced for the EKF. Thus, the attitude and angular rates of the satellite can be estimated accurately in spite of any change in the process noise covariance.

## 8. References

- [1] H. E. Soken and C. Hajiyeve, "Estimation of Pico Satellite Attitude Dynamics and External Torques via Unscented Kalman Filter," *J. Aerosp. Technol. Manag.*, volume 6, no. 2, pp. 149–157, May 2014, doi: 10.5028/jatm.v6i2.352.
- [2] J. C. Springmann and J. W. Cutler, "Flight Results of a Low-cost Attitude Determination Systems," *Acta Astronaut.*, volume 99, pp. 201–214, 2014, doi: 10.1016/j.actaastro.2014.02.026.
- [3] C. Hajiyeve and D. Cilden-Guler, "Review on Gyroless Attitude Determination Methods for Small Satellites," *Prog. Aerosp. Sci.*, volume 90, pp. 54–66, 2017, doi: 10.1016/j.paerosci.2017.03.003.

- [4] C. Hajiyeve and M. Bahar, "Attitude Determination and Control System Design of the ITU-UUBF LEO1 Satellite," *Acta Astronaut.*, volume 52, no. 2–6, pp. 493–499, 2003, doi: 10.1016/S0094-5765(02)00192-3.
- [5] B. Y. Mimasu and J. C. Van der Ha, "Attitude Determination Concept for QSAT," *Trans. Japan Soc. Aeronaut. Sp. Sci. Aerosp. Technol. Japan*, volume 7, pp. 63–68, 2009, doi: 10.2322/tstj.7.Pd\_63.
- [6] C. Hajiyeve, D. Cilden, and Y. Somov, "Gyro-free attitude and rate estimation for a small satellite using SVD and EKF," *Aerosp. Sci. Technol.*, volume 55, pp. 324–331, 2016, doi: 10.1016/j.ast.2016.06.004.
- [7] D. Cilden, H. E. Soken, and C. Hajiyeve, "Nanosatellite attitude estimation from vector measurements using SVD-AIDED UKF algorithm," *Metrol. Meas. Syst.*, volume 24, no. 1, pp. 113–125, 2017, doi: 10.1515/mms-2017-0011.
- [8] H. E. Soken, D. Cilden, and C. Hajiyeve, "Attitude Estimation for Nanosatellites Using Singular Value Decomposition and Unscented Kalman Filter," 2015.
- [9] D. Cilden-Guler, H. E. Soken, and C. Hajiyeve, "SVD-Aided UKF for Estimation of Nanosatellite Rotational Motion Parameters," *WSEAS Trans. Signal Procesing*, volume 14, pp. 27–35, 2018, Accessed: Apr. 07, 2021. [Online]. Available: <https://www.wseas.org/multimedia/journals/signal/2018/a085914-062.php>.
- [10] B. Zheng, P. Fu, B. Li, and X. Yuan, "A Robust Adaptive Unscented Kalman Filter for Nonlinear Estimation with Uncertain Noise Covariance," *Sensors*, volume 18, no. 3, p. 808, Mar. 2018, doi: 10.3390/s18030808.
- [11] J. Wang, T. Xu, and Z. Wang, "Adaptive Robust Unscented Kalman Filter for AUV Acoustic Navigation," *Sensors*, volume 20, no. 1, p. 60, Dec. 2019, doi: 10.3390/s20010060.
- [12] C. Hajiyeve and H. E. Soken, "Robust Adaptive Unscented Kalman Filter for Attitude Estimation of Pico Satellites," *Int. J. Adapt. Control Signal Process.*, volume 28, no. 2, pp. 107–120, 2014, doi: 10.1002/acs.2393.
- [13] G. Wahba, "Problem 65-1: A Least Squares Estimate of Satellite Attitude," *Soc. Ind. Appl. Math. Rev.*, volume 7, no. 3, p. 409, 1965, [Online]. Available: [http://www.malcolmdshuster.com/FC\\_Wahba\\_1965\\_J\\_SIAM.pdf](http://www.malcolmdshuster.com/FC_Wahba_1965_J_SIAM.pdf).
- [14] E. Thébault, C. C. Finlay, C. D. Beggan, P. Alken, and E. Al., "International Geomagnetic Reference Field: the 12th generation," *Earth, Planets Sp.*, volume 67:69, 2015, doi: 10.1186/s40623-015-0228-9.
- [15] D. A. Vallado, *Fundamentals of Astrodynamics and Applications (3rd Ed.)*, volume 21. USA: Microcosm Press/Springer, 2007.
- [16] J. R. Wertz, *Spacecraft Attitude Determination and Control*, no. 73. Dordrecht, Holland: D.Reidel Publishing Company, 2002.
- [17] C. Hajiyeve, H. E. Soken, and D. Cilden-Guler, "Q-Adaptation of SVD-aided UKF Algorithm for Nanosatellite Attitude Estimation," *IFAC-PapersOnLine*, volume 50, no. 1, pp. 8273–8278, Jul. 2017, doi: 10.1016/j.ifacol.2017.08.1399.
- [18] C. Hajiyeve, H. E. Soken, and D. Cilden-Guler, "Nontraditional Attitude Filtering with Simultaneous Process and Measurement Covariance Adaptation," *J. Aerosp. Eng.*, volume 32, no. 5, p. 04019054, Sep. 2019, doi: 10.1061/(ASCE)AS.1943-5525.0001038.
- [19] C. Hajiyeve and H. E. Soken, "Robust Estimation of UAV Dynamics in the Presence of Measurement Faults," *J. Aerosp. Eng.*, volume 25, no. 1, pp. 80–89, Jan. 2012, doi: 10.1061/(ASCE)AS.1943-5525.0000095.

# Modeling the Dynamics of Information Panic in Society. COVID-19 case

Sergiy Bushuyev<sup>a</sup>, Victoria Bushuieva<sup>a</sup>, Svitlana Onyshchenko<sup>b</sup> and Alla Bondar<sup>b</sup>

<sup>a</sup> Kyiv National University of Construction and Architecture, Povitroflotsky Avenue, 31, 03680 Kyiv, Ukraine

<sup>b</sup> Odesa National Maritime University, Mechnikova 34, 65029, Odesa, Ukraine

## Abstract

Global processes and their trends that occur in various fields of knowledge with a significant acceleration affect the development of information technology. The modern world is a world in which the media and social networks play a dominant role in the formation of a certain model of the world in the minds of individuals. Naturally, the consciousness of each person has a certain "information filter", the functional ability of which varies greatly depending on the characteristics of the person. The "strength" and capacity of this "filter" are influenced by many factors, which are the subject of a separate study for psychologists. Thus, let us accept the fact that this "filter" exists, and it is he who is responsible for the fact that a person filters all information coming to him in a certain way. And this applies to any information received by a person, including negative information. Most people, having received a portion of negative information - about politics, economics, natural disasters, pandemics, react in a certain way, primarily on an emotional level. For some, the first emotional reaction is gradually extinguished, and the action of the information negatively ends without causing any consequences. For others, a weak information filter allows a reaction in the form of fears, which may exist up to a certain moment, and then either the fear disappears, or, receiving a new informational "feed", that is, a new portion of negativity, increases.

## Keywords 1

Information entropy; panicdemic; energy; external environment

## 1. Introduction

The existing "information filter" for some people may not "survive" with certain portions of negative information and a certain frequency. But with an increase in the intensity of the information negative, even in people "stable" in this aspect, the filter may not cope.

At the same time, it should be noted that for some people, information negative causes a desire to relay it further - to acquaintances, relatives in live communication, and to anyone on social networks. Part of the environment reacts to this information message, which plays the role of a catalyst in the dissemination of negative information. It's no secret that positive information often causes less reaction than negative information. Thus, the process of further movement and strengthening of negative information is launched.

So, for some people, during the spring lockdown of 2020, the information in the news and social networks about the mortality rate from COVID 19 in Italy, videos, and reports about the dire consequences of this virus in the north of this country (where the highest and unexplained mortality rate was observed in that period) caused serious psychological and psychological consequences in the form of panic and a desire to "infect" your environment with it.

---

CMIS-2021: The Fourth International Workshop on Computer Modeling and Intelligent Systems, April 27, 2021, Zaporizhzhia, Ukraine  
EMAIL sbushuyev@ukr.net (S. Bushuyev); bushuieva.v@gmail.com (V. Bushuieva); onyshenko@gmail.com (S. Onyshchenko);  
ocheretyankaalla@gmail.com (A. Bondar)  
ORCID: 0000-0002-7815-8129 (S. Bushuyev); 0000-0001-7298-4369 (V. Bushuyeva) 0000-0002-728-4939 (S. Onyshchenko); 0000-0003-2228-2726 (A. Bondar)



© 2020 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

Thus, negative information plays a destructive role in society. Taking into account the current situation, the problem of studying the impact of the growth of negative information in society on its state is urgent.

The issues of the impact of information negativity on society are considered in works [1,2], where this impact was studied in the context of a modern pandemic, generated by COVID 19.

Entropy is a universal category for assessing the state of a system in the process of its life. And for social systems, this category has found wide application. Examples of such studies are [3-14]. In works [15-19], the basic provisions of a new, entropy concept at the organization level are presented and formulas are obtained for assessing entropy and studying its dynamics. These results (in particular, [16,17]) can serve as a basis for further research related to assessing the state of society in conditions of negative information impact and identifying critical conditions.

We will introduce the concept of "information panic" as Panicdemia, by which we mean panic in society generated by information sources (media, social networks, and other types of human communications) and leading to negative consequences for the entire society. The nature of information panic is a fear of prolonged action, generated and amplified by the media and human communications throughout society. Panic is based on fear associated with certain events (for example, a pandemic), but it is the informational impact that turns the fears of individual individuals into information panic at the level of the entire society.

Note that information panic should be distinguished from panic, which, for example, occurs in a certain region during natural disasters or during emergencies in a certain place at a certain time. In this situation, the media and social networks, of course, spread information negatively, but it is "justified", so to speak, a real threat to human lives at a given time in a particular region or place. Thus, in such cases, as a rule, panic occurs only among those who are involved in the specified situation. Note that for individual individuals with a weak information filter, negative information about such an event can also cause fears and lead to panic, but this is the psycho-specificity of these individuals and does not apply to society as a whole.

The emergence of information panic in society is a consequence of a complex information impact on individuals, in the process of which their fears are transformed into social psychosis - information panic. The scheme for increasing the volume of negative information in the process of communication and the impact of the media is shown in Fig. 1. When the total amount of negative information associated with a certain situation (for example, a pandemic) in society exceeds a certain critical value, it can be stated that society is in a state of information panic.

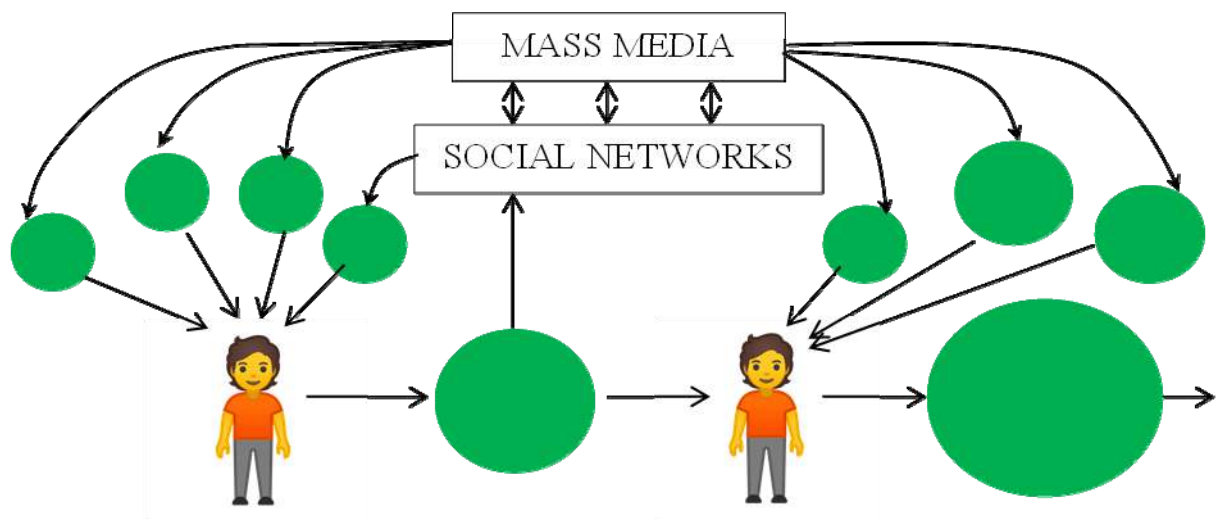


**Figure 1:** The influence of media and human communications on the formation of information panic

Such an increase in negative information, the formation of general fear, and, as a consequence, information panic can be carried out according to various scenarios (Fig. 2).



In the figure, the critical level is set in a conditional volume. The zero points is the amount of negative information that is observed in the normal state of society.

Scenario 1 - negative information in society gradually increased and at the same rate returned to its original value.

Scenario 2 - there is a slow growth of negative information, which may further go beyond the critical value.

Scenario 3 - the amount of negative information grows rapidly, reaching a critical level, and at this peak point it also rapidly decreases.

Scenario 4 - negative energy grows with acceleration, rapidly passing over the critical level, and information panic has arisen in society.

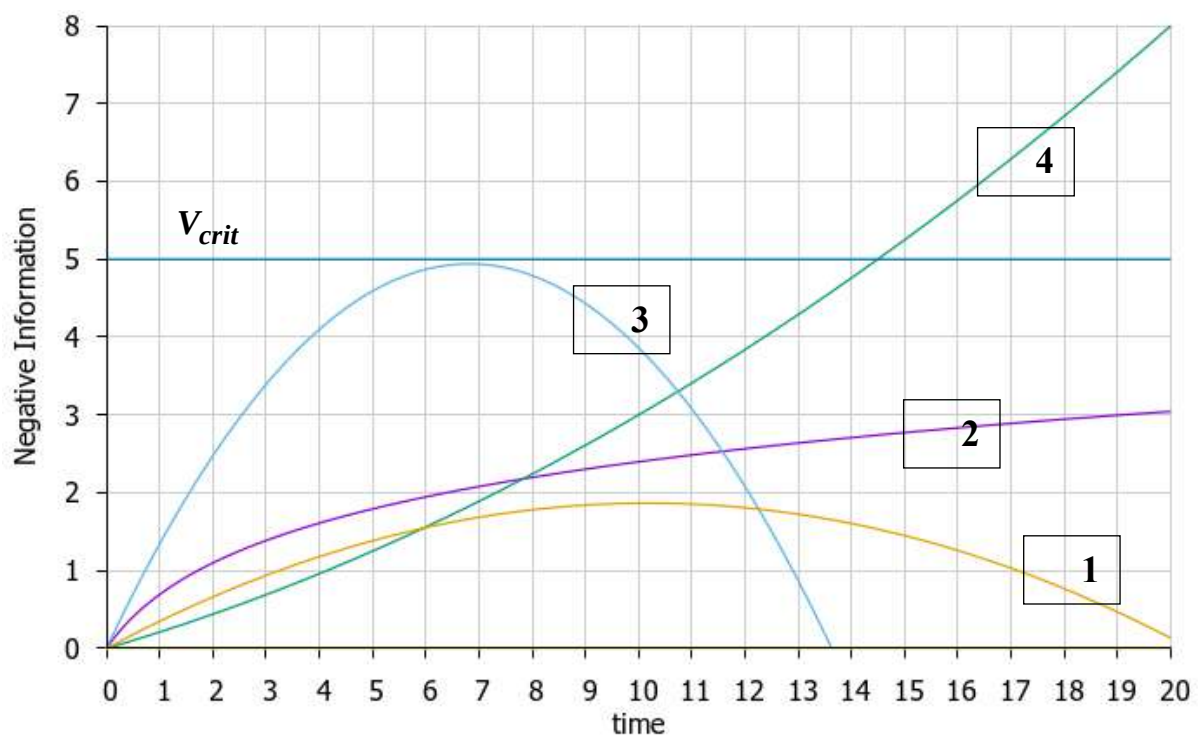


**Figure 2:** Scenarios for the development of information panic (negative information is presented in a conditional volume)

What are the consequences of information panic?

How do you assess the power of information panic?

Since information panic is the state of society or society as a whole, an indicator is needed that would assess this change in the state of society under the influence of information negativity. At the same time, the essence of information panic should play the role of a black box - this study does not consider the processes of its formation, but examines the result of the impact, which allows us to conclude about its presence and strength.

The most suitable for this purpose is the use of entropy, which, having initially received its place in thermodynamics within the framework of the second principle, gradually began to be used in other aspects and for systems of different nature. Entropy in thermodynamics characterizes the uselessness of energy, and it is the change in entropy that gives an idea of the state of the system under consideration.

In [17, 18], the main provisions of the entropy theory of organization were presented, within the framework of which the connection of entropy as a measure of energy uselessness with Shannon's informational entropy was substantiated. Since the organization is a mini-society, some of the results [17, 18] can be developed and extended to society in a broad sense to solve the problem under consideration.



So, by the entropy of society, we mean the measure of the uselessness of its energy. The entropy concept is based on the universality of the energy balance, which can be represented as:

$$U = A + \Delta U + Q = A + \Delta U + T \cdot S, \quad (1)$$

- where is the total energy of society; - energy gain; A - work; Q - heat, which is formed under the influence of temperature T and S entropy (energy entropy).

From the standpoint of the entropy concept, we define the following categories for society:

- "energy" is the aggregate resources of society (goods, estimated in monetary terms);
- "work" - performance of the main function of society for the reproduction of goods;
- "energy gain" - an increase in benefits as a result of the performance of the main function of society;
- "heat" - energy consumption for maintaining systemic connections in society (according to [20] - payment for the structure);
- "temperature" is a relative indicator of the state of society in comparison with a certain ideal state;
- "entropy" - an indicator of the state of society in the context of its energy losses (dissipation) in the process of energy turnover.

So, the main function of society is to ensure the reproduction and growth of benefits - the energy of society.

The temperature of a society reflects its state in comparison with a certain ideal state, for which we take the following:

- complete certainty (society reliably knows the value of future benefits, which are formed as a result of its main function);
- maximum possible energy efficiency - an increase in benefits as a result of the performance of the main function by society.

Therefore, as the temperature of society, we take the measure of its energy efficiency according to

[17]:  $\eta = \frac{U + \Delta U}{U}$  concerning information entropy, which characterizes order and certainty:

$$T = \frac{\mu}{H} = \frac{U + \Delta U}{U \cdot \eta^{id} \cdot H}, \quad (2)$$

$$\mu = \frac{\eta}{\eta^{id}}, \quad H = - \sum_{k=1}^K p(A_k) \cdot \ln(p(A_k)), \quad (3)$$

where  $A_k$  are the options for the state of society from the point of view of obtaining benefits,  $p(A_k)$  are the probabilities of these states;  $\mu$  reflects the relative level of efficiency in comparison with the "ideal" level  $\eta^{id}$  (for example, the level of increase in benefits in economically successful and developed countries). reflect the reproduction and growth of goods, that is, options.

Expression of entropy, taking into account the designations following [18]:

$$S = \frac{(U - E^{in}) \cdot U \cdot \eta^{id} \cdot H}{U + E^{in} - E^{ex}}, \quad (4)$$

where  $A + \Delta U = E^{in} - E^{ex}$  is the difference between the expended  $E^{ex}$  and received  $E^{in}$  benefits.

What happens to the entropy components in the process of increasing information panic? Analyzing the situation in the global society, starting with the spring 2020 lockdown related to COVID-19, the following conclusions can be drawn:

1) information entropy  $H$  is growing rapidly since variants  $A_k$  with a sufficiently high probability cease to prevail in their structure. In this case, the number of options becomes larger, and their probabilities tend to equalize;

2)  $E^{ex}$  begins to decrease: the business has already partially suffered losses during the pandemic, therefore, it is reducing operating costs, also carefully investing in development, many projects are frozen or their scale is being revised. Consumers are also holding on to their savings and spending less on a traditional set of goods, especially when a significant portion of the population works remotely;

3)  $E^{in}$  also decreases: business income decreases, therefore, accordingly, incomes of individual individuals associated with this business decrease.

Summing up, we can summarize the following: with the growth of information panic, the uncertainty of the future benefits of society grows, its spending and income decrease, that is, the energy stops circulating, providing a certain level of efficiency, and acquires the property of "uselessness".

So, with a decrease  $E^{ex}$ , there is a decrease  $E^{in}$ , and,  $\Delta U$  consequently, the efficiency of the benefits of society decreases  $\mu$ . Thus, in the expression of the temperature of society (2), the numerator decreases, and the denominator, taking into account the growth of information entropy  $H$ , increases. This leads to an increase in the temperature of society  $T$ . In this case, entropy (4) also grows.

Thus, to assess information panic in society, it is proposed to use the change in entropy  $\Delta S$ , namely, the rate of its growth in comparison with the base level, which is the essence of the information panic indicator:

$$I_{panic} = \frac{\Delta S}{S(t_0)} = \frac{S(t_i) - S(t_0)}{S(t_0)}, i=1,2,3.. \quad (5)$$

where  $t_i$  these are the time points for measuring the entropy,  $t_0$  is the initial (base) moment. When  $\Delta S > 0$  the entropy grows,  $I_{panic} > 0$ . Accordingly,  $I_{panic}$  at the entropy decreases, and. The information panic indicator demonstrates how much the entropy of the society has changed under the influence of negative information impact, leading to information panic.

In the normal state of society, the entropy  $S$  should gradually decrease due to a decrease in the information entropy  $H$  and an increase in the efficiency of the benefits of this society. In an extreme case, the entropy  $S$  can remain at the same level, its insignificant fluctuations are permissible. Accordingly, in the normal state of society  $I_{panic} \leq 0$ , and a positive value  $I_{panic} > 0$  indicates the presence of information panic in society.

The question arises - which level  $I_{panic}$  is critical, and what happens when this indicator goes beyond this critical value? As you know, the growth of entropy accompanies the appearance and growth of chaos in the system, therefore, it is obvious that the growth of entropy in society also indicates the approaching chaos (Fig. 3). Therefore  $I_{panic}$ , it is a kind and indicator of chaos in society under the influence of information panic. Note that determining the critical level requires separate studies based on the processing of a significant amount of statistical data, which can be done in the future. In any case, if,  $I_{panic} \rightarrow I_{panic}^{crit}$  then chaos is brewing in society.

The chain of events in Fig. 3 can be interpreted as a Markov chain. Note that the states "Growth of negative information" and "Reduction of negative information" can be represented in more detail in the form of several events with different levels of negative information and  $I_{panic}$ . This chain can serve as the basis for the Markov decision process, based on which modeling decision making in situations where outcomes are partly random and partly under the control of a decision-maker is carried out. The role of the latter, in this case, is played by the media, social networks, and those who can influence them. Having probabilistic assessments of transitions from state to state and their changes as a result of the influence of a decision-maker, one can assess the state of society after a certain period.

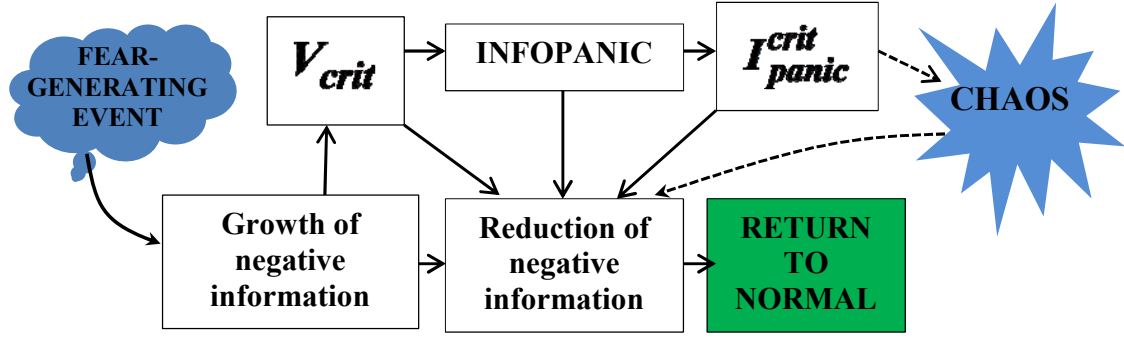


**Figure 3:** The logical relationship of various states of society from the position of information panic

Let us consider, using a design example, how an increase in informational entropy  $H$  and a decrease  $E^{ex}$ ,  $E^{in}$  affect the level  $I_{panic}$ .

An increase in information entropy  $H$  under the influence of information panic can occur both linearly and according to a polynomial pattern or even exponential, therefore, consider the following options for the growth of  $H$ :

$$H = a + b \cdot t, \quad (6)$$

$$H = a \cdot t^2 + b \cdot t + c, \quad (7)$$

$$H = d \cdot e^{f \cdot t} \quad (8)$$

where  $a, b, c, d, f$  are numerical coefficients,  $0 < d, f < 1$ . A decrease  $E^{ex}, E^{in}$  can also occur at different rates with similar  $H$  dependences, only in this case  $d > 0, f < 0$ . Figure 4 shows the variants of the dynamics of the entropy of society for different dynamics of the information entropy  $H$  with other fixed conditions. Figure 5 shows the dynamics of the entropy  $S$  with a simultaneous change in the information entropy  $H$  and energy flows,  $E^{ex}$ ,  $E^{in}$ .

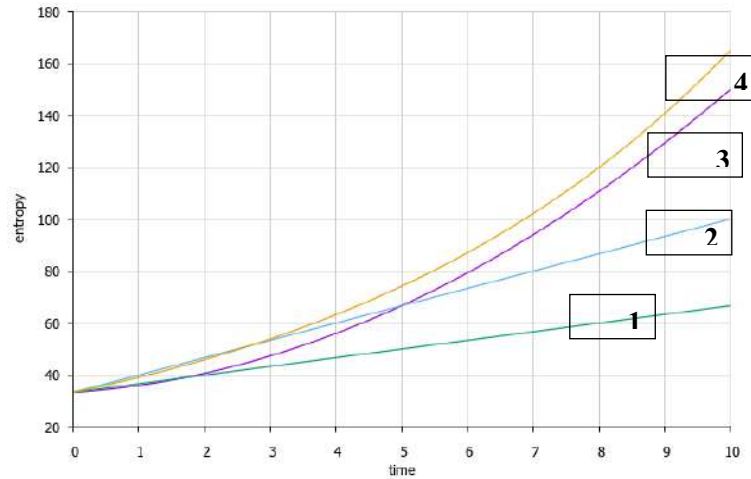


**Figure 4:** Dynamics of entropy  $S$  for different dynamics of information entropy  $H$ : 1 option;  $H = 1 + 0,1 \cdot t$  Option 2  $H = 1 + 0,2 \cdot t$ ; Option 3  $H = 1 + 0,05 \cdot t + 0,03 \cdot t^2$ ; Option 4  $H = e^{0,16t}$ .

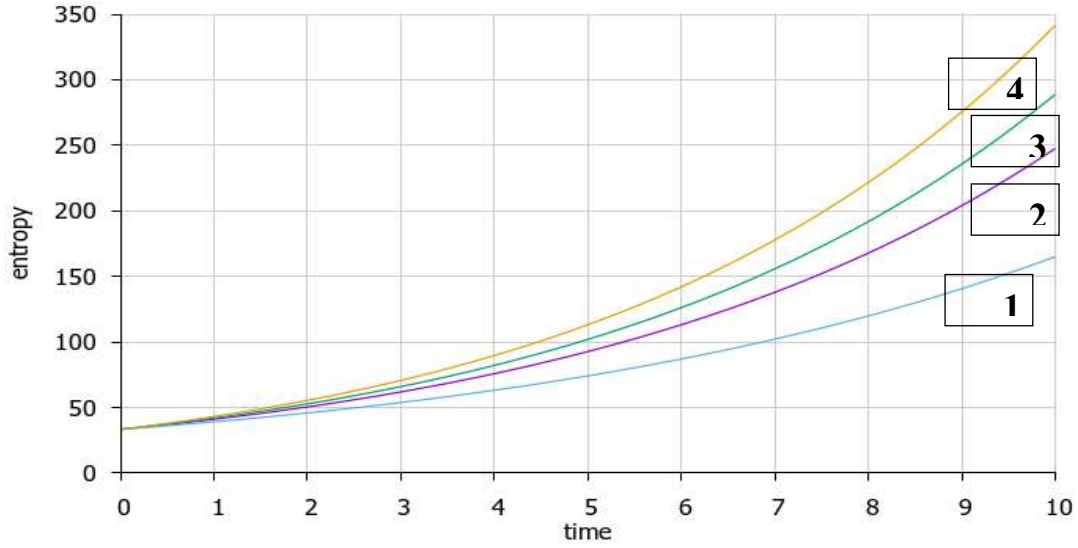


**Figure 5:** Dynamics of entropy  $S$  for different dynamics  $E^{ex}$ ,  $E^{in}$ , and information entropy  $H=e^{0,16t}$ : 1 option,  $E^{ex} = const; E^{in} = const$ , Option  $E^{ex} = 20-t; E^{in}=30-t$ ; Option 3  $E^{ex} = 20-1,5t; E^{in}=30-1,5t$ ; Option 4  $E^{ex} = 20-1,8t; E^{in}=30-2t$

As you can see, only under the influence of the growth of information entropy there is a significant increase in the entropy of society (Fig. 4), but several times its growth becomes faster and more significant with a decrease in the incoming and outgoing energy of the society (Fig. 5).

The calculation of the information panic index  $I_{panic}$  was carried out for the 1st and 4th scenarios of the entropy dynamics shown in Fig. 5. The calculation results are shown in Fig. 6.

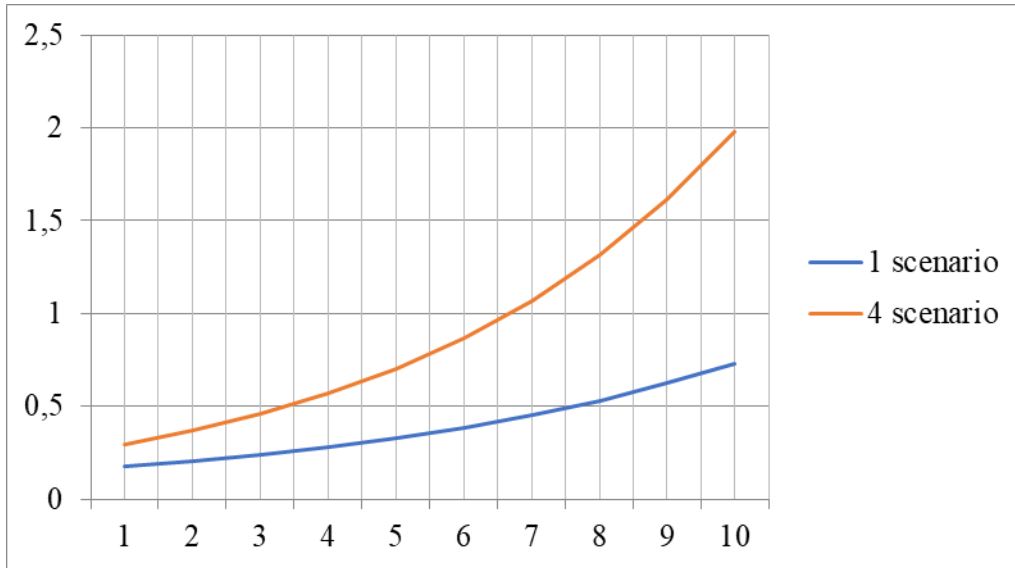


**Figure 6:** Dynamics  $I_{panic}$  for different dynamics of entropy.

As you can see, the information panic index for the 4th scenario increases significantly. Even if we assume that the critical value  $I_{panic}^{crit} = 1,5$  (that is, a little more than a three-fold increase compared to the initial level), then in the fourth scenario, Panicdemia definitely sets in.

## 2. Acknowledgments

In this paper, it is proposed to use the entropy concept to identify and analyze information panics in society. Under information panic, it is proposed to understand panic in society generated by information sources (media, social networks, and other types of human communications) and leading to negative consequences for the entire society.

Entropy as a measure of the uselessness of energy is adapted to assess the state of society in the process of increasing negative information impact. Introduced into consideration an indicator - an indicator of information panic, based on the change in entropy.

The influence of various dynamics of information entropy and energy flows of society (movement and growth of goods) on entropy and the indicator of information panic is analyzed using a calculated example in conventional figures.

The main states of society in the process of entropy growth are identified, their relationship is presented in the form of a Markov chain, the initial event of which is an event that generates fear in society; the final event with the endless growth of information panic is chaos.

The results obtained are the basis for further research on the impact of negative information on society and identifying ways to return to its normal state.

## 3. References

- [1] D. Bushuyev. Immune memory as a management tool enterprise development program. *Management of Development of Complex Systems*, 25, (2016): 11 – 16
- [2] S. Bushuyev, D. Bushuyev, V. Bushuyeva. Project management during Infodemic of the Covid-19 Pandemic. *Innovative Technologies and Scientific Solutions for Industries*. (2020): 13-21. <https://doi.org/10.30837/2522-9818.2020.12.013>
- [3] D. Bahr, E. Passerini,. Statistical Mechanics of Opinion Formation and Collective Behavior: Micro-Sociology. *Journal of Mathematical Sociology* 23, (1998): 1–27,
- [4] N. Sharma, S. Acharya, (2019): The Social Entropy and Social Ecology.
- [5] M. O'Connor, Entropy, structure, and organisational change, *Ecological Economics*, Volume 3, Issue 2, (1991): 95-122, [https://doi.org/10.1016/0921-8009\(91\)90012-4](https://doi.org/10.1016/0921-8009(91)90012-4) .
- [6] N. Petric. Application of the thermodynamic theory to social events. *International Review of Modern Sociology*, 21(1), (1991): 121-129. Retrieved February 18, 2021, from <http://www.jstor.org/stable/41420991>
- [7] K. Bailey, Social Entropy Theory: An overview. *Systems Practice* 3, (1990): 365–382 <https://doi.org/10.1007/BF01063441>
- [8] J. Needham. Evolution and Thermodynamics: A Paradox with Social Significance. *Science & Society*, 6(4), (1942): 352-375. Retrieved February 18, 2021, from <http://www.jstor.org/stable/40399498>
- [9] J. Moore. Entropy in a Social Science Context 26, (2006): <https://ssrn.com/abstract=886162> or <http://dx.doi.org/10.2139/ssrn.886162>
- [10] R. Fistola, C. Gargiulo, R. La Rocca.. Rethinking vulnerability in city-systems: A methodological proposal to assess "urban entropy". *Environmental impact assessment review*, 85, (2020) 106464. <https://doi.org/10.1016/j.eiar.2020.106464>
- [11] A. Mayer, R. Donovan, C. Pawlowski. Information and entropy theory for the sustainability of coupled human and natural systems. *Ecology and Society* 19(3), (2014): 11. <http://dx.doi.org/10.5751/ES-06626-190311>
- [12] Y. Neuman, D. Vilenchik, Y. Cohen. From physical to social interactions: The relative entropy model. *Sci Rep* 10, (2020): 1565. <https://doi.org/10.1038/s41598-020-58549-8>
- [13] V. Netto, J. Meirelles, F. Ribeiro. Cities and Entropy: Assessing Urban Sustainability as a Problem of Coordination. In C. Binder, R. Wyss, & E. Massaro (Eds.), *Sustainability Assessment of Urban Systems* (2020): 438-459. Cambridge: Cambridge University Press.
- [14] T. Mavrofides, A. Kameas, D. Papageorgiou, A. Los On the Entropy of Social Systems: A Revision of the Concepts of Entropy and Energy in the Social Context. *Systems Research and Behavioral Science* Volume 28, Issue 4, (2011): 353-368 <https://doi.org/10.1002/sres.1084>

- [15] A. Bondar, S. Bushuyev, S. Onyshchenko, H. Tanaka. Entropy Paradigm of Project-Oriented Organizations Management // Proceedings of the 1st International Workshop IT Project Management (ITPM 2020) Volume 1. Lviv, Ukraine, February 18-20, (2020); CEUR Workshop Proceedings (CEUR-WS.org), (2020): 233-243 <http://ceur-ws.org/Vol-2565/paper20.pdf>
- [16] Bondar, A., Onyshchenko, S., Vishnevskaya, O., Vishnevskiy, D., Glovatska, S., & Zelenskyi, A. (2020). Constructing and investigating a model of the energy entropy dynamics of organizations. Eastern-European Journal of Enterprise Technologies, 3(3), 105. <https://doi.org/10.15587/1729-4061.2020.206254>
- [17] A. Bondar, N. Bushuyeva, S. Bushuyev, S. Onyshchenko. Modelling of Creation Organisational Energy Sergey -Entropy, (2020): IEEE 15th International Conference on Computer Sciences and Information Technologies (CSIT), Zbarazh, Ukraine, (2020): 141-145, <https://doi.org/10.1109/CSIT49958.2020.9321997>.
- [18] A. Bondar, S. Bushuyev, V. Bushuieva, N. Bushuyeva, S. Onyshchenko. Action-Entropy Approach to Modeling of 'Infodemic-Pandemic' System on the COVID - 19 Cases. In: Shakhovska N., Medykovsky M.O. (eds) Advances in Intelligent Systems and Computing V. CSIT 2020. Advances in Intelligent Systems and Computing, vol 1293. (2021): Springer, Cham. [https://doi.org/10.1007/978-3-030-63270-0\\_61](https://doi.org/10.1007/978-3-030-63270-0_61)
- [19] A. Bondar, S. Onyshchenko. Experimental studies of a model for optimizing the portfolio of a project-oriented organization based on the entropy concept. Innovative Technologies and Scientific Solutions for Industries, 4 (14), (2020) 21-30. <https://doi.org/10.30837/ITSSI.2020.14.021>
- [20] J. Stepanić, G. Sabol, S. Žebec, Describing social systems using social free energy and social entropy, Kybernetes, Vol. 34 No. 6. (2005);, 857-868. <https://doi.org/10.1108/03684920510595535>

# Parameters identification for fractional-fractal model of filtration-consolidation using GPU

Vsevolod Bohaienko<sup>a</sup>, Anatolij Gladky<sup>a</sup>

<sup>a</sup> V.M. Glushkov Institute of Cybernetics of NAS of Ukraine, Glushkov ave., 40, Kyiv, 03187, Ukraine

## Abstract

The paper considers some computational problems arising in the important practical field of the determination of safe operation conditions of engineering facilities that pollute soils and groundwater. In the case of complex geological and hydrological conditions, such problems are widely considered using mathematical modeling of deformation and consolidation processes in water-saturated soils, particularly, in the foundations of hydraulic structures. To simulate the dynamics of such processes, we use a fractional-fractal approach that allows considering temporal non-locality of transfer processes in media of fractal structure. The used one-dimensional differential model contains a non-local Caputo derivative with respect to the time variable and a local fractal derivative with respect to the space variable. Some of model's parameters, namely the orders of fractional derivatives, can only be determined fitting them to the measured data related to the state of a process. We propose to use particle swarm optimization algorithm to perform an identification of fractional derivatives' orders and present the results of its testing on noised subsets of direct problem solutions. In this context, we have determined that the order of space-fractal derivative is restored with a relative error of not more than 1% while the order of time-fractional derivative is restored with higher errors of not more than 10%. The lowest number of observation points that ensures stable restoration of the orders was equal to 25. As high computational complexity is combined with highly independent computational blocks while applying evolutionary optimization algorithms to the problems of differential models' parameters identification, we implemented the proposed algorithm on graphical processing units (GPU) using OpenCL framework and on multi-threaded systems using OpenMP. The results of performance testing showed up to 4-times lower GPU execution time compared to the case of multi-threaded execution on 6 cores of central processing unit (CPU).

## Keywords

filtration-consolidation processes, fractional-fractal model, parameters identification, particle swarm optimization, GPU, OpenCL.

## 1. Introduction

The determination of the conditions for safe functioning of numerous engineering facilities that pollute soils and groundwater are among the most important and relevant in the connection with environment protection issues. This makes urgent the development of effective and reliable methods for mathematical modeling of deformation and consolidation dynamics in saturated soils, particularly, in the foundations of hydraulic structures. Such methods are usually based on the solution of initial-boundary value problems for the corresponding systems of partial differential equations (e.g. [1-3]). Recently, in order to take into account the effects of temporal and spatial correlations, a series of models have been developed by introducing into them local and non-local operators of fractional-order differentiation (e.g. [4-6]).

---

CMIS-2021: The Fourth International Workshop on Computer Modeling and Intelligent Systems, April 27, 2021, Zaporizhzhia, Ukraine

EMAIL: sevab@ukr.net (V. Bohaienko); gladky@ukr.net (A. Gladky)

ORCID: 0000-0002-3317-9022 (V. Bohaienko); 0000-0002-3839-9550 (A. Gladky)



© 2021 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

In this context, we continue the studies presented in [6] on the simulation of anomalous processes of filtration-consolidation in water-saturated soils under the influence of salt transfer. The model developed in [6] uses the fractional-fractal approach [7-10] that enables modeling of temporal non-locality of processes in soils of fractal structure. Comparing to the previously developed mass transfer models within such an approach this model widens its application taking into consideration soil compaction and chemical osmosis.

Practical usage of fractional-order differential models to predict the dynamics of mass transfer and soil compaction requires identification of their parameters. However, the orders of fractional derivatives usually do not have technical means of measurement. The only approach that can be used to determine their values is to select them the way to make the model best describe available measurements.

In general, two approaches used to identify the parameters of fractional-differential models can be singled out. Within the first one, corresponding inverse problems are solved analytically or using numerical-analytical methods (e.g. [11, 12]). Such an approach is usually applied to relatively simple, mostly one-dimensional models. Despite some problems of restoration of fractional derivative orders are solved by least-squares techniques [13] or the Tikhonov method [14], the second approach that is based on metaheuristic optimization techniques (e.g. [15-17]) have less limitation on its usage. One of such algorithms that were efficiently applied [17] to some models of mass transfer processes in soils is the particle swarm optimization (PSO) algorithm [18]. Regarding most metaheuristic optimization algorithms it should also be noted that data parallelism can be easily exploited in them.

Hence, in this paper we study the accuracy and performance of the PSO algorithm applying it to solve parameters identification problem for the fractal-fractional model of filtration-consolidation proposed recently in [6]. We also consider PSO algorithm's implementation on graphical processors (GPU) and study its performance comparing it with the case of multi-threaded implementation on CPU.

## 2. Direct problem

Considering the non-local in time isothermal filtration-consolidation process in a soil of fractal structure saturated with a salt solution, we study the following mathematical model [6]:

$$D_t^{(\beta)} H = \frac{\partial}{\partial x^\alpha} \left( \frac{\partial}{\partial x^\alpha} (C_v H - \mu C) \right) \quad (1)$$

$$\sigma D_t^{(\beta)} C = d_* \frac{\partial}{\partial x^\alpha} \left( \frac{\partial C}{\partial x^\alpha} \right) + \frac{\partial}{\partial x^\alpha} (kH - \nu C) \cdot \frac{\partial C}{\partial x^\alpha} \quad (2)$$

where  $H(x,t)$  is the water head ( $m$ ),  $C(x,t)$  is the concentration of salts in the liquid phase ( $kg/m^3$ ),  $k$  is the filtration coefficient in fractal dimension ( $m^{2\alpha-1}/s^\beta$ ),  $\nu$  is the coefficient of chemical osmosis in fractal dimension ( $\frac{m^{3+2\alpha}}{kg \cdot s^\beta}$ ),  $d_*$  is the coefficient of convective diffusion in

fractal dimension ( $m^{2\alpha}/s^\beta$ ),  $C_v$  is the consolidation coefficient in fractal dimension ( $m^{2\alpha}/s^\beta$ ),

$\mu = \frac{\nu C_v}{k}$ ,  $\sigma$  is the porosity of a medium,  $D_t^{(\beta)} f(t) = \frac{1}{\Gamma(1-\beta)} \frac{\partial}{\partial t} \left( \int_0^t (t-\tau)^{-\beta} f(\tau) d\tau - \tau^{-\beta} f(0) \right)$  is the

regularized fractional Caputo derivative of the order  $\beta$ ,  $0 < \beta \leq 1$  with respect to the variable  $t$

[19-21],  $\frac{\partial}{\partial x^\alpha}$  is the operator of fractal derivative [7-9], which can be represented under the known

conditions as  $\frac{\partial}{\partial x^\alpha} f(x) = \frac{d}{dx} f(x) \frac{1}{\alpha x^{\alpha-1}}$ ,  $\alpha > 0.5$  is the fractal dimension.

The equations (1),(2) contains coefficients in fractal dimension in space and time. To make them consistent with measurable values, we, following [22], introduce parameters  $\sigma_t(s)$ ,  $\sigma_x(m)$  of time



and space dimensions and represent fractal dimensional filtration coefficient as  $k = \bar{k} \frac{\sigma_x^{2-2\alpha}}{\sigma_t^{1-\beta}}$  where  $\bar{k}$

is the measurable value of the coefficient. The same is performed for the other coefficients in fractal dimension. While the choice of the values of dimensional parameters  $\sigma_t$ ,  $\sigma_x$  strongly influence simulation results, there are significant difficulties of their direct determination. Thus, we consider the model (1),(2) as semi-empirical and set  $\sigma_t = \sigma_x = 1$  (one of the cases studied in [22]) further focusing on the identification of fractional derivatives' orders only.

Let us note, that from equations (1), (2) when  $\alpha \rightarrow 1$  we obtain a system of equations [23] of the corresponding fractional-differential model that does not consider fractal properties of a soil. When  $\alpha, \beta \rightarrow 1$  we obtain the integer-order classical model [1, 2].

Using in (1), (2) the representation of the fractal derivative operator on the base of integer-order derivative [7-9] we obtain the system of equations in the following form [6]:

$$D_t^{(\beta)} H = C_v \left[ s_\alpha(x) \frac{\partial^2 H}{\partial x^2} + r_\alpha(x) \frac{\partial H}{\partial x} \right] - \mu \left[ s_\alpha(x) \frac{\partial^2 C}{\partial x^2} + r_\alpha(x) \frac{\partial C}{\partial x} \right] \quad (3)$$

$$\sigma D_t^{(\beta)} C = d_* \left[ s_\alpha(x) \frac{\partial^2 C}{\partial x^2} + r_\alpha(x) \frac{\partial C}{\partial x} \right] + s_\alpha(x) \frac{\partial v}{\partial x} \cdot \frac{\partial C}{\partial x} \quad (4)$$

where

$$v(x, t) = kH(x, t) - vC(x, t), r_\alpha(x) = \frac{1-\alpha}{\alpha^2} x^{1-2\alpha}, s_\alpha(x) = \frac{1}{\alpha^2} x^{2(1-\alpha)}.$$

To describe the dynamics of the considered process in the domain  $\Omega = \{(x, t): 0 < x < l, t > 0\}$  with permeable boundaries we complement the equations (3), (4) with boundary conditions

$$H(0, t) = 0, \quad H(l, t) = 0, \quad H(x, 0) = H_0 \quad (5)$$

$$C(0, t) = C_0, \quad C(l, t) = 0, \quad C(x, 0) = 0 \quad (6)$$

where  $H_0$  is the initial value of water head,  $C_0$  is the specific value of salts concentration at the inlet of filtration flow.

### 3. Numerical method for the direct problem

We numerically solve the boundary value problem (3)-(6) using a finite-difference technique briefly described below.

In the grid domain  $\omega_{hr} = \{(x_i, t_j): x_i = ih \ (i = \overline{0, m+1}), t_j = j\tau \ (j = \overline{0, n})\}$  where  $h, \tau$  are the grid steps with respect to the spatial variable and time the considered problem can be discretized using the linearized Crank–Nicholson scheme written using the notations from [24] as [6]

$$\sigma \Delta_t^{(\beta)} C = 0.5d_* \left[ s_\alpha(\hat{C}_{\bar{x}\bar{x}} + C_{\bar{x}\bar{x}}) + r_\alpha(\hat{C}_x + C_x) \right] + 0.5s_\alpha v_x (\hat{C}_x + C_x) \quad (7)$$

$$\Delta_t^{(\beta)} H = 0.5C_v \left[ s_\alpha(\hat{H}_{\bar{x}\bar{x}} + H_{\bar{x}\bar{x}}) + r_\alpha(\hat{H}_x + H_x) \right] - 0.5\mu \left[ s_\alpha(\hat{C}_{\bar{x}\bar{x}} + C_{\bar{x}\bar{x}}) + r_\alpha(\hat{C}_x + C_x) \right] \quad (8)$$

where  $\Delta_t^{(\beta)} u$  is the discrete analogue of the fractional derivative  $D_t^{(\beta)} u$  defined as

$$\Delta_{t_{j+1}}^{(\beta)} u \approx \frac{u^{j+1} - u^j}{\tau^\beta \Gamma(2-\beta)} + \sum_{v=0}^{j-1} \omega_v^{(j)} \frac{u^{v+1} - u^v}{\tau}, \quad \omega_v^{(j)} = \frac{\tau^{1-\beta}}{\Gamma(2-\beta)} [(j-v+1)^{1-\beta} - (j-v)^{1-\beta}], \quad (9)$$

and  $\Gamma(z)$  is the Euler's gamma function.

Substituting (9) into (7), (8) we reduce the solution of the problem (3)-(6) on the  $(j+1)$ -th time step to the solution of the systems of linear algebraic equations

$$A_i^j C_{i-1}^{j+1} - S_i^j C_i^{j+1} + B_i^j C_{i+1}^{j+1} = F_i^j \quad (i = \overline{1, m}; j = \overline{0, n}), \quad (10)$$

$$\tilde{A}_i^j H_{i-1}^{j+1} - \tilde{S}_i^j H_i^{j+1} + \tilde{B}_i^j H_{i+1}^{j+1} = \tilde{F}_i^j \quad (i = \overline{1, m}; j = \overline{0, n}), \quad (11)$$

$$C_0^{j+1} = C_0, \quad C_{m+1}^{j+1} = 0, \quad C_i^0 = 0 \quad (i = \overline{0, m+1}; j = \overline{0, n}), \quad (12)$$

$$H_0^{j+1} = 0, \quad H_{m+1}^{j+1} = 0, \quad H_i^0 = H_0 \quad (i = \overline{0, m+1}; \quad j = \overline{0, n}) \quad (13)$$

where

$$\begin{aligned} A_i^j &= \frac{0.5}{h} \left[ d_* \left( \frac{s_\alpha^i}{h} - \frac{r_\alpha^i}{2} \right) - \frac{s_\alpha^i}{4h} (\nu_{i+1}^j - \nu_{i-1}^j) \right], \quad B_i^j = \frac{0.5}{h} \left[ d_* \left( \frac{s_\alpha^i}{h} + \frac{r_\alpha^i}{2} \right) + \frac{s_\alpha^i}{4h} (\nu_{i+1}^j - \nu_{i-1}^j) \right], \\ S_i^j &= A_i^j + B_i^j + \frac{\sigma}{\tau^\beta \Gamma(2-\beta)}, \\ F_i^j &= \sigma \left[ \sum_{v=0}^{j-1} \omega_v^{(j)} \frac{C_i^{v+1} - C_i^v}{\tau} - \frac{C_i^j}{\tau^\beta \Gamma(2-\beta)} \right] - \frac{0.5d_*}{h} \left[ \frac{s_\alpha^i}{h} (C_{i-1}^j - 2C_i^j + C_{i+1}^j) + \frac{r_\alpha^i}{2} (C_{i+1}^j - C_{i-1}^j) \right] - \\ &\quad - \frac{0.5s_\alpha^i}{4h^2} (\nu_{i+1}^j - \nu_{i-1}^j) (C_{i+1}^j - C_{i-1}^j), \\ \tilde{A}_i^j &= \frac{0.5C_v}{h} \left( \frac{s_\alpha^i}{h} - \frac{r_\alpha^i}{2} \right), \quad \tilde{B}_i^j = \frac{0.5C_v}{h} \left( \frac{s_\alpha^i}{h} + \frac{r_\alpha^i}{2} \right), \quad \tilde{S}_i^j = \tilde{A}_i^j + \tilde{B}_i^j + \frac{1}{\tau^\beta \Gamma(2-\beta)}, \\ \tilde{F}_i^j &= \sum_{v=0}^{j-1} \omega_v^{(j)} \frac{H_i^{v+1} - H_i^v}{\tau} - \frac{H_i^j}{\tau^\beta \Gamma(2-\beta)} - \frac{0.5C_v}{h} \left[ \frac{s_\alpha^i}{h} (H_{i-1}^j - 2H_i^j + H_{i+1}^j) + \frac{r_\alpha^i}{2} (H_{i+1}^j - H_{i-1}^j) \right] + \\ &\quad + \frac{0.5\mu}{h} \left[ \frac{s_\alpha^i}{h} (C_{i-1}^{j+1} - 2C_i^{j+1} + C_{i+1}^{j+1} + C_{i-1}^j - 2C_i^j + C_{i+1}^j) + \frac{r_\alpha^i}{2} (C_{i+1}^{j+1} - C_{i-1}^{j+1} + C_{i+1}^j - C_{i-1}^j) \right], \\ s_\alpha^i &= s_\alpha(x_i), \quad r_\alpha^i = r_\alpha(x_i), \quad \nu_i^j = kH_i^j - \nu C_i^j. \end{aligned}$$

Equations (10)-(13) can be effectively solved by the Thomas algorithm [24].

#### 4. Inverse problem and particle swarm optimization algorithm

Having the direct problem and numerical method stated, we further consider the inverse problem of identification of derivatives' orders  $\alpha, \beta$  in the model (3)-(6) on the base of some subset of possibly noised measurements. The inverse problem is posed the following way:

- assume there are  $N$  known values  $C_i, H_i, i=1, \dots, N$  of concentration and water head measured in the moments of time  $T_i$  in the depths  $x_i$ ;
- assuming that the values of other model's parameters are known, find such values  $\bar{\alpha}, \bar{\beta}$  of orders  $\alpha, \beta$  that minimize the following goal function:

$$F(\hat{\alpha}, \hat{\beta}) = \sum_{i=1}^N \left[ \left( H(x_i, T_i, \hat{\alpha}, \hat{\beta}) - H_i \right)^2 + \left( C(x_i, T_i, \hat{\alpha}, \hat{\beta}) - C_i \right)^2 \right] \quad (14)$$

where  $C(x, t, \hat{\alpha}, \hat{\beta}), H(x, t, \hat{\alpha}, \hat{\beta})$  are the solutions of the direct problem (3)-(6) with  $\alpha = \hat{\alpha}, \beta = \hat{\beta}$  in the point  $x$  at the time  $t$ .

Taking into account the complexity of the inverse problem and the fact that the parameters to be identified are floating-point numbers, we propose to solve it by the particle swarm optimization (PSO) algorithm [18], which can be briefly described as follows:

- denote  $S$  as the number of particles in the swarm;  $\bar{x}_i = (\hat{\alpha}, \hat{\beta}), v_i$  as the coordinates and velocity of the particle  $i$ ;  $p_i$  as the coordinates of the particle  $i$  that corresponds to the best goal function (14) value obtained by it;  $g$  as the coordinates that corresponds to the best goal function value obtained by the swarm;
- introduce the parameters of the algorithm  $\omega, \varphi_p, \varphi_g$ ;
- at the initialization stage, the coordinates of the particles are generated randomly and velocities are set to zero. The values of the goal function are calculated for each particle along with the initial values of  $p_i$  and  $g$ .
- on the iteration  $j$  for the particle  $i$

- o generate random numbers  $r_p, r_g \in [0,1]$ ;
- o modify the velocity:  $v_i^{(j+1)} = \omega \cdot v_i^{(j)} + \varphi_p r_p (p_i^{(j)} - \bar{x}_i^{(j)}) + \varphi_g r_g (g^{(j)} - \bar{x}_i^{(j)})$ ;
- o modify particle's coordinates:  $\bar{x}_i^{(j+1)} = \bar{x}_i^{(j)} + v_i^{(j)}$ ;
- o calculate goal function value and modify  $p_i$  and  $g$ .

- Iteration process is finished when the given maximal number of iterations  $I_{\max}$  is reached, the zero best value of the goal function is achieved, or the difference between the best and the worst values of the goal function for the particles in the swarm becomes lower than the given constant.

In our case, the computation of direct problem's solutions is the most time-consuming part when identifying parameters values. As the changes in particles' states are independent, the corresponding direct problems can be efficiently solved on shared-memory parallel systems.

## 5. Parallel implementation

The following scheme is used to implement the solution of parameters' identification problem on GPU.

As the solution of direct problems while calculating goal function values forms the main part of algorithm's computational complexity, we choose it as the only part that is executed on GPU. We allocate a group of threads for each particle and, correspondingly, for a set of derivatives' orders for which the problem (3)-(6) has to be solved.

The main share of time while solving the direct problem is formed by the calculation of sums derived from the discretization (9) of the Caputo fractional derivative because the number of terms in them grows linearly with the increase of time step number. These sums must be calculated on each time step for each grid node, thus, within a thread block we use thread to grid node mapping. So, on each time step, at the first stage all threads compute the values of sums in the right parts of (10),(11) and then the threads with local ids 0 perform the solution of (10),(11). Solution on GPU is performed up to the moment of time  $\max_i T_i$  without interaction with CPU. Then the solutions for the time steps

$j: t_j \leq T_i < t_{j+1}$  are transferred into the memory of CPU and the values of  $H(x_i, T_i, \hat{\alpha}, \hat{\beta})$  and  $C(x_i, T_i, \hat{\alpha}, \hat{\beta})$  are determined using linear interpolation both in space and time.

In such a computational scheme, the amount of used GPU memory is  $S \cdot (m+2) \cdot \left\lceil \max_i T_i / \tau \right\rceil$  and the number of running GPU threads is  $S \cdot (m+2)$ . Assuming that the time spent on calculations on CPU can be neglected comparing with the time spent on GPU and recalling that the Thomas algorithm has the complexity order  $O(m)$ , the execution time of the presented algorithm is  $T(S, m, n) = k \cdot m \cdot \left\lfloor S \cdot (m+2) / N_c \right\rfloor \cdot n \cdot (n-1) / 2$  where  $k$  is the parameter of system's performance,  $N_c$  is the number of GPU cores,  $n$  is the number of time steps.

Further, we compare GPU implementation's performance with the performance of multi-threaded implementation in which for the same procedure CPU core to particle mapping is used.

## 6. Accuracy testing results

The described solution schemes for direct and inverse problems were implemented in C++ language. The source codes are publicly available through [https://github.com/sevaboh/cons\\_chem\\_osm](https://github.com/sevaboh/cons_chem_osm). To experimentally study the accuracy of the inverse problem's solution algorithm, we generate testing datasets as described below.

We fix the following values of physical parameters [6]:  $C_v = 0.34$ ,  $l = 25$ ,  $\mu = 0.00095$ ,  $\sigma = 0.38$ ,  $d = 0.02$ ,  $k = 0.01$ ,  $\nu = 2.8 \cdot 10^{-5}$ ,  $C_0 = 200$ ,  $H_0 = 10$ . With  $\tau = 0.01$  and fixed  $m, \alpha, \beta$ , the problem (3)-(6) is solved up to the time step for which  $T_e = 1$ . Then, the given per cent  $P$  of grid nodes that simulate the number of observation points are randomly selected as  $x_i$ . Setting all  $T_i = T_e$ ,

we further set  $H_i = H(x_i, T_e, \hat{\alpha}, \hat{\beta}) + r$ ,  $C_i = C(x_i, T_e, \hat{\alpha}, \hat{\beta}) + r$  where  $r$  is the random variable that is uniform in  $[-0.5R, 0.5R]$ ,  $R$  is the given noise level. The solution to the problem (3)–(6) obtained for  $T_e = 2$  and noised values in observation points for  $R = 2$ ,  $P = 0.5$ ,  $\beta = 0.6, \alpha = 0.8$  are depicted in Figure 1.

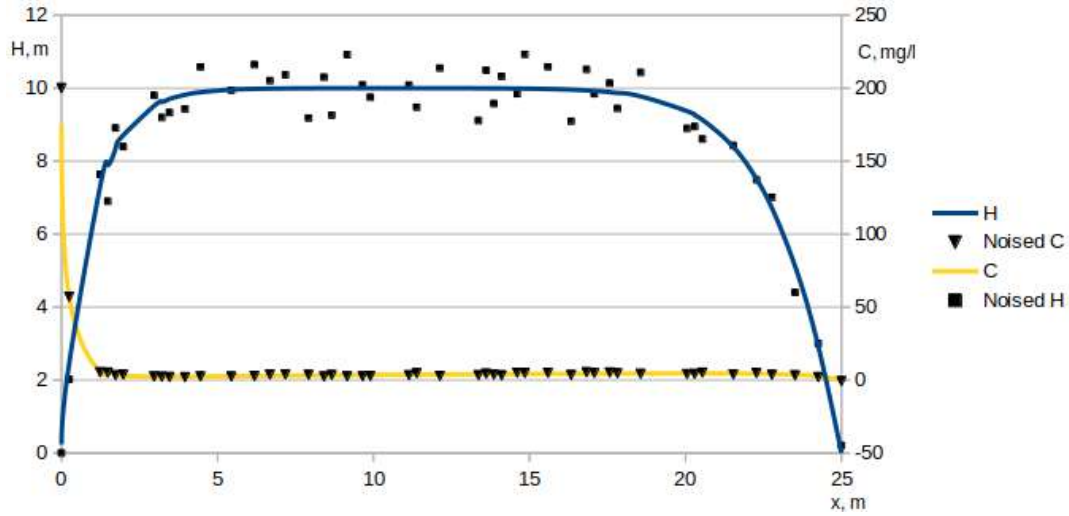


**Figure 1:** Solutions and noised values in observation points

Several series of computational experiments were performed for  $m = 100$  and different values of  $\alpha, \beta, P, R$ . In the first one with fixed  $\beta = 0.6, \alpha = 0.8$  and  $S = 10, \omega = \varphi_p = \varphi_g = 0.5$ , we tested the influence of noise level and number of observation points on the accuracy of fractional derivatives orders restoration. The values of  $\alpha, \beta$  were restricted to the range  $[0.5, 1]$ ;  $P$  was equal to 0.75, 0.5, 0.25, and 0.1;  $R$  was equal to 1, 2, and 3. We performed 10 runs of the algorithm with 30 iterations in each run.

The value of  $\alpha$  was in all conducted experiments restored with the maximal relative error of 0.65%. Standard deviation within the performed 10 runs was not more than 0.29% of the obtained average values of  $\alpha$ .

The relative errors for the case of  $\beta$  are given in Table. 1. Here the errors increase with the increase of noise level  $R$ . For  $P \geq 0.25$  (more than 25 observation points) changes in the errors are small but its decrease to 0.1 (10 observation points) up to 3 times decrease the accuracy. Standard deviation behave the similar way being 4% of the average values for  $P \geq 0.25$  and becoming up to 8% for  $P = 0.1$ .

**Table 1**  
Relative errors of the restoration of  $\beta$

| $R$ | $P = 0.75$ | $0.5$ | $0.25$ | $0.1$  |
|-----|------------|-------|--------|--------|
| 1   | 4.98%      | 1.85% | 4.36%  | 16.66% |
| 2   | 6.78%      | 6.86% | 7.10%  | 26.13% |
| 3   | 9.13%      | 8.74% | 10.16% | 34.16% |

To determine the influence of initial values of  $\alpha$  and  $\beta$  on the accuracy of their restoration we change  $\alpha$  in the range  $[0.55, 0.95]$  for the fixed  $\beta = 0.75$  and vice versa change  $\beta$  in the same range for  $\alpha = 0.75$ . The obtained relative errors are given for  $P = 0.5, R = 2$  in Tables 2 and 3. In these computational experiments,  $\alpha$  was as in the previous one restored with rather high accuracy (relative error not more than 0.6% in all cases). The error of  $\beta$  restoration being not more than 8.43% was higher when  $\beta$  tends to the edges of the range.

The last series of experiments with variable  $T_e$  showed that better accuracy of parameters identification can be achieved for higher values of  $T_e$ : relative error of  $\beta$  restoration with  $R = 2$ ,  $P = 0.5$  lowered from  $\sim 7\%$  for  $0.1 \leq T_e \leq 1$  to  $\sim 1.5\%$  for  $T_e = 2$ .

**Table 2**

Relative errors  $e_\alpha, e_\beta$  of  $\alpha$  and  $\beta$  restoration for the fixed  $\alpha = 0.75$

| $\beta$ | $e_\beta$ | $e_\alpha$ |
|---------|-----------|------------|
| 0.55    | 6.96%     | 0.30%      |
| 0.6     | 6.67%     | 0.28%      |
| 0.7     | 3.26%     | 0.25%      |
| 0.8     | 0.94%     | 0.32%      |
| 0.95    | 8.43%     | 0.60%      |

**Table 3**

Relative errors  $e_\alpha, e_\beta$  of  $\alpha$  and  $\beta$  restoration for the fixed  $\beta = 0.75$

| $\alpha$ | $e_\beta$ | $e_\alpha$ |
|----------|-----------|------------|
| 0.55     | 0.48%     | 0.18%      |
| 0.6      | 0.23%     | 0.16%      |
| 0.7      | 4.73%     | 0.22%      |
| 0.8      | 0.96%     | 0.43%      |
| 0.95     | 5.52%     | 0.55%      |

## 7. Solution time testing results

GPU implementation of the PSO algorithm was performed using the OpenCL framework while multi-threaded implementation used OpenMP.

Time spent on parameters' identification using GPU and multi-threaded implementations was measured solving the problem with  $\beta = 0.6, \alpha = 0.8, P = 0.5, R = 2$  on one node of SCIT-4 cluster of VM Glushkov Institute of cybernetics of NAS of Ukraine (NVidia RTX 2080 Ti GPU, 2 Intel(R) Xeon(R) Bronze 3104 CPUs, 1 CPU with 6 cores was allocated for OpenMP).

During the tests we changed the number  $m$  of nodes in the finite-difference grid and the number  $S$  of particles in PSO swarm. Average solution time among 10 runs is given in Table 4. Regarding the accuracy, it increased with the increase of the number of observation points as  $m$  increases with fixed  $P$  reaching the maximal relative error of 0.24% for  $m = 200$ .

**Table 4**

Average solution time, ms

| $S$ | $m = 50$ |       | $m = 100$ |       | $m = 200$ |        |
|-----|----------|-------|-----------|-------|-----------|--------|
|     | GPU      | CPU   | GPU       | CPU   | GPU       | CPU    |
| 10  | 10524    | 11819 | 19944     | 23586 | 38791     | 46907  |
| 20  | 10539    | 23162 | 20010     | 46407 | 38930     | 92550  |
| 30  | 10590    | 29163 | 20082     | 58477 | 39008     | 116521 |
| 40  | 10652    | 40834 | 20089     | 80818 | 39160     | 162622 |

As can be seen from the data in Table 4, execution time of GPU algorithm does not depend on the number of particles and linearly increases with the increase of grid size. This can be due to the non-complete allocation of GPU's computational resources. On the other hand, CPU execution time linearly increases both with the increase of  $m$  and  $S$  that proves the efficiency of GPU implementation, which in the conducted experiments gave up to 4-times decrease of solution time.

The efficiency of GPU implementation slightly increased with the increase of  $m$ : average solution time per one grid node per particle decreased from 21.1 ms for  $m = 50$  to 19.4 ms for  $m = 200$ . The corresponding indicator for CPU implementation did not change significantly with the change of  $m$  but decreased from 23.6 ms for  $S = 10$  to 20.4 ms for  $S = 40$ . Such behavior reflects the general heuristic that parallel implementation's efficiency increase with the increase of the volume of computations.

The series of computational experiments conducted for  $S = 10$ ,  $N = 100$  with variable  $T_e = 0.5, 1, 2, 3, 4$  confirmed quadratic dependency of execution time on the number of time steps (Figure 2). GPU implementation here allowed obtaining slower increase of solution time.

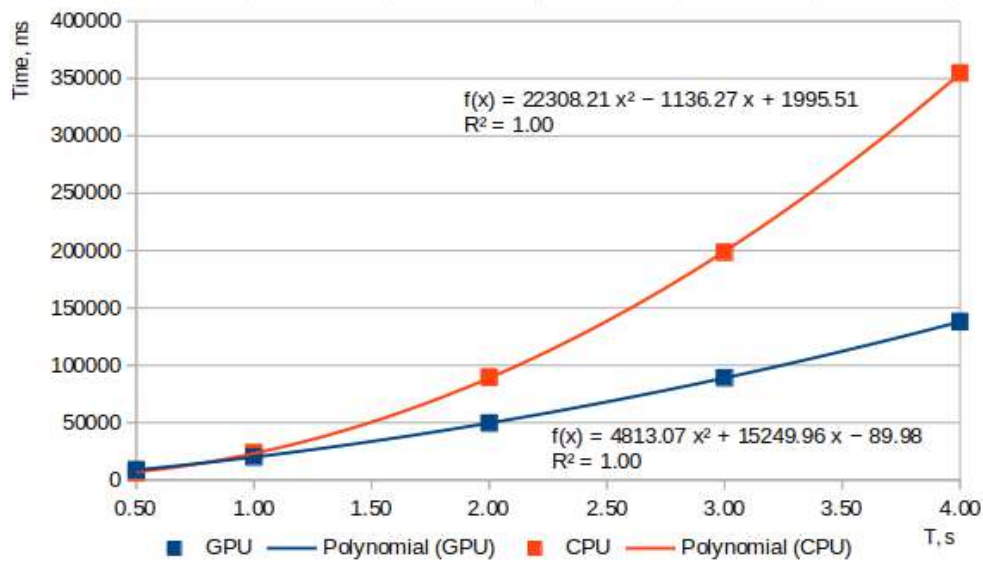


**Figure 2:** Execution time of GPU and CPU implementations subject to the ending time of simulation

## 8. Conclusions

In the paper we presented the computational scheme for identifying the values of fractional derivatives orders in the fractional-fractal model of filtration-consolidation based on the available measurements. The scheme is based on the particle swarm optimization technique and is implemented in multi-threaded mode and for the execution on GPU.

The results of computational experiments show that the order of the local fractal derivative with respect to the spatial variable can be restored with significantly higher accuracy than the order of non-local time-fractional derivative. Expectedly, relative errors increase with the increase of noise level. Considering the number of observation points, errors less than ~10% were obtained in all cases for their number more or equal to 25.

OpenCL GPU implementation of the inverse problem's solution algorithm performed closely to the OpenMP CPU implementation that used 6 CPU cores on the minimal tested volume of computations ( $m = 50$ ,  $S = 10$ ). With the increase of the volume, GPU overperformed CPU by up to 4 times.

Thus, the proposed computational scheme can be effectively applied with regard to both accuracy and speed to adapt the fractional-fractal model to the observed dynamics of salt transfer and compaction of water-saturated soils in complex geological and hydrological conditions.

## 9. References

- [1] V.M. Bulavatsky, Iu.H. Kryvonos, V.V. Skopetsky, Non-classical mathematical models of heat and mass transfer (in Ukrainian), Naukova Dumka, Kyiv, Ukraine, 2005.
- [2] A.P. Vlasniuk, P.M. Martyniuk, Mathematical modeling of consolidation in soils in the conditions of salt solution filtration (in Ukrainian), Publishing house of UDUVHP, Rivne, Ukraine, 2004.
- [3] V.A. Florin, Fundamentals of Soil Mechanics, National Technical Information Service, Moscow, USSR, 1961.
- [4] V.M. Bulavatsky, Some modelling problems of fractional-differential geofiltration dynamics within the framework of generalized mathematical models. *Journal of Automation and Information Science* 48(5) (2016) 27–41. doi:10.1615/JAutomatInfScien.v48.i5.30.
- [5] V.M. Bulavatsky, Bohaienko V.O., Numerical simulation of fractional-differential filtration-consolidation dynamics within the framework of models with non-singular kernel. *Cybernetics and Systems Analysis* 54(2) (2018) 193–204. doi:10.1007/s10559-018-0020-5.
- [6] V. Bohaienko, V. Bulavatsky, Fractional-Fractal Modeling of Filtration-Consolidation Processes in Saline Saturated Soils. *Fractal and Fractional* 4 (4) 2020. doi:10.3390/fractalfract4040059.
- [7] A. Allwright, A. Atangana, Fractal advection-dispersion equation for groundwater transport in fractured aquifers with self-similarities. *The European Physical Journal Plus* 133(2) (2018) 1–14. doi:10.1140/epjp/i2018-11885-3.
- [8] W. Chen, Time-space fabric underlying anomalous diffusion. *Chaos, Soliton. Fract.* 28(4) (2006) 923–929. doi:10.1016/j.chaos.2005.08.199.
- [9] W. Cai, W. Chen, F. Wang, Three-dimensional Hausdorff derivative diffusion model for isotropic/anisotropic fractal porous media. *Thermal Science* 22(1) (2018) S1-S6.
- [10] A. Atangana, A. Akgül, K.M. Owolabi, Analysis of fractal fractional differential equations. *Alexandria Engineering Journal* 59(3) 2020 1117-1134. doi:10.1016/j.aej.2020.01.005.
- [11] L.D. Long, N.H. Luc, Y. Zhou, C. Nguyen, Identification of Source Term for the Time-Fractional Diffusion-Wave Equation by Fractional Tikhonov Method. *Mathematics* 7 (2019). doi:10.3390/math7100934.
- [12] V.M. Bulavatsky, V.O. Bohaienko, Some Consolidation Dynamics Problems within the Framework of the Biparabolic Mathematical Model and its Fractional-Differential Analog. *Cybernetics and Systems Analysis* 56(5) (2020) 770-783. doi:10.1007/s10559-020-00298-7.
- [13] K. Liao, T. Wei, Identifying a fractional order and a space source term in a time-fractional diffusion-wave equation simultaneously. *Inverse Problems* 35(11) (2019). doi:10.1088/1361-6420/ab383f.
- [14] M. Krasnoschok, S. Pereverzyev, S.V. Siryk, N. Vasylyeva, Determination of the fractional order in semilinear subdiffusion equations. *Fractional Calculus and Applied Analysis* 23(3) (2020) 694-722. doi:10.1515/fca-2020-0035.
- [15] F. Gao, X.J. Lee, H.Q. Tong, F.X. Fei, H.L. Zhao, Identification of Unknown Parameters and Orders via Cuckoo Search Oriented Statistically by Differential Evolution for Noncommensurate Fractional-Order Chaotic Systems. *Abstract and Applied Analysis* 2013 (2013). doi:10.1155/2013/382834.
- [16] L.G. Yuan, Q.G. Yang, Parameter identification and synchronization of fractional-order chaotic systems. *Commun. Nonlinear Sci. Numer. Simul.* 17(1) (2012) 305–316. doi:10.1016/j.cnsns.2011.04.005.
- [17] V. Bohaienko, A. Gladky, M. Romashchenko, T. Matiash, Identification of fractional water transport model with  $\psi$ -Caputo derivatives using particle swarm optimization algorithm. *Applied Mathematics and Computation* 390 (2021). doi:10.1016/j.amc.2020.125665.
- [18] Y. Zhang, Sh. Wang, G. Ji, A Comprehensive Survey on Particle Swarm Optimization Algorithm and Its Applications. *Mathematical Problems in Engineering*. 2015 (2015). doi:10.1155/2015/931256.
- [19] I. Podlubny, Fractional differential equations, Academic Press, New York, 1999.
- [20] A.A. Kilbas, H.M. Srivastava, J.J. Trujillo, Theory and applications of fractional differential equations, Elsevier, Amsterdam, The Netherlands, 2006.
- [21] T. Sandev, Z. Tomovsky, Fractional equations and models. Theory and applications, Springer Nature Switzerland AG, Cham, Switzerland, 2019.

- [22] J.F. Gómez-Aguilar, M. Miranda-Hernández, M.G. López-López, V.M. Alvarado-Martínez, D. Baleanu, Modeling and simulation of the fractional space-time diffusion equation. *Communications in Nonlinear Science and Numerical Simulation* 30 (2016) 115–127. doi:10.1016/j.cnsns.2015.06.014.
- [23] V.M. Bulavatsky, Mathematical Model of Geoinformatics for Investigation of Dynamics for Locally Nonequilibrium Geofiltration Processes. *Journal of Automation and Information Sciences* 43(12) (2011) 12–20. doi:10.1615/JAutomatInfScien.v43.i12.20.
- [24] A. Samarskii, *The Theory of Difference Schemes*, CRC Press, New York, 2001.



# Methodology for determining the level of process control in complex systems taking into account risk-oriented factors from safe time to pandemics

Volodymyr Polishchuk<sup>a</sup>, Miroslav Kelemen<sup>b</sup>, Martin Kelemen jr.<sup>b</sup>

<sup>a</sup> *Uzhhorod National University, Narodna Square, 3, Uzhhorod, 88000, Ukraine*

<sup>b</sup> *Technical University of Kosice, Rampova 7, Kosice, 04121, Slovak republic*

## Abstract

Research has been conducted on the current problem of developing a methodology for determining the level of process control in complex systems, taking into account risk-oriented factors influencing it. For the first time, the study proposes stages of risk management in the process of assessing the level of controllability of complex systems, formalizes the input data used to assess risks using fuzzy models for various complex systems, constructs generalized step-by-step algorithms for risk-oriented assessment of controllability in complex systems. An experimental approbation has been made of the research on the problem of determining the level of process control in the airline system, taking into account risk-oriented factors influencing it. The study will be a useful tool to support decision-making to improve process control in various complex systems by taking into account the risks and threats to its operation for managers from safe time to pandemics as COVID-19.

## Keywords <sup>1</sup>

Process control, risk factors, fuzzy set, decision making.

## 1. Introduction

The level of security of human activity is getting better every day due to the improvement of technology, data analysis, management systems, and others. Standards of personnel training and safety management have also become significantly higher. Nevertheless, we face many risks that could potentially jeopardize the success of our operations if not adequately managed. One of the key components of measuring the risk of the functioning of complex systems is the generation of scenarios for the development of key risk factors. At present, the preparation of scenarios and risk management strategies is especially appreciated by the security and crisis managers, who usually work in conditions of extensive constraints and even an emergency situation in the companies and the national economy.

Each complex system of operation has a different basic level, in relation to the uniqueness of its data. This is a certain amount and quality of data that can be processed. Big data in complex systems can be created from a large set of sensors, databases, information systems, social networks, etc.

Therefore, the operation of a complex system requires certain decisions, and this, in turn, involves assessing the future. This assessment is implemented by maximizing uncertainties and risks. Effective risk management will improve the quality of decisions. Risks are a necessary component of human activity when there is uncertainty about the results of a decision process. When making management decisions, the problem of disclosing the uncertainty of the data on which decisions are made and adequate assessment to minimize risks is especially relevant.

---

CMIS-2021: The Fourth International Workshop on Computer Modeling and Intelligent Systems, April 27, 2021, Zaporizhzhia, Ukraine  
EMAIL: volodymyr.polishchuk@uzhnu.edu.ua (V. Polishchuk); miroslav.kelemen@tuke.sk (M. Kelemen); martin.kelemen@tuke.sk (M. Kelemen, jr)

ORCID: 0000-0003-4586-1333 (V. Polishchuk); 0000-0001-7459-927X (M. Kelemen); 0000-0003-1015-1112 (M. Kelemen, jr)



© 2021 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

The aim of the study is the urgent task of developing a methodology for determining the level of process control in the system, taking into account risk-oriented factors influencing it. The methodology is represented by a set of fuzzy methods and models based on risk-oriented factors using: the experience of experts in their assessment, the linguistic level of risk assessment criteria, the consequences of various threats to the system. The developed methodology will allow assessing the controllability of processes for a complex system taking into account the various risks that affect the functioning of the system.

This methodology can be used to solve a number of applied problems in various complex systems, for example: for socio-economic systems, it is a modeling of financial risks [1-2]; for sociotechnical it is the assessment of risks and security incidents of airport network and information systems [3]; for humanistic systems, it is an assessment of the risk of performing individual parachutists' jumps to increase their safety [4]; for technical systems, it is the assessment of the risks of drone flights to study the environment in the mountains [5], and many others.

The logic of the study is as follows: if the overall assessment of the system is high, risk-oriented factors influencing the controllability of a complex system, then we can talk about a high level of controllability of processes in the system, and competent management decisions will lead to system goals. Also, ensure the appropriate level of security of the system operation environment.

## **2. Overview of Domestic and Foreign Research Studies**

The Oxford Dictionary of English defines risk as "danger, the likelihood of adverse effects, loss or propensity to the accident". McNeill, Frey, and Embrechts [6] define risk as any event or action that may adversely affect an organization's ability to achieve its goals and implement its strategies. Although all of these definitions cover some key elements of risk, it is not possible to cover all aspects of it in a single definition from a single sentence. However, it is clear that the risk is strongly associated with uncertainty about future events. In the general case, the concept of risk can be considered a situation of the danger of the opportunity of an adverse event or the threat of possible losses, for which it is unknown whether they will occur in the future.

According to [7, 8], there are three main approaches to risk assessment in solving practical problems:

1. Intuition, experience, and knowledge of experts who make judgments about the nature of the peculiarities of the processes of functioning of a complex system. For new innovative research objects, this is the only available way to analyze processes that did not exist or had a different structure in the past, and statistical history for which is missing or meaningless.
2. Observations of actual situations and losses from risks over a long period of time. In many practical problems, such data are absent. In addition, in critical technical, social, economic, and environmental systems, in which a very variable process environment has a significant impact, such data describe the dynamics of the process in past environmental conditions, and this is often out of date.
3. Use of observations of groups of individual risks with similar characteristics that can be made in a short period of time. This approach is the most suitable and promising in complex risk management problems in all types of uncertainties [9]. Individual risk-oriented factors form risk classes, which generally constitute a certain classification of risks for tasks of a given type.

To date, no general principles and features of risk classification have been formed, there are virtually no developments for the generalization and formalization of risk classification that can be used in interdisciplinary research.

In [10], Zgurovsky considers the problem of formalized risk classification as information-interconnected problems of classification of risk situations and the problem of recognizing risk situations, based on many factors of their occurrence. It is noted that the fundamental influence on the quality of classification, which is expressed in the accuracy and certainty of risk classification, causes the choice of many risk factors and many features of each risk factor.

In [11], Panyagin identifies two main approaches to subject and management classification. Subject classification is carried out according to the meaning and content of each type and type of risk, it makes it possible to identify risks and characterize the possible consequences of risks

(operational, innovative, non-return of investment, etc.) [12]. Management classification involves the allocation of risk classes by sources and stages of risk and by methods of risk management (technological, environmental, personnel errors, etc.).

The approach of grouping risks by resources that are at risk of loss, allows you to predict or the occurrence of a particular risk situation to classify the risk and quickly pre-determine a group of adequate measures and management models for a particular class of risk [13].

H. Felix Kloman [14] under risk management defines the following. For many analysts, politicians, and scientists, environmental and nuclear risk management are the technologies that generate macro risks that threaten our very existence. For bankers and financial officials, it is prudent to use methods such as currency hedging and interest rate swaps. For insurance agents, it is the coordination of insurance risks and reduction of insurance costs. For hospital administrators, this can mean "quality assurance." For safety professionals, this is a reduction in accidents and injuries.

Thus, Kloman believes that risk management is a range of methods used to more effectively manage their lives and organizations in conditions of "unprecedented uncertainty." Or it is a discipline to ensure resilience to future events that may cause adverse consequences.

Analyzing the sources, we obtain that the following researchers described the risk management process. For example, Furley [15] presented seven steps: (1) identify risk factors; (2) assess the likelihood and consequences of the risk; (3) develop strategies to reduce identified risks; (4) control risk factors; (5) apply a contingency plan; (6) manage the crisis; (7) get out of the crisis. Bem [16] described the process with two main stages: risk assessment includes identification, analysis, prioritization, and risk control, which includes risk management planning, risk management and planning, tracking, and corrective action. Similar to the Deming quality improvement cycle (plan, do, test, act), Kliem and Ludin [17] proposed a four-phase process (identification, analysis, control, and reporting). According to the International Organization for Standardization 31000, risk management creates and protects value [18]. Several popular methods of risk management analysis have been reported in the literature, including Monte Carlo modeling [19], the process of analytical hierarchy [20], and fuzzy set theory [21].

The identification of risk factors is influenced by the field of activity of the object of study. For example, key risk factors for public-private partnership projects fall into two categories. The first includes risk factors that have powerful, independent effects, such as delays in government approval, public credit, and imperfections in the legal and regulatory systems. The second category includes risk factors that are highly volatile and easily influenced, such as downtime risks, insufficient market returns, and fee changes [22].

Ameyau and Chan [23] mentioned other risk factors, such as market/income risks, financial risks, relationship risks, and social risks. According to Lesard [24], risk management requires systematic management of the risks generated within each link in the chain and, more importantly, in the interfaces between the links to limit disruptions and their spread throughout the system.

All this proves that effective risk management requires a systems thinking approach - understanding how systems affect each other as a whole.

According to Neeman [25], the risk management process has become an integral part of defense project management procedures, for which uncertainty management is one of the main problems of current planning and project management. Moreover, in response to dangerous events, such as plane crashes or takeoffs, security requirements in the defense industry are strict and demanding.

The classification of risks in solving problems of assessing the controllability of processes in complex systems is important, and often crucial for the formalization and formulation of the problem to be solved, and, accordingly, the quality of the result. Choosing an inefficient or incorrectly classified risk classification for a particular problem can lead to a narrowing and incomplete description of risks in the problem, and, as a consequence, can lead to incorrect solutions of the problem, or even prevent adequate solutions [26].

In addition, our study uses expert information that reflects the skills, experience, and knowledge of experts on risk indicators and is given in natural language, which is appropriate to use the theory of fuzzy sets [27-28]. For example, [29-30] discusses the general ideas and benefits on which modern views on the use of fuzzy logic in decision support systems are based. In [31-32] the use of fuzzy logic in different areas of application is presented, which allows determining the optimal parameters

under conditions of uncertainty of the input data. And in [33-35] the advantages of research of complex objects of functioning in different modes and system analysis are scientifically substantiated.

The above, argues and confirms the relevance of our study: the development of a methodology for determining the level of process control in complex systems, taking into account risk-oriented factors influencing it. The relevance of this study is proved by the need to understand the controllability of processes in various complex research systems, taking into account the risks of its operation, to achieve the goals of the system and the formalization of such processes.

### **3. Conceptual and methodological bases of constructions of risk-oriented level of process control in complex systems**

#### **3.1. Formal formulation of the class of tasks and input data**

Let us know some objects of study that we will consider as a complex poorly structured system  $S$ . There are many known system goals and many factors that affect the controllability of complex systems. Also known indicators of risk-oriented factors influencing the system within the objectives of the system. Fuzzy models of system evaluation are built on the basis of known indicators. Within this, it is necessary to assess the controllability of processes in the object of study for quality decision-making depending on risk-oriented factors influencing the system.

Thus, the stages of risk management in the process of assessing the level of controllability of complex systems are proposed.

1. Formation of the list of risk factors:

- Identification of goals and objectives for identifying risk-oriented factors;
- Identification of the most risk-oriented factors of influence;
- Analysis of risk-oriented factors of influence.

The first stage is to understand the specificity of the threat and the place of its possible manifestation.

Under the identification and analysis of risks means the study of its specifics and features, which are due to their nature and other features characteristic of this occasion. It is important to study future losses, as well as changes in risks over time, the degree of threat to a particular period.

2. Building models for assessing the risks of process control in complex systems.

The main purpose of this stage is to study and develop models, methods, tools, and instruments that will assess risk factors in a complex system, as well as to study their negative impact on the controllability of complex systems.

3. Analysis, selection, and decision-making to reduce the threat of risk-oriented factors:

- Choice of risk management strategy and tactics;
- Selection of an action program (scenario) for risk reduction;
- Decision-making and organization of the developed program.

At this stage, the decision-maker (DM) forms and selects an individual approach to the risk of a complex system. The need for this selection procedure is related to the different effectiveness of risk management methods and the different size of resources required for their implementation. When choosing risk and the method of its management, you should always take into account resource constraints and try to optimize their losses. To assess all resources can be summarized on one scale - financial.

Therefore, depending on the complex system under consideration, it is necessary to adequately select risk factors in relation to the system resource that may suffer losses in the event of risk; determine risk assessment scales; build evaluation methods; perform analysis, selection and decision-making on proactive risk reduction management to increase process control in the system under study.

Examining the system, we determine the sets of input data  $K = (K_1, K_2, \dots, K_m)$ , according to which the level of process control will be assessed in a complex system  $S$ . Indicators can be a whole system of risk criteria, factors, and models, based on which a single aggregate assessment is derived. For example, the level of controllability of the aircraft is influenced by indicators that depend on risk-oriented situations, or the level of safe financing of innovative projects is influenced by risk management and forecasting factors.

To assess the level of controllability of the system operation processes according to risk-oriented factors influencing the system, the class of problems  $Z$  is proposed - fuzzy evaluation of the system taking into account risk-oriented factors of influence. Solving problem  $Z$  provides an understanding of process control in complex systems, taking into account and disclosing various risks.

A system set-theoretic model of the problem of estimating process control in complex systems, taking into account risk-oriented factors of influence can be represented as follows:

$$Z = \{S, K_R, M_R | Y\}, \quad (1)$$

Where:  $S$  is a complex poorly structured system;  $K_R$  is information models of criteria (groups of criteria) for assessing risk-oriented factors affecting the system, or models for assessing threats to the system, which will potentially lead to a weakening of controllability of processes in it;  $M_R$  is type of model for assessing control processes in complex systems, taking into account system threats, indicators of various risks, and experience of experts.

As a result, we obtain an initial estimate of  $Y$ , which carries the content of process control in the system, taking into account risk-oriented factors. This assessment will be called an aggregate risk-based assessment of the level of process control in the system.

Fig. 1. illustrates the algorithm for selecting a model of risk-oriented assessment of the level of process control in a complex system, based on available input data, taking into account system threats, indicators of various risks, and the experience of experts. At the output of each model will be aggregated risk-oriented assessment of the level of process control in the system, based on the assessment of the risk of system operation.

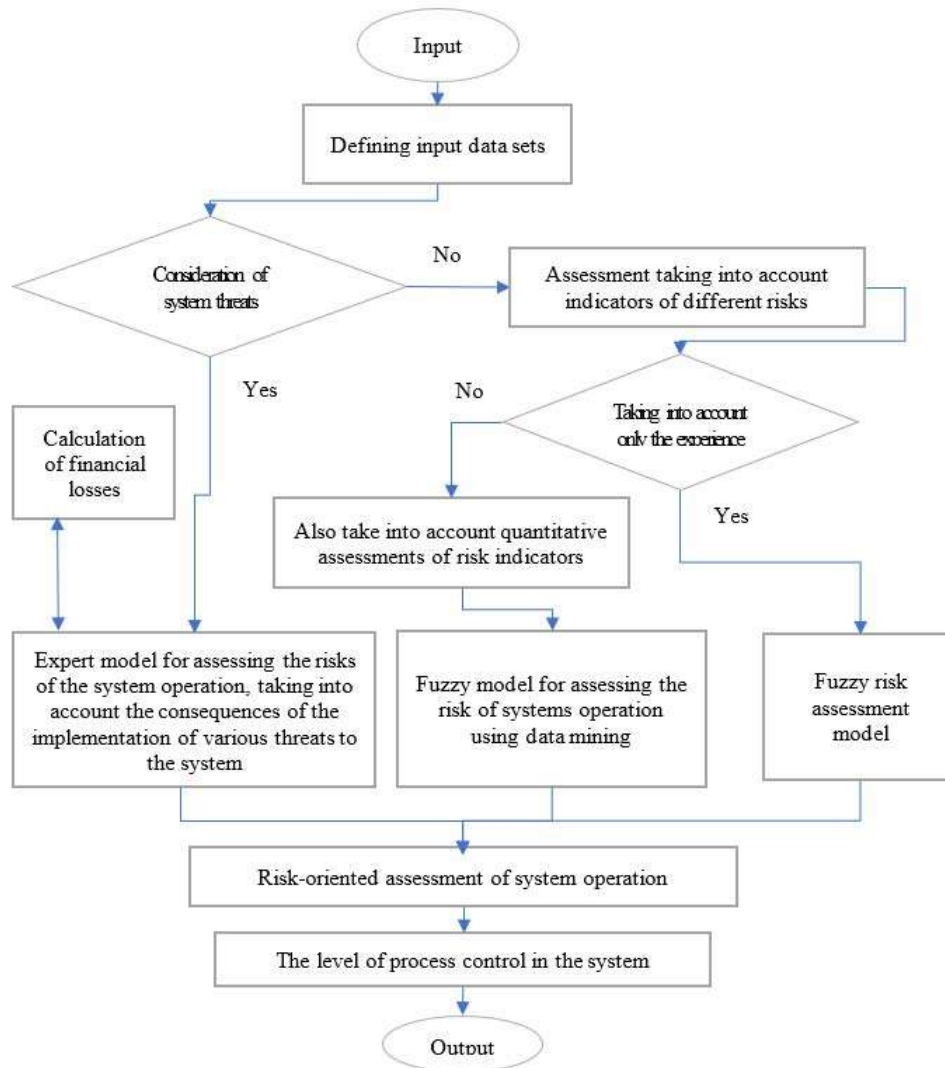


**Figure 1:** Algorithm for selecting systems risk functioning assessment models

Since this study focuses on various complex systems, therefore, based on the theoretical study, we can formalize the input data used to assess risks using fuzzy models.

4. Risk-oriented impact factors are indicators of risk that are assessed by some expert using a linguistic variable. To do this, based on experience, skills, and knowledge about the impact of risks on the system  $S$ , a group of experts (or an expert) draws conclusions and makes a linguistic assessment of each indicator  $K$ , from some term set  $T = \{T_1; T_2; \dots; T_l\}$ . We present the term set of linguistic variables as the level of probability of occurrence of a risk event described by the indicator  $K$  to achieve the target needs of the system. Also, to each assessment, the expert puts or number of "reliability"  $\mu(T)$  their reasoning from the interval  $[0; 1]$ , or normalized quantitative risk assessment. To obtain a quantitative assessment of risk, separately for each criterion, intelligent analysis of these values, which generate risk, based on the theory of fuzzy sets and membership functions [6]. The membership function is constructed and investigated depending on the type of data, their structure, the frequency of receipt, the subjectivity of receipt, and other characteristics. This will allow comparing the obtained estimates, by translating into a normalized scale, to reveal the vagueness and uncertainty of the obtained data, which will improve the quality of decision-making made using the intellectual analysis of such data.

5. Risk-oriented factors are a set of threats  $K$  of some assets of the system, which, if possible, the occurrence of such a phenomenon or event, which may result in undesirable effects on the system, which will weaken the controllability of processes in it. The input data in this concept are defined as follows:

- $T$  – the consequences of the implementation of threats  $K$ . This indicator is determined using a linguistic variable from some term set, for example,  $T = \{M; A; H; C\}$ , where  $M$  - the minimum consequences of the threat,  $A$  - the average consequences of the threat,  $H$  - maximum consequences of the threat,  $C$  - critical consequences of the threat;
- $\mu$  – the degree of possibility of realization of the threat in the system. The degree is quantified from the interval  $[0; 1]$ , for example, putting different meanings: it is impossible to realize the threat; the minimal possibility of threat realization; the average possibility of threat realization; the high possibility of threat realization; critical threat capability;
- $L$  – the severity of the consequences of the incident on the asset. An incident is a realized threat. This indicator is evaluated expertly using a linguistic variable from some term set  $L$ , where the variables can be, for example, vital consequence (if it makes it impossible to achieve global goals by the system), decisive consequence (if it makes it impossible to achieve local goals of the system), and to some extent affects the quality of achieving global goals), useful or not applicable (other cases).

Next, to present the methodology for obtaining an aggregate risk-oriented assessment of the level of process control in the system, we present generalized algorithms of fuzzy models: a two-level fuzzy mathematical model for establishing the risk of system operation and the level of process control in it; fuzzy risk assessment model for data mining systems; an expert model for assessing the risks of the system operation, taking into account the consequences of the implementation of various threats to the system and assessing financial losses.

### **3.2. Two-level fuzzy mathematical model for establishing the risk of system operation and the level of process control in it**

A two-level mathematical model given, on the basis of which we can take into account the processes of controllability in complex systems and to assess the level of risk of systems functioning [2]. The model is based on the use of fuzzy mathematics, linguistic variables, and the assessment of the confidence of assigning such variables by an expert.

Suppose that at the input we have some complex system  $S$ , consisting of many research objects (subsystems)  $X = \{x_1, x_2, \dots, x_n\}$ , which need to assess the level of risk of the system and draw a conclusion about the level of process control system taking into account risk-oriented factors influencing it. The objects of the study will be evaluated according to the input expert assessments,

the proposed set of risk criteria  $K = (K_{g1}, K_{g2}, \dots, K_{gm})$ , which is classified into groups  $g$ . Each criterion is evaluated by an expert: linguistic evaluation of the term set of linguistic variables  $T = \{t_1; t_2; \dots; t_l\}$ .

For example, such a term set of linguistic variables can be denoted as follows:  $T = \{L; BA; A; AA; H\}$ . Where: L – "low risk"; BA – "below average risk"; A – "average risk"; AA – "above average risk"; H – "high risk".

Based on the fuzzy estimation model, we construct a generalized step-by-step algorithm for obtaining an aggregate estimate, and the reliability and adequacy of the model are proved in [2].

**Step 1.** Introduction of expert input data and determination of the resulting thermal evaluation.

Based on the knowledge entered on the object of research and the constructed knowledge base, the resulting thermal evaluation is determined by the formula (2) for the relevant groups of criteria.

Let the object with  $m$  inputs and one output be analyzed:

$$T_{gj} = L(T_{g1j}, T_{g2j}, \dots, T_{gmj}). \quad (2)$$

Where:  $T_{gj}$  – initial resulting term estimate from the term set  $T$  for the group of criteria  $g$ ;  $T_{g1j}, T_{g2j}, \dots, T_{gmj}$  – input linguistic estimates by a group of criteria  $g$ .  $L$  – is an operator that matches the initial resulting thermal estimate  $T_{gj}$  for a group of criteria, with the input variables  $T_{g1j}, T_{g2j}, \dots, T_{gmj}$  (inference rule),  $j = \overline{1, n}$ .

Next, the expert, or group of experts, for each  $T$  builds the rules of belonging of the resulting terms. These rules can be constructed in the percentage of belonging of these or those in terms of an input variable. Formally, the rules of belonging represent a system of logical expressions - "If, Then, Else", which associate the values of the input variables  $T_{g1j}, T_{g2j}, \dots, T_{gmj}$  with one of the possible values  $T_{gj}$ ,  $j = \overline{1, n}$ :

**If** ( $K_{g1j} = L$  and  $K_{g2j} = L$  and ... and  $K_{gmj} = L$ ) **Or** ( $K_{g1j} = L$  and  $K_{g2j} = L$  and ... and  $K_{gmj} = BA$ ) **Or** ... **Or** ( $K_{g1j} = BA$  and  $K_{g2j} = L$  and ... and  $K_{gmj} = L$ ) **Then**  $T_{gj} = L$  **Else** ....

**Step 2.** An aggregate assessment of the reliability of the expert's opinions is determined.

The aggregate estimate of the reliability  $\mu(T_{gj})$ ,  $j = \overline{1, n}$  calculated by the following formula:

$$\mu(T_{gj}) = \frac{1}{k} \sum_{i=1}^m \mu(T_{gij}), \quad j = \overline{1, n}, \quad (3)$$

**Step 3.** Obtaining one generalized risk assessment of research objects by groups of criteria  $g$ .

For each group of criteria  $g$ , the risk level  $\chi_{gj}$ , is calculated, relative to the percentage scale  $[a; b]$  and the resulting thermal estimate  $T_{gj}$ , according to the formula:

$$\chi_{gj} = \begin{cases} \sqrt{\frac{\mu(T_{gj})}{2}}(b-a) + a, & 0 \leq \mu(T_{gj}) \leq 0,5; \\ b - \sqrt{\frac{1-\mu(T_{gj})}{2}}(b-a), & 0,5 < \mu(T_{gj}) \leq 1. \end{cases} \quad (4)$$

A generalized estimate of the risk  $O_{gj}$  for each group of criteria  $g$  is obtained by the formula:

$$O_{gj} = (b - \chi_{gj}) / b, \quad j = \overline{1, n}. \quad (5)$$

**Step 4.** Introduction of weights by groups of risk criteria.

For each group of criteria, DM sets the weights  $p_g$ , after which the normalized weights are calculated by the formula:

$$a_g = \frac{p_g}{\sum_g p_g}, \quad \sum_g p_g = 1. \quad (6)$$

**Step 5.** Reckoning of risk-oriented assessment of the system for all groups of criteria.

The aggregate score is determined by the convolutions, for example, the average will look like:

$$O_{R3}(x_j) = \sum_g a_g O_{gj}; j = \overline{1, n}. \quad (7)$$

**Step 6.** Derivation of the level of process control in the system.

Estimates of  $O_R(x)$  with the initial variable  $Y$  are equated to establish the level of process control in the system taking into account risk-oriented factors. For example, the original variable may take the value: "low risk of system operation, which determines a high level of process control in the system".

The model reveals the vagueness of input estimates, increases the objectivity of expert judgments. The formulated knowledge base does not depend on the number of criteria by groups, so they can be increased if necessary, as well as change the levels of decision-making.

### 3.3. Fuzzy model for assessing the risk of systems operation using data mining

A fuzzy model for assessing the risk of functioning of data mining systems is developed, proven and tested for the tasks of assessing the security risks of network systems, assessing the risks of project financing, assessing the risk of the municipal system and others [4].

Suppose we have a complex system of functioning  $S$ , consisting of a set of objects of study (subsystems)  $X = \{x_1, x_2, \dots, x_n\}$ . Each object of study has its own set of criteria for assessing the risk of functioning of the object in the system  $K_i = \{K_{i1}; K_{i2}; \dots; K_{im_i}\}$ ,  $i = \overline{1, n}$ . The evaluation of the criteria is proposed in a hybrid way, based on the experience of experts for the functioning of the system and the intellectual analysis of quantitative data [36].

Each indicator of the functioning system is evaluated by an expert using the linguistic variable  $T$ , for example: high level of criterion; the level of the criterion is above average; average level of criterion and others. In addition, for each criterion, a quantitative assessment of the risk  $\delta$  from the interval  $[0; 1]$ , for example within the analysis of "Big data", or assigned a quantitative risk assessment in an expert manner, such as the confidence of the expert. Input data, risk assessment of the system are as follows:  $t_{ij}$  – variable from the term set  $T$  in the  $i$ -th object of study by the  $j$ -th criterion;  $\delta_{ij}$  – quantitative estimate from the interval  $[0; 1]$ , according to the  $i$ -th object of research and the  $j$ -th criterion,  $i = \overline{1, n}$ ,  $j = \overline{1, m_i}$ .

Formally, we present a fuzzy model for assessing the risk of system operation in the following form [4]:

$$\Delta(t; \delta) \rightarrow \tau(R(S); Y). \quad (8)$$

$\Delta$  operator that matches the set of output values of  $\tau$ , with input variables  $t; \delta$ . The input data of the model are:  $t$  – expert level of the criterion on the object of study of the functioning of the system;  $\delta$  is a quantitative assessment of the considered criterion. At the output of the evaluation model we have:  $R(S)$  risk-oriented assessment of system operation;  $Y$  the level of process control in the system.

Based on the fuzzy estimation model, we construct a generalized step-by-step algorithm.

**Step 1.** Fuzzification of input hybrid data.

Let the term set of linguistic variables  $T$  be represented on some numerical interval to distinguish the terms  $[a_1; a_{k+1}]$ ,  $k$  is the number of linguistic variables. The values of the partitioning of the intervals can be adjusted and changed in the process of using the real data of the operating system. The dependence of the level of the criterion of the object of study  $t_{ij}$  and the quantitative assessment of the considered criterion  $\delta_{ij}$  ( $i = \overline{1, n}$ ;  $j = \overline{1, m_i}$ ) we propose to consider using membership functions [4]. This will allow you to adequately compare the data. Since the quantitative values of  $\delta_{ij}$  are known and the intervals of numerical values for  $t_{ij}$ , are known, then for each criterion for subsystems, it is necessary to express the dependence on the membership function, similar to formula (4). Since, the terms are defined on some range of values, then the obtained values of  $\chi_{ij}$  are normalized:

$$\theta_{ij} = \chi_{ij} / a_{k+1}. \quad (9)$$



The obtained values of  $\theta_{ij}$ , ( $i = \overline{1, n}$ ;  $j = \overline{1, m_i}$ ) are interpreted as a function of revealing the uncertainty of fuzzy expert considerations regarding the level of the criterion in the object of study and quantification of the considered criterion. Thus, in the first step, the input of hybrid data was fuzzification and moved from linguistic and quantitative estimates to a single estimate.

**Step 2.** Aggregation of fuzzification input hybrid data within a complex system.

The second step displays the aggregate value for the objects of study  $x_1; x_2; \dots; x_n$ , according to the following formula:

$$\varphi_i = \frac{1}{m_i} \sum_{j=1}^{m_i} \theta_{ij}. \quad (10)$$

Where  $m_i$ ,  $i = \overline{1, n}$  the number of criteria for the object of study  $x_i$ . Estimates  $\varphi_i \in [0; 1]$  are quantitative, normalized indicators of the corresponding object of study  $x_i$  in the system  $S$ . The greater the value of  $\varphi_i \in [0; 1]$ , the worse the level of control of the corresponding object of study.

**Step 3.** Considering the importance of research objects.

Suppose that in order to make further decisions regarding the risk assessment of the system operation, there is a need to set weighting factors  $\{w_1, w_2, \dots, w_n\}$  from the interval  $[1; 10]$  for each object of research. If experts do not need it, without reducing the generality, the objects of study will be considered balanced. Since all calculations are reduced to one measurement scale, then determine the normalized weights  $v_i$  using different approaches [36].

**Step 4.** Defuzzification data.

The aggregate risk-oriented assessment of the functioning of the system  $S$  is calculated. The obtained estimates  $\varphi_i, i = \overline{1, n}$  are normalized, but it is necessary to change the direction of the target, because the smaller  $\varphi_i$  the lower the risk functioning of the system. To do this, use one of the convolutions, changing the direction of the target [4]. For example, a risk-based assessment based on average expert judgment will take the form:

$$R_3(S) = \sum_{i=1}^n v_i (1 - \varphi_i). \quad (11)$$

Estimates of  $R(S)$  are normalized, so DM, based on experience and knowledge in the applied application problem can leave a distinction between the levels of process control in the system  $Y$ .

### 3.4. Expert model for assessing the risks of the system operation, taking into account the consequences of the implementation of various threats to the system and assessing financial losses

An expert model for assessing the risks of the system operation, taking into account the consequences of the implementation of various threats to the system and assessing financial losses, which is tested for Evaluation of Airport NIS Security for Safe and Sustainable Air Transport [3]. The model is able to assess the risks of a complex system, uses intelligent analysis of knowledge of experts, reveals the vagueness of input estimates, increases the degree of validity of further management decisions based on the results.

For this model, risk-oriented factors are the set of threats  $K$ , the consequences of the implementation of threats  $T$ , the degree of possibility of the threat in the system  $\mu$  and the severity of the consequences of the incident on the asset  $L$ .

Based on the expert evaluation model, we construct a generalized step-by-step algorithm.

**Step 1.** Fuzzification of input data.

In the first step, we consider the dependence of the consequences of the threat of the system and the degree of possibility of such implementation. To do this, we use an approach based on fuzzy set theories [3, 36].

The dependence of the consequences of the realization of the threat  $T_{ij}$  and the degree of

possibility of such realization  $\mu_{ij}$ , ( $i = \overline{1, n}$ ;  $j = \overline{1, m}$ ) is naturally considered as a statement of the membership functions "value  $x$  is greater". This is due to the fact that the higher the possibility of a threat and the more critical the consequence, the more dangerous the incident in the system, which entails a high risk of process control in the system. Therefore, one of the functions of belonging to the type "value  $x$  is greater" is constructed. Since the value of the degree of possibility of realization of threat  $\mu_{ij}$  is known and the intervals of numerical values for  $T_{ij}$ , are known, then for each threat on the considered assets, the dependence of  $\chi_{ij}$  on the formula of the constructed membership function is expressed, similarly to formula (4).

Since, risk-oriented factors of influence are associatively compared with the percentage scale, so it is natural to determine on the percentage scale and in terms of the consequences of the implementation of threats  $K$ . The obtained values of  $\chi_{ij} \in [0; 100]$  are normalized, similarly (9):

$$f_{ij} = \chi_{ij} / 100. \quad (12)$$

The obtained value  $f_{ij}$ , ( $i = \overline{1, n}$ ;  $j = \overline{1, m}$ ) is called the "incident factor" and will be interpreted as a function of revealing the uncertainty of fuzzy expert considerations regarding the consequences of the system threat and the degree of possibility of such implementation  $i$ -th asset,  $j$ -th threat. The greater the value of  $f_{ij} \in [0; 1]$ , the greater the  $j$ -th threat to system security within the  $i$ -th asset.

**Step 2.** Aggregation of the "incident rate" within the asset.

In the second step, we derive the aggregate value for the asset  $A_i$ , by a formula similar to (10):

$$a_i = \frac{1}{k_i} \sum_{j=1}^m f_{ij}. \quad (13)$$

Where  $k_i$ ,  $i = \overline{1, n}$  the number of system threats to the asset  $A_i$ . Estimates  $a_i \in [0; 1]$  are quantitative, normalized security indicators of the system corresponding to the asset  $A_i$ .

The greater the value of  $a_i \in [0; 1]$  the greater the possibility of a phenomenon or event, which may result in undesirable effects on the system, which will lead to a weakening of process control within the  $i$ -th asset.

**Step 3.** Risk assessment of system assets.

The risk of system operation will depend on the "incident rate" and the severity of the consequences of  $L$ . For each indicator, a group of experts makes an assessment from some interval, for example  $[0, 1; 1]$ , where:  $V$  - vital  $[0, 8; 1]$ ,  $C$  - decisive  $[0, 4; 0, 8]$ ,  $U$  - useful or not used  $[0, 1; 0, 4]$ . Risk assessment is defined as multiplication:

$$R_i = a_i \cdot L_i, \quad i = \overline{1, n}. \quad (14)$$

The more severe the consequence for the system asset, the greater the risk, and the  $R$ -score is close to one. The obtained risk-oriented estimate  $R$  is normalized, then on its basis we will determine the degree of risk of the system asset, for example: "insignificant degree of risk of the system on the  $i$ -th asset".

**Step 4.** Considering the importance of the system asset.

Suppose that for further decisions on the risks of the system, there is a need for each asset to set the weights  $\{w_1, w_2, \dots, w_n\}$ , or the assets will be considered equilibrium. If the weights of some interval  $[a; b]$ , then you need to reduce to one scale of measurement, by normalizing them –  $v_i$ .

**Step 5.** Defuzzification data.

Next, we derive an aggregate risk-oriented assessment of the functioning of the system for the assets under consideration, to understand the level of process control in the system. It is necessary to change the direction of the goal, because the less  $R$  the less risk, and the more  $v$  the more important the asset. To do this, use one of the convolutions, for example, the middle will look like:

$$Y_3 = 1 - \sum_{i=1}^n v_i (1 - R_i). \quad (15)$$

**Step 6.** Calculation of financial losses of the incident for assets.

The realized risk of some asset of the system leads to a decrease in the level of process control in

the system, failure to achieve systemic goals, and this usually leads to financial losses. The performance of an asset has its price over a period of time. The cost of recovery depends on the asset. In this case, you can calculate the financial loss separately for assets, this can be done using the approach described in [3].

Based on a risk-oriented assessment of the functioning of the system can be compared with the  $Y$  linguistic representation of the level of controllability of processes in the system, taking into account financial losses or without them.

## 4. Results

Consider an example of a developed methodology for determining the level of process control in the system of an airline, taking into account risk-oriented factors influencing it. For example, risk-oriented factors influencing the controllability of the processes of the airline system, the following set [36]  $F_R = \{R_1, R_2, \dots, R_{12}\}$ :

- $R_1$  – risk in the breaks of aviation services;
- $R_2$  – competition risks;
- $R_3$  – risks of stagnation in the market of aviation services;
- $R_4$  – risks of natural disasters;
- $R_5$  – risks of changes in the regulator and in the legislation;
- $R_6$  – risks of political and social crises;
- $R_7$  – risks of changes in the price of raw materials;
- $R_8$  – risks of technological innovations;
- $R_9$  – risks of cyber threats;
- $R_{10}$  – risks of terrorism;
- $R_{11}$  – risks of environmental pollution;
- $R_{12}$  – explosion hazards.

Have a group of experts, based on experience and knowledge of the airline system, analyze and insert one linguistic estimate for each  $F_R$  indicator. We propose the following term set  $T = \{T_1; T_2; T_3; T_4; T_5\}$ , where:  $T_1$  – "minor risk";  $T_2$  – "low risk";  $T_3$  – "medium risk";  $T_4$  – "high risk";  $T_5$  – "marginal risk". In addition, to each assessment, the expert puts the number of "reliability"  $\mu(T)$  of their reasoning from the interval  $[0; 1]$ . Let us represent the training test input data and weight coefficients  $w$  (interval  $[1; 10]$ ) in Fig. 2.

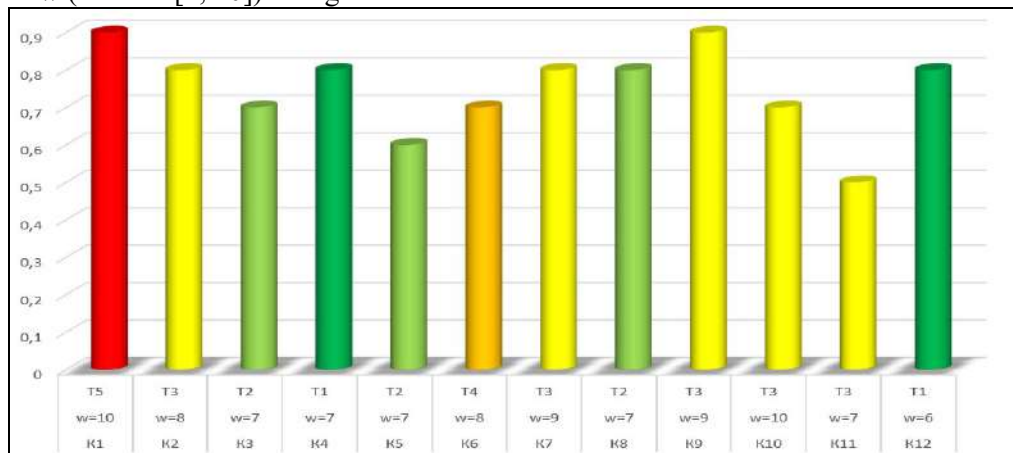


Figure 2: Visualization of input data

Let's evaluate the level of process control in the system of the airline, according to the fuzzy model of risk assessment of the operation of systems using data mining and based on the above algorithm.

**Step 1.** Fuzzyfication of hybrid input data.

First of all, let's fuzzyfication the input hybrid data. Terms are defined on a percentage scale, to distinguish terms  $[0; 100]$ , where  $T_1 - [0; 20]$ ,  $T_2 - [20; 40]$ ,  $T_3 - [40; 60]$ ,  $T_4 - [60; 80]$ ,  $T_5 - [80; 100]$ .

Next, the quadratic S-spline membership function is selected, and based on the input data  $x$  is calculated (according to the approach from formula (4)):

$$\chi = (32,91; 36,75; 40,83; 36,75; 45,22; 40,83; 36,75; 36,75; 32,91; 40,83; 50; 36,75).$$

Next, according to formula (9), we normalize the obtained values:

$$\theta = (0,33; 0,37; 0,41; 0,37; 0,45; 0,41; 0,37; 0,37; 0,33; 0,41; 0,50; 0,37).$$

**Step 2.** Aggregation of fuzzyfication input hybrid data within a complex system.

Because in the studied system of the airline risk-oriented factors are not classified by groups of objects of study or subsystems, they are already separate objects, so this step is skipped.

**Step 3.** Considering the importance of risk-oriented factors.

Let the influencing factors be considered separately as objects of research that have their importance  $w$ .

Then determine the normalized weights, respectively:  $v_i = w_i / \sum_{i=1}^{12} w_i, i = \overline{1,12}$ .

$$v = (0,105; 0,084; 0,074; 0,074; 0,074; 0,084; 0,095; 0,074; 0,095; 0,105; 0,074; 0,063).$$

**Step 4.** Deffuzification data.

The aggregate risk-oriented assessment of the airline's operation is calculated according to the formula (11):  $R_3(S) = 0,613$ .

The distinction between the levels of process control in the system, taking into account risk-oriented factors, is proposed as follows:  $R \in (0,8;1]$  high;  $R \in (0,6;0,8]$  above average;  $R \in (0,4;0,6]$  average;  $R \in (0,2;0,4]$  below average;  $R \in [0;0,2]$  low.

The distinctions between levels are determined either expertly or by teaching methods in the presence of a reliable, large enough sample of data.

As a result, on the test example we can conclude that the airline has a level of process control given the risk-oriented factors influencing its activities – above average.

## 5. Discussion

To apply the developed methodology, determine the level of process control in the system, to solve the class of applied problems of fuzzy system evaluation taking into account risk-oriented factors, it is necessary to adequately determine the set of evaluation risks, select and configure the proposed fuzzy models, configure fuzzyfication of input data. All these tasks are assigned to system analysts who develop an information system within the application problem. Thus, for a qualitative comparison of data, the delimitation of terms must be carried out separately for each indicator, as different indicators have their own numerical meaning. To qualitatively obtain input quantitative estimates and data mining (knowledge) using fuzzy set theory, membership functions and a systems approach, you can use the models described in [2-4, 36, 38].

The study proposes a class of tasks for fuzzy evaluation of the system taking into account risk-oriented factors of influence, the solution of which provides an understanding of process control in complex systems, taking into account and disclosing various risks. The built methodology, consisting of a set of methods and models based on risk-oriented factors has a number of advantages, namely: increases the objectivity of expert assessments using input hybrid data: linguistic variables and quantitative assessments; derives an aggregate risk-oriented assessment of the functioning of the system taking into account the considerations of DM; displays the level of controllability of processes in the system.

For the applied application of the developed methodology the generalized algorithms which choice depends on the available input data, the account of threats of system, indicators of various risks and experience of experts are offered. The proposed algorithms are a tool with which we can adequately solve the innovation problem. In the algorithm for fuzzyfication of input data we determine what type and type of membership function you want to choose for proper operation.

The use of different types of convolutions can lead to ambiguity of the final results. This can be attributed to the disadvantage of this approach.

## 6. Conclusions

Research of the current problem of developing a methodology for determining the level of process control in complex systems, taking into account risk-oriented factors influencing it. This is the first time the following results have been obtained:

- The stages of risk management in the process of assessing the level of controllability of complex systems on the basis of the conducted theoretical research are offered;
- Formalized input data used to assess risks using fuzzy models for various complex systems, namely: risk indicators assessed by some expert using a linguistic variable; quantitative assessment of the "reliability" of experts on considerations of the risk indicator; quantitative assessment of the risk criterion on the basis of intellectual analysis of the data that generate risk, using the theory of fuzzy sets and membership functions; linguistic variable consequences of the implementation of threats to the system; the degree of possibility of realization of the threat in the system; the severity of the consequences of the incident on the asset of the system, which is estimated by some expert using a linguistic variable;
- A set-theoretic model is proposed, which solves a class of problems of process control assessment in complex systems taking into account risk-oriented factors of influence and an algorithm for choosing a risk-oriented assessment model;
- Generalized step-by-step algorithms are constructed: a two-level fuzzy mathematical model for establishing the risk of the system functioning and the level of process control in it; fuzzy model of risk assessment of systems operation using data mining; expert model for assessing the risks of the system operation, taking into account the consequences of the implementation of various threats to the system and assessing financial losses;
- An example of the developed methodology for determining the level of process control in the airline system taking into account risk-oriented factors influencing it is demonstrated.

The rationality of the obtained aggregate risk-oriented assessments of the system operation to determine the level of process control in the system is proved by significant approbations of models for various applied tasks. The reliability of the obtained results is ensured by the correct use of fuzzy set theory for the development of expert knowledge, a systematic approach, which is confirmed by research results.

We see further research in the developed methodologies for assessing the level of process control in complex systems using the effects of human factors on a complex system. We also plan research of expert database and creating an information model based on fuzzy logic for quantitative and qualitative evaluation of physical data of a new chemical working substance for decontamination, disinfection, and deactivation for transport services, respectively also for operational risk and data assessment and other project activities.

The study will be a useful tool to support decision-making to improve process control in various complex systems by taking into account the risks and threats to its operation, and a platform for exploring new application areas for the use of knowledge.

## 7. Acknowledgements

This work was supported by the Slovak Research and Development Agency (APVV) under the contract no. PP-COVID-20-0002.

## 8. References

- [1] S.M. Shehata, A.M. Abdeljawad, L.A. Mazouz, L.Y.K. Aldossary, M.Y. Alsaeed and M. Noureldin Said. "The Moderating Role of Perceived Risks in the Relationship between

- Financial Knowledge and the Intention to Invest in the Saudi Arabian Stock Market”. *Int. J. Financial Stud.*, Vol. 9, 9 (2021). doi: 10.3390/ijfs9010009
- [2] O. Voloshyn, M. Malvar, M. Sharkadi and V. Polishchuk, Fuzzy Mathematical Modeling Financial Risks, in: *Proceedings of Second International Conference on Data Stream Mining & Processing (DSMP)*, IEEE, Lviv, Ukraine, 2018, pp. 65-69. doi: 10.1109/DSMP.2018.8478604.
  - [3] M. Kelemen, V. Polishchuk, B. Gavurová, R. Andoga, S. Szabo, W. Yang, J. Christodoulakis, M. Gera, J. Kozuba, P. Kaľavský, and M. Antoško. “Educational Model for Evaluation of Airport NIS Security for Safe and Sustainable Air Transport”. *Sustainability*, Vol. 12, (2020): 6352. doi: 10.3390/su12166352.
  - [4] M. Malyar, V. Polishchuk, M. Sharkadi and A. Polishchuk, Model of Operation Management Systems Risk Assessment, in: *Proceedings of 15th International Conference on Computer Sciences and Information Technologies (CSIT)*, IEEE, Zbarazh, Ukraine, 2020, pp. 190-193. doi: 10.1109/CSIT49958.2020.9321930.
  - [5] V. Polishchuk, A. Polishchuk, J. Jevčák, L. Choma and M. Kelemen jr, Criteria for the Information model for assessing the risks of UAV flights in environmental research on mountain terrain, in: *Proceedings of XXth International Multidisciplinary Scientific GeoConference SGEM*, Varna, Bulgaria, 2020, 2.1. pp. 97-102. doi:10.5593/sgem2020/2.1/s07.013
  - [6] A. J. McNeil, R. Frey, and P. Embrechts, *Quantitative risk management: Concepts, techniques and tools*, 2nd ed., Princeton university press, 2015.
  - [7] A. Chaudhuri, S.K Ghosh, *Need for Risk Classification*, in: *Bankruptcy Prediction through Soft Computing based Deep Learning Technique*, Springer, Singapore, 2017, pp. 19-20. doi: 10.1007/978-981-10-6683-2\_5
  - [8] A. Dicke, *On Risk Classification. A public policy Monograph*, American Academy of Actuaries, New York, NY, 2011.
  - [9] Adam M. Finkel, *Confronting uncertainty in risk management. A Guide for Decision-Makers*. Washington, DC, Center for Risk Management, Resources for the Future, 1990.
  - [10] M. Zgurovsky, and N. Pankratova, *Osnovy systemnoho analizu*, [Fundamentals of systems analysis]. Kiev, VNV, 2007. (In Ukrainian).
  - [11] A. Panyagina. “Podkhody k ponimaniyu i klassifikatsii riskov”. [Approaches to understanding and classifying risks]. *Modern Economics: Problems, Trends, Prospects*, 6 (2012). (In Russian).
  - [12] K.J. Crocker, A. Snow, *The Theory of Risk Classification*, in: G. Dionne (Eds.), *Handbook of Insurance*, Springer, New York, 2013, pp. 281-313. doi: 10.1007/978-1-4614-0155-1\_11.
  - [13] D. Porrini. “Risk Classification Efficiency and the Insurance Market Regulation”. *Risks*, 3(4) (2015). doi: <https://doi.org/10.3390/risks3040445>.
  - [14] H. F. Kloman. “Risk management agonists”. *Risk Analysis*, 10 (1990).
  - [15] R. Fairley. “Risk Management for software projects”. *IEEE Softw.*, 11 (1994).
  - [16] B.W. Boehm. “Software risk management: Principles and practices”. *IEEE Softw.*, 8 (1991).
  - [17] R.L. Kliem, and I.S. Ludin, *Reducing Project Risk*, UK, Gower: Farnham, 1997.
  - [18] *Risk Management-Principles and Guidelines ISO 31000:2018*, 2018. URL: <https://www.iso.org/standard/65694.html>
  - [19] S. Ye, and R.L.K. Tiong. “NPV-at-risk method in infrastructure project investment evaluation”. *J. Constr. Eng. Manag.*, 126 (2000).
  - [20] M. Hastak, and A. Shaked. “ICRAM-1: Model for international construction risk assessment”. *J. Manag. Eng.*, 16 (2000).
  - [21] Y. Wu, Z. Song, L. Li, and R. Xu. “Risk management of public-private partnership charging infrastructure projects in China based on a three-dimension framework”. *Energy*, 165 (2018).
  - [22] Y. Wang, Y. Wan, Y. Wu, and J. Li. “Exploring the risk factors of infrastructure PPP projects for sustainable delivery: A social network perspective”. *Sustainability*, 12 (10), 4152 (2020). doi: 10.3390/su12104152
  - [23] E.E. Ameyaw, and A.P.C. Chan. “Evaluation and ranking of risk factors in public-private partnership water supply projects in developing countries using fuzzy synthetic evaluation approach”. *Expert Syst. Appl.*, 42 (2015). doi: 10.1016/j.eswa.2015.02.041
  - [24] D.R. Lessard, *Uncertainty and Risk in Global Supply Chains*; Research Paper No. 4991-13, MIT Sloan School, Cambridge, MA, USA, 2013. doi: 10.2139/ssrn.2240274

- [25] A. Naaman. Establishment of the Armed Force—Research Development in Air Force, Between the Poles, Part 3, IDF, Jerusalem, Israel, 2016.
- [26] Risk Classification. Statement of Principles, 2014. URL: <http://www.actuarialstandardsboard.org/wpcontent/uploads/2014/07/riskclassificationSOP.pdf>
- [27] G.J. Klir, and B. Yuan, Fuzzy Sets and Fuzzy Logic: Theory and Applications, NJ, USA, Prentice Hall, Englewood Cliffs, 1995.
- [28] M. Sugeno, G. T. Kang. “Structure identification of fuzzy model”. Fuzzy sets and systems. 1. 28 (1988): 15-33.
- [29] E. Lughofer, Evolving fuzzy systems-methodologies, advanced concepts and applications. Vol. 53. Berlin, Springer, 2011.
- [30] R.R. Yager, and L.A. Zadeh, An introduction to fuzzy logic applications in intelligent systems. Vol. 165. NY, Springer Science & Business Media, 2012.
- [31] V. Polishchuk, M. Kelemen, B. Gavurová, C. Varotsos, R. Andoga, M. Gera, J. Christodoulakis, R. Soušek, J. Kozuba, P. Blišťan, S. and Szabo Jr. “A Fuzzy Model of Risk Assessment for Environmental Start-up Projects in the Air Transport Sector”. Int. J. Environ. Res. Public Health, Vol. 16, 3573 (2019). doi: 10.3390/ijerph16193573\_2
- [32] A. Sobrino. “Fuzzy Logic and Education: Teaching the Basics of Fuzzy Logic through an Example (by Way of Cycling)”. Education Sciences, Vol. 3 (2) (2013). doi: 10.3390/educsci3020075.
- [33] K. Olejnik, A. Fabijanska, P. Pelczynski, and A. Stanislawski. “Controllability of low-consistency chemical pulp refining process”. Computers & Chemical Engineering, 128, (2019). doi: 10.1016/j.compchemeng.2019.06.024
- [34] E. Johann, M. Franceschetti, and J. Köpke. “Controllability of business processes with temporal variables”. 34th. ACM/SIGAPP Symposium on Applied Computing, 2019, 40–47.
- [35] M. Zattereri, C. Combi, and L. Viganò. “Resource controllability of workflows under conditional uncertainty”. In: C. Di Francescomarino, R. Dijkman, U. Zdun (Ed.), Business Process Management Workshops, Springer, Cham, 2019, pp. 68-80. doi: 10.1007/978-3-030-37453-2\_7
- [36] O. Voloshyn, M. Malyar, V. Polishchuk, and M. Sharkadi. “Informatsiyne modelyuvannya nechitkykh znan”. [Information modeling of fuzzy knowledge]. Radioelektronika, informatyka, upravlinnya, 2018/4 (2018). (In Ukrainian). doi: 10.15588/1607-3274-2018-4-8
- [37] V.P. Kharchenko, T.F. Shmelova, and YU.V. Sikirda, Pryynyattya rishen' v sotsiotekhnichnykh systemakh: monohrafiya, [Decision making in sociotechnical systems: a monograph]. Kyiv, NAU, 2016. (In Ukrainian)
- [38] Z. Dvorak, D. Rehak, A. David, and Z. Cekerevac. “Qualitative Approach to Environmental Risk Assessment in Transport”. Int. J. Environ. Res. Public Health, 17, 5494, (2020). doi: 10.3390/ijerph17155494

# Bi-objective Circular-Hole Based Topology Optimization Problem in Additive Manufacturing

Georgiy Yaskov<sup>a,b</sup>, Andrey Chugay<sup>a,c</sup>, Tatiyana Romanova<sup>a,d</sup> and Sergey Shekhovtsov<sup>c,d</sup>

<sup>a</sup> Institute for Mechanical Engineering Problems of the National Academy of Sciences of Ukraine, 2/10 Pozharskogo st., Kharkiv 61046, Ukraine

<sup>b</sup> Kharkiv Aviation Institute, National Aerospace University, 17 Chkalov st., Kharkiv 61070, Ukraine

<sup>c</sup> National University of Internal Affairs, L. Landau av., 27, Kharkiv, 61080, Ukraine

<sup>d</sup> Kharkiv National University of Radioelectronics, 14 Nauky ave., Kharkiv 61166, Ukraine

## Abstract

The paper deals with the circle packing problem which arises in topology optimization for additive manufacturing. The problem consists in packing a number of circles of radii within a particular range imposed by technical limitations, the packing factor being maximized. A bi-objective formulation for the problem compromising the packing factor and the maximum mechanical stress of parts is suggested. The  $\varepsilon$ -constraint method is applied to search for a trade-off solution of the problem. A new packing approach based on a modified Apollonian circle packing and nonlinear optimization is developed. Numerical examples and graphical illustration of results are given.

## Keywords 1

Additive manufacturing, topology optimization, mechanical stress, bi-objective optimization,  $\varepsilon$ -constraint method, packing, circle, polygon, Apollonian circle packing, nonlinear optimization

## 1. Introduction

One of the most important topics of modern mechanical engineering is reducing weight while maintaining specific characteristics of structures designed. Such problems with conflicting criteria are related to decision making in a formidable manufacturing process. A key objective of the problems is finding optimal geometric shape and topology of the designed product ensuring minimum weight at specified strength. To this end, topology optimization methods [1] are used to find the best design parameters satisfying the technological and strength constraints and thus providing the objective function extremum. When optimizing topology of a structure, the stress level in a particular part of the structure can be used as an indicator of ineffective material usage. Ideally, the stress level in the structure should be uniform, close to the limiting, but safe value [2].

Applying topology optimization methods in mechanical engineering is a newish development in the design procedure. The methods received the greatest impetus in their development when it became possible to use additive technologies in the manufacturing process [3] instead of classical subtractive methods. Additive technologies enable to expand the range of designs for the same product.

In the last two decades, topology optimization became an active field for scientific research. This led to the use of a multidisciplinary approach in the search for solutions to the problems, which use the methods of solid mechanics, thermodynamics, biology simultaneously [4].

Paper [2] reviews modern topology optimization methods.

---

CMIS-2021: The Fourth International Workshop on Computer Modeling and Intelligent Systems, April 27, 2021, Zaporizhzhia, Ukraine  
EMAIL: yaskov@ukr.net (G. Yaskov); chugay.andrey80@gmail.com (A. Chugay); tarom27@yahoo.com (T. Romanova) ; tarom7@yahoo.com (S. Shekhovtsov)  
ORCID: 0000-0002-1476-1818 (G. Yaskov); 0000-0002-4079-5632 (A. Chugay); 0000-0002-8618-4917 (T. Romanova) ; 0000-0003-2381-7999 (S. Shekhovtsov)



© 2020 Copyright for this paper by its authors.  
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).  
CEUR Workshop Proceedings (CEUR-WS.org)



Currently, the following main methods of topology optimization can be distinguished: SIMP (solid isotropic material with penalization), ESO (evolutionary structural optimization), Level-Set (level setting method) and their various combinations. One of the most effective applications of these methods is in optimizing the topology of continuous structures, i.e. finding the best placement location and geometry for holes (cavities) within the modeling area.

Circle packing problems (CPP) have a wide range of industrial and engineering applications, for example, facility layout design, cutting stock problems in the glass, metal, paper, textile and wood industries etc.

As a rule, practical optimization problems have several conflicting objectives. Such problems are called multi-objective. Optimization methods for multi-objective optimization problems and its applications are reviewed in [5, 6].

In this paper we first propose a bi-objective formulation for circular-hole based topology optimization which takes into account both the packing factor and the maximum mechanical stress of parts. A new approach for packing circles based on a modified finite Apollonian circle packing (ACP) and optimized homothetic transformations of circles is suggested. Both techniques allow to advantageously arrange circles within polygonal parts minimizing material expenses. The  $\varepsilon$ -constraint method [7] is applied to find out a trade-off circular arrangement.

In accordance with ACP (see, for example [8]), starting from three mutually tangent circles, next circles are added to be tangent to the three circles forming four mutually tangent circles (or tangent to the frontier of the container). We modify ACP for taking into account the upper bound of circle radii and the minimal allowed distance between the circles. Some circles are allowed to be moved within the container and located to positions which give a possibility to enlarge radii of next circles being packed.

The main contributions of the paper are as follows:

- A bi-objective model of non-standard circular packing problem considering both the maximum packing factor and mechanical stress of parts for a topology optimization
- A new approach for constructing starting points based on ACP
- New benchmark instances for packing circles with variable radii.

## 2. Related papers

The number of papers concerning CPP goes up and up each year. Paper [9] reviews selected works dealing with packing methods and applications for CPP. We make mention of some recent papers [10, 11, 12]. A powerful tool of solving CPP with unequal circles is treating additional variable metric characteristics of circles and/or containers (see, for example, [11, 12, 13, 14]). The best results for a set of benchmark CPP instances are continuously updated in the well known website [15].

Paper [16] proposes a circular packing model and a fast heuristic algorithm to optimize part geometries subject to the DMLS constraints. A feature of the model is that radii of circles are preassigned.

In [9] a model and a numerical solution approach to packing egg-shaped objects with arbitrary size and orientation into optimized convex containers is presented. The area of the container is minimized. The solution strategy is based on the analysis of embedded Lagrange multipliers and nonlinear optimization. Within the universal model, circles and ellipses are considered to be special cases of egg.

The  $\varepsilon$ -constraint method introduced in [7] is often used for solving multi-objective optimization problems. The method optimizes only one objective while the other objectives are imposed by limits. In [5] the Pareto optimality of solutions obtained by the  $\varepsilon$ -constraint method is investigated.

A bi-objective function for packing in a spacecraft module is proposed in [17]. The dimensions of the module are minimized and other criteria such as desired adjacency between items and packing costs are also taken into account. Paper [18] uses a genetic algorithm for bi-objective topology optimization: minimization mass and maximization effective flexural and torsional rigidities. Methods for analysis of stress state in parts with different geometries are discussed in [16, 19]. Effective methods of nonsmooth optimization for allocation problems are considered in [20].

### 3. Problem statement

Let  $P_0$  be a polygon given by its vertices  $v_{0i} = (x_{0i}, y_{0i})$ ,  $i \in I_0 = \{1, 2, \dots, s_0\}$ . We define a disconnected polygonal domain of the form

$$P = \bigcup_{l \in I_p} P_l, \quad P_l \subset P_0, \quad l \in I_p, \quad P_l \cap P_j = \emptyset, \quad l < j \in I_p = \{1, 2, \dots, n_p\},$$

where

$$P_l = \{(x, y) \in \mathbf{R}^2 : \varphi_{ml}(x, y) \geq 0, \quad m = 1, 2, \dots, M_l\}$$

are convex polygons given by vertices  $v_{li} = (x_{li}, y_{li})$ ,  $i \in I_l = \{1, 2, \dots, s_l\}$ ;  $\varphi_{ml}(x, y) = a_{ml}x + b_{ml}y + c_{ml} = 0$ ,  $l \in I_p$ , are normal equations of the edges of  $P_l$ ,  $l \in I_p$ . Let us also define a collection of circles

$$C_q = C_q(v_q, r_q) = \{(x, y) \in \mathbf{R}^2 : (x - x_q)^2 + (y - y_q)^2 \leq r_q^2\}$$

of variable radii  $r_q$  and translation vectors  $v_q = (x_q, y_q)$  for  $q \in I_N = \{1, 2, \dots, N\}$ ,  $0 < r^- \leq r_q \leq r^+ \leq r^{++}$ ,  $r^-$ ,  $r^+$  are given lower and upper allowed values for  $r_q$ ,  $r^{++}$  is the maximal possible radius of the circle which can be inscribed into  $P_l$ ,  $l \in I_p$ .

Conditions of packing the circles  $C_q(v_q, r_q)$ ,  $q \in I_N$ , into the domain  $P$  are determined as follows.

- Containment the circle  $C_q(v_q, r_q)$  into the domain  $P$

$$C_q(v_q, r_q) \subset P \Leftrightarrow \text{int } C_q(v_q, r_q) \cap P^* = \emptyset, \quad q \in I_N, \quad (1)$$

where  $P^* = \mathbf{R}^2 \setminus \text{int } P$

- Minimal allowed distance between the circles  $C_q(v_q, r_q)$  and  $C_g(v_g, r_g)$

$$\text{dist}(C_q(v_q, r_q), C_g(v_g, r_g)) \geq \rho, \quad (q, g) \in \Theta_l, \quad l \in I_p, \quad \rho > 0 \quad (2)$$

where

$$\text{dist}(C_q(v_q, r_q), C_g(v_g, r_g)) = \min_{a \in C_q(v_q, r_q), b \in C_g(v_g, r_g)} \rho(a, b),$$

$\rho(a, b)$  is Euclidean distance between points  $a, b \in \mathbf{R}^2$ ,

$$\Theta_l = \{(q, g) : C_q(v_q, r_q) \subset P_l, \quad C_g(v_g, r_g) \subset P_l, \quad q < g\}.$$

A problem of bi-objective circle topology optimization into the polygonal domain is formulated as follows.

Pack circles  $C_q(v_q, r_q)$ ,  $q \in I_N$ , into the domain  $P$  providing containment  $\lambda_l$  circles into the polygon  $P_l$ ,  $l \in I_p$ ,  $\sum_{l \in I_p} \lambda_l = n \leq N$  and the distance constraints (2) and maximizing the packing factor while minimizing the maximum mechanical stress.

The packing conditions (1), (2) are analytically described using the phi-function technique [21], which makes it possible presenting a mathematical model as the following bi-objective problem:

$$\begin{aligned} & \max \{\kappa(\omega), -\sigma(\omega)\} \\ & \text{subject to } \omega \in W, \end{aligned} \quad (3)$$

where  $\omega = (v, r)$  is a vector of variables;  $v = (v_1, v_2, \dots, v_n)$  is a vector of variable packing parameters;  $r = (r_1, r_2, \dots, r_n)$  is a vector of variable radii of circles; function

$$\kappa(\omega) = \pi \sum_{q \in I_n} r_q^2 \quad (4)$$

is the total sum of the circle areas;  $\sigma(\omega)$  is an implicit function which defines the maximum mechanical stress depending on the vector  $v$  of center coordinates of circles and the vector  $r$  of radii of circles;

$$W = \{\omega \in \mathbf{R}^{3n} : \Phi_{qg}(v_q, v_g, r_q, r_g) - \rho \geq 0, (q, g) \in \Theta_l, l \in I_p, \Phi_q(v_q, r_q) \geq 0, q \in I_n, \\ r_q - r_q^- \geq 0, q \in I_n, -r_q + r_q^+ \geq 0, q \in I_n\} \quad (5)$$

is a feasible region;

$$\Phi_{qg}(v_q, v_g, r_q, r_g) = (x_q - x_g)^2 + (y_q - y_g)^2 - (r_q + r_g + \rho)^2,$$

is an adjusted phi-function of the circles  $C_q$  and  $C_g$ ,  $(q, g) \in \Theta_l$ ;

$$\Phi_q(v_q, r_q) = \max_{l \in I_p} \left\{ \min_{m=1,2,\dots,M_l} \{\phi_{ml}(v_q) - r_q\} \right\}$$

is a phi-function of the circle  $C_q$ ,  $q \in I_n$ , and the set  $P^* = \mathbf{R}^2 \setminus \text{int } P$ .

The inequality  $\Phi_{qg}(v_q, v_g, r_q, r_g) \geq 0$  ensures the condition (2) for packing the circles  $C_q$  and  $C_g$ ,  $(q, g) \in \Theta_l$ , on the minimal allowed distance  $\rho$  and the inequality  $\Phi_q(v_q, r_q) \geq 0$  provides the condition (1) for packing the circle  $C_q$  into the polygon  $P$ .

We note that  $\kappa(\omega)$  and  $-\sigma(\omega)$  are compromising objectives. Therefore, our aim is to find a trade-off solution of the bi-objective problem (3)-(5).

## 4. Solution approach

We use the multistart strategy [22] for solving the circular problem (3)-(5).

### 4.1. Solution strategy

On the assumption that an accepted value of the maximum mechanical stress is tolerable we solve the problem by means of the  $\varepsilon$ -constraint method [7] which reduces the problem (3)-(5) to a single-objective one

$$\begin{aligned} & \max \kappa(\omega) \\ & \text{subject to } \omega \in W' \\ & W' = \{\omega \in W \mid -\sigma(\omega) + \varepsilon \geq 0\}. \end{aligned} \quad (6)$$

The main objective is maximizing sum of squares of circles to be packed. Continual review of mechanical stress is a formidable problem. Of importance is the threshold value  $\varepsilon$  of mechanical stress.

Let us temporarily relax the inequality

$$-\sigma(\omega) + \varepsilon \geq 0 \quad (7)$$

Then we can evaluate a posteriori the maximum mechanical stress for the obtained topology of the domain  $P_0 \setminus \text{int } \bigcup_{i=1}^n C_i(\omega^*)$  and verify the condition (7) with respect to the local maximum point  $\omega^*$ .

If the condition (7) holds true, i.e.  $-\sigma(\omega^*) + \varepsilon \geq 0$ , then a feasible point of the problem (6)-(7) is found, otherwise we calculate another local maximum point  $\omega^{**}$  and repeat the procedure until (7) will be met.

We decompose the problem (6)-(7) into  $l$  subproblems separately for each polygon  $P_l$ ,  $l \in I_p$ . Our multistart strategy involves the following main stages.

Stage 1. Define a lower bound  $\lambda_l^-$  of the number  $\lambda_l$  of circles which can be packed into the appropriate polygon  $P_l$ ,  $l \in I_p$  with the maximum packing factor, using Algorithm 1 based on packing equal circles [14]. Form a point  $u_l^*$ .

Stage 2. Generate feasible points for the problem (6) starting from the point  $u_l^*$  by use of Algorithm 2 based on ACP (Apollonian Circle Packing) [8]. Form a point  $w_l^*$ .

Stage 3. Search for a local maximum point of the problem (6)-(7) starting from the point  $\omega^0 = (w_1^*, w_2^*, \dots, w_{n_p}^*)$  using IPOPT [23].

Stage 4. Choose the best local maximum point satisfying the inequality (7) found at Stage 3.

## 4.2. Solution algorithm

Let us consider our solution strategy in details.

### 4.2.1. Algorithm 1. Searching for a lower bound of the number of circles

Packing results depend heavily on the number of circles packed and starting points. We make use the idea of ACP. According to ACP circles are packed repeatedly touching three other circles [8]. To start ACP we first construct an initial configuration of tangent circles. To this end we realize a preliminary packing of equal circles with radius  $r^+$  (if possible). The problem is known as IIPP (Identical Item Packing Problem) [24] The circle layout algorithm is based on the models described in [14].

We consider packing circles into the polygon  $P_l$ ,  $l \in I_p$ . Let  $n_{l-1} = \sum_{j=1}^{l-1} \lambda_j$  be packed into the polygons  $P_j$ ,  $j = 1, 2, \dots, l-1$ . Now we pack the circle  $C_q$ ,  $q = n_1 + 1$ . By condition,  $r_q$  is bounded above by  $r^+$ , that can be, in general, less than the maximal possible radius  $r_l^+$ ,  $l \in I_p$ , the circle  $C_q(v_q, r_l^+)$  being inscribed into  $P_l$ . In this case the center  $v_q$  of the circle  $C_q(v_q, r^+)$  is free to move within a polygonal set  $\hat{P}_l \subset P_l$  (Figure 1) ensuring feasible locations of  $C_q(v_q, r^+)$  in  $P_l$ .

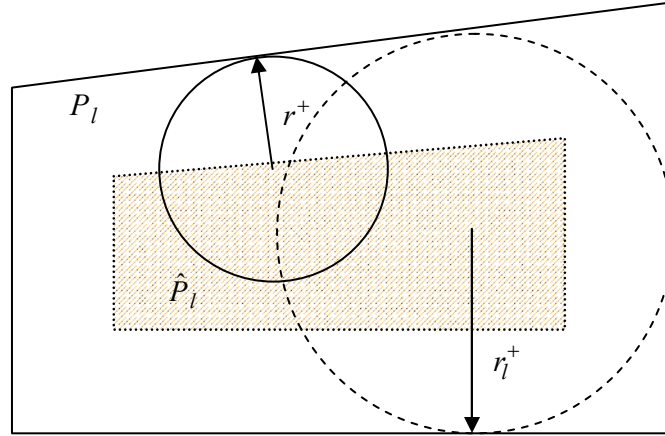


**Figure 1:** Packing the first circle into  $P_l$

A random position of the circle center is chosen and then a step-by-step procedure of sequential addition of next circles according to the algorithm [14, 25] is carried out. If a circle with radius  $r^+$  cannot be packed into  $P_l$ , we go to Stage 2 (ACP) starting from the circle with the maximal possible radius  $r_l^*$ ,  $l \in I_p$ . If there are several possible locations for the circle, we choose a location for which the circle is tangent to the maximal number of the polygon edges.

To this end we solve consequently the problems.

$$\max r_{q+k}, \quad k=0,1,2,\dots,k_l^* \quad (8)$$

$$\text{subject to } u_l^{(k)} \in W_l^{(k)}$$

where  $u_l^{(k)} = (x_q, y_q, r_q) \in W_l^{(k)} \subset \mathbf{R}^3$  if  $k=0$  and  $u_l^{(k)} = (x_q, y_q, \dots, x_{q+k}, y_{q+k}, r_{q+k}) \in W_l^{(k)} \subset \mathbf{R}^{2k+3}$  for  $k=1,2,\dots,k_l^*$ ;

$$W_l^{(k)} = \{ u_l^{(k)} \in \mathbf{R}^{2k+3} : \Psi_{ml}(x_q, y_q) - r_q \geq 0, \quad m=1,2,\dots,M_l, \quad l \in I_p, \quad (9)$$

$$(x_{q+j-1} - x_{q+t-1})^2 + (y_{q+j-1} - y_{q+t-1})^2 - (2r^+ + \rho)^2 \geq 0,$$

$$j, t = 1, 2, \dots, k, \quad j < t, \quad k \in \{2, \dots, k_l^*\},$$

$$(x_{q+j-1} - x_{q+k})^2 + (y_{q+j-1} - y_{q+k})^2 - (r_q^+ + r_{q+k} + d)^2 \geq 0,$$

$$j = 1, 2, \dots, k, \quad k \in \{1, 2, \dots, k_l^*\} \}.$$

A local maximum of the problem (8)-(9) is calculated. If  $r_{q+k}^* = r^+$ , then we continue to solve problem (8)-(9) for the next  $k$ , considering  $u_l^{(k)*} = (x_q^*, y_q^*, \dots, x_{q+k-1}^*, y_{q+k-1}^*, x_{q+k}, y_{q+k}, 0)$  where  $x_{q+k}, y_{q+k}$  are randomly chosen ( $u_l^{(k)*} \in W_l^{(k)}$ ) as a starting point. If  $r_{q+k}^* < r^+$ , we stop the iterative process. The point

$$u_l^* = (v_l^*, r_l^*) = (x_q^*, y_q^*, \dots, x_{q+k-1}^*, y_{q+k-1}^*, x_{q+k}, y_{q+k}, \underbrace{r_{q+k}^*}_{k_l^*-1}, r_{q+k_l^*}^*)$$

is the algorithm output. Then  $k = k_l^*$ , the circle  $C_{q+k_l^*}$  touches at least three circles (or edges of  $P_l$ ) and can be considered as a first circle packed according to ACP (Figure 2).

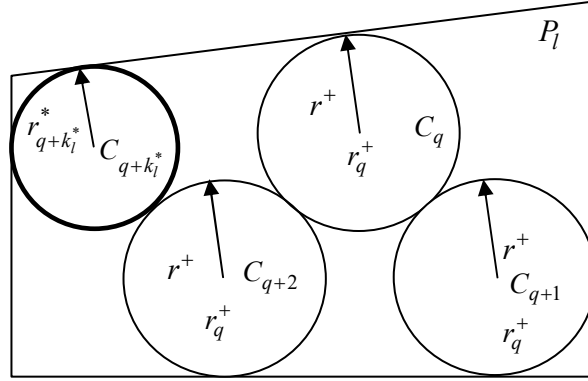


**Figure 2:** Construction of a starting circle configuration for ACP

#### 4.2.2. Algorithm 2. Generating starting feasible points

If  $r_{q+k}^*$  is a strict local maximum of the problem (6)-(7), then the circles  $C_q, C_{q+1}, \dots, C_{q+k_l^*}$  are rigidly fixed forming a “jammed” packing [26]. We pack next circles  $C_{q+k_l^*+1}, C_{q+k_l^*+2}, \dots, C_{q+\lambda_l-1}$  according to ACP until the current radius  $r_{q+k_l^*+k_{ASP}+1}^*$  where  $k_{ASP}$  is the number of additional circles following ACP becomes less than  $r^-$  (Figure 3). A point

$$w_l^* = (x_q^*, y_q^*, \dots, x_{q+\lambda_l-1}^*, y_{q+\lambda_l-1}^*, \underbrace{r_{q+k_l}^*, \dots, r_{q+\lambda_l-1}^*}_{k_l^*-1})$$

is obtained. The total number of circles packed into  $P_l$  is  $\lambda_l = k_l^* + k_{ASP}$ , the radii values being within  $[r^-, r^+]$ .

### 4.2.3. Local optimization

After having constructed the starting packing for each polygon  $P_l$ ,  $l \in I_p$ , we have the total number  $n = \sum_{l \in I_p} \lambda_l$  of circles to be packed into  $P$ .

We turn to problem (6)-(7), calculate a local maximum point and then construct a starting point  $\omega^0 = (w_1^*, w_2^*, \dots, w_{n_p}^*)$ . To solve the nonlinear programming problem we make use of IPOPT solver [23] together with the decomposition strategy [15].

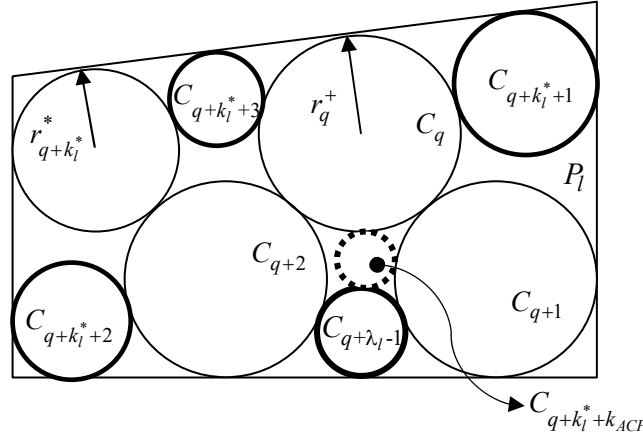


**Figure 3:** Circle arrangement according to ACP

Several local maximum points are computed and ordered in decreasing order of  $\kappa(\omega)$  (see problem (6)-(7)) and then one can choose a point meeting (7). The first point in the ordering for which the inequality (7) holds true is a trade-off solution of the problem (3)-(5). If there is no a point satisfying the threshold value  $\sigma_a$  (see (7)), new local maximum points are computed or a compromise value  $\sigma_a - \Delta\sigma$  is selected. Calculation of mechanical stress values is not considered in this paper. We refer to paper [10].

As computations show, local maximum points are close to or coincide with the starting points constructed according to ACP.

## 5. Computational results

We consider the benchmark geometry presented in [10] and provide new benchmark examples for packing circles with variable radii. The dependence of the number of circles packed and the packing density on the minimal allowed distance between circles is studied. All experiments were running on an Intel Core I5 750 computer. We use the Delphi programming language and the Windows 10 operation system. Software library IPOPT [23, 27] for nonlinear optimization problems exploiting first and second derivative (Hessians) information is utilized.

Example 1.  $P_0$  is given by 11 vertices  $v_{01} = (0,0)$ ,  $v_{02} = (0,9)$ ,  $v_{03} = (18,9)$ ,  $v_{04} = (5,28)$ ,  $v_{05} = (0,33)$ ,  $v_{06} = (0,40)$ ,  $v_{07} = (60,40)$ ,  $v_{08} = (76,34)$ ,  $v_{09} = (98,11)$ ,  $v_{0,10} = (100,6)$ ,

$v_{0,11} = (100,0)$ .  $P = \bigcup_{l \in I_p} P_l$ , where  $I_p = \{1,2,\dots,5\}$ , vertices of  $P_l$ ,  $l \in I_p$ , are  $v_{11} = (22,31)$ ,  $v_{12} = (35,27)$ ,  $v_{13} = (6,8)$ ;  $v_{21} = (40,25)$ ,  $v_{22} = (54,9)$ ,  $v_{23} = (12,7)$ ;  $v_{31} = (44,30)$ ,  $v_{32} = (68,12)$ ,  $v_{33} = (59,8)$ ,  $v_{34} = (42,28)$ ;  $v_{41} = (42,36)$ ,  $v_{42} = (63,36)$ ,  $v_{43} = (89,30)$ ,  $v_{44} = (73,13)$ ;  $v_{51} = (74,37)$ ,  $v_{52} = (95,37)$ ,  $v_{53} = (91,31)$ ,  $v_{54} = (74,35)$ . The minimal allowed distance between circles is  $\rho = 0$ ;  $r^- = 0.5$ ;  $r^+ = 5$ ;

The total number of circles packed into  $P$  is  $n = 91$  ( $\lambda_1 = 15, \lambda_2 = 23, \lambda_3 = 14, \lambda_4 = 31, \lambda_5 = 8$ ). Value of the objective in the starting point (ACP) is  $\kappa(\omega^0) = 361.3287$  and one in the local maximum point is  $\kappa^* = 362.0946$ . Illustration of the circles packed is shown in Figure 4. This example is relevant to maximizing the packing factor and meaningless in relation to mechanical stress being infinite at  $\rho = 0$ .

Example 2. See Example 1.  $\rho = 0.5$ . The starting objective value is  $\kappa(\omega^0) = 317.9666$  and the local maximum point is  $\kappa^* = 319.3224$ . Illustration is shown in Figure 5.  $n = 66$  ( $\lambda_1 = 11, \lambda_2 = 19, \lambda_3 = 10, \lambda_4 = 21, \lambda_5 = 5$ ).

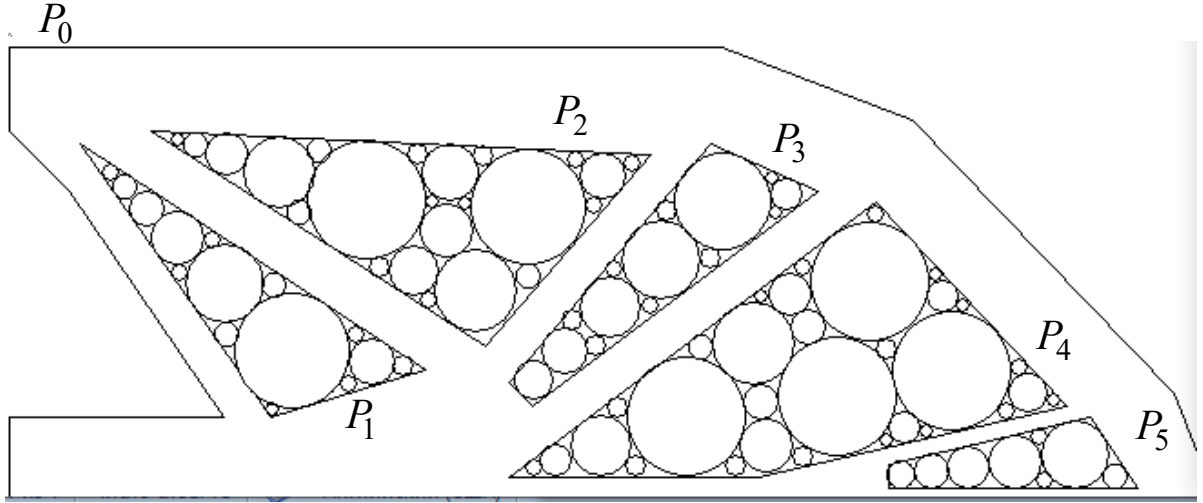


**Figure 4:** Circle arrangement according to Example 1

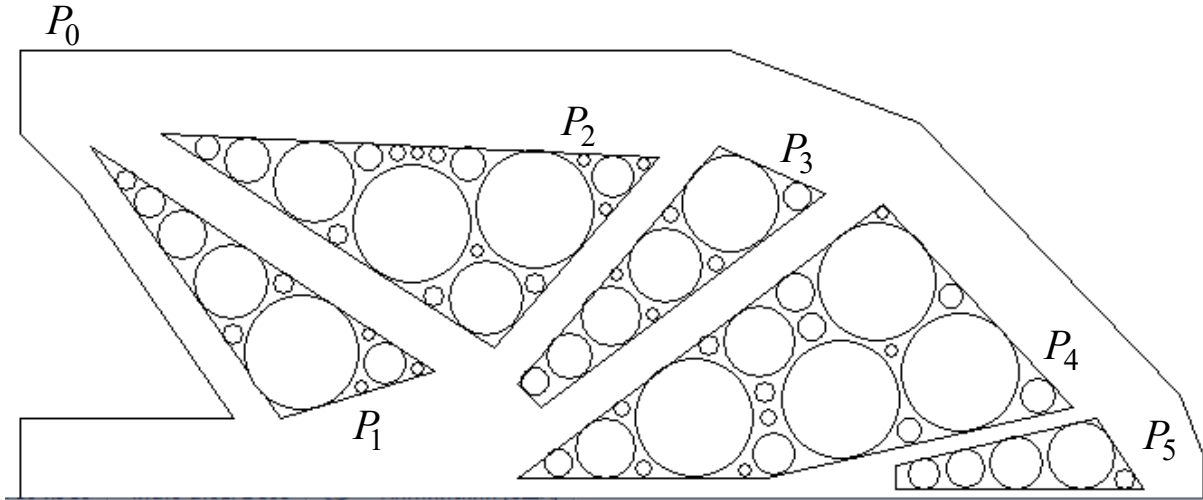


**Figure 5:** Circle arrangement according to Example 2

Example 3. See Example 1.  $\rho = 0.75$ . The starting objective value is  $\kappa(\omega^0) = 301.0078$  and the local maximum point is  $\kappa^* = 301.7119$ . Illustration is shown in Figure 6.  $n = 50$  ( $\lambda_1 = 8, \lambda_2 = 12, \lambda_3 = 8, \lambda_4 = 17, \lambda_5 = 5$ ).

Example 4. See Example 1.  $\rho = 1$ . The starting objective value is  $\kappa(\omega^0) = 288.9095$  and the local maximum point is  $\kappa^* = 289.7335$ . Illustration is shown in Figure 7.  $n = 47$  ( $\lambda_1 = 8, \lambda_2 = 12, \lambda_3 = 8, \lambda_4 = 14, \lambda_5 = 5$ ).

The runtime is within 30 seconds for all examples.

As computational results show, with an increase in the minimal allowed distance between circles, the contribution of large circles to the total circle area increases, since it is proportional to the squares of the circle radii. Obviously, this leads to a decrease in the number of circles with small radii and an increase in the radii of other circles.

In this regard, the choice of  $\rho$  is of importance. A compromise value is chosen: a too low value can lead to poor quality printing due to the technical features of the process, while a too large one causes additional material consumption. Furthermore, the secondary objective  $\sigma(\omega)$  can be influenced by varying the values of  $\rho$ ,  $r^-$  and  $r^+$ .

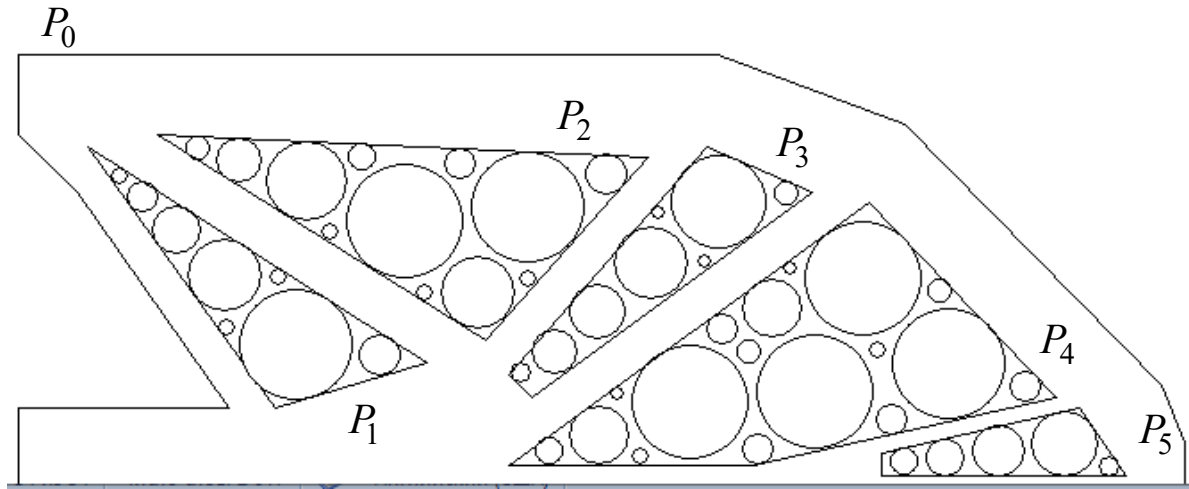


**Figure 6:** Circle arrangement according to Example 3

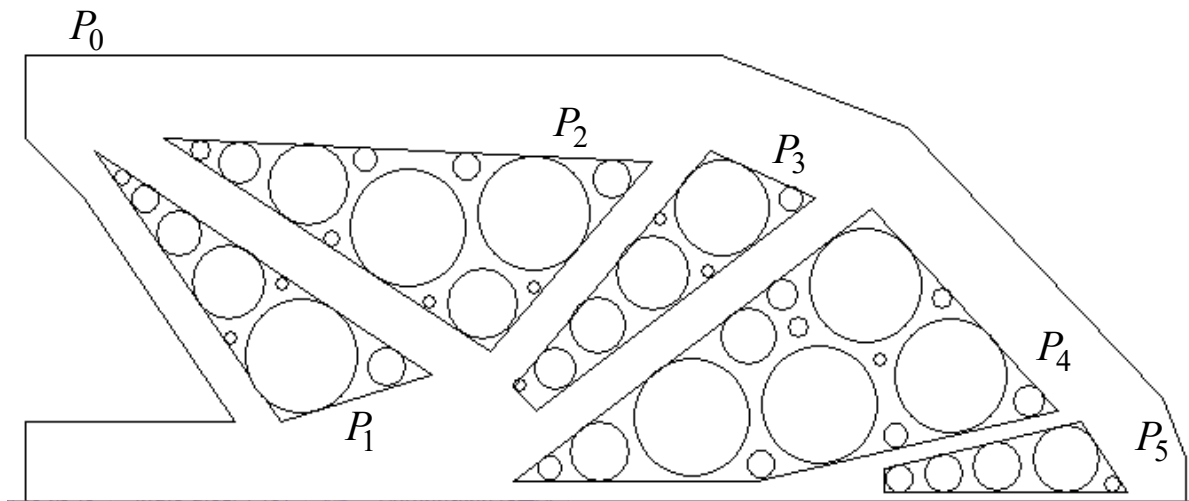


**Figure 7:** Circle arrangement according to Example 4



## 6. Conclusions

Topology optimization is a key point in additive manufacturing. Circular-hole based layout models help to cut down material expenses while meeting strength requirements.

We formulate non-standard packing problem of circles with variable metric characteristics. The proposed bi-objective model takes into account both the maximum packing factor and mechanical stress of parts. An efficient packing algorithm based on Apollonian circle packing, nonlinear programming and the  $\varepsilon$ -constraint method has been developed. The approach allows estimating the number of circular perforations needed and search for an approximate solution of the problem maximizing the packing factor and following a threshold value of mechanical stress.

Further research is directed to layout of circular perforations inside a 3D part considering balancing conditions [22, 28]. Therewith a more nuanced approach to multi-objective optimization should be constructed.

## 7. Acknowledgements

The investigation is 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) and National Research Foundation of Ukraine (#02.2020/167).

## 8. References

- [1] O. Sigmund, K. Maute, Struct topology optimization approaches, *Structural and Multidisciplinary Optimization* 48 (2013) 1031–1055. doi:10.1007/s00158-013-0978-6.
- [2] J. Liu, Y. Ma, A survey of manufacturing oriented topology optimization methods, *Advances in Engineering Softwar* 100 (2016) 161–175. doi:10.1016/j.advengsoft.2016.07.017.
- [3] A. Ramya, S. I. Vanapalli, 3D printing technologies in various applications, *International Journal of Mechanical Engineering and Technology* 7 (2016) 396–409. Available from: <http://www.iaeme.com/currentissue.asp?JType=IJMET&VType=7&IType=3>.
- [4] J. D. Deaton, R. V. Grandhi, A survey of structural and multidisciplinary continuum topology optimization: post 2000, *Structural and Multidisciplinary Optimization* 49 (2014) 1–38. doi:10.1007/s00158-013-0956-z.
- [5] K. Miettinen, *Nonlinear Multiobjective Optimization*, Springer Science & Business Media, 2012.
- [6] N. Gunantara, Q. Ai, A review of multi-objective optimization: Methods and its applications, *Cogent Engineering*, 5:1 (2018). doi:10.1080/23311916.2018.1502242.
- [7] Y. Y. Haimes, L. S. Lasdon, D. A. Wismer, On a bicriterion formulation of the problems of integrated system identification and system optimization, *IEEE Trans. Syst. Man. Cybern.* 1 (1971) 296–297. doi:10.1109/TSMC.1971.4308298.
- [8] W. Chen, M. Jiao, C. Kessler, A. Malik, X. Zhang, Spatial Statistics of Apollonian Gaskets, *Experimental Mathematics* 28 (2019) 263270. doi:10.1080/10586458.2017.1385037.
- [9] F. J. Kampas, J. D. Pintér, I. Castillo, Packing ovals in optimized regular polygons, *J Glob Optim* 77 (2020) 175–196. doi:10.1007/s10898-019-00824-8.
- [10] S. P. Fekete, S. Morr, C. Scheffer, Split Packing: Algorithms for Packing Circles with Optimal Worst-Case Density, *Discrete Comput Geom* 61 (2019) 562–594. doi:10.1007/s00454-018-0020-2.
- [11] Y. Stoyan, G. Yaskov, T. Romanova, I. Litvinchev, S. Yakovlev, J. M. Velarde Cantú, Optimized packing multidimensional hyperspheres: a unified approach, *Math Biosci Eng.* 17 (2020) 6601–6630. doi:10.3934/mbe.2020344.
- [12] S. Yakovlev, O. Kartashov, K. Korobchynskiy, B. Skripka, Numerical Results of Variable Radii Method in the Unequal Circles Packing Problem, in: *Proceedings of 2019 IEEE 15th International Conference on the Experience of Designing and Application of CAD Systems (CADSM)*, Polyana, Ukraine, 2019, pp. 1–4. doi:10.1109/CADSM.2019.8779288.

- [13] Y. Stoyan, G. Yaskov, Packing equal circles into a circle with circular prohibited areas, *International Journal of Computer Mathematics* 89 (2012) 1355–1369. doi:10.1080/00207160.2012.685468.
- [14] S. Yakovlev, The Expanding Space Method in Sphere Packing Problem, in: S. Babichev, V. Lytvynenko, W. Wójcik, S. Vyshemyrskaya (Eds.), *Lecture Notes in Computational Intelligence and Decision Making, ISDMCI 2020, Advances in Intelligent Systems and Computing*, volume 1246, Springer, Cham, 2021, pp. 151–163. doi:10.1007/978-3-030-54215-3\_10.
- [15] E. Specht, [www.packomania.com](http://packomania.com), 2020. Available from: <http://packomania.com>.
- [16] I. Yanchevskyi, R. Lachmayer, I. Mozgova, R-B. Lippert, G. Yaskov, T. Romanova, I. Litvinchev, Circular packing for support-free structures, *EAI Endorsed Transactions* 7 (2020). doi:10.4108/eai.13-7-2018.164561.
- [17] R. A. Alaimo, Overlap Packing Optimization for Spacecraft Layout Design. M.S. thesis. The University of North Carolina, Charlotte, NC, 2018.
- [18] J Lim, C. You, I. Dayyani, Multi-objective topology optimization and structural analysis of periodic spaceframe structures, *Materials & Design* 190 (2020) 108552. doi:10.1016/j.matdes.2020.108552.
- [19] T. Romanova, Y. Stoyan, A. Pankratov, I. Litvinchev, I. Yanchevsky, I. Mozgova, Optimal Packing in Additive Manufacturing, *IFAC-PapersOnLine* 52 (2019) 2758–2763. doi:10.1016/j.ifacol.2019.11.625.
- [20] E. M. Kiseleva, Y. E. Kadochnikova, Solving a Continuous Single-product Problem of Optimal Partitioning with Additional Conditions, *Journal of Automation and Information Sciences* 41 (2009) 48–63. doi: 10.1615/JAutomatInfScien.v41.i7.30.
- [21] N. Chernov, Y. Stoyan, T. Romanova, Mathematical model and efficient algorithms for object packing problem, *Computational Geometry: Theory and Applications* 43 (2014) 535–553. doi:10.1016/j.comgeo.2009.12.003.
- [22] P. I. Stetsyuk, T. E. Romanova, G. Scheithauer, On the global minimum in a balanced circular packing problem, *Optim Lett* 10 (2016) 1347–1360. doi:10.1007/s11590-015-0937-9.
- [23] A. Wächter, L. T. Biegler, On the implementation of a primal-dual interior point filter line search algorithm for large-scale nonlinear programming, *Math. Program.* 106 (2006) 25–57. doi:10.1007/s10107-004-0559-y.
- [24] G. Waescher, H. Haussner, An improved typology of cutting and packing problems. *European Journal of Operational Research* 183 (2007) 1109–1130. doi:10.1016/j.ejor.2005.12.047.
- [25] G. Yaskov, A. Chugay, Packing equal spheres by means of the block coordinate descent method, *CMIS* (2020) 156–168.
- [26] S. Torquato, A. Donev, F. H. Stillinger, Breakdown of elasticity theory for jammed hard-particle packings: conical nonlinear constitutive theory, *International Journal of Solids and Structures* 40 (2003) 7143–7153. doi:10.1016/S0020-7683(03)00359-7.
- [27] <http://www.coin-or.org/download/source/Ipopt>.
- [28] I. V. Grebennik, A. A. Kovalenko, T. E. Romanova, I. A. Urniaieva, S. B. Shekhovtsov, Combinatorial Configurations in Balance Layout Optimization Problems, *Cybern Syst Anal* 54 (2018) 221–231. doi:10.1007/s10559-018-0023-2.

# Optimization of data transmission system information parameters for complex technical system's state diagnosing

Vladimir Vychuzhanin<sup>a</sup>, Natalia Shibaeva<sup>a</sup>, Nickolay Rudnichenko<sup>a</sup>, Alexey Vychuzhanin<sup>a</sup>

<sup>a</sup> Odessa National Polytechnic University, Shevchenko Avenue 1, Odessa, 65001, Ukraine

## Abstract

Complex technical systems are often operated in tense, abnormal conditions, with a large number of correlated parameters. Timely and high-quality diagnostics, including remote diagnostics, of complex technical systems components can increase their reliability. To successfully solve the problem of such systems trouble-free operation in abnormal operating modes, it is necessary to maximize the use of information technologies with software and hardware diagnostic modules. In order to make a reasonable choice of system's information parameters for diagnosing complex technical system's state it is necessary to solve the problems of multi-parameter, multi-criteria optimization data transmission systems. Such task aimed for increasing the speed, minimizing the error and risk of system's elements failure, and maximizing the transmitted information protection. The objective function of optimizing the information parameters of the data transmission systems for remote diagnostics is its multi-parameter optimization of the variables that affect the remote diagnostics of all system's efficiency. Solving the problem of optimizing the remote data transmission systems for remote diagnostics information parameters based on the developed model allows us to find several Pareto-optimal solutions for quality indicators, which are affecting the efficiency of the data transmission systems for remote diagnostics. Basic parameters and characteristics of the data transmission system are the optimal number of devices types for monitoring complex technical systems technological parameters, topology efficiency, throughput, speed, error of the data transmission system, the risk of component failures and the information protection effectiveness in the remote diagnostics system.

## Keywords

Complex technical systems, diagnostics, information parameters, data transmission system, multi-parameter, multi-criteria optimization, Pareto-optimal solutions, throughput, error.

## 1. Introduction

Considering the problem of diagnosing complex technical systems (CTS), these are objects often with multimodality, uncertainty of behavior, a large number of correlated parameters, functioning in tense, abnormal modes [1–3]. The purpose of CTS state diagnosing is to improve reliability, prevent accidents that cause shutdown and damage to systems, in many cases accompanied by human casualties and damage to the environment [4–6].

## 2. Description of Problem

The analysis of technical solutions that improve the reliability of the CTS operation showed that timely and high-quality diagnostics, including remote diagnostics, of the CTS components can

CMIS-2021: The Fourth International Workshop on Computer Modeling and Intelligent Systems, April 27, 2021, Zaporizhzhia, Ukraine  
EMAIL: vint532@yandex.ua (V. Vychuzhanin); nati.sh@gmail.com (N. Shibaeva); nickolay.rud@gmail.com (N. Rudnichenko); vint532@gmail.com (A. Vychuzhanin)  
ORCID: 0000-0002-6302-1832 (V. Vychuzhanin); 0000-0002-7869-9953 (N. Shibaeva); 0000-0002-7343-8076 (N. Rudnichenko); 0000-0001-8779-2503 (A. Vychuzhanin)



© 2021 Copyright for this paper by its authors.  
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).  
CEUR Workshop Proceedings (CEUR-WS.org)

improve the reliability of systems and their operation efficiency [4,7]. To successfully solve problem of CTS accident-free operation in abnormal operating modes, it is necessary to maximize the use of information technology with software and hardware diagnostic modules [4,8].

These can be elements and devices in Distributed Industrial Control Systems (DICS) serving technological objects of "critical infrastructure" (energy, transport, mining and processing of minerals, communications, national security, life support systems, etc.).

The system allows combining into a single system the lower level of technological subsystems for collecting, storing, transmission and distributing information with the upper level of control, monitoring and diagnostics.

To effectively perform their functions to maintain the high quality of processes and the specified characteristics of channels and paths for transmitting technological and other information, the CTS state diagnostic system must provide [9]:

- large flows information processing generated by a significant number of controlled parameters;
- increasing the reliability and reducing the redundancy of information at its processing stages;
- the speed and functioning of the system in real time, namely, the processing of information arriving at its inputs (reception and transmission of data between system's functional elements, synchronization of the information further analysis process) at a rate not lower than the rate of events development in the CTS.

In order to make a reasonable choice of the data transmission system (DTS) topology for diagnosing, its high-quality operational parameters, it is necessary to solve the problems of multi-parameter, multi-criteria optimization of information parameters aimed at increasing the speed, minimizing errors and the risk system elements failures, and maximizing the transmitted information protection [10]. The use of well-known optimization methods for solving practical problems of multi-criteria, multi-parameter optimization [10 - 12] should take into account: large dimension of problems (tens and hundreds of variables and constraints); topological complexity of the optimized function; significant computational costs; the need to solve the problem in a multicriteria formulation, the use of unrelated models. Thus, ensuring the quality of the functioning of the diagnostic information system (including the remote one) with optimal parameters is an urgent scientific and technical problem. The aim of the work is to maximize CTS remote diagnostics system efficiency by optimizing its information parameters.

### **3. Data transmission system for complex technical system state remote diagnostics development**

Operation of a single open platform for CTS remote diagnostics systems should be based on ISO standards in the field of "Condition monitoring and diagnostics of machines" [13]. The use of a standard base allows for the unification of approaches to obtaining information from primary measuring instruments and its transmission.

In Fig. 1 shows the block diagram of the DTS is shown for diagnosing the state of CTS in a distributed information management system.

Any ship's CTS can be selected as an object of remote diagnostics.

The remote diagnostic (RD) system consists of two subsystems, one of which is installed on the ship, the other in the cloud or the coastal dispatch center.

The main method for diagnosing the state of CTS is the collection, processing and analysis of measurement information.

In order to optimize the information parameters of the DTS system of CTS ship's state, data transmission system model shown in Fig. 2.

Information transmission is carried out based on the use of one of the standards for digital information transmission systems IEEE 802.15, WiMax, IEEE 802.22, UMTS, LTE.

The system parameters are selected in such a way as to ensure a certain level of information transmission quality.

The accumulated diagnostic data of the CTS condition is sent in the form of reports to the coastal diagnostic center.

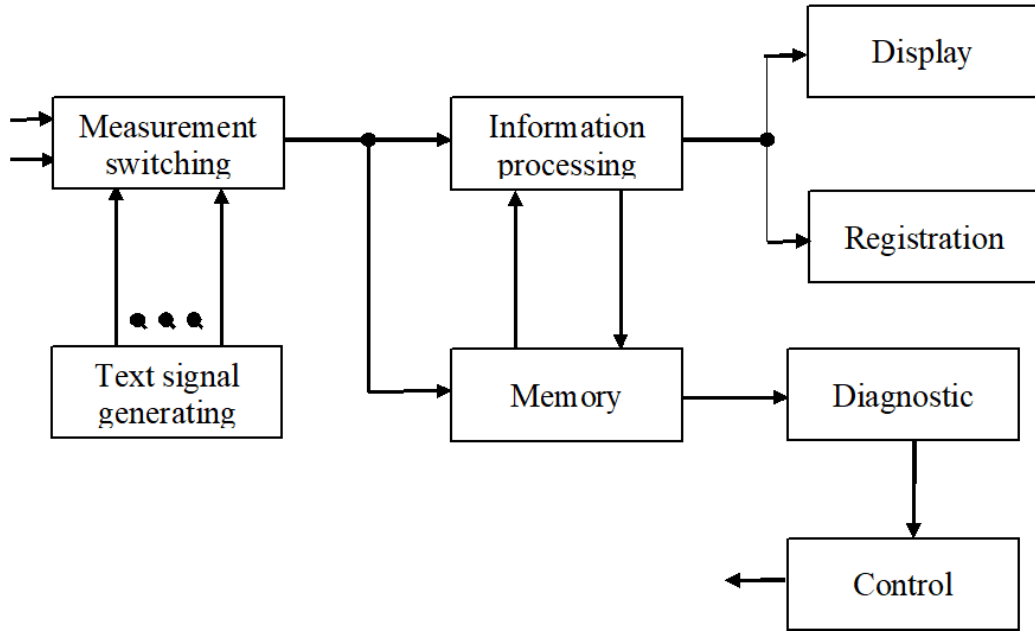


**Figure 1:** Block diagram DTS for diagnosing the state of CTS in a distributed information management system

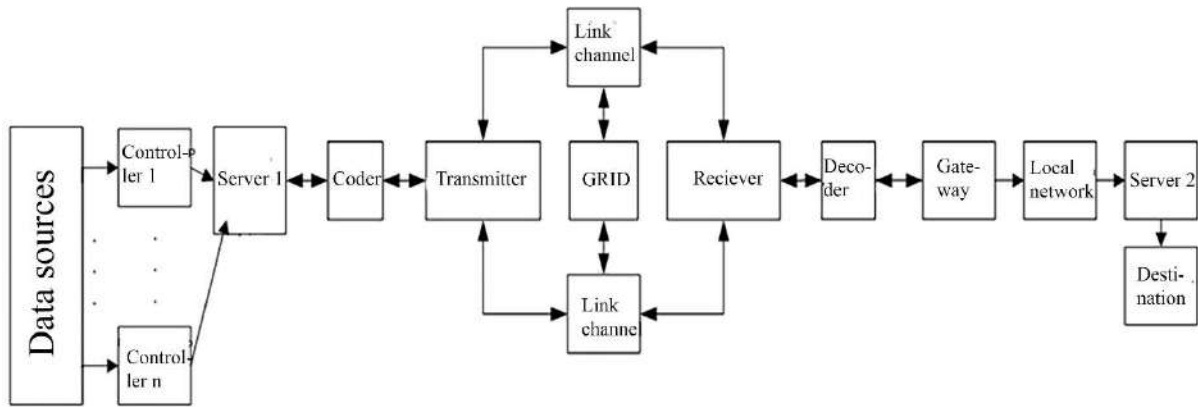


**Figure 2:** Block diagram of the DTS model for diagnosing the state of the ship's CTS

The quality of the DTS for RD system model is define by a set of characteristics that affect the efficiency of its operation: the topology ( $F_o$ ); throughput ( $B$ ); performance ( $T_\Sigma$ ); permissible data transmission error ( $\sigma$ ); efficiency of information protection in the system ( $Z_\Sigma$ ), risk of DTS device failures in RD( $R_\Sigma$ ).

We assume that the model describing the RD system is linear, with both deterministic and stochastic lumped parameters.

The set of quality indicators of the model can be written in the vector form, the coordinates of which are particular indicators, and their specified values must be ensured (increased, improved to the desired level)

$$Q = \{\{F_o\}, \{T_\Sigma\}, \{B\}, \{\delta\}, \{R_\Sigma\}, \{Z_\Sigma\}\} \quad (1)$$

The generalized indicator of the quality of the functioning of the DTS in RD system is determined by the results of information parameters optimization, as well as its reliability characteristics. The objective function of optimizing the information parameters of the DTS in RD model of the CTS state is its variables multiparameter optimization that affect system's efficiency.

Data transmission system topology indicators optimization objective function can be presented is a such form

$$\varphi(F) = \max \varphi(F_o) = \max \varphi(D, B, C), \quad (2)$$

$$D = \{D_o \in D | D_{\min} \angle D \angle D_{\max}\}, \quad K = \{K_o \in K | K_{\min} \angle K \angle K_{\max}\}, \quad C = \{C_o \in C | 0 \angle C \angle C_o\}$$

where  $D$  is length of data transmission paths;

$K$  is compact structure of the system;

$C$  is degree of system structure centralization;

$D_0, K_0, C_0$  are normalized partial criteria indicators of the system topology efficiency obtained by reducing the indicators  $D, B, C$  to dimensionless form.

Objective function of optimizing the response time of the data transmission system RD

$$\varphi(T) = \max \varphi(T_\Sigma) = \max \varphi\left(\sum_{l=1}^p T_{S_{no_l}}, T_{KO}, T_{C1}, T_K, T_{PD}, T_{KC}, T_{PK}, T_{DK}, T_S, T_{LC}, T_{C2}\right), \quad (3)$$

$$T = \{T_{S_{no_l}} \in T | T_{S_{o_{\min}}} \angle T_{S_{o_l}} \angle T_{S_{o_{\max}}}\}, \quad T = \{T_{KO} \in T | T_{KO_{\min}} \angle T_{KO} \angle T_{KO_{\max}}\}, \quad T = \{T_{C1} \in T | T_{C1_{\min}} \angle T_{C1} \angle T_{C1_{\max}}\},$$

$$T = \{T_{PK} \in T | T_{PK_{\min}} \angle T_{PK} \angle T_{PK_{\max}}\}, \quad T = \{T_{DK} \in T | T_{DK_{\min}} \angle T_{DK} \angle T_{DK_{\max}}\}, \quad T = \{T_{C2} \in T | T_{C2_{\min}} \angle T_{C2} \angle T_{C2_{\max}}\},$$

$$T = \{T_K \in T | T_{K_{\min}} \angle T_K \angle T_{K_{\max}}\}, \quad T = \{T_{LC} \in T | T_{LC_{\min}} \angle T_{LC} \angle T_{LC_{\max}}\}, \quad T = \{T_S \in T | T_{S_{\min}} \angle T_S \angle T_{S_{\max}}\},$$

$$T = \{T_{KC} \in T | T_{KC_{\min}} \angle T_{KC} \angle T_{KC_{\max}}\}$$

$$\varphi(T) = \min \varphi(T_Z),$$

$$T = \{T_Z \in T | T_{Z_{\min}} \angle T_Z \angle T_{Z_{\max}}\}$$

$$\varphi(T_K) = \max \varphi(n, T_L),$$

$$T = \{T_L \in T | T_{L_{\min}} \angle T_L \angle T_{L_{\max}}\},$$

$$\varphi(T_{LC}) = \max \varphi(B_{LC}, f), \quad \varphi(T_S) = \max \varphi(B_S, f), \quad \varphi(T_{KC}) = \max \varphi(B_{KC}, f),$$

$$\varphi(B) = \max \varphi(B_{LC}, B_S, B_{KC}),$$

$$B = \{B_{LC} \in B | B_{LC_{\min}} \angle B_{LC} \angle B_{LC_{\max}}\}, \quad B = \{B_S \in B | B_{S_{\min}} \angle B_S \angle B_{S_{\max}}\}, \quad B = \{B_{KC} \in B | B_{KC_{\min}} \angle B_{KC} \angle B_{KC_{\max}}\},$$

where  $T_{S_{no_i}}$  is technological parameters control devices performance;

$T_{KO}$  is controller performance;

$T_{C1}, T_{C2}$  are server performance;

$T_K, T_{DK}$  are encoder and decoder performance;

$T_{PD}$  is transmitter performance;

$T_{PK}$  is receiver performance;

$T_{LC}$  is processing speed of network streams in the receiving local network;

$T_S$  is processing speed of network packets in the receiving transit gateway;

$T_Z$  is signal delay at redundant nodes;

$f$  is network packet size;

$B_{LC}$  is host local network bandwidth;

$B_S$  is receiving transport gateway throughput;

$B_{KC}$  is communication channel capacity;

$n$  is code combination length;

$T_L$  is time required to transmit the codeword.

Objective function of RD system data transmission error

$$\varphi(\delta) = \min \varphi(\delta_{S_{n0}}, \delta_{SPD}), \quad (4)$$

$$\delta_{S_{n0}} = \frac{\delta}{X_{MV} \cdot 100}, \quad \delta_{SPD} = \frac{|y(t) - y'(t)|}{Y}, \quad \delta = [\delta_{S_{n0}} \in 0;1], \quad \delta = [\delta_{SPD} \in 0;1],$$

where  $\delta$  is relative measurement error;

$\delta_{S_{n0}}$  are technological parameters control devices primary converters errors;

$\delta_{SPD}$  are errors arising from data loss of the data transmission system;

$X_{MV}$  is measured value;

$y(t)$  is perfect signal;

$y'(t)$  is real signal;

$Y$  is signal variation range;

$t$  is signal observation time ( $t_o \angle t \angle t_N$ )

The task of optimizing information protection is to ensure the maximum level of security with a minimum risk from the likelihood of hacking - the data transmission system, i.e.

$$\varphi(Z_\Sigma) = \min \varphi(R_{HACK}, T_a), \quad \varphi(Z_\Sigma) = \max \varphi(N_f) \quad (5)$$

at  $N_f \geq N_{f0}, \quad T_a \leq T_{a0}, \quad R_{HACK} = p_{HACK} \cdot H, \quad p_{HACK} = p_i \cdot p_a \cdot p_b \cdot p_c \cdot p_d,$

where  $R_{HACK}$  is multiplicative criterion of risk from the probability of hacking the data transmission system;

$p_{HACK}$  is the probability of an information system being hacked, determined on the basis of expert data;

$p_i$  is the probability that this information system falls into the list of possible targets and such on the information system;

$p_a$  is the probability that the system will be selected from the list and attacked;

$p_b$  is the probability that bordering technologies will be hacked;

$p_c$  is the probability that attacks will intensify;

$p_d$  is the probability that system will be harmed;

$H$  – potential damage from information security breaches;

$N_f$  is the number of functions that characterize the functionality of the data transmission system;

$T_a$  is average time of access to data transmission system protection objects;

$N_{f0}$  and  $T_{a0}$  are specified restrictions on the functionality and performance of the data transmission system.

The risk of RD structure failure is determined by the risk of CTS ( $R_{S_{n0}}$ ) parameter control device failure, as well as the average risk of the data transmission system ( $R_{SPD}$ ).

Objective function of the system RD devices failure risk

$$\varphi(R) = \min \varphi(R_{\Sigma}, R_{S_{n0}}, R_{SPD}, p_{S_{n0}}, H_{S_{n0}}, R_i, P(S_i)) \quad (6)$$

CTS parameter monitors failure risk

$$R_{S_{n0}} = p_{S_{n0}} \cdot H_{S_{n0}}, \quad (7)$$

where  $p_{S_{n0}}$  is the probability of CTS parameter control devices failure, determined on the basis of expert estimates;

$H_{S_{n0}}$  is damage from failure of CTS parameter control devices;

$R_i$  is contingent transmission risk;

$P(S_i)$  is conditional transmission error probability.

Conditional risk in data transmission is determined by the sum of error probabilities

$$R_i = \sum_{i=1}^m l_{ij} \cdot p(S_j / S_i) \quad (8)$$

where  $S_j$  is erroneous signal;

$S_i$  is original (correct signal);

$l_{ij}$  are losses arising from an erroneous decision due to a system error;

$p(S_j / S_i)$  is error probability.

#### 4. Experiments and results analysis

The solution to the optimization problem of the CTS state information parameters DTS in RD is based on the use of the developed optimization models. The goal is information DTS in RD effectiveness indication.

The model is designed to optimize the information parameters of the DTS in RD of CTS system in terms of the diagnostic system efficiency.

The method used, which makes it possible to study the developed model and optimize the information parameters of the DTS in state RD system, is based on the presence of many conflicting requirements for such system.

Optimal solutions to a multicriteria problem should be sought only among the elements of alternatives set. In this area, no criterion can be improved without deteriorating at least one of the others.

The complexity of solving multicriteria optimization problems is that the criteria compete with each other.

In most practical problems, the search for a more preferable solution according to one criterion leads to the fact that the solution becomes less preferable according to another criterion, i.e. solutions are incomparable.

The problem can be solved by applying the Pareto optimality principle [14,15]. The property of the optimal Pareto set is the ability to "reject" from the set of alternatives obviously unsuccessful ones that are inferior to others in all criteria.

As a result of solving the problem of optimizing information parameters, a set of alternative solutions is determined that correspond to the Pareto optimality principle and satisfy the set constraints.

The statement of the problem of multicriteria optimization includes three components, namely, the set of possible solutions, the vector criterion and the decision maker preference relation [10]. This



strategy for solving optimization problems significantly differs from the known approaches of nonlinear programming, has a higher efficiency and provides significantly wider possibilities.

The sequence of the RD system information parameters optimization includes the following stages:

- determination of a set of independent parameters, as well as conditions that determine the permissible values of the variables;
- obtaining the objective function as a measure of the quality of the optimization object with the given variables;
- choice of method and solution of the optimization problem.

Thus, the formulation of the optimization problem for the information parameters of the RD system will make it possible to develop a model for multi-criteria, RD system multi-parameter optimization.

To study the model of multicriteria and multiparametric optimization of DTS in RD information parameters, one can use algorithms implemented in the free software IOSO 3.3 [16], based on the technology of constructing a response surface.

A distinctive feature of this technology is the high efficiency of finding the optimal solution in the study of the DTS in RD system, modeled at high levels of complexity and hierarchy, including the achievements of mathematical modeling (2- and 3-dimensional problems) and the ability to quickly integrate with such computational packages as ANSYS, NASTRAN, TaskFlow, Star - CD, FineDesign, etc.

The program implements an optimization algorithm that does not belong either to gradient methods of nonlinear programming, or to genetic ones.

Therefore, unlike genetic algorithms, it has a high convergence rate (the minimum required number of calls to the user model) and is characterized by the fact that:

- no starting points are needed to start work;
- the algorithm allows to successfully solve optimization problems in the presence of non-differentiability, noisiness, local non-computability of objective functions and limited parameters;
- it allows us to find several Pareto-optimal solutions within the framework of one task;
- has good ability to find the global extremum on multi-extreme problems.

The program is designed for the numerical solution of problems of multicriteria parametric optimization of complex functional dependencies in the presence of functional constraints and operates taking into account the objective functions (1) - (6). The ranges of the numerical values of the RD system quality indicators are given in Table 1.

**Table 2**

Ranges of numerical values of the DTS in RD system quality indicators

| Level of quality                                  | Limitations |          |
|---------------------------------------------------|-------------|----------|
|                                                   | Min         | Max      |
| Bandwidth (B)                                     | 10 MB/s     | 100 MB/s |
| Effectiveness of information protection ( $Z_z$ ) | 0.95        | 1.0      |
| Structure risks ( $R_z$ )                         | 0.2         | 0.37     |
| Data transmission error tolerance ( $\sigma$ )    | 0.5 byte    | 1.0 byte |

The solution to the optimization problem of information parameters DTS in RD is to find the maximum efficiency of the such system under certain states of its indicators.

In order to optimize the information parameters of the ship CTS RD system state the structure of the software for the functioning of the RD system has been developed (Fig. 3).

Data transmission system capacity is determined by the maximum bandwidth of the DTS in RD system, the receiving LAN, and the receiving DTS transport gateway.

The maximum throughput of a communication channel with additive noise is determined by Shannon

$$B = W \log_2 \left( 1 + \frac{S}{W \cdot N} \right) \quad (9)$$

DTS bandwidth objective function

$$\varphi(B) = \max \varphi(W, S/N) \quad (10)$$

As an example, Fig. 4 shows the results when solving the DTS bandwidth optimization problem for  $W$  is channel bandwidth,  $S/N_0$  is signal-to-noise ratio at the destination receiver.

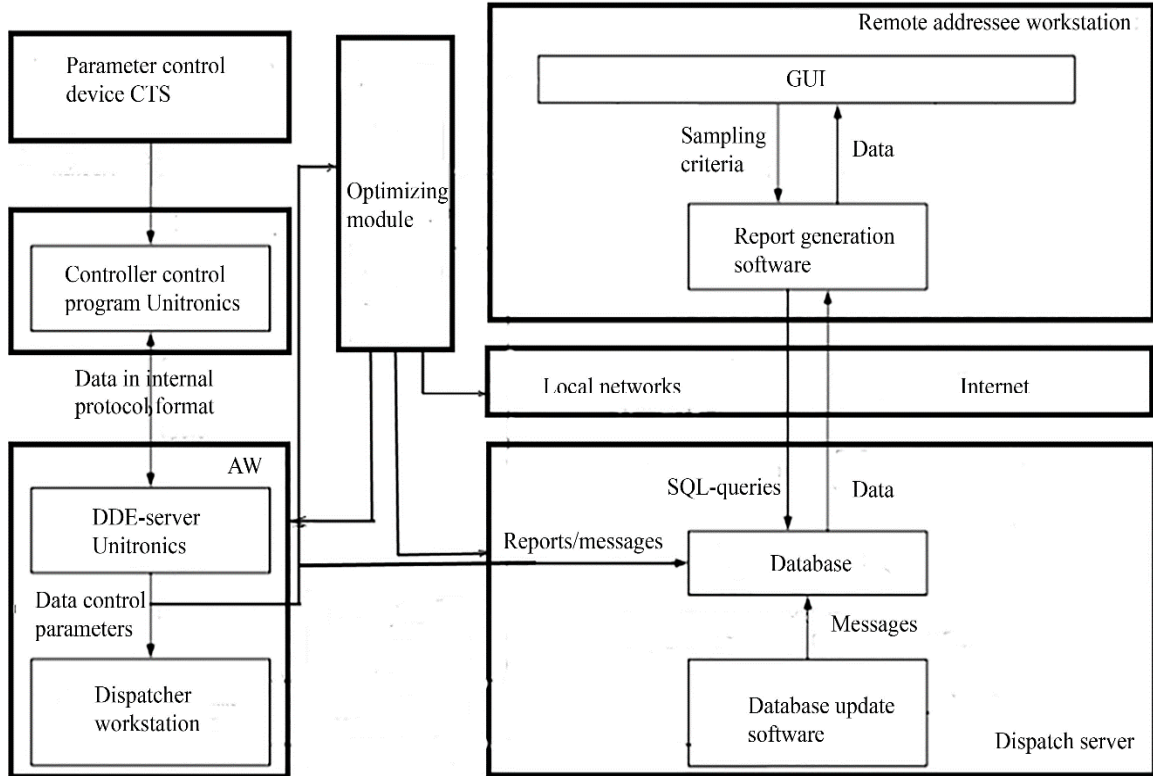


**Figure 3:** The structure of the DTS in RD software

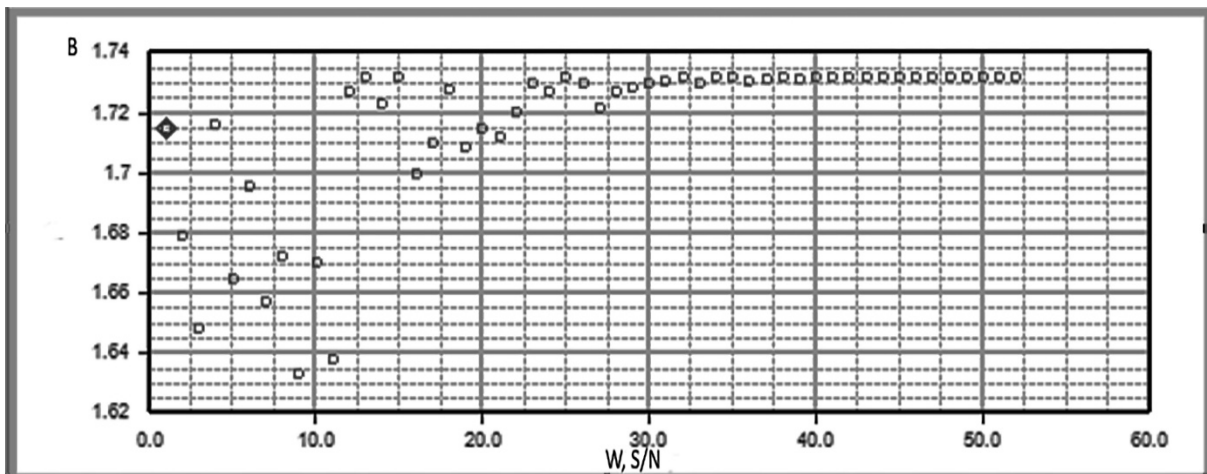


**Figure 4:** The set of Pareto-optimal solutions for the capacity  $B$  relative to the aggregate  $W$  and  $S/N$

The solution to the optimization problem of the information parameters of the DTS in RD system based on the developed model allows us to find several Pareto-optimal solutions for the quality indicator (criterion) - the DTS throughput, which affects the efficiency of the RD system.

## 5. Conclusion

In the formulation and solution of the optimization problem for the information parameters DTS in RD of the CTS state, a set of independent parameters and conditions determining their admissible values were determined, objective functions were obtained, and a method for solving optimization problems was selected.

The developed model of optimization of information characteristics of the CTS state DTS in RD system allows:

- exclude the influence of the human factor when taking into account the CTS various technological parameters;
- generate reports on CTS states at any time for any reporting period for subsequent analysis;
- monitor the state of CTS in real time, which will avoid accidents during their operation;
- to reduce the risk of DTS in RD elements failures;
- improve the DTS efficiency.

The solution to the optimization problem of information parameters DTS in RD on the basis of the developed model makes it possible to find several Pareto-optimal solutions for indicators (criteria) of quality that affect the DTS in RD system efficiency, the main of which are: the optimal number of types of devices for controlling CTS technological parameters; topology efficiency; bandwidth; speed; system error; the risk of component failures; efficiency of information protection.

## 6. References

- [1] V.V. Vychuzhanin, N.D. Rudnichenko, Assessment of risks structurally and functionally complex technical systems, *Eastern-European Journal of Enterprise Technologies* 2 (2014) 18–22. doi:10.15587/1729-4061.2014.19846
- [2] A.N. Dorokhov, V.A. Kernozhitskiy, A.N. Mironov, O.L. Shestopalova, *Ensuring the reliability of complex technical systems BHV*, Saint Petersburg, 2011.
- [3] V.Y. Tsvetkov, *Complex technical systems, Educational resources and technologies* 20 (2017) 86–92.
- [4] V.V. Vychuzhanin, N.D. Rudnichenko, *Information technology methods in diagnosing the state of complex technical systems, Ecology*, Odessa, 2019.
- [5] V. Vychuzhanin, N. Shibaeva, 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* 3 (2016) 205–214.
- [6] V.A. Smirnov, Intelligent decision support system for the control of complex technical systems, *IOP Conf. Series: Journal of Physics* 1327 (2019) 1–9.
- [7] V. Freyman, I. Bezukladnikov, The application of soft decision making on decoding and assessment of test diagnosing results within control systems elements, *Proceedings of the 2017 XX IEEE International Conference on Soft Computing and Measurements*, 2017, pp.124–128. doi: 10.1109/SCM.2017.7970515
- [8] V.V. Vychuzhanin, N.D. Rudnichenko, A.V. Vychuzhanin, A.Y. Kozlov, Information support of risk management systems for automotive equipment failures, *Informatics and mathematical methods in modeling* 3 (2019) 121–133.
- [9] V.A. Fuks, Universal system for remote vehicle diagnostics, *Young scientist* 12 (2019) 40–44. doi:10.1051/mateconf/201930201019
- [10] V.D. Nogin, *Multi-criteria decision making, BHV*, Saint Petersburg, 2007.
- [11] N. Vakhania, F. Werner, A brief look at multi-criteria problems - multi-threshold optimization versus Pareto-Optimization. *Multicriteria Optimization - Pareto-optimal and Related Principles*, IntechOpen, Zagreb (2020). doi:10.5772/intechopen.91169

- [12] R. Steuer, Multiple Criteria Optimization: Theory, Computation, and Application, Wiley, New York, NY, 1986.
- [13] ISO 14830-1:2019 Condition monitoring and diagnostics of machine systems - Tribology-based monitoring and diagnostics (2019).
- [14] J. Horn, N. Nafpliotis, D. Goldberg, A niched pareto genetic algorithm for multiobjective optimization, Proceedings of the 1st IEEE Congress of Evolutionary Computation, 1994, pp. 82–87.
- [15] M.A. Rosenman, J.S. Gero, Reducing the pareto optimal set in multicriteria optimization, Engineering Optimization 8 (1985) 189–206.
- [16] PC optimization program IOSO, 2020. URL: <http://www.aec-service.ru/articles/10.html>.

# Algorithm of analysis and conversion of input data of a two-factor multi-variative transport problem with weight coefficients

Oleksii Chyzhmotria<sup>a</sup>, Olena Chyzhmotria<sup>a</sup> and Tetiana Vakaliuk<sup>a</sup>

<sup>a</sup> Zhytomyr Polytechnic State University, Chudnivska str., 103, Zhytomyr, 10005, Ukraine

## Abstract

The article is devoted to the analysis of input data of a two-factor multivariate transport problem with weighting factors. The article aims to develop and describe an algorithm for bringing this problem to a form suitable for the application of one of the existing methods of solving the classical transport problem. The developed algorithm should be such that it can be relatively easily programmed in one of the existing programming languages. The input data for the development of the algorithm is the presence of two independent optimization criteria; different values of weighting factors of two factors for each pair "supplier-consumer"; a different number of options for transportation of goods for each pair "supplier-consumer" with the corresponding values of both factors. According to the goal, the algorithm must choose the objectively best of several options for transportation of goods for each pair "supplier-consumer", taking into account the two-factor and the presence of weights. The issues related to the choice of the best of the options for transportation of goods for a single pair "supplier-consumer" taking into account the weight coefficients are considered on the examples. An analysis of the influence of the values of the factors of one pair "supplier-consumer" on the resulting criteria of other pairs. Developed an algorithm for bringing the initial data to a single numerical range, calculating the resulting criteria, and determining the best transportation option for each pair "supplier-consumer".

## Keywords 1

Two-factor transport problem, transport transportation, quality criteria, weight factors, weight coefficients, multivariate, algorithms.

## 1. Introduction

One of the types of transport problem is a two-factor problem, in which it is necessary to minimize costs simultaneously for two factors. There is a common method of pairwise multiplication of the corresponding values of factors with subsequent selection of the smallest of the obtained values.

In the case of several variants of pair wise values of two factors for the pair "supplier-consumer", the mentioned method of pair wise multiplication of the corresponding values of factors with the subsequent selection of the smallest of the obtained values remains valid. The selected value will be used in the future as the best option for the existing pair "supplier-consumer".

Suppose we have two factors ( $C$ ,  $T$ ) and three transport options from supplier A to consumer B with different values for each factor ( $(c_1, c_2, c_3$  and  $t_1, t_2, t_3)$ ). For the minimization problem, it will be enough to choose the minimum value from three pairwise products of the values of factors  $C$  and  $T$ :

$$r = \min\{c_1 \times t_1; c_2 \times t_2; c_3 \times t_3\} \quad (1)$$

CMIS-2021: The Fourth International Workshop on Computer Modeling and Intelligent Systems, April 27, 2021, Zaporizhzhia, Ukraine  
EMAIL: 4ov.ztu@gmail.com (O. V. Chyzhmotria); ch-o-g@ztu.edu.ua (O.G.Chyzhmotria); tetianavakaliuk@gmail.com (T. Vakaliuk)  
ORCID: 0000-0002-5515-6550 (O. V. Chyzhmotria); 0000-0001-8597-1292 (O.G.Chyzhmotria); 0000-0001-6825-4697 (T. Vakaliuk)



© 2020 Copyright for this paper by its authors.  
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).  
CEUR Workshop Proceedings (CEUR-WS.org)

The value of  $r$  and will be the resulting criterion, and the corresponding transportation option - the best of the proposed.

In real conditions, different suppliers and different consumers may have different priorities when transporting goods: someone needs to receive the goods as quickly as possible, someone needs to minimize the cost of transportation, someone is looking for a balanced solution, etc. Thus, there is a need to vary the "weight" of factors for each pair "supplier-consumer" within one transport task.

Assuming the presence of weight coefficients for each of the factors ( $k_c, k_t$ ), there is a problem of choosing an objectively better option from the proposed for each pair of "supplier-consumer". Note that for each variant of a certain pair "supplier-consumer" weight coefficients are constant values.

Taking into account the weight coefficients, formula 1 takes the following form:

$$r = \min\{c_1 \times k_c \times t_1 \times k_t; c_2 \times k_c \times t_2 \times k_t; c_3 \times k_c \times t_3 \times k_t\} \quad (2)$$

or after mathematical transformations:

$$r = k_c \times k_t \times \min\{c_1 \times t_1; c_2 \times t_2; c_3 \times t_3\} \quad (3)$$

From formula 3 it can be seen that the resulting best option is still chosen as the minimum of three pair wise products of the values of factors  $C$  and  $T$ . The weight coefficients only equally increase each of the products in  $k_c \times k_t$  times and in no way affect the choice of the best option. This method eliminates the very essence of weight coefficients as levers of influence when choosing the best option.

Simple pairwise multiplication of the values of the factors and their weight coefficients with the subsequent choice of the smallest of the obtained values does not allow to choose the objectively best of the proposed options.

## 2. Review of the literature

Nowadays mathematical methods solve many problems of operational planning for transportation. Many scientists dedicate their work to the topic of transport optimization.

Chyzmotria O. et al. [4] on an example considered the problems connected with the search of the optimum plan of transportations simultaneously on two quality criteria. They conducted a comparative analysis of the four transportation plans according to the condition of the example.

Burduk A., Musial K. [2; 3] solved the problem of optimization using genetic algorithms. They described genetic algorithms, their properties, and their capabilities in solving computational problems. To solve the problem under study, the authors used the program MATLAB.

Prifti V. et al. [6] considered a real problem of linear programming in detail by taking an example in an Albania company. For the company under consideration the problem of minimizing transportation costs was solved by solving 3 methods: The North West Corner Method, the Least Cost Method, and the Vogel's Approximation Method. The calculations showed that based on the demand from the 9 geographical sites (destinations) and the capacity offered by the two manufacturing plants (sources), the most optimal solution turns out to be the one obtained by Vogel's method.

Sun Y. et al. [8] in their study presented a systematic review of the problem of planning route transportation of goods in a multimodal transport network. In this study, the formulation characteristics are divided and classified into six aspects, and optimization models in recent studies are determined based on the respective formulation characteristics.

Gunantara N. [5] in his work considered two methods of multi-objective optimization (MOO) that do not require complicated mathematical equations. These two methods are Pareto and scalarization.

In the article by Zhang Y. et al. [10] an optimization method for multiple batches of express freight demands is proposed for the shippers of railway express freight to select the most suitable transportation products to transport, considering the priority of shippers and capacity constraints. Five transport attributes, the most common concerns of express freight shippers, including freight transit time, transport cost, convenience, safety, and reliability, are selected as main indexes. Furthermore, a solution algorithm is designed by considering important clients who prioritize choosing transportation products.

Badica A. et al. [1] proposed a method for declarative modeling and optimization of freight transportation brokering using agents and constraint logic programming.

Stoilova S. [7] in her study proposed a step-by-step approach to determining the transport plan of passenger trains. In the first step, criteria for optimizing the transport plan were defined. In the second stage, variants of the transport plan were formulated. In the third stage, the weight coefficients of the criteria were determined. Multi-purpose optimization was performed in the fourth step. The impact of changes in passenger traffic on the choice of the optimal transport plan was studied in the fifth stage.

In the article of the authors Zabolotnii S. and Mogilei S. [9] a study of existing methods for constructing support plans for the transportation problem with several means of cargo delivery is conducted, and the task itself is defined as multimodal. Based on the criterion of reducing the number of numerous iterations in finding solutions to such a problem, a more perfect method of constructing its support plans, the so-called Steiner method, is proposed. And also a general formulation of the multimodal transport problem is implemented - its objective function (criterion) of optimization and an admissible set of solutions are formalized.

The **article aims** to develop and describe an algorithm for bringing this problem to a form suitable for the application of one of the existing methods of solving the classical transport problem. The developed algorithm should be such that it can be relatively easily programmed in one of the existing programming languages.

### 3. Results

Consider and analyze the following method of choosing the best option, taking into account the weight coefficients  $k_c$  and  $k_t$ . The essence of the method will be to add the values of the two factors C and T, multiplied by the corresponding weight coefficients.

The following example 1 should provide an answer as to the feasibility or inadmissibility of this method. The input data will be the values of the factors  $c_1 = 900$ ,  $c_2 = 850$ ,  $t_1 = 1$ ,  $t_2 = 5$ . In example 1, we will deal with two options for transporting goods from point A to point B. The weight coefficients will change in the range from 0 to 1 in steps of 0.1. The sum of the coefficients will always be equal to 1.

The best option for transportation and the value of the resulting criterion will be sought by the formula:

$$r = \min\{c_1 \times k_c + t_1 \times k_t; c_2 \times k_c + t_2 \times k_t\} \quad (4)$$

The results of the calculations are summarized in table 1:

**Table 1**

Calculation of values and selection of the resulting criterion

| $k_c$ | $k_t$ | $r_1 = c_1 \times k_c + t_1 \times k_t$ | $r_2 = c_2 \times k_c + t_2 \times k_t$ | $r$   |
|-------|-------|-----------------------------------------|-----------------------------------------|-------|
| 1     | 0     | 900                                     | 850                                     | $r_2$ |
| 0,9   | 0,1   | 810,1                                   | 765,5                                   | $r_2$ |
| 0,8   | 0,2   | 720,2                                   | 681                                     | $r_2$ |
| 0,7   | 0,3   | 630,3                                   | 596,5                                   | $r_2$ |
| 0,6   | 0,4   | 540,4                                   | 512                                     | $r_2$ |
| 0,5   | 0,5   | 450,5                                   | 427,5                                   | $r_2$ |
| 0,4   | 0,6   | 360,6                                   | 343                                     | $r_2$ |
| 0,3   | 0,7   | 270,7                                   | 258,5                                   | $r_2$ |
| 0,2   | 0,8   | 180,8                                   | 174                                     | $r_2$ |
| 0,1   | 0,9   | 90,9                                    | 89,5                                    | $r_2$ |
| 0     | 1     | 1                                       | 5                                       | $r_1$ |

Let us analyze the data in Table 1, according to which it is seen that with a decrease in the influence of factor C and a corresponding increase in the influence of factor T, the difference between the values of the resulting criteria  $r_1$  and  $r_2$  decreases. But even when the weight of

the factor T acquires a conditional 90% ( $k_t = 0.9$ ), still the smaller of the two values of the resulting criterion remains  $r_2$ . And this even though under the condition of the example, the value of  $t_2$  is five times greater than the value of  $t_1$ . This result can be considered distorted and biased, and the method of adding the values of the two factors C and T multiplied by the corresponding weight coefficients should be considered unacceptable as such.

It should be noted that in this example we were dealing with two different numerical ranges, one of which was hundreds of times larger than the other. From the very beginning of the calculations, this range of values dominated and, accordingly, the corresponding factor dominated. This had a direct impact on the value of the resulting criterion and the choice of the best freight option. Even the use of weight coefficients could not significantly affect the final result. The difference in the values of the two ranges was so great that the weight coefficients could not perform their direct function in terms of influencing the choice of the best option, for which they were generally introduced into the mathematical model. Thus, the reason for the biased results was a large difference in the values of the numerical ranges of the two factors. It can also be argued that even a small difference can also negatively affect the objectivity of the result.

In this study, we consider an algorithm that will get rid of the distortion of the results due to the difference in the values of the two numerical ranges. The essence of the algorithm will first be represented schematically:



**Figure 1:** Scheme of the algorithm for bringing data to a single numerical range

At the first stage of the algorithm, it is necessary to determine the largest value for each factor.  
For factor C:

$$C_{\max} = \max\{c_1, c_2, \dots, c_m\} \quad (5)$$



For factor T:

$$t_{\max} = \max\{t_1, t_2, \dots, t_m\} \quad (6)$$

When programming to find the maximum value, as in the case of finding the minimum value, use arrays, cyclic structures with a precondition or postcondition, and branching structures.

In the second stage of the algorithm, it is necessary to calculate the least common multiple (LCM) for the values

$$LCM(c_{\max}, t_{\max}) = \frac{c_{\max} \times t_{\max}}{GCD(c_{\max}, t_{\max})} \quad (7)$$

where GCD is the greatest common divisor.

To find the largest common divisor, as an option, you can use the well-known Euclidean algorithm. The algorithm contains a loop with a premise and several branches and can be easily programmed.

It should be noted that at the beginning of the Euclidean algorithm, all input data must be integers, so, if necessary, it is necessary to simultaneously increase the input data by  $10^n$  times.

In the third stage, additional factors are determined for each of the factors.

For factor C:

$$c_{am} = LCM(c_{\max}, t_{\max}) / c_{\max} \quad (8)$$

For factor T:

$$t_{am} = LCM(c_{\max}, t_{\max}) / t_{\max} \quad (9)$$

In the last, fourth stage, each value of each of the factors must be multiplied by the corresponding additional factor ( $c_{am}$  or  $t_{am}$ , respectively).

At the end of the algorithm, you can proceed to search by formula 4 the value of the resulting criterion and choose the best option.

Let's return to the above example 1. The input data in the example were the values of the factors

$$c_1 = 900, c_2 = 850, t_1 = 1, t_2 = 5.$$

According to the algorithm for bringing data to a single numerical range (see Fig. 1) step by step we get:

- $c_{\max} = \max\{900, 850\} = 900, t_{\max} = \max\{1, 5\} = 5.$
- $LCM(c_{\max}, t_{\max}) = LCM(900, 5) = 900.$
- $c_{am} = LCM(c_{\max}, t_{\max}) / c_{\max} = 900 / 900 = 1;$   
 $t_{am} = LCM(c_{\max}, t_{\max}) / t_{\max} = 900 / 5 = 180.$
- $c_1' = c_1 \times c_{am} = 900 \times 1 = 900;$   
 $c_2' = c_2 \times c_{am} = 850 \times 1 = 850;$   
 $t_1' = t_1 \times t_{am} = 1 \times 180 = 180;$   
 $t_2' = t_2 \times t_{am} = 5 \times 180 = 900.$

The value of the resulting criterion and the best option for transportation will be sought by formula 4. The weight coefficients will be changed in the range from 0 to 1 in steps of 0.1.

The results of the calculations are summarized in table2:

The results shown in table 2 are radically different from the results in table 1. From table 2 we see that after reducing the initial data to a single numerical range, even the minimum effect of factor T ( $k_t = 0.1$  or 10%) was sufficient to the resulting criterion was the criterion  $r_1$ . This choice is logical, because the factor T in its values differs from the minimum to the maximum 5 times, while the factor C - only  $\approx 1.06$  times. Accordingly, the effect of factor C should be minimal. At the same time, it is the T factor that should play a key role in choosing the best transportation options, as demonstrated by the proposed algorithm.

The resulting criterion  $r'$  will thus participate in the further solution of the two-factor transport problem.

**Table 2**

Adjusted calculation of values and selection of the resulting criterion

| $k_c$ | $k_t$ | $r_1' = c_1' \times k_c + t_1' \times k_t$ | $r_2' = c_2' \times k_c + t_2' \times k_t$ | $r'$   |
|-------|-------|--------------------------------------------|--------------------------------------------|--------|
| 1     | 0     | 900                                        | 850                                        | $r_2'$ |
| 0,9   | 0,1   | 828                                        | 855                                        | $r_1'$ |
| 0,8   | 0,2   | 756                                        | 860                                        | $r_1'$ |
| 0,7   | 0,3   | 684                                        | 865                                        | $r_1'$ |
| 0,6   | 0,4   | 612                                        | 870                                        | $r_1'$ |
| 0,5   | 0,5   | 540                                        | 875                                        | $r_1'$ |
| 0,4   | 0,6   | 468                                        | 880                                        | $r_1'$ |
| 0,3   | 0,7   | 396                                        | 885                                        | $r_1'$ |
| 0,2   | 0,8   | 324                                        | 890                                        | $r_1'$ |
| 0,1   | 0,9   | 252                                        | 895                                        | $r_1'$ |
| 0     | 1     | 180                                        | 900                                        | $r_1'$ |

It will be recalled that before that it was a question of transportation of cargo from supplier A to consumer B. However, the transport task assumes the presence of  $m$  departure points  $A_1, A_2, \dots, A_m$  and  $n$  consumers  $B_1, B_2, \dots, B_n$ . For each pair "supplier-consumer" within the research topic, the variability of the values of each of the two factors is allowed. Also, each pair "supplier-consumer" for the problem may have different values of weights for these factors. Thus, there is a task of developing an algorithm for bringing a two-factor multivariate transport problem with weight coefficients to a unified form, suitable for the application of one of the existing methods for solving problems of the corresponding type. Separate questions are the possible influence of factor values of one "supplier-consumer" pair on the resulting criteria of other pairs and the probable different numerical ranges of factor values for different "supplier-consumer" pairs.

For further analysis and development of the algorithm as example 2 consider a fragment of a two-factor multivariate transport problem, given in the tabular form:

**Table 3**

A fragment of a two-factor multivariate transport problem

| Suppliers   | Consumers |       |       |       | Cargo stocks |
|-------------|-----------|-------|-------|-------|--------------|
|             | $B_1$     | $B_2$ | $B_3$ | $B_4$ |              |
| $A_1$       | ...       | ...   | $C$   | $T$   | $a_1$        |
|             |           |       | 15    | 6     |              |
|             |           |       | 10    | 9     |              |
| $A_2$       | ...       | ...   | ...   | ...   | $a_2$        |
|             |           |       |       |       |              |
| $A_3$       | ...       | $C$   | $T$   | ...   | $a_3$        |
|             |           | 4     | 6     |       |              |
|             |           | 5     | 5     |       |              |
|             |           | 2     | 13    |       |              |
| Cargo needs | $b_1$     | 3     | 9     | $b_4$ |              |
|             |           | $b_2$ | $b_3$ |       |              |

To analyze the possible influence of factor values of one pair "supplier-consumer" on the resulting criteria of other pairs and to address the issue of different numerical ranges of factor values for different pairs "supplier-consumer" it will suffice to use two pairs "supplier-consumer" (in our example pairs  $A_1$ - $B_3$  and  $A_3$ - $B_2$ ). The same two pairs "supplier-consumer" will develop an algorithm for bringing a two-factor multivariate transport problem with weight coefficients to a unified form, suitable for the application of one of the existing methods of solving problems of the corresponding type.

To begin with, we determine the best transportation option and the value of the resulting criterion separately for each pair A1-B3 and A3-B2. To do this, we use formula 4 and the algorithm for bringing data to a single numerical range (see Fig. 1, formulas 5-9).

For the A1-B3 pair we have two options for cargo transportation. According to the algorithm step by step we get:

- $c_{13max} = \max\{15, 10\} = 15$ ,  $t_{13max} = \max\{6, 9\} = 9$ .
- $LCM(c_{13max}, t_{13max}) = LCM(15, 9) = 45$ .
- $c_{13am} = LCM(c_{13max}, t_{13max}) / c_{13max} = 45 / 15 = 3$ ;  
 $t_{13am} = LCM(c_{13max}, t_{13max}) / t_{13max} = 45 / 9 = 5$ .
- $c_{131}' = c_{131} \times c_{13am} = 15 \times 3 = 45$ ;  
 $c_{132}' = c_{132} \times c_{13am} = 10 \times 3 = 30$ ;  
 $t_{131}' = t_{131} \times t_{13am} = 6 \times 5 = 30$ ;  
 $t_{132}' = t_{132} \times t_{13am} = 9 \times 5 = 45$ .

Weight coefficients will be changed in the range from 0 to 1 in steps of 0.1. The results of calculations according to formula 4 are summarized in table 4:

**Table 4**

Calculation of values and selection of the resulting criterion for the pair A1-B3

| $k_{13c}$ | $k_{13t}$ | $r_{131}'$ | $r_{132}'$ | $r_{13}'$             |
|-----------|-----------|------------|------------|-----------------------|
| 1         | 0         | 45         | 30         | $r_{132}'$            |
| 0,9       | 0,1       | 43,5       | 31,5       | $r_{132}'$            |
| 0,8       | 0,2       | 42         | 33         | $r_{132}'$            |
| 0,7       | 0,3       | 40,5       | 34,5       | $r_{132}'$            |
| 0,6       | 0,4       | 39         | 36         | $r_{132}'$            |
| 0,5       | 0,5       | 37,5       | 37,5       | $r_{131}' = r_{132}'$ |
| 0,4       | 0,6       | 36         | 39         | $r_{131}'$            |
| 0,3       | 0,7       | 34,5       | 40,5       | $r_{131}'$            |
| 0,2       | 0,8       | 33         | 42         | $r_{131}'$            |
| 0,1       | 0,9       | 31,5       | 43,5       | $r_{131}'$            |
| 0         | 1         | 30         | 45         | $r_{131}'$            |

For the pair A3-B2 we have four options for cargo transportation. According to the algorithm step by step we get:

- $c_{32max} = \max\{4, 5, 2, 3\} = 5$ ,  $t_{32max} = \max\{6, 5, 13, 9\} = 13$ .
- $LCM(c_{32max}, t_{32max}) = LCM(5, 13) = 65$ .
- $c_{32am} = LCM(c_{32max}, t_{32max}) / c_{32max} = 65 / 5 = 13$ ;  
 $t_{32am} = LCM(c_{32max}, t_{32max}) / t_{32max} = 65 / 13 = 5$ .
- $c_{321}' = c_{321} \times c_{32am} = 4 \times 13 = 52$ ;  
 $c_{322}' = c_{322} \times c_{32am} = 5 \times 13 = 65$ ;  
 $c_{323}' = c_{323} \times c_{32am} = 2 \times 13 = 26$ ;  
 $c_{324}' = c_{324} \times c_{32am} = 3 \times 13 = 39$ ;  
 $t_{321}' = t_{321} \times t_{32am} = 6 \times 5 = 30$ ;  
 $t_{322}' = t_{322} \times t_{32am} = 5 \times 5 = 25$ ;  
 $t_{323}' = t_{323} \times t_{32am} = 13 \times 5 = 65$ ;  
 $t_{324}' = t_{324} \times t_{32am} = 9 \times 5 = 45$ .

The results of calculations according to formula 4 are summarized in table 5.

The results of the calculations for the pair A3-B2, shown in table 5, indicate that each of the options for transporting goods from supplier A3 to consumer B2 may be the best option depending on the weight coefficients.

**Table 5**

Calculation of values and selection of the resulting criterion for the pair A3-B2

| $k_{32c}$ | $k_{32t}$ | $r_{321}'$ | $r_{322}'$ | $r_{323}'$ | $r_{324}'$ | $r_{32}'$  |
|-----------|-----------|------------|------------|------------|------------|------------|
| 1         | 0         | 52         | 65         | 26         | 39         | $r_{323}'$ |
| 0,9       | 0,1       | 49,8       | 61         | 29,9       | 39,6       | $r_{323}'$ |
| 0,8       | 0,2       | 47,6       | 57         | 33,8       | 40,2       | $r_{323}'$ |
| 0,7       | 0,3       | 45,4       | 53         | 37,7       | 40,8       | $r_{323}'$ |
| 0,6       | 0,4       | 43,2       | 49         | 41,6       | 41,4       | $r_{324}'$ |
| 0,5       | 0,5       | 41         | 45         | 45,5       | 42         | $r_{321}'$ |
| 0,4       | 0,6       | 38,8       | 41         | 49,4       | 42,6       | $r_{321}'$ |
| 0,3       | 0,7       | 36,6       | 37         | 53,3       | 43,2       | $r_{321}'$ |
| 0,2       | 0,8       | 34,4       | 33         | 57,2       | 43,8       | $r_{322}'$ |
| 0,1       | 0,9       | 32,2       | 29         | 61,1       | 44,4       | $r_{322}'$ |
| 0         | 1         | 30         | 25         | 65         | 45         | $r_{322}'$ |

For both pairs A1-B3 and A3-B2, the algorithm for calculating the value of the resulting criterion and finding the best option yielded results. They can be considered objective, but only separately for each couple. If we analyze and compare the calculations for both pairs, we see that the values of the factors in these pairs were in different numerical ranges. For the pair A1-B3, the numerical range with the largest value of 45 was obtained, and for the pair A3-B2, the numerical range with the largest value of 65 was obtained. The case with different numerical ranges has already been considered and described above. It has also been concluded that it is inadmissible to use the obtained values in this form for further calculations.

To solve this problem in the study it is proposed to combine the values of each of the factors of all options from both pairs "supplier-consumer" of the current problem. This option is perfectly acceptable, because under the condition of the transport problem, the load is homogeneous, and the factors together with the units of measurement are the same for all pairs of "supplier-consumer" of the current problem. In this case, the algorithm for bringing data to a single numerical range, shown in Fig. 1, remains unchanged. Also, we have the opportunity to identify and analyze the possible influence of the values of the factors of one pair "supplier-consumer" on the resulting criteria of other pairs.

For the current example, after performing the first three steps of the algorithm, we obtain:

- $c_{max} = \max\{15, 10, 4, 5, 2, 3\} = 15$ ;
- $t_{max} = \max\{6, 9, 6, 5, 13, 9\} = 13$ ;
- $LCM(c_{max}, t_{max}) = LCM(15, 13) = 195$ ;
- $c_{am} = LCM(c_{max}, t_{max}) / c_{max} = 195 / 15 = 13$ ;
- $t_{am} = LCM(c_{max}, t_{max}) / t_{max} = 195 / 13 = 15$ .

For pair A1-B3 the fourth step of the algorithm:

- $c_{131}' = c_{131} \times c_{am} = 15 \times 13 = 195$ ;
- $c_{132}' = c_{132} \times c_{am} = 10 \times 13 = 130$ ;
- $t_{131}' = t_{131} \times t_{am} = 6 \times 15 = 90$ ;
- $t_{132}' = t_{132} \times t_{am} = 9 \times 15 = 135$ .

For the pair A3-B2, the fourth step of the algorithm:

- $c_{321}' = c_{321} \times c_{am} = 4 \times 13 = 52$ ;
- $c_{322}' = c_{322} \times c_{am} = 5 \times 13 = 65$ ;
- $c_{323}' = c_{323} \times c_{am} = 2 \times 13 = 26$ ;
- $c_{324}' = c_{324} \times c_{am} = 3 \times 13 = 39$ ;
- $t_{321}' = t_{321} \times t_{am} = 6 \times 15 = 90$ ;
- $t_{322}' = t_{322} \times t_{am} = 5 \times 15 = 75$ ;
- $t_{323}' = t_{323} \times t_{am} = 13 \times 15 = 195$ ;
- $t_{324}' = t_{324} \times t_{am} = 9 \times 15 = 135$ .

The results of calculations according to formula 4 for the pair A1-B3 are summarized in table 6.

**Table 6**

Adjusted calculation of values and selection of the resulting criterion for the pair A1-B3

| $k_{13c}$ | $k_{13t}$ | $r_{131}'$ | $r_{132}'$ | $r_{13}'$  |
|-----------|-----------|------------|------------|------------|
| 1         | 0         | 195        | 130        | $r_{132}'$ |
| 0,9       | 0,1       | 184,5      | 130,5      | $r_{132}'$ |
| 0,8       | 0,2       | 174        | 131        | $r_{132}'$ |
| 0,7       | 0,3       | 163,5      | 131,5      | $r_{132}'$ |
| 0,6       | 0,4       | 153        | 132        | $r_{132}'$ |
| 0,5       | 0,5       | 142,5      | 132,5      | $r_{132}'$ |
| 0,4       | 0,6       | 132        | 133        | $r_{131}'$ |
| 0,3       | 0,7       | 121,5      | 133,5      | $r_{131}'$ |
| 0,2       | 0,8       | 111        | 134        | $r_{131}'$ |
| 0,1       | 0,9       | 100,5      | 134,5      | $r_{131}'$ |
| 0         | 1         | 90         | 135        | $r_{131}'$ |

The results of calculations according to formula 4 for the pair A3-B2 are summarized in table 7:

**Table 7**

Adjusted calculation of values and selection of the resulting criterion for the pair A3-B2

| $k_{32c}$ | $k_{32t}$ | $r_{321}'$ | $r_{322}'$ | $r_{323}'$ | $r_{324}'$ | $r_{32}'$  |
|-----------|-----------|------------|------------|------------|------------|------------|
| 1         | 0         | 52         | 65         | 26         | 39         | $r_{323}'$ |
| 0,9       | 0,1       | 55,8       | 66         | 42,9       | 48,6       | $r_{323}'$ |
| 0,8       | 0,2       | 59,6       | 67         | 59,8       | 58,2       | $r_{324}'$ |
| 0,7       | 0,3       | 63,4       | 68         | 76,7       | 67,8       | $r_{321}'$ |
| 0,6       | 0,4       | 67,2       | 69         | 93,6       | 77,4       | $r_{321}'$ |
| 0,5       | 0,5       | 71         | 70         | 110,5      | 87         | $r_{322}'$ |
| 0,4       | 0,6       | 74,8       | 71         | 127,4      | 96,6       | $r_{322}'$ |
| 0,3       | 0,7       | 78,6       | 72         | 144,3      | 106,2      | $r_{322}'$ |
| 0,2       | 0,8       | 82,4       | 73         | 161,2      | 115,8      | $r_{322}'$ |
| 0,1       | 0,9       | 86,2       | 74         | 178,1      | 125,4      | $r_{322}'$ |
| 0         | 1         | 90         | 75         | 195        | 135        | $r_{322}'$ |

Let's analyze the results.

For pair A1-B3 we compare the data of tables 4 and 6. In table 4 at weight coefficients of 0,5 / 0,5 both variants of transportation of freight have identical result:  $r_{13}' = r_{131}' = r_{132}'$ .

After combining the values of each of the factors of all options from both pairs "supplier-consumer" of the current problem at the same weight coefficients of 0.5 / 0.5, the resulting criterion was chosen  $r_{132}'$ :  $r_{13}' = r_{132}'$ .

Given the invariance of the initial data in the pair A1-B3, it is possible to draw an unambiguous conclusion about the direct impact on the result in the pair A1-B3 values of the second pair A3-B2.

For the pair A3-B2, we compare the data from Tables 5 and 7. Here we see even greater differences in the results: for six of the eleven pairs of weight coefficients, the best option for transporting cargo from the four existing ones has changed. Here, too, with constant initial data, the significant influence of the values of the pair A1-B3 is obvious.

## 4. Conclusions

For a two-factor multivariate transport problem with weight coefficients, the method of pair wise multiplication of factor values and their weight coefficients with subsequent selection of the smallest

of the obtained values does not allow to choose the objectively best of the proposed options, as in this case weight coefficients do not affect the choice of the best option.

The difference in the values of the numerical ranges of the two factors leads to biased results, as the weight coefficients, in this case, can not fully perform their direct function to influence the choice of the best option. Therefore, it is mandatory to bring the values of the numerical ranges of the two factors to a single range both within a single pair "supplier-consumer" and within the entire transport task.

The proposed algorithm for bringing the initial data to a single numerical range, calculating the resulting criteria, and determining the best option for transportation for each pair "supplier-consumer" has fully performed its function. The obtained resulting criteria are ready for use in the further solution of the transport problem by one of the existing methods.

The algorithm allowed us to draw an important conclusion about the influence of the values of the factors of a single pair "supplier-consumer" on the resulting criteria of other pairs.

The developed algorithm can be relatively easily programmed in one of the existing programming languages. To write a program according to the given algorithm will require knowledge, skills, and abilities to work, in particular, with one-dimensional and multidimensional arrays, cycles of different types, branching design.

## 5. References

- [1] Badica A., Badica C., Leon F., Danciulescu D.: Optimization of Freight Transportation Brokerage Using Agents and Constraints. International Conference on Engineering Applications of Neural Networks, vol. 744. (2017). 451-464. DOI: [https://doi.org/10.1007/978-3-319-65172-9\\_38](https://doi.org/10.1007/978-3-319-65172-9_38)
- [2] Burduk A., Musiał K.: Genetic Algorithm Adoption to Transport Task Optimization. International Joint Conference SOCO'16-CISIS'16-ICEUTE'16, vol. 527. (2017) 366-375.
- [3] Burduk A., Musiał K.: Optimization of Chosen Transport Task by Using Generic Algorithms, vol. 9842. (2016) 197-205. DOI: 10.1007/978-3-319-45378-1\_18
- [4] Chyzhmotria O., Chyzhmotria O., Bolotina V.: Analysis of the Two-Criteria Transport Task. Znanstvena Misel Journal, vol. 32(1) (2019) 52-55.
- [5] Gunantara N.: A review of multi-objective optimization: Methods and its applications. Cogent Engineering, vol. 5(1). (2018) 1-16. DOI: <https://doi.org/10.1080/23311916.2018.1502242>
- [6] Prifti V., Dervishi I., Dhoska K., Markja I., Pramono A.: Minimization of transport costs in an industrial company through linear programming. International Conference on Advanced Mechanical and Industrial engineering, vol. 909(1), 012040 (2020) doi:10.1088/1757-899X/909/1/012040
- [7] Stoilova S.: An Integrated Multi-Criteria and Multi-Objective Optimization Approach for Establishing the Transport Plan of Intercity Trains. Sustainability, vol. 12(2). (2020) 687. DOI: <https://doi.org/10.3390/su12020687>
- [8] Sun Y., Lang M., Wang D.: Optimization Models and Solution Algorithms for Freight Routing Planning Problem in the Multi-Modal Transportation Networks: A Review of the State-of-the-Art. Open Civil Engineering Journal, vol. 9(1). (2015) 714-723. DOI: <http://dx.doi.org/10.2174/1874149501509010714>
- [9] Zabolotnii S., Mogilei S.: Optimization of the Method of Constructing Reference Plans of Multimodal Transport Problem. Technology Audit and Production Reserves, vol. 1/2(45). (2019) 15-20. DOI: 10.15587/2312-8372.2019.154561
- [10] Zhang Y., Wang J., Hu Z.: Optimization Model of Transportation Product Selection for Railway Express Freight. Journal of Engineering Science and Technology Review, vol. 9(5). (2016) 104-110.

# Improvement of Grayscale Images in Orthogonal Basis of the Type-2 Membership Function

Lyudmila Akhmetshina<sup>a</sup> and Artyom Yegorov<sup>a</sup>

<sup>a</sup> Dnieper National University Named by Oles Honchar, Gagarin Avenue, house 72, Dnieper, 49010, Ukraine

## Abstract

While analyzing different images, it is very important to identify similar and/or homogeneous areas and boundaries of the objects of interest. Certain ambiguity occurred at this step can be caused by physical characteristics of the used equipment and noise in the process of the image formation, on the one part, and by inaccuracy and fuzziness introduced during digital representation and by processing algorithms, on the other part. It is shown that transition of the features to a fuzzy space, followed by the use of orthogonal transformation and visualization of characteristics synthesized on the basis of their eigenvalues, improves reliability of the objects of interest identification during analyzing of the grayscale images. Informational capabilities of characteristics synthesized with the use of the method of singular decomposition of the type-2 features in a fuzzy space are considered from the aspect of improvement of the grayscale image quality. The obtained experimental results are shown on the example of the real microscopic images.

## Keywords 1

image processing, orthogonal transformations, singular decomposition, fuzzy logic, Membership Function Type-2.

## 1. Introduction

The number of practical problems associated with digital image processing, which are obtained with the help of standard research methods, and which are used, for example, in materials science, or medicine, or flaw detection, is constantly growing. Images formed by various information, tracking or diagnostic systems very often have a quality that is not sufficient for performing a reliable analysis. As a rule, these images contain distortions caused by physical system of their creation and process of their formation (heterogeneity of detectors, lighting, background, dynamic distortions), on the one hand, and by the methods of their representation and displaying in the processing system (discretization and quantization errors, changes in color reproduction, gray ambiguity, etc.), on the other hand.

In order to improve reliability of image analysis performed either by unaided eye or with the help of automated systems, it is desirable, first, to improve brightness characteristics of the images. Their characteristics such as contrast, brightness and resolution are of especial importance for visual perception, for example, in medical applications. For the cases when automated processing systems are used for solving a specific problem and for obtaining quantitative indicators, it is recommended to begin with determining and identifying the necessary value parameters on the basis of initial data for further identifying of the objects of interest [1].

The inherent inconsistency of the image transformation process lies in the fact that it is essential to ensure, on the one hand, maximum sensitivity of the used methods to insignificant local variations in brightness, and, on the other hand, resistance to the effects of the structure- and measurement-



generated noises. In this case, in addition to randomness, which is described in accordance with the theory of probability, it is necessary to take into consideration an uncertainty present in the images [2, 3, 4], which is an attribute of information [5]. The currently used approach to solving the problem of image analysis is based on the use of fuzzy methods [6, 7, 8] because of inaccuracy and incompleteness of initial data and ambiguity of processing algorithms (for example, when one determines classes, regions/boundaries of the objects).

## 2. Review of Literature

One of the ways to solve the problem of improving quality of the images and reliability of their analysis assumes transition of informative features formed by the method of spatial transformations [10] to a new space. From the number of methods used for forming informative features, a specific one should be chosen by physical nature and brightness characteristics of initial data and in accordance with the task set [11]. The difficulty in determining an effective solution method is explained by the fact that characteristics of the imaging system, as well as presence and parameters of the object of interest are often unknown a priori.

Fuzzy sets of the type-1 (T1) make it possible to transform an uncertainty into the membership function that has a numerical value within the range  $[0,1]$ . Nonlinearity of the fuzzy processing methods increases influence of variations in the brightness properties of the analyzed images and eliminates ambiguity of gray. However, the fuzzy sets of the type-1 (T1) do not take into consideration uncertainties in the membership functions, since they are characterized by clear values [12, 13, 14].

Fuzzy logic of the type-2 (T2) was introduced by Zadeh L. as a generalized concept of the theory of ordinary fuzzy sets and make it possible to consider problems with a higher degree of uncertainty, which is typical, in particular, for the methods of image representation and for the algorithms of the image processing [15, 16]. The corresponding membership functions T2 are determined as a generalized fuzzy set by introducing fuzzy intervals; such approach correlates with inaccuracy perception by humans [6]. Fuzzy sets of the T2 feature fuzzy membership functions and are able to simulate such uncertainties [8, 17]. Fuzzy membership functions T2 is characterized by upper  $\overline{h(x)}$  and lower  $\underline{h(x)}$  boundaries, each of them is determined by the lower (LMF) and upper (UMF) membership functions T1.

The usage of fuzzy logic makes it possible to minimize uncertainties: removing randomness in fuzzy sets T1 leads to unambiguity, and removing uncertainty in T2 leads to the fuzzy sets T1.

Today, image processing algorithms with using T2 are proposed for solving problems of clustering [18], filtering [19], boundary detection [20] and in the image classification [21]. In [22], informational capabilities of the method of segmentation of low contrast grayscale images based on the fuzzy sets T2 were investigated.

## 3. Problem statement

Purpose of this work was to study a method for improving quality of the grayscale images basing on orthogonalization of fuzzy membership functions of the type-2 with using a method of singular value decomposition and selective filtering of combined noise in conditions of a priori uncertainty of the imaging system characteristics and in the absence of a priori information about the location of a possible objects of interest.

## 4. Materials and Methods

The ideas of methods of projection into eigen subspaces as one of the tools for mathematical processing of experimental data were presented in [23, 24]. From the theory, it is well known that decorrelation makes it possible to separate information and improve accuracy of estimates, which,



from a fundamental point of view, opens up additional possibilities for heightening sensitivity of the analysis procedure.

The Principal Component Analysis (PCA) algorithm for image processing problems representing two-dimensional structures was first applied in practice in the 2000s [25]. Currently, the PCA and other orthogonalization methods are used for solving such problems as compression of visual information and feature extraction during object recognition and video image search.

In the tasks of processing an ensemble of  $K$  images represented by a matrix of brightness values  $I$  and dimension  $[dy \times dx]$  with the help of orthogonalization methods, the main purpose is to transform initial data into a new coordinate system, for which the following condition is satisfied: the sample variance of the data along the values of the  $k$ -th coordinate should be maximum on the assumption of orthogonality to the first  $k-1$  coordinates.

In [26], an algorithm for improving sensitivity and reliability of the image segmentation is described, which is based on the process of multi-stage processing, which includes: expansion of space of the input features based on the initial data by using the fuzzy clustering method; orthogonalization of the obtained fuzzy membership functions; and, on their basis, formation and visualization of new informative features.

In [27, 28], informational capabilities of the method of automorphic imaging with implementation of orthogonal decomposition are considered in relation to solving the problems of filtering of low contrast grayscale images. The method is based on the use of dimension-increasing informative features of locally adaptive transformations, and makes it possible to apply methods of multidimensional data processing to the grayscale images.

Processing of images with the help of the fuzzy methods can be considered as a nonlinear type of the initial data transformation, the feature of which is that it is performed for the functions of pixel membership with predefined clusters [2].

One of the most common types of images formed by systems of various physical nature are the grayscale images represented by the level of quantization of their energy characteristics  $I(x, y)$ , where  $x$  and  $y$  are the coordinates of a pixel, which usually takes a value within the range  $[0, 1]$  or  $[0, 255]$ .

An image of the size of  $[dy \times dx]$  with gray levels  $L$  can be considered as an array of the fuzzy singletons (i.e. fuzzy sets with a single reference point), which display a value of the membership of the fuzzy set T1  $u_{x,y}$  for each point of the image  $I(x, y)$  relative to its property to be analyzed (for example, brightness, homogeneity, noise, etc.). Then, in order to describe an ambiguity of the type-2 inherent to the image, one can introduce the fuzzy membership functions of the type-2 (MFT2) basing on the expression:

$$a_{k,i}^t = u_{k,i}^h - u_{k,i}^l, \quad (1)$$

where  $u^h$  is the "upper" value of the membership function T1 (MFT1), and  $u^l$  is the "lower" value of the MFT1.

In Figure 1 graphical image of the membership functions T2 is shown, which is characterized by upper and lower boundaries, each of them is determined by the lower (LMF) and upper (UMF) membership functions of the type-1. The shaded area in right part of Figure 1 is called a footprint of uncertainty (FOU) and displays uncertainty of the solution made in the range between the upper and the lower boundaries. The distribution form is determined by nature of the problem uncertainty. The membership estimation for each element of the T2 type is a fuzzy set with the value of the T2 within the range  $[0, 1]$ .

For the case of a gray-scale image, this approach allows forming a multidimensional matrix  $X$  with dimensions  $[dy \times dx \times K]$ , which makes it possible to use an orthogonal singular transformation during its processing.

From mathematical point of view, matrix  $X$  is decomposed into the product of three matrices:

$$X = U \cdot S \cdot V^T. \quad (2)$$

Here,  $U$  (left singular vectors) is a matrix formed by the orthonormal eigenvectors  $u_r$  of the matrix  $X \cdot X^T$ , which correspond to the values  $\lambda_r$ :

$$X \cdot X^T \cdot u_r = \lambda_r \cdot u_r, \quad (3)$$

$V$  (right singular vectors) is a matrix formed by the orthonormal eigenvectors  $v_r$  of the matrix  $X^T \cdot X$ :

$$X^T \cdot X \cdot v_r = \lambda_r \cdot v_r, \quad (4)$$

and  $S$  (eigenvalues) is a positively-defined diagonal matrix whose elements are singular values  $\sigma_1 \geq \dots \geq \sigma_r \geq 0$  equal to the square roots of the eigenvalues  $\lambda_r$ .

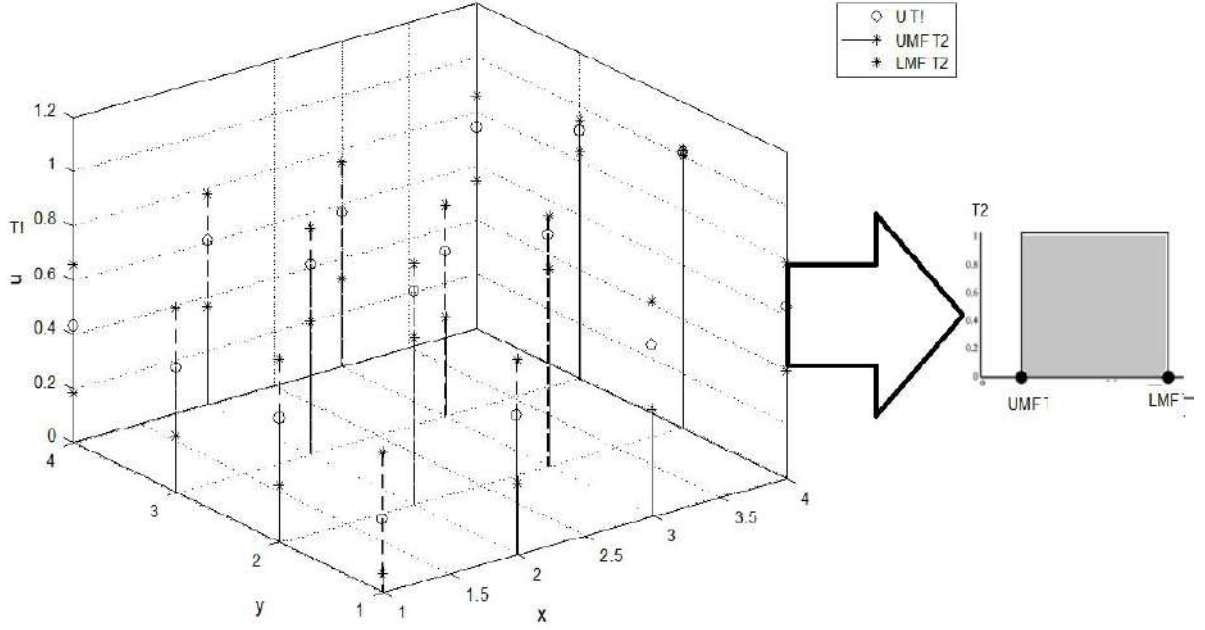


**Figure 1:** Type-2 fuzzy set: lower (LMF) and upper (UMF) membership functions and its interval footprint of uncertainty (FOU).

As a result, this transformation provides a transition of informative features to the new orthogonal space. The first singular vector is low-frequency and has a constant component not equal to zero. The rest of the singular vectors can be interpreted as noise components.

The combination of singular eigenvectors obtained in the space of membership functions T2 for gray levels of the grayscale image allows performing selective filtering of the noise.

Therefore, the proposed in this work algorithm for improving quality of the grayscale images consists of the following steps.

1. Scaling of initial range of the input image brightness to the range  $[0,1]$ .
2. Preprocessing of initial image The need for this step arises due to the possible irregularity of the background component in the image, which, in particular, leads to the appearance of light-stuck or darkened areas. In this research, we performed this procedure by the methods of adaptive power-law correction of brightness and local background subtraction [29]. In the first method, the initial grayscale image  $I$  undergoes the following transformations:

$$I_{x,y}^1 = I_{x,y} - 1/256, I_{x,y} > (\bar{I} + 0.5)/2, x \in [1, dy], y \in [1, dx], \quad (5)$$

where  $\bar{I}$  is a mean of the brightness of image  $I$ . The need of applying the expression (5) is explained by necessity to use the power transformations. As a result, a slight decrease in brightness occurs, and values equal to 1 are deleted. Then, the image  $I^1$  is subjected to the following transformations:

$$I_{x,y}^2 = \left( I_{x,y}^1 \right)^{1 + \text{sgn}(I_{x,y}^1 - \bar{I})} \left( I_{x,y}^1 \right)^{\left( 1 - \text{sgn}(I_{x,y}^1 - \bar{I}) \right) \left( \frac{\bar{I} + 0.5}{2} - I_{x,y}^1 \right)}, \quad (6)$$

$$I_{x,y}^3 = \left(I_{x,y}^2\right)^{1+\left(I_{x,y}^2\right)^{-I_{x,y}^2/2}}. \quad (7)$$

Transformation (6) proportionally reduces brightness of the light-stuck areas and increases brightness of the dark areas. Transformation (7) proportionally reduces brightness of the entire image.

For the case when brightness correction is made by the method of local background subtraction [29], transformation of the initial image is performed with the help of non-overlapping windows (in our experiments window size was  $[15 \times 15]$ ), and brightness of all pixels in each window is determined by the expression:

$$w_{x,y}^1 = (\bar{w} + \bar{I}) / 2. \quad (8)$$

So, this is the way of how values of the image brightness  $I^1$ . On its basis, the image  $I^3$  is formed by the formula:

$$I_{x,y}^3 = I_{x,y} - I_{x,y}^1. \quad (9)$$

After applying expression (9) the image  $I^3$  is scaled to the range  $[0,1]$  and adaptive histogram equalization is applied to the obtained image.

3. The image  $I^3$  is interpreted as a fuzzy membership function. And for it, the MFT2 ( $I^4$ ) is formed on the basis of the "upper" ( $I_h^3$ ) and "lower" ( $I_l^3$ ) values of the MFT1 according to the formula (1). In this case,  $I_h^3$  и  $I_l^3$  are calculated by the method of power transformations with using the following formulas:

$$\left(I_{x,y}^3\right)_h = \left(I_{x,y}^3\right)^{1-\left(I_{x,y}^3\right)^{-I_{x,y}^3/2}}, \quad (10)$$

$$\left(I_{x,y}^3\right)_l = \left(I_{x,y}^3\right)^{1+\left(I_{x,y}^3\right)^{-I_{x,y}^3/2}}. \quad (11)$$

This step is necessary for the subsequent application of the singular value decomposition.

4. An ensemble of images is formed ( $I^3$ ,  $I^4$ ,  $I_l^3$  when performing the step 2 with the use of power-law correction, or  $I^4$ ,  $I_h^3$ ,  $I_l^3$  in the case of preprocessing by the method of local background subtraction). Various methods are required for forming an ensemble of images because when the processed images are light-stuck the preprocessing method affects the brightness level  $I_h^3$ . In case of the use of adaptive power-law correction, brightness level  $I_h^3$  слишком высок, is too high, which makes this image uninformative in comparison with the preprocessed initial picture.

For this ensemble, singular value decomposition is applied. The resulting matrix of the left singular vectors  $U$  is interpreted as a multidimensional image  $I^5$  with dimensions  $[dy \times dx \times K]$ , each spectral component of which is scaled to the range  $[0,1]$ .

5. On the basis of the matrix of right singular vectors  $V$ , a vector of coefficients  $C$  is calculated (it is used for estimating significance of the components of the matrix of left singular vectors) by the following formula:

$$C_i = \left( \left| \sum_{j=1}^K V_{i,j} \right| + \left| \sum_{j=1}^K V_{j,i} \right| \right) / 2, i \in [1, K]. \quad (12)$$

This vector is ordered in descending order, and its elements are normalized so that their sum is equal to 1.

6. The vector  $dC$  is formed, which contains differences for each pair of adjacent elements of the vector  $C$ .

7. The value  $dC_a$  is calculated by the formula:

$$dC_a = \left( \left( \sum_{j=1}^{K-1} dC_j / (K-1) \right) + (dC_{\min} + dC_{\max}) / 2 \right) / 2, \quad (13)$$

where  $dC_{\min}$  and  $dC_{\max}$  are the minimum and maximum elements of the vector  $dC$ , respectively.

8. On the basis of the value  $dC_a$ , when scanning the elements of the vector  $dC$  starting from the end such index  $i_{\max}$ , is selected, for which the following condition should be satisfied:

$$dC_{i_{\max}} \geq dC_a. \quad (14)$$

Then elements of the vector  $dC$  with indices from 1 to  $i_{\max} + 1$  are normalized so that their sum is equal to 1.

9. The final image  $I^6$  is formed as a weighted sum of the most significant components of the matrix  $U$  according to the following formulas:

$$I_{y,x}^6 = \sum_{j=1}^{i_{\max}+1} I_{x,y,j}^5 \cdot dC_j \cdot S_j, \quad (15)$$

$$S_j = \text{sgn} \left( \sum_{i=1}^{n_i} V_{i,j} \cdot \sum_{j=1}^{n_j} V_{j,i} \right). \quad (16)$$

For the final image, histogram equalization is applied when its preprocessing is performed by the method of adaptive power-law correction, and inversion, which is followed by adaptive histogram equalization, is applied when the preprocessing is based on the local background subtraction method.

## 5. Results and Discussion

Quantitative metallography is widely used for specifying characteristics of the alloy microstructure, namely: volumetric content of phases, grain size, specific surface area of grain boundaries, distance between similar particles or phases, and others.

To obtain high quality materials, experimental alloys are smelted and mechanism of forming their structure and morphology, which determine their properties, is researched. For achieving this goal, among others, methods that allow to form images of the test samples are used (for example, micro X-ray or X-ray structural, metallographic analyzes, microscopy). Based on analysis of the images brightness characteristics it is possible to determine such quantitative parameters as the average size of phases, geometrical value of the external specific surface of the phase, and statistical data.

However, used digital images often feature an inadequate quality due to irregular background, noise, aberration artifacts, poor contrast, etc. In order to obtain reliable quantitative information from the pixel intensity values, it is necessary to apply correction methods to ensure good accuracy of photometry and to eliminate common defects of the images.

In Figure 2a, an example of grayscale image of the phosphorus-containing alloy Fe-2%P-0.042%C is shown. This image was obtained with magnification of x250 on a metallographic microscope GX51 with a digital image analysis system of the company "Olympus". To separate the individual phases of the alloy during determining its structure and properties, chemical etching was used at first. After that thermal etching at temperatures of 400-600 °C with natural air circulation was performed. The atoms that make up the phases interact with oxygen; therefore, an oxide membrane of different thickness is formed in different phases. When observing through an optical microscope and, accordingly, in the image, the phases have different levels of intensity [30].

The analysis of image in Figure 2a is difficult, in particular, due to the nonuniformity of the background and the presence of the light-stuck area.

To obtain experimental results we used Matlab 6.1. Source code for proposed methods was written in internal Matlab language (except for standard functions such as svd, histeq, adapthisteq, mat2gray).

In Figure 2b and Figure 2c, the results of preprocessing of the initial picture by means of the adaptive power-law correction and by method of local background subtraction are shown, respectively. In both cases, we managed to reduce brightness of the light-stuck area and to preserve the overall intensity level at the level acceptable for visual analysis. In the latter case, the image is more detailed, but at the same time it contains a blur effect.

In Figure 3 and Figure 4, the results of MFT2 formation are shown, as well as the "lower" and "upper" values of the MFT1 interpreted as grayscale images, for both preprocessing methods. It can

be easily seen that preprocessing based on the local background subtraction method gives higher brightness in the generated images.

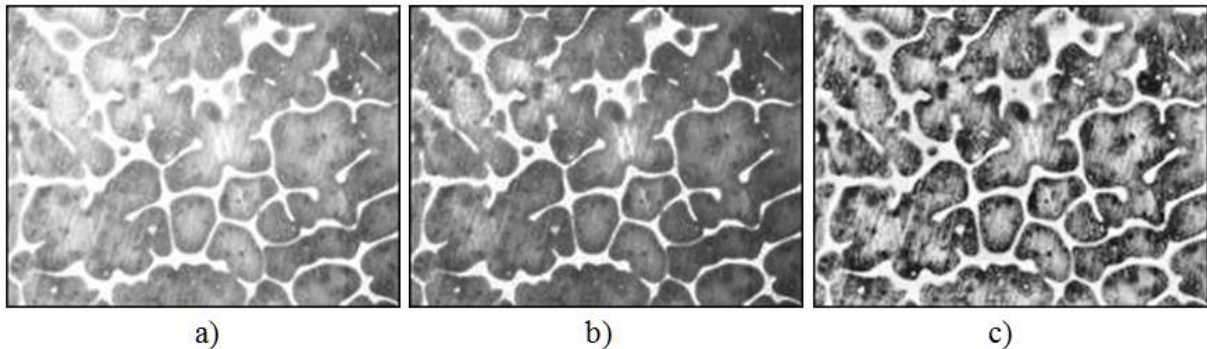


**Figure 2:** Image of the phosphorus-containing alloy Fe–2%P–0,042%C: a – initial grayscale image (142x186); preprocessing based on the b – adaptive power-law brightness correction; c – method of local background subtraction.

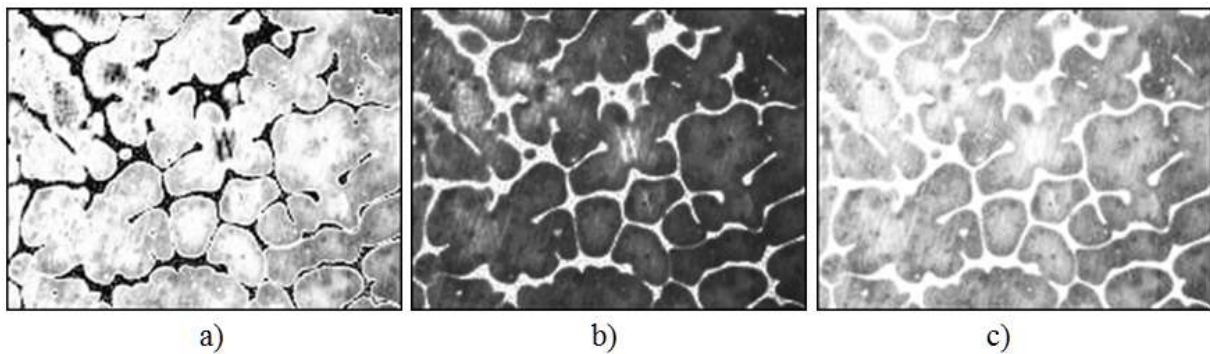


**Figure 3:** Calculation of the MFT2 for preprocessing based on the adaptive power-law brightness correction: a – MFT2; b – "lower" value of the MFT1; c – "upper" value of the MFT1.

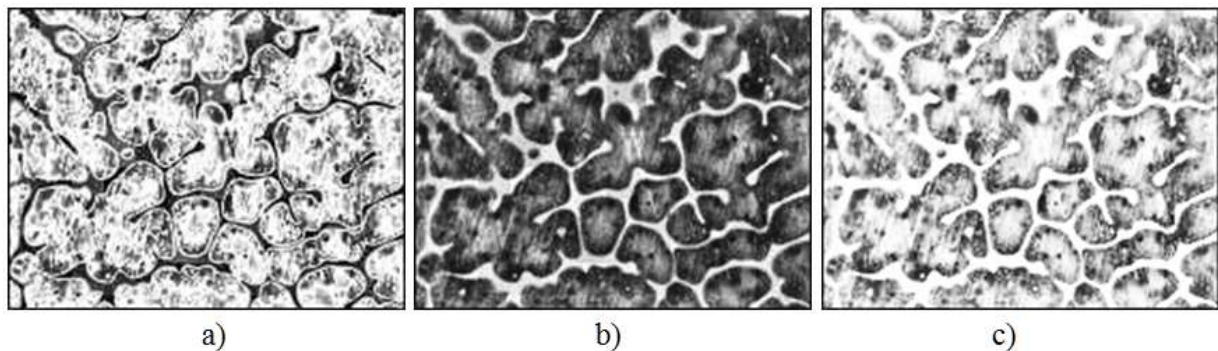


**Figure 4:** Calculation of the MFT2 for preprocessing based on the method of local background subtraction: a – MFT2; b – "lower" value of the MFT1; c – "upper" value of the MFT1.

In Figure 5, formation of the resulting image for both preprocessing methods is shown. It should be noted about different levels of detailing in the results. Use of the adaptive power-law correction led to the lower level of the resulting image brightness, but, at the same time, gives better image definition, which is preferable when determining quantitative parameters and highlighting the contours of the objects and various areas within one object. Preprocessing based on the method of local background subtraction resulted in a brighter and more contrasting image, which simplifies its visual analysis, but there the blur effect is seen.



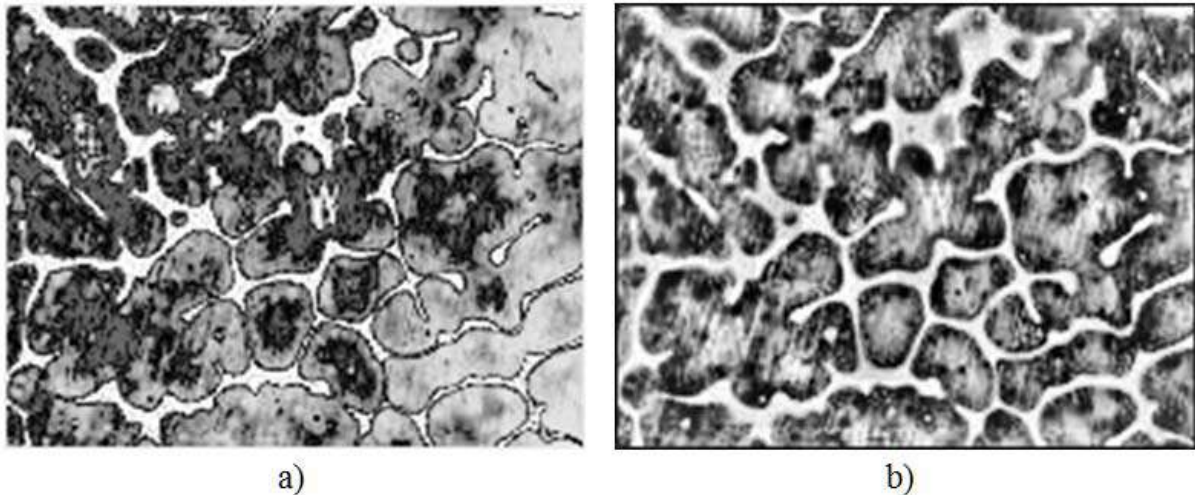


**Figure 5:** The resulting image after preprocessing by the: a – adaptive power-law intensity correction; b – local background subtraction method.

## 6. Conclusions

On the basis of analysis of the experimental results obtained, the following conclusions are made:

- the usage of fuzzy sets of the T2 type allows to synthesize additional parameters on the basis of the grayscale initial image with using of nonlinear functions based on the local transformation of the brightness levels, and to obtain an ensemble of data, to which methods of multidimensional information processing can be applied;
- orthogonalization, the method of singular value decomposition in particular, applied to the ensemble of images with taking into account components of the T2 set, makes it possible to take into consideration original ambiguity and uncertainty of the initial data, to analyze the ensemble as a whole and, at the same time, to interpret each new component as a result of anisotropic filtering in two-dimensional plane of the spatial frequencies;
- visualization of the parameters synthesized on the basis of the eigenvalues of the singular value decomposition, makes it possible to increase the level of detailing, contrast and resolution of the resulting image and, hence, to improve reliability of visual and automated analysis;
- promising areas for further researches are the usage of different orthogonal transformations; and applying of various methods for preprocessing of initial data and formation of sets of the T2 type.

## 7. Acknowledgements

The authors thank for the images used in experiments, which were kindly provided by candidate of physical and mathematical sciences, docent of Dnieper National University Named by Oles Honchar (Faculty of Physics, Electronics and Computer Systems) Nadezhda Karpenko.

## 8. References

- [1] W.K. Pratt, Digital Image Processing, John Wiley and Sons Inc., New York, NY, 2001.
- [2] J.C.Bezdek, J. Keller, R. Krishnapuram, N.R. Pal, Fuzzy Models and Algorithms for Pattern Recognition and Image Processing, Handbooks of Fuzzy Sets series, Kluwer Academic Publisher, Boston, 1999.
- [3] H.R. Tizhoosh, H. HauBecker, Fuzzy Image Processing: An Overview, in: B. Jähne, H. HauBecker. and P. GeiBler, (Ed.). Handbook on Computer Vision and Applications, Academic Press, Boston, 1999, volume 2, pp. 683-727.

- [4] H. Bustince et al., (Ed.), *Fuzzy Sets and Their Extensions: Representation, Aggregation and Models*, Springer, 2008.
- [5] L.A. Zadeh, The concept of a linguistic variable and its application to approximate reasoning, *Information Science* 8 (1975) 199–249. doi: 10.1016/0020-0255(75)90036-5.
- [6] Z. Chi, H. Yan, T. Pham, *Fuzzy algorithms: With Applications to Image Processing and Pattern Recognition*, Singapore, New Jersey, London, Hong Kong, Word Scientific, 1998.
- [7] L.A. Zadeh, Fuzzy sets and their application to pattern recognition and clustering analysis, in: J. Van. Ryzin (Ed.), *Classification and Clustering*, Academic Press, 1977, pp. 251-299. doi: 10.1016/C2013-0-11644-3.
- [8] L. Cinquea, G. Forestib, L. Lombardic, Clustering fuzzy approach for image segmentation. *Pattern Recognition* 37 (2004) 1797-1807. doi: 10.1016/j.patcog.2003.04.001.
- [9] L. Pham Dzung, Jerry L. Prince, Adaptive Fuzzy Segmentation of Magnetic Resonance Images, *IEEE Transactions on Medical Imaging* 18 (1999) 737-751. doi: 10.1109/42.802752.
- [10] D. Forsyth, J. Pons, *Computer vision: a modern approach*, 2-nd ed., Prentice Hall, 2012.
- [11] A. Yegorov, L. Akhmetshina, *Optimizatsiya yarkosti izobrazheniy na osnove neyro-fazzi tekhnologiy*, Lambert, 2015.
- [12] O. Castillo, P. Melin, *Type-2 Fuzzy Logic: Theory and Applications*, Springer-Verlag, 2008.
- [13] Wu Haoyang, Wu Yuyuan, Luo Jinping, An Interval Type-2 Fuzzy Rough Set Model for Attribute Reduction, *IEEE Transactions on Fuzzy Systems* 17.2 (2009) 301–315. doi: 10.1109/TFUZZ.2009.2013458.
- [14] J.M. Mendel, R. John, Type 2 Fuzzy Sets Made Simple, *IEEE Transactions On Fuzzy Systems* 10.2 (2002) 117-127. doi: 10.1109/91.995115.
- [15] J.M. Mendel et al, Interval Type 2 Fuzzy Logic Systems Made Simple, *IEEE Transactions on Fuzzy Systems* 14.6 (2007) 808-821. doi: 10.1109/TFUZZ.2006.879986.
- [16] W.B. Zhang, W.J. Liu, Fuzzy Clustering for Rule Extraction of Interval Type-2 Fuzzy Logic System, in: *Proceedings of 46th IEEE Conference on Decision and Control*, 2007, pp. 5318-5322.
- [17] C. Hwang, F. Rhee, An interval type-2 fuzzy C spherical shells algorithm, in: *Proceedings of IEEE International Conference on Fuzzy Systems*, 2004, pp. 1117–1122. doi: 10.1109/FUZZY.2004.1375568.
- [18] Z. Ji, Y. Xia, Q. Sun, G. Cao, Interval-valued possibilistic fuzzy C-means clustering algorithm, *Fuzzy Sets and Systems* 253 (2014) 138–156. doi: 10.1016/j.fss.2013.12.011.
- [19] M. Yuksel, A. Basturk, Application of Type-2 Fuzzy Logic Filtering to Reduce Noise in Color Images, *IEEE Computational Intelligence Magazine* 7.3 (2012) 25–35. doi: 10.1109/MCI.2012.2200624.
- [20] P. Melin, C.I. Gonzalez, J.R. Castro, O. Mendoza, O. Castillo, Edge-Detection Method for Image Processing Based on Generalized Type-2 Fuzzy Logic, *IEEE Transactions on Fuzzy Systems* 22 (2014) 1515–1525.
- [21] L. Lucas, T. Centeno, M. Delgado, Land cover classification based on general type-2 fuzzy classifiers, *International Journal of Fuzzy Systems* 10 (2008) 207–216.
- [22] L. Akhmetshina, A. Yegorov, Low-Contrast Image Segmentation by using of the Type-2 Fuzzy Clustering Based on the Membership Function Statistical Characteristics In: *International Scientific Conference Lecture Notes in Computational Intelligence and Decision Making. AISC*, 2019, volume 1020. pp 689-700.
- [23] K. Pearson, On lines and planes of closest fit to systems of points in space. *The London, Edinburgh and Dublin Philosophical Magazine and Journal of Sciences* 6.2 (1901) 559-572.
- [24] H. Hotelling, Analysis of complex variables into principal components, *Journal of Educational Psychology* 24.6 (1933) 417-441.
- [25] J. Yang, D. Zhang, A.F. Frangi, J.-Y. Yang, Two-dimensional PCA: A new approach to appearance-based face representation and recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 26.1 (2004) 131-137. doi: 10.1109/TPAMI.2004.1261097.
- [26] L.G. Akhmetshina, A.A. Yegorov, Adaptivnaya nechetskaya segmentatsiya izobrazheniy na osnove kombinirovannogo singulyarnogo razlozheniya, *Vestnik KHNTU* 3 (2016) 3 (2016) 198-202.
- [27] L.G. Akhmetshina, Povysheniye razreshayushchey sposobnosti izobrazheniy geofizicheskikh

- poley na osnove metoda mnogomernoy ortogonal'noy adaptivnoy klasterizatsii, Naukovyy Visnyk Natsional'noho hirnychoho universytetu 10 (2003) 35-38.
- [28] L.G. Akhmetshina Adaptivnaya fil'tratsiya shumov v signalakh i izobrazheniyakh: metod selektivnogo singulyarnogo razlozheniya avtomorfnogo otobrazheniya, Iskusstvennyy intellekt 3 (2005) 328-335.
- [29] A.A. Luk'yanitsa, A.G. Shishkin, Tsifrovaya obrabotka videoizobrazheniy, Ay-Es-Es Press, Moskva, 2009.
- [30] O.B. Sukhova, N.V. Karpenko, The peculiarities in contact interaction processes and technology to fabricate multi-layered composites, Metallofizika i Noveishie Tekhnologii 30 (2008) 585-594.



# Neuro-fuzzy Control System for a Non-deterministic Object in Real Time

Sergey Ivanov<sup>a</sup>, Nataliia Maksyshko<sup>a</sup> and Mykola Ivanov<sup>a</sup>

<sup>a</sup> Zaporizhzhia National University, Zhukovsky str., 66, Zaporizhzhia, 69063, Ukraine

## Abstract

In this paper discusses the use of neuro-fuzzy control systems as a tool for managing non-deterministic objects in real time. This paper discusses modern control tools structural models of a discrete quasi-invariant automated control system. In this paper the analysis of an automated control system is presented, which is based on the use of typical models of discrete automated control systems. According to the proposed solution in the automated control system in real time it is proposed to use a neuro-fuzzy control system as a function of the object and the system's transfer ratio. The neuro-fuzzy control system is based on the learning process of an artificial neural network (ANN), which allows to define the rules of fuzzy inference (FIS). The paper proposes the ANFIS model, which is implemented by using the fuzzy system Takagi T., Sugeno M., also is considered an algorithm based on seven fuzzy rules. In this paper is presented a technique for implementing a neuro-fuzzy control system for non-deterministic objects by using Matlab. The use of Matlab made possible to create a model of an adaptive neuro-fuzzy inference system. The paper describes the process of training a neural network, where a hybrid method is chosen, which is a combination of the least squares method and the method of decreasing the inverse gradient. The results of testing a neuro-fuzzy control system of a non-deterministic object are presented, which confirmed the possibility of using a neuro-fuzzy control model. It is constructed the structure and the result in the form of a control surface of the neuro-fuzzy ANFIS model. The results presented in this paper allow to conclude about the possibility of using a neuro-fuzzy control system for non-deterministic objects in real time.

## Keywords

Neuro-fuzzy control systems, adaptive neural network (ANN), fuzzy inference system (FIS), ANFIS model

## 1. Introduction

One of the methods for constructing modern control systems is the synthesis of intelligent systems based on neuro-fuzzy systems [1]. The peculiarity of systems of this class is the use of neural networks and fuzzy logic to control complex dynamic objects that function under conditions of uncertainty and conflict. Uncertainty in this case it exists a lack of information, which is necessary to obtain a quantitative description of the processes occurring in the system, and the complexity of the control object. The use of classical methods, the description of the control system assumes that the control objects are described by linear dynamic links of a low order. This assumption often leads to the fact that classical control systems in practice do not provide the desired indicators of fast and efficient control. Therefore, the neuro-fuzzy control system, using the procedures of artificial neural networks and fuzzy logic, makes it possible to identify complex processes both in technology and in economics. The use of neuro-fuzzy systems makes it possible to solve the problem of constructing control systems in conditions of uncertainty on the basis of the available statistical and experimental

CMIS-2021: The Fourth International Workshop on Computer Modeling and Intelligent Systems, April 27, 2021, Zaporizhzhia, Ukraine  
EMAIL: flydaiver@gmail.com (S. Ivanov); maxishko@ukr.net (N. Maksyshko); nn\_iva@ukr.net (M. Ivanov)  
ORCID: 0000-0003-1086-0701 (S. Ivanov); 0000-0002-0473-7195 (N. Maksyshko); 0000-0002-1908-0763 (M. Ivanov)



© 2020 Copyright for this paper by its authors.  
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).  
CEUR Workshop Proceedings (CEUR-WS.org)

data obtained about the control object. It should be noted that the use of only one neural network in the tasks of automated control has a number of disadvantages. This is how the neural network receives information about the control object in the learning process, and this requires statistical data. Therefore, this disadvantage can be eliminated by using fuzzy set structures, which allow for the formalization of fuzzy variables. Therefore, this article is devoted to the actual problem of constructing a neuro-fuzzy control system for non-deterministic objects in real-time systems.

## 2. The purpose of the article

The paper is devoted to the construction of a neuro-fuzzy control system for a non-deterministic object in real time. The paper presents an approach to constructing a model of a neuro-fuzzy control system for non-deterministic objects, as well as the process of implementing a model of an automated control system ANFIS.

## 3. Literature review

Today, in control systems fuzzy sets, the authors Ivanov M., Maksyshko N., Ivanov S. and Terentieva N. applied a method of modeling multidimensional processes [2].

In the works of Casillas J., Cordyn O. [3], Cordyn O., Herrera F. [4], and Espinosa J., Vandewalle J. [5], the problems of tuning fuzzy systems based on rules for linguistic variables and an algorithm for extracting rules are considered.

Authors Wang, D., He, H., Zhao, B. and Liu, D. [6] identified the main advantages of using optimal control based on a neural network (NN) with feedforward.

In the articles of the authors Guillaume S. [7, 8] and Herrera F. [9], the problems of constructing a fuzzy inference system (FIS) for modeling and control process are considered, as well as the use of genetic algorithms for the design of fuzzy systems.

In the work, Anzaklis P.J. [10] considered the issues of hybridization of fuzzy logic systems. The identified shortcomings in the use of a fuzzy inference system are solved on the basis of a neural network that is able to learn and take into account previous knowledge.

In the article, Cui R., Yang C., Li Y., and Sharma S. [11] solved the problem of trajectory tracking and object control in the area of discrete time based on two neural networks (NN). Therefore, the data during the development of the control system must be specified explicitly.

Using a neuro-fuzzy approach eliminates the above problems. Jang J.S.R. in [12] presented an architecture and training procedure, which is based on ANFIS.

The results of the study by Boyacioglu M.A. and Avci D. in [13] showed the ability of the ANFIS system to predict the efficiency of the stock market.

In the works of Takagi T., Sugeno M. [14], a mathematical tool is presented for constructing a fuzzy model of a system in which fuzzy inferences are used. The authors in their works have shown a method for identifying a system using its input-output data.

In the work of Zhang H. and Liu D. [15], the methodology of fuzzy logic is presented and efficiency is proved when working with complex nonlinear systems that contain uncertainties.

New solutions for the use of an artificial neural network and an adaptive neuro-fuzzy inference system were considered in the work of Suparta, W., Alhasa, K. M. [16]. This paper presents the theoretical foundations and explains in detail this method, as well as emphasizes its importance for evaluating the model under study.

Saugat B., Debabrota B., Amit K. and Tibarewala D. [17] examined neural networks and showed that the Adaptive Neural Fuzzy Inference System (ANFIS) works effectively to deal with uncertainties.

The problems of assessing the quality of input signals were considered in the work of Kumarl A. and Qureshi M.F [18]. This paper provides an overview and analyzes of active filter methods. The main goal of this work is to develop a highly efficient system that is integrated with the network. This solution is based on the use of ANFIS, which allows the system to work quickly and give improved results.

Karaboga, D., Kaya, E. [19] considered two groups of ANFIS parameters: premises and consequences. These studies explored hybrid ANFIS training approaches to assess ANFIS training.

#### 4. Building a model of a neuro-fuzzy object control system

To determine the input components of the neuro-fuzzy object control system, the method of system analysis is applied. This method allows you to analyze the mechanisms of interaction of an economic object with the environment. Typical procedures correspond to business processes of non-deterministic discrete objects. The decision to use marketing, resource and production procedures is made based on an analysis of the degree of compliance, in this case, with business processes in an economic entity. This approach combines the advantages of the principle of using typical subsystems of automated control systems and the process approach. To build automated control systems that operate under conditions of random processes, the automated systems approach can be applied. However, for automated control systems it is not always possible to obtain a system of equations that fully described it. Existing automated control systems are conditionally invariant (quasi-invariant) to external factors. Therefore, the classical structural model of a discrete automated control system is usually presented in the following form (Fig. 1.).

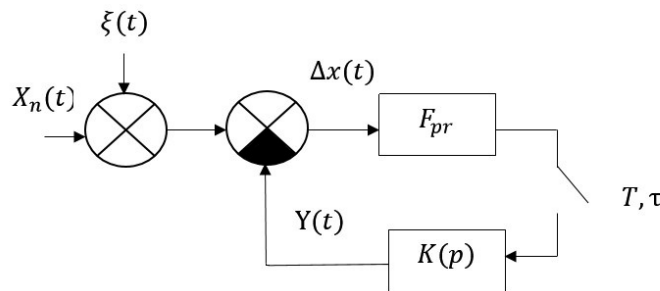


**Figure 1:** Structural model of a discrete quasi-invariant automated control system

The structural model of an automated control system includes a production function and vectors of controlled variables. Automated control system data depends on:

$X_n(t)$  – vectors of input parameters of the automated system;

$F_{pr}$  – object (enterprise) functions;

$\xi(t)$  – vectors of external disturbances on the control system.

It should be noted that during the time  $\Delta t$ , the vector of mismatch  $\Delta x(t)$ , is generated, which is necessary for the analysis and processing of the input data. All components of the structural model of the control system depend on time, which change in accordance with the period of the switch (time sampling). The switch is displayed in a discrete automated control system in the form of a switch with a period  $T$ . This period of time in the system is necessary for the analysis and processing of the initial data from the moment of their arrival to the control of the object. Therefore,  $\tau$  - the duration of pauses will be determined by the time between the arrival of the input data and the control of the object. The closed position of the switch of the automated system characterizes the state when information is provided for making management decisions in accordance with the transmission ratio. In the case of an open switch, the automated system is in watch mode. It should be noted that fluctuations of the input parameters fall into the control system mismatch block and determines the need to transform managerial decisions. In general, the system of automated control of the object should be considered as a set of tasks to be solved. According to the proposed solution in an automated control system in real time, it is proposed to use a neuro-fuzzy control system as a function of the object ( $F_{pr}$ ) and the system's transfer ratio ( $K(p)$ ). The model of a neuro-fuzzy system for automated control of non-deterministic objects is shown in the Figure 2.

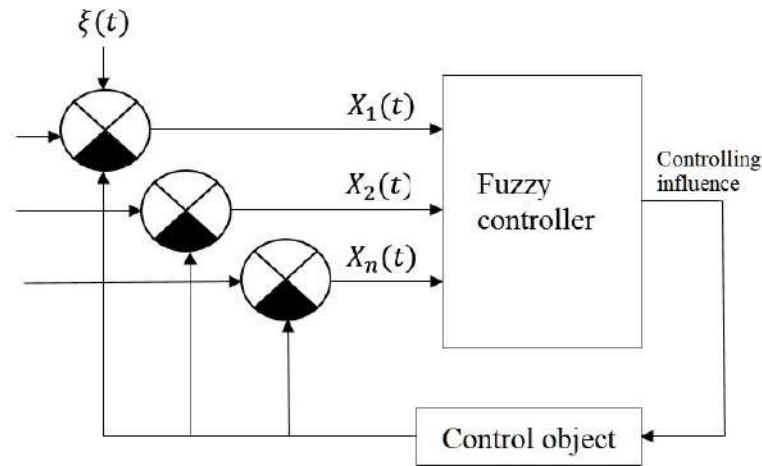


**Figure 2:** Model of a neuro-fuzzy control system for non-deterministic objects

The system ensures the adoption of  $n$  managerial decisions under the condition of fuzzy input values. The adaptation of the system is provided at the stage of training the neural network. The neuro-fuzzy control system is based on the learning process of an artificial neural network (ANN), which allows to define the rules of fuzzy inference (FIS). As soon as the parameters of the fuzzy inference are determined, the neural networks operate as usual. This is an integrated model in which a neural network (ANN) training algorithm is used to determine the parameters of a fuzzy inference system (FIS). Fuzzy inference system and corresponding membership functions. On the other hand, the learning mechanism of a neural network does not depend on statistical information, but is standard for the chosen architecture of an artificial neural network.

The ANFIS model (Adaptive-Neuro-Based Fuzzy Inference System) ANFIS with the implementation of the Takagi T., Sugeno M. fuzzy system, which is a five-layer feedforward neural network, is shown in the Figure 3.

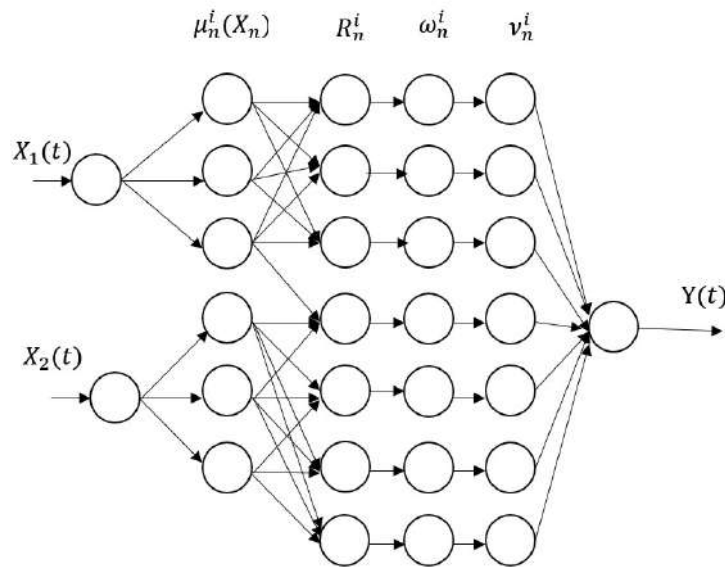


**Figure 3:** Structural view of ANFIS system.

The input values of the model  $X_1$  and  $X_2$  allow you to determine the mismatch between the current and planned value of the variable. The output variable  $Y$  is the control value of the automated system.

The first layer of the ANFIS system allows the definition of fuzzy sets of a set of input quantities. The outputs of the nodes of the layers of this level are the domains of membership functions for certain input values  $\mu_1(X_n)$ .

The second level of the system makes it possible to determine the generated fuzzy rules. At this level, one fuzzy rule will correspond to each layer. The layer of the second level of the system is connected with the nodes of the first level, which form the formation of the corresponding rules. The outputs of the layers in the system are calculated as ratios of the input quantities  $\omega_1$ .

The third level allows you to normalize the degree of rule compliance

$$\overline{\omega_i} = \frac{\omega_i}{\sum \omega_i}, \quad i = 1, \dots, n. \quad (1)$$

In the system, the non-adaptive layer determines the weight value of the execution of the fuzzy rule.

The fourth level of the adaptive system determines the contribution of each fuzzy rule to the value of the output value of the network. The layer of the fourth level determines the contribution of the fuzzy rule to the value of the output value of the system.

The fifth layer forms the value of the magnitude of the control system

$$Y = \sum Y_{t,n}^i. \quad (2)$$

Therefore, the ANFIS automated control system determines that only one fuzzy represents each value set. The training procedure from the ANFIS neural network has no restrictions on the modification of membership functions. To ensure the speed of training the neural network and the adaptability of the software implementation, the model Takagi T., Sugeno M.

The Takagi T., Sugeno M model is based on a high-performance neural network training procedure. The following indicators are selected to build the model:

$Y_{t,n}$  – output value of the control system per year by months  $t : t = 0, 1, \dots, 11$ ;

$x_{t,1}$  – the quantity of the first product sold per week (units);

$x_{t,2}$  – the quantity of the second item sold per week (units);

$x_{t,HR}$  – the amount of human resources that were used to produce goods.

This choice of variables allows you to determine the position of goods on the market and track the current changes in the market. As the output value of the control system, the model can be written in the following form:

$$x_{t,3} = \sum x_{t,HR}.$$

The choice of variables allows determine the position of goods on the market and current changes in the market. As the output value of the control system, the model can be written in the following form:

$$Y_{t,n}^i = a_0 + \beta_1 Y_{t-1,1} + \beta_2 x_{t,1} + \beta_3 x_{t,2} + \beta_4 x_{t,HR}. \quad (3)$$

Further, it can be seen that all variables affect the formation of fuzzy rules. The rules for the Takagi T., Sugeno M. model were constructed according to the following algorithm. The algorithm, which is built of seven fuzzy rules, is as follows:

*The Rule<sup>1</sup> :*

*If ( $x_{t,2} = A_1$ ) and ( $Y_{t-1,1} = B_1$ ) and ( $x_{t,1} = C_1$ ) and ( $x_{t,HR} = D_1$ ) Then*

$$Y_t^1 = a_0^1 + \beta_1^1 Y_{t-1,1} + \beta_2^1 x_{t,1} + \beta_3^1 x_{t,2} + \beta_4^1 x_{t,HR}$$

*The Rule<sup>2</sup> :*

*If ( $x_{t,2} = A_1$ ) and ( $Y_{t-1,1} = B_1$ ) and ( $x_{t,1} = C_1$ ) and ( $x_{t,HR} = D_2$ ) Then*

$$Y_t^2 = a_0^2 + \beta_1^2 Y_{t-1,1} + \beta_2^2 x_{t,1} + \beta_3^2 x_{t,2} + \beta_4^2 x_{t,HR}.$$

*The Rule<sup>3</sup> :*

*If ( $x_{t,2} = A_1$ ) and ( $Y_{t-1,1} = B_1$ ) and ( $x_{t,1} = C_2$ ) Then*

$$Y_t^3 = a_0^3 + \beta_1^3 Y_{t-1,1} + \beta_2^3 x_{t,1} + \beta_3^3 x_{t,2} + \beta_4^3 x_{t,HR}.$$

*The Rule<sup>4</sup> :*

If  $(x_{t,2} = A_1)$  and  $(Y_{t-1,1} = B_1)$  and  $(x_{t,1} = C_3)$  Then

$$Y_t^4 = a_0^4 + \beta_1^4 Y_{t-1,1} + \beta_2^4 x_{t,1} + \beta_3^4 x_{t,2} + \beta_4^4 x_{t,HR}.$$

The Rule<sup>5</sup> :

If  $(x_{t,2} = A_1)$  and  $(Y_{t-1,1} = B_2)$  Then

$$Y_t^5 = a_0^5 + \beta_1^5 Y_{t-1,1} + \beta_2^5 x_{t,1} + \beta_3^5 x_{t,2} + \beta_4^5 x_{t,HR}.$$

The Rule<sup>6</sup> :

If  $(x_{t,2} = A_2)$  Then

$$Y_t^6 = a_0^6 + \beta_1^6 Y_{t-1,1} + \beta_2^6 x_{t,1} + \beta_3^6 x_{t,2} + \beta_4^6 x_{t,HR}.$$

The Rule<sup>7</sup> :

If  $(x_{t,2} = A_3)$  Then

$$Y_t^7 = a_0^7 + \beta_1^7 Y_{t-1,1} + \beta_2^7 x_{t,1} + \beta_3^7 x_{t,2} + \beta_4^7 x_{t,HR}.$$

In the presented rules  $A_1, A_2, A_3, B_1, B_2, C_1, C_2, C_3$  are fuzzy sets, where membership functions are constructed using the built-in algorithm in Matlab. Also, Matlab calculates the coefficients of the equations  $\alpha_0^i, \alpha_1^i, \beta_1^i, \beta_2^i, \beta_3^i, \beta_4^i, i = 1, \dots, 7$ .

Implementation of the ANFIS automated control system model

Matlab applied the implementation of a neuro-fuzzy control system for non-deterministic objects. Matlab allows you to create a model of an adaptive system of neuro-fuzzy inference, as well as perform its training is shown in the Figure 4.

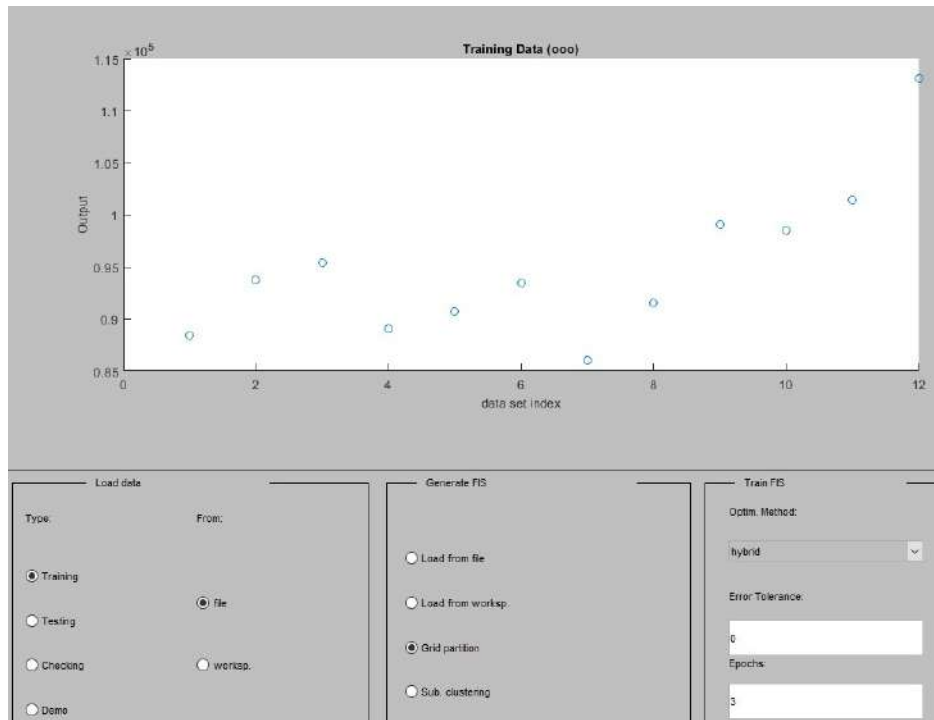


**Figure 4:** Data Entry in ANFIS.

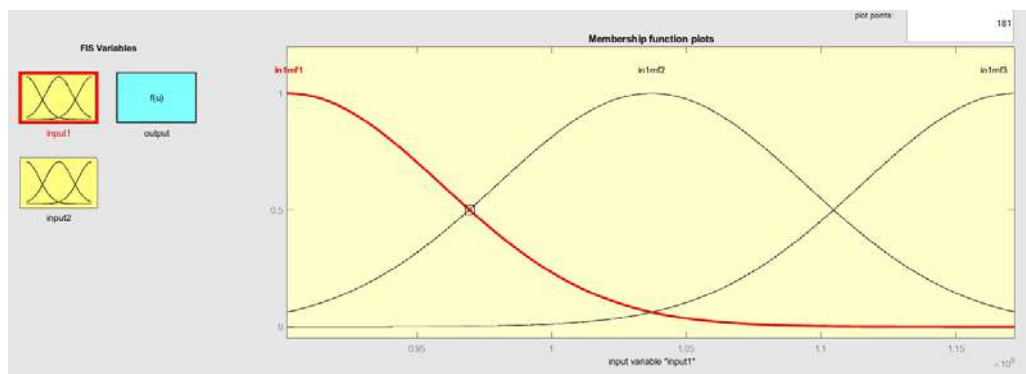
In the process of creating a management system in ANFIS, the statistical values of the demand for goods in the market for the year were used, as well as the optimistic and pessimistic scenarios of the production of goods (Table 1).

At the stage of FIS formation, membership functions were selected and the parameters of the fuzzy inference system were determined. When constructing, Gauss functions (gaussmf) were selected, which made it possible to display the input and output membership functions of the FIS, shown in Figure 5.

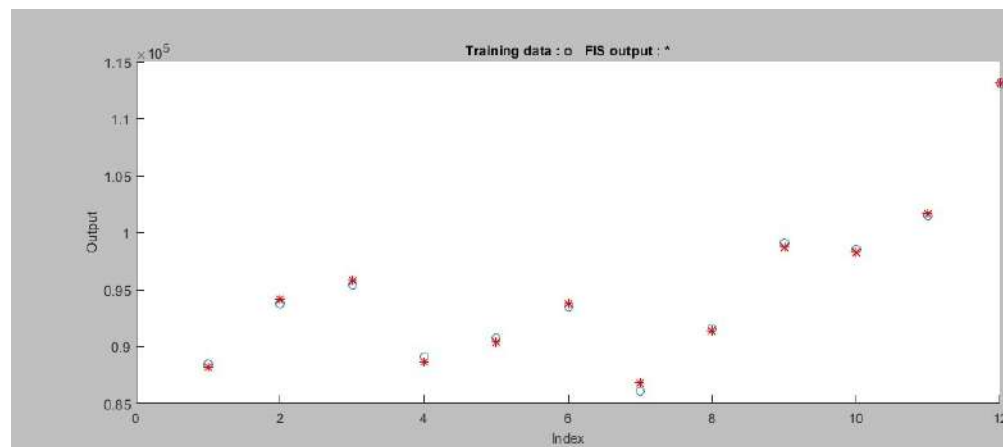
**Table 1**

Matching Statistics to Output Values

| The months | Statistical values of the demand for goods on the market for the year, quantity | Optimistic scenario for the production of goods, quantity | Pessimistic scenario for the production of goods, quantity |
|------------|---------------------------------------------------------------------------------|-----------------------------------------------------------|------------------------------------------------------------|
| January    | 92619                                                                           | 96760                                                     | 88478                                                      |
| February   | 97948                                                                           | 102089                                                    | 93807                                                      |
| March      | 99593                                                                           | 103734                                                    | 95452                                                      |
| April      | 93251                                                                           | 97392                                                     | 89110                                                      |
| May        | 94839                                                                           | 98980                                                     | 90698                                                      |
| June       | 97591                                                                           | 101732                                                    | 93450                                                      |
| July       | 90192                                                                           | 94333                                                     | 86051                                                      |
| August     | 95690                                                                           | 99831                                                     | 91549                                                      |
| September  | 103223                                                                          | 107364                                                    | 99082                                                      |
| October    | 102629                                                                          | 106770                                                    | 98488                                                      |
| November   | 105591                                                                          | 109733                                                    | 101450                                                     |
| December   | 117227                                                                          | 121368                                                    | 113086                                                     |

**Figure 5:** Input and output membership functions of FIS.

The next stage is the network training procedure, where a hybrid method is chosen, which is a combination of the least squares method and the inverse gradient decay method. Also, the number of training cycles (Epochs) is set equal to 100, after which the testing procedure follows (Fig. 6).

**Figure 6:** The result of the ANFIS system testing process.

The results of constructing the system are presented in the form of the structure of the ANFIS model as follows (Fig. 7).

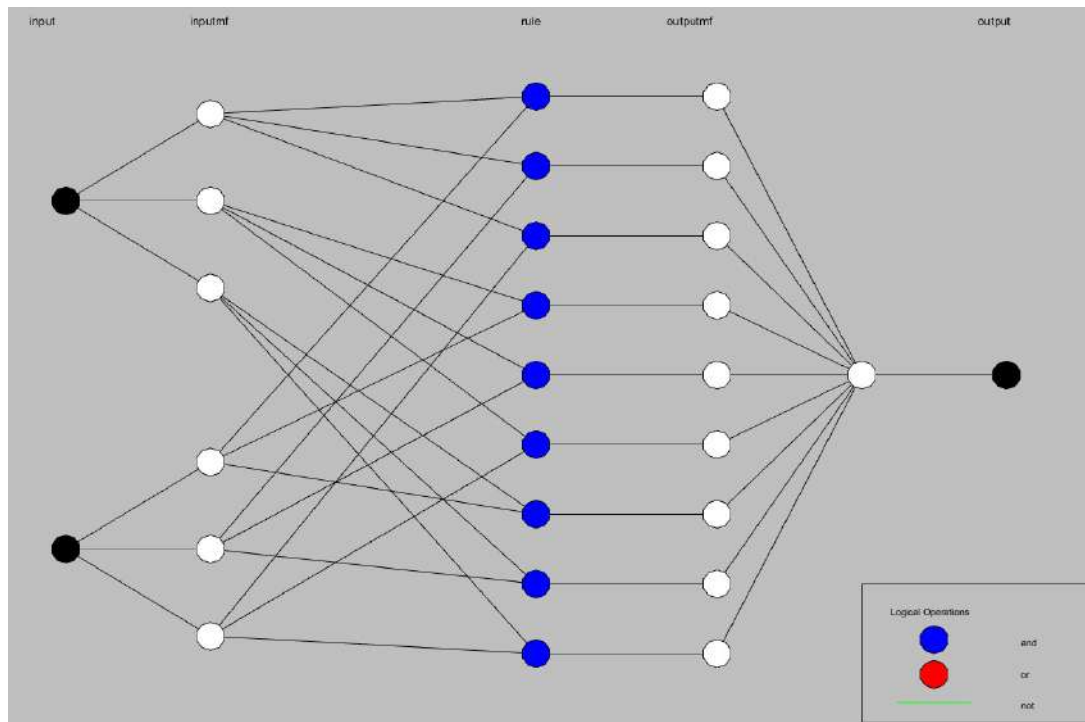


**Figure 7:** The constructed structure of the neuro-fuzzy model ANFIS.

In the process of studying the ANFIS model, the built-in Matlab procedure was used to view the obtained rules of fuzzy inference (Fig. 8).

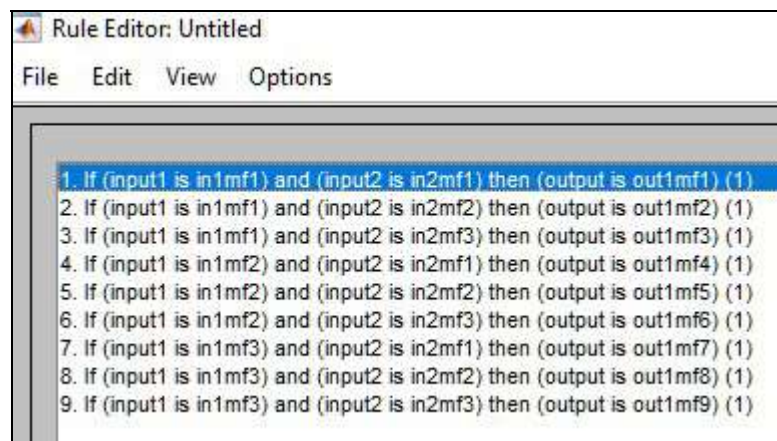


**Figure 8:** Procedure for viewing rules in a fuzzy inference system.

To analyze the results obtained, the following values of the input variables were set: the statistical value of the demand for goods on the market for the year  $1,0 \cdot e^{+5}$  and the optimistic scenario  $1,04 \cdot e^{+5}$ . To obtain this result, the control effect on the production of goods in the ANFIS system will be equal to  $1,07 \cdot e^{+4}$  (Fig. 9).

In addition, Matlab obtained control surfaces of the neuro-fuzzy model ANFIS, which is shown in the Figure 10.



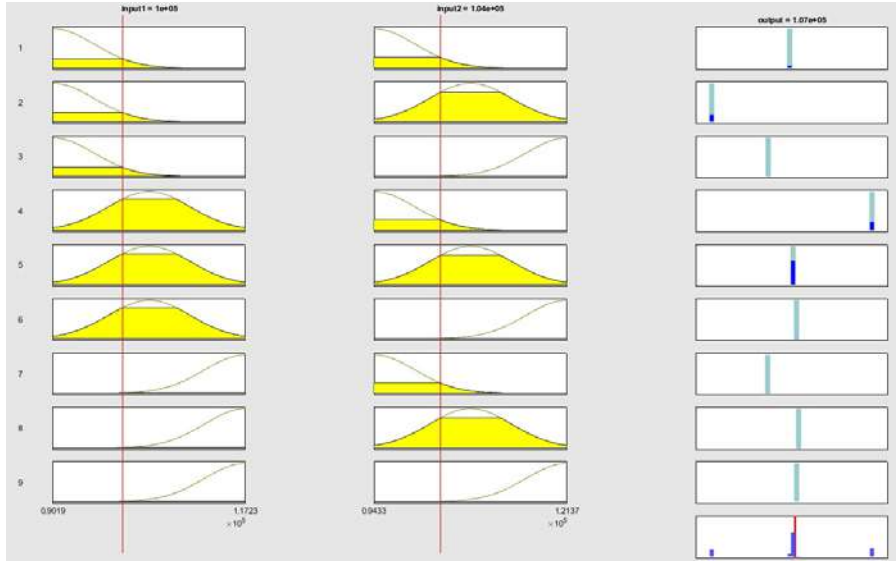


**Figure 9:** The results of the neuro-fuzzy control system.

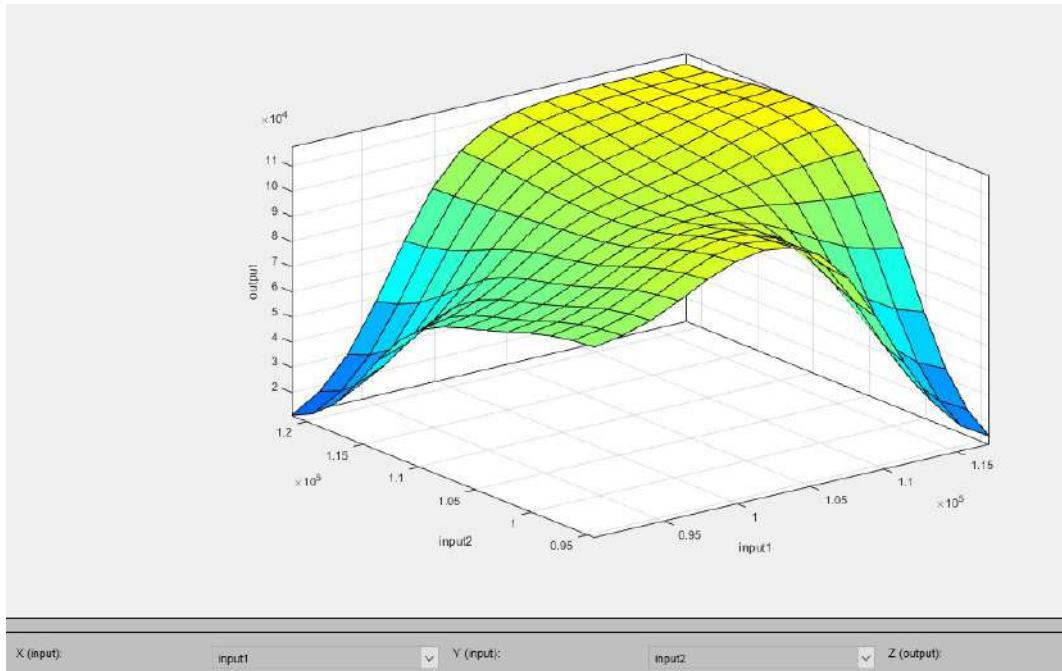


**Figure 10:** ANFIS neuro-fuzzy model control surface.

## 5. Conclusion

A model of a neuro-fuzzy control system for a non-deterministic object in real time is proposed. Structural models of a discrete quasi-invariant automated control system and a neuro-fuzzy control system for non-deterministic objects are considered and analyzed.

The analysis of the ANFIS model, carried out using the fuzzy system Takagi T., Sugeno M. An algorithm constructed from 7 fuzzy rules is considered. The article presents a technique for implementing a neuro-fuzzy control system for non-deterministic objects using Matlab.

Matlab made it possible to create a solution for an adaptive system of neuro-fuzzy inference, as well as carry out procedures for its training. The paper presents the results of the work of an adaptive neuro-fuzzy control system for non-deterministic objects.

## 6. References

- [1] P.C. Chang, C.Y. Fan, A Hybrid System Integrating a Wavelet and TSK Fuzzy Rules for Stock Price Forecasting, *IEEE Transactions on Systems, Man, and Cybernetics* 38(6) (2008) 802–815. doi:10.1109/TSMCC.2008.2001694.
- [2] M. Ivanov, N. Maksyshko, S. Ivanov, N. Terentieva, Intelligent Data Analysis in HR Process Management, in: *Proceedings of The Third International Workshop on Computer Modeling and Intelligent Systems (CMIS-2020)*, volume 2608 of *CEUR Workshop Proceedings CEUR Workshop Proceedings*, CEUR-WS, Zaporizhzhia Ukraine, 2020, pp. 754-768.
- [3] J. Casillas, O. Cordyn, M.J. del Jesus, F. Herrera, Genetic tuning of fuzzy rule deep structures preserving interpretability and its interaction with fuzzy rule set reduction, *IEEE Transactions on Fuzzy Systems* 13 (2005) 13–29. doi:10.1109/TFUZZ.2004.839670.
- [4] O. Cordyn, F. Herrera, P. Villar, Generating the knowledge base of a fuzzy rule-based system by the genetic learning of data base, *IEEE Transactions on Fuzzy Systems* 9 (2001) 667–674. doi:10.1109/91.940977.
- [5] J. Espinosa, J. Vandewalle, Constructing fuzzy models with linguistic integrity from numerical data-AFRELI algorithm, *IEEE Transactions on Fuzzy Systems* 8 (2000) 591–600. doi:10.1109/91.873582.
- [6] D. Wang, H. He, B. Zhao, D. Liu, Adaptive near-optimal controllers for nonlinear decentralised feedback stabilisation problems, *IET Control Theory Appl.* 11(6) (2017) 799–806. doi:10.1049/iet-cta.2016.1383.
- [7] S. Guillaume, Designing fuzzy inference systems from data: an interpretability-oriented review, *IEEE Transactions on Fuzzy Systems* 9 (2001) 426–443. doi:10.1109/91.928739.
- [8] S. Guillaume, B. Charnomordic, Generating an interpretable family of fuzzy partitions from data, *IEEE Transactions on Fuzzy Systems* 12 (2004) 324–335. doi:10.1109/TFUZZ.2004.825979.
- [9] F. Herrera, Genetic fuzzy systems: taxonomy current research trends and prospects, *Evolutionary Intelligence* 1 (2008) 27–46. doi:10.1007/s12065-007-0001-5.
- [10] P.J. Antsaklis, Neural Networks in Control Systems, Guest Editor's Introduction, *IEEE Control Systems Magazine* 10(3) (1990) 3-5. doi:10.1109/72.80237.
- [11] R. Cui, C. Yang, Y. Li, S. Sharma, Adaptive Neural Network Control of AUVs With Control Input Nonlinearities Using Reinforcement Learning, *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 47(6) (2017) 1019–1029. doi:10.1109/TSMC.2016.2645699
- [12] J.S.R. Jang, ANFIS: Adaptive-Network-Based Fuzzy Inference System, *IEEE Transactions on Systems, Man, and Cybernetics* 23(3) (1993) 665–685. doi:10.1109/21.256541.
- [13] M.A. Boyacioglu, D. Avci, An Adaptive Network-based Fuzzy Inference System (ANFIS) for the Prediction of Stock Market Return: The Case of the Istanbul Stock Exchange, *Expert Systems with Applications* 37(12) (2010) 7908–7912. doi:10.1016/j.eswa.2010.04.045.
- [14] T. Takagi, M. Sugeno, Fuzzy Identification of Systems and its Applications to Modeling and Control, *IEEE Transactions on Systems, Man, and Cybernetics* 1 (1985) 116–132. doi:10.1109/TSMC.1985.6313399
- [15] H. Zhang, D. Liu, *Fuzzy Modeling and Fuzzy Control*, Springer Science & Business Media, 2006. doi:10.1007/978-0-8176-4539-7.
- [16] W. Suparta, K. M. Alhasa., *Modeling of Tropospheric Delays Using ANFIS*, Springer International, 2016. doi:10.1007/978-3-319-28437-8
- [17] B. Saugat, B. Debabrota, K. Amit, D. N. Tibarewala, Interval type-2 fuzzy logic based multiclass ANFIS algorithm for realtime EEG based movement control of a robot arm, *Robotics and Autonomous Systems* 68 (2015) 104–115. doi:10.1016/j.robot.2015.01.007.
- [18] A. Kumar1, M. F. Qureshi, Power quality improvement in pv grid connected system by using active filter and ANFIS based MPPT, *International Journal of Recent Scientific Research* 9 (2018) 29074-29081. doi:10.24327/ijrsr.2018.0910.2788.
- [19] D. Karaboga, E. Kaya, Adaptive network based fuzzy inference system (ANFIS) training approaches: a comprehensive survey, *Artif Intell Rev* 52 (2019) 2263–2293. doi:10.1007/s10462-017-9610-2.

# Topology Properties of Hierarchical Honeycomb Meshes

Burhan Selcuk<sup>a</sup>, Ayse Nur Altintas Tankul<sup>a</sup> and Ali Karci<sup>b</sup>

<sup>a</sup> Karabuk University, Department of Computer Engineering, Karabuk, Turkey

<sup>b</sup> Inonu University, Department of Computer Engineering, Karabuk, Turkey

## Abstract

Honeycomb meshes can be seen widely in nature and are using in many different areas because of its properties. Using honeycomb meshes for constructing hierarchical structures has some advantages. In this study, hierarchical honeycomb meshes (HHM) are investigated. The construction of HHM is introduced with an example, topological properties of HHM explained in detail, a labeling algorithm in the process of the construction phase and also routing algorithms are given. This study shows a HHM(n) has a fractal structure and its graph is a Hamiltonian graph.

## Keywords 1

Hierarchical honeycomb meshes, Interconnection network, Hamilton graph, Gray code, Network topology, Routing

## 1. Introduction

A network is formed with nodes and links that connect them. A hierarchical network is a type of network consists clusters that are disjoint group of nodes. Every cluster has one or more hub that is connected to other hubs by backbone links and similarly inside of these clusters nodes are also connected by cluster links. Different type of topologies can be used for cluster and backbone structure for hierarchical networks such as cluster networks are sparse like a star or a tree while backbone networks are denser like mesh, ring or fully interconnected [1].

Lia et al. worked on diagnosability and connectivity of hierarchical star network for forbidden faulty sets [2]. Kwak et al. studied on torus ring on multiprocessor systems and increased performance of an interconnecting network by altering a hierarchical ring. Hierarchical spanning tree is used for network designing by Kim et al. with Nash genetic algorithm and this work is more about designing backbone topology structure to increase overall routing performance [3]. Hierarchical mesh tree is used as a protocol for multi hop data collection efficiently by Rabarijaona et al. and had great results at MAC layer via efficient routing for gathering multi-hop data in intelligent transportation system networks, wireless sensor networks, ad-hoc networks, etc. [4]. On the other hand, Dehghani and RahimiZadeh used mesh tree for designing hierarchical wireless network on chip for multicore systems and evaluated its performance that has a competitive potential [5]. Hierarchical cubic network is also used for designing network. Zhao et. al used hierarchical cubic network for an interconnection network design [6] while Yun and Park shows hierarchical hypercube network is more suitable for building massively parallel computers as an interconnection network than hierarchical cubic network [7]. Honeycomb structure is also used for hierarchical network studies [8][9][10][11].

Honeycomb structure is a nature inspired pattern that can be found in biological systems and organic materials. It plays an important role in the functionality and survival of the thing it contains. In engineering, it is also important for designing materials because of its structural properties such as rigid, strong, lightweight structure for materials and also topological properties for network and

connections. Engelmayer et al. used honeycomb for tissue engineering for cardiac anisotropy and get better results than previous studies [12]. It is also used in chemical engineering [13], satellite component research [14], computer graphics [15], station positioning for cellular phones [16], multiprocessor interconnection networks design [17][18] and in many different engineering area. Honeycomb mesh properties are also studied such as Hamilton properties [19], fractal properties [20][21], honeycomb networks in higher dimension [22][23], topological properties and communication algorithm of honeycomb networks [24].

In hierarchical honeycomb network (HHN) studies each of them have different way to form a hierarchical honeycomb network. Oftadeh et al. construct a HHN by changing all tree edge vertex of a main honeycomb network cell with a smaller one and repeating the procedure [10]. Ajdari et al. also replace every three edge vertex of a base hexagonal network with a smaller hexagon but the construction of hierarchical network results differently [9] can be seen in Figure 1. Fang et al. uses different strategy 5. In the study of Xu et al., a self-similar hierarchical honeycomb structure is created by iteratively inserting sub-hexagons at the corners of the base hexagon and with an internal rib, the vertex of two adjacent sub-hexagons are connecting [11]. The 1st and 2nd order hierarchical structures are created by repeating this process.

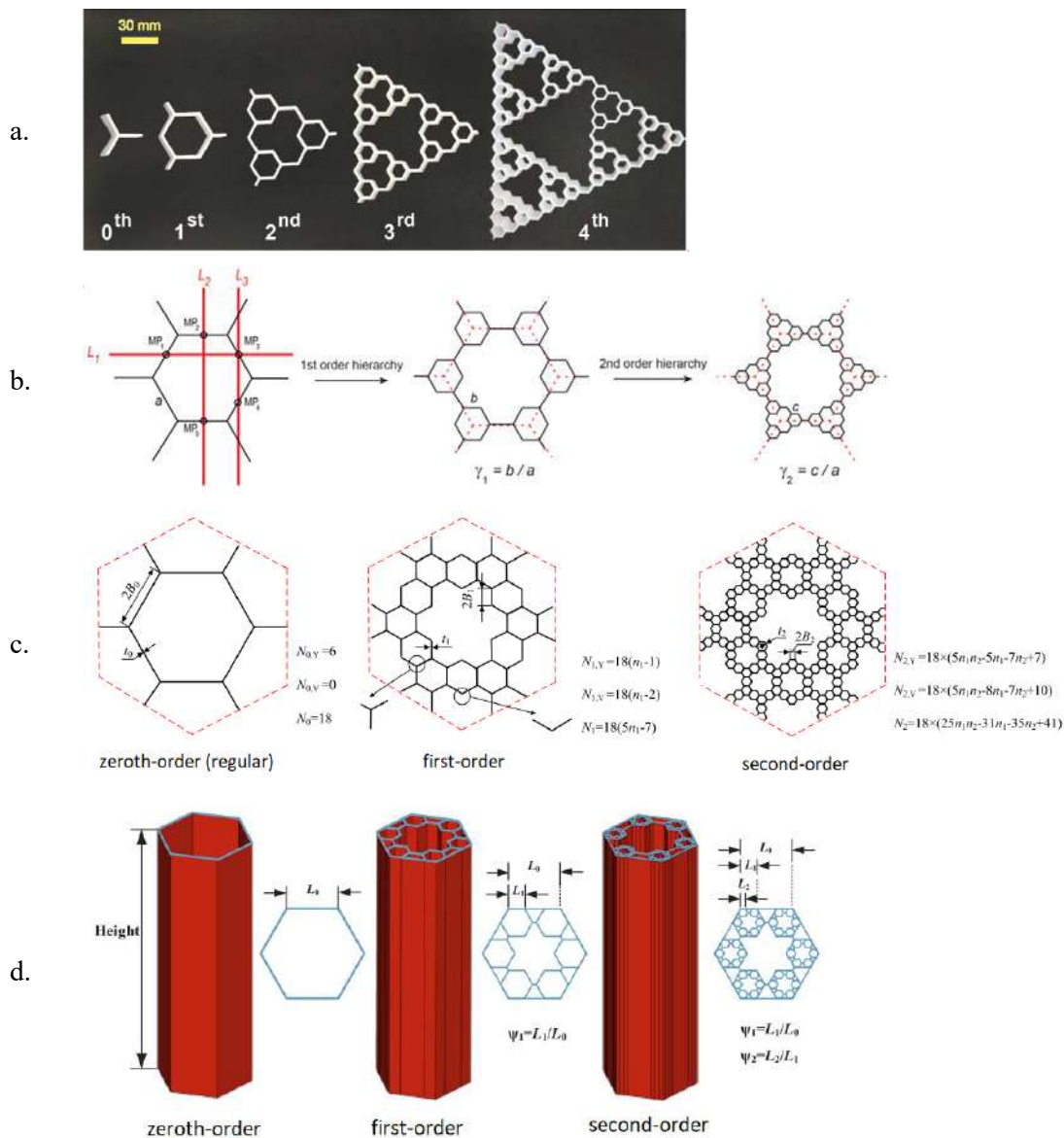


**Figure 1:** Different type of Hierarchical Honeycomb Network structure **a.** Oftadeh et al. [10], **b.** Ajdari et al. [9], **c.** Fang et al. [8] and **d.** Xu et al. [11].



## 2.2. Hierarchical Honeycomb Meshes (HHM)

Honeycomb Networks can be used for hierarchical fractal structure. In this section, we consider two variants of Hierarchical Honeycomb Meshes called HHM. HHM is suitable for creating this structure.

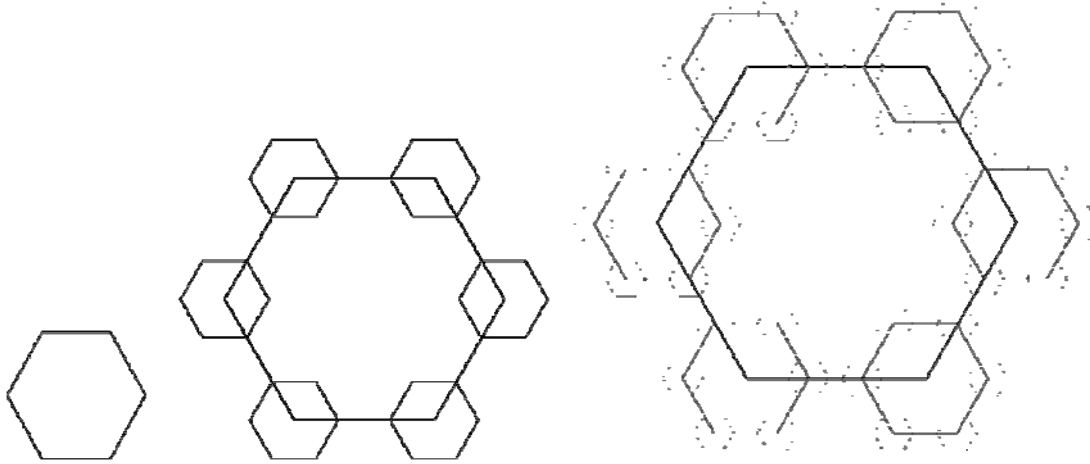


Figure 4: Hierarchical Honeycomb Meshes; zero order, 1<sup>st</sup> order, 2<sup>nd</sup> order.

## 2.3. Hierarchical Extended Fibonacci Cubes (HEFC)

**Definition 2.** (Extended Fibonacci Cubes [26]) Assume  $EFC(n) = (V(n), E(n))$ ,  $EFC(n-1) = (V(n-1), E(n-1))$ , and  $EFC(n-r) = (V(n-r), E(n-r))$ ,  $r < n$ . Then,  $V(n) = \{0 \parallel V(n-1) \cup 10 \parallel V(n-2)\}$ . In  $EFC(n)$ , two different nodes can be connected to each other with an edge  $E(n)$  only if the label of those two nodes are similar except for one bit. As a starting point of recursion,  $V(3) = \{0, 1\}$  and  $V(4) = \{00, 10, 11, 01\}$ .

**Definition 3.** (Hierarchical Extended Fibonacci Cubes) The resulting interconnection network was given the name Hierarchical Fibonacci Cube ( $HFC(n)$ ). The edges inside of a cluster named horizontal edges and the edges connect two different cluster named diagonal edges. Figure 5 depicts some  $HFC(n)$ s, with dashed lines reflecting diagonal connections. In terms of edges and nodes, an interconnection network can be described hierarchically if that is between hypercube and  $EFCK(n)$ .

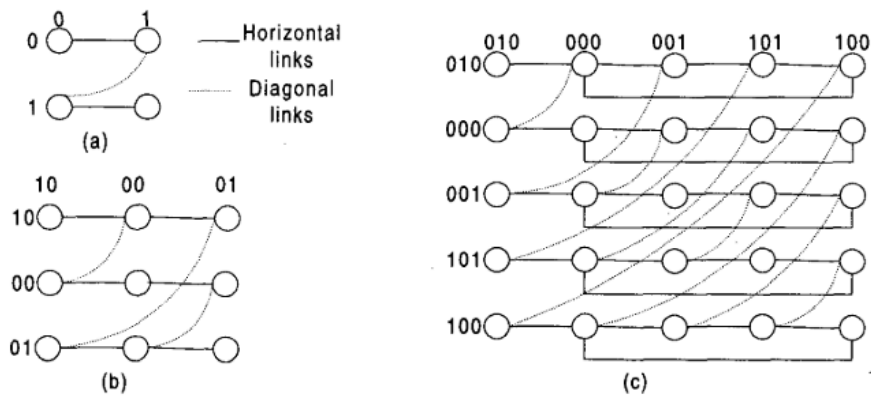


Figure 5: a.  $HFC(3)$ , b.  $HFC(4)$  and c.  $HFC(5)$  [26].

An EFC(n) is cluster that we select and let say  $|V(n)| = m$ , in any cluster any node can be represented with  $(i, j)$  that  $i$  is the cluster's node address and the node address inside of the cluster is  $j$  and the node  $(i, j)$  in the  $i^{\text{th}}$  cluster can be connected directly to the node  $(j, i)$  in the  $j^{\text{th}}$  cluster in that way  $m$  cluster can be connected. Every cluster in EFC(n) is addressed with using node addressing. In EFC(n) every edge is named horizontal edge and between clusters every edge is named diagonal edge. The interconnection network that is constructed with using  $m^2$  nodes is called Hierarchical Extended Fibonnaci Cube (HEFC(n)). Figure 6 shows some of the HEFC(n).

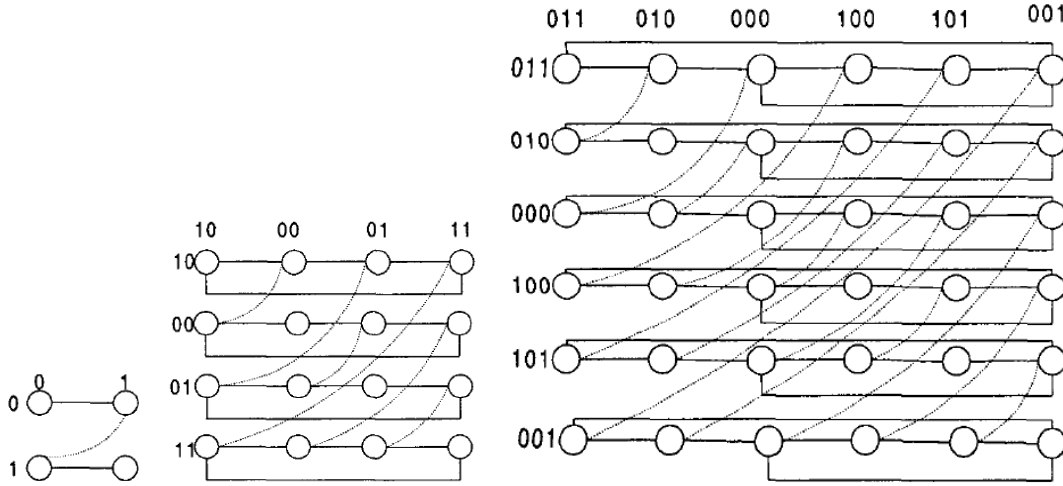


**Figure 6.** Hierarchical Extended Fibonnaci Cubes (HEFC(n)) [26].

### 3. Construction of HHM

$HHM(n) = (V(n), E(n))$  can be constructed hierarchical using six  $HHM(n-1)$  graphs  $n=1,2,3, \dots$

The symbol "||" is used for the concatenation of two strings in this paper. In the Hamming distance

$\sum_{i=0}^{n-1} (a_i \oplus b_i)$  definition, summation is  $a_i \oplus b_i$  (bitwise-XOR operation).

**Definition 4 (HHM(n)).** Hierarchical Honeycomb Meshes have the following description. Then, for  $n > 0$ ,  $HHM(n) = (V(n), E(n))$  can be designed in the following way for  $n > 0$

$$HHM(n) = HHM(n-1) || HHM(0)$$

where  $V(n) = V(n-1) || V(0)$ ,  $V(0) = \{000, 010, 011, 001, 101, 100\}$  and

$$E(n) = \begin{cases} (000 || E(n-1) \cup 010 || E(n-1) \cup 011 || E(n-1) \cup \\ 001 || E(n-1) \cup 101 || E(n-1) \cup 100 || E(n-1) \cup E' \end{cases}$$

where  $E' = \{(e_i, e_j) | e_i = abc \text{ and } e_j = dfg \text{ and } (abc) \oplus (dfg) = 1\}$ , the labels of  $e_i$  and  $e_j$  are same except  $abc$  of  $e_i$  and  $dfg$  of  $e_j$  where  $m=0,3,6, \dots$ . Assume that  $v_1$  and  $v_2$  are two nodes whose labels are  $a_{(3k+2)}a_{(3k+1)}a_{3k} \dots a_1a_0$  and  $b_{(3k+2)}b_{(3k+1)}b_{3k} \dots b_1b_0$ , respectively. By using one of the following conditions, we can get the  $i^{\text{th}}$  order external edges:

$$(a_{(3k+2)} \oplus b_{(3k+2)}) \oplus (a_{(3k+1)} \oplus b_{(3k+1)}) \oplus (a_{3k} \oplus b_{3k}) = 1, 1 \leq i \leq k.$$

**Example 1.** HHM(1) can be constructed with integration of six HHM(0) graphs. There will be 36 nodes by concatenating 000,010,011,001,101,100 partial binary labels to all labels of nodes of HHM(0) where  $V(0) = \{000,010,011,001,101,100\}$ . Then there will be 6 edges. The labels  $\overbrace{a_5 a_4 a_3 a_2 a_1 a_0}^{c_1 c_0}$  and  $\overbrace{b_5 b_4 b_3 b_2 b_1 b_0}^{d_1 d_0}$  can be illustrated by  $c_1 c_0$  and  $d_1 d_0$ , respectively. The edges  $c_i \oplus d_i = 1$  for  $1 \leq i \leq k$  are added to  $E(1)$ . Figure 4 shows an example of HHM(1).

We get labelling of nodes and edges of HHM(1) as the follow:

$$\begin{aligned} V(1) &= V(0) \parallel V(0) = \{000,010,011,001,101,100\} \parallel \{000,010,011,001,101,100\} \\ &= 000 \parallel \{000,010,011,001,101,100\} \cup 010 \parallel \{000,010,011,001,101,100\} \cup \\ &\quad 011 \parallel \{000,010,011,001,101,100\} \cup 001 \parallel \{000,010,011,001,101,100\} \cup \\ &\quad 101 \parallel \{000,010,011,001,101,100\} \cup 100 \parallel \{000,010,011,001,101,100\}, \end{aligned}$$

and

$$E(1) = \begin{cases} (000 \parallel E(0) \cup 010 \parallel E(0) \cup 011 \parallel E(0) \cup \\ 001 \parallel E(0) \cup 101 \parallel E(0) \cup 100 \parallel E(0) \cup E') \end{cases}$$

$$\begin{aligned} &000 \parallel \{(000,010), (010,011), (011,001), (001,101), (101,100), (100,000)\} \cup \\ &010 \parallel \{(000,010), (010,011), (011,001), (001,101), (101,100), (100,000)\} \cup \\ &011 \parallel \{(000,010), (010,011), (011,001), (001,101), (101,100), (100,000)\} \cup \\ &001 \parallel \{(000,010), (010,011), (011,001), (001,101), (101,100), (100,000)\} \cup \\ &101 \parallel \{(000,010), (010,011), (011,001), (001,101), (101,100), (100,000)\} \cup \\ &100 \parallel \{(000,010), (010,011), (011,001), (001,101), (101,100), (100,000)\} \cup E' \end{aligned}$$

$$\begin{aligned} &(000000,000010), (000010,000011), (000011,000001), (000001,000101), (000101,000100), \\ &(000100,000000), (010000,010010), (010010,010011), (010011,010001), (010001,010101), \\ &(010101,010100), (010100,010000), (011000,011010), (011010,011011), (011011,011001), \\ &(011001,011101), (011101,011100), (011100,011000), (001000,001010), (001010,001011), \\ &(001011,001001), (001001,001101), (001101,001100), (001100,001000), (101000,101010), \\ &(101010,101011), (101011,101001), (101001,101101), (101101,101100), (101100,101000), \\ &(100000,100010), (100010,100011), (100011,100001), (100001,100101), (100101,100100), \\ &(100100,100000) \cup E' \end{aligned}$$

where the first order external edges are  $E' = \{(000000,010000), (010000,011000), (011000,001000), (001000,101000), (101000,100000), (100000,000000)\}$ .



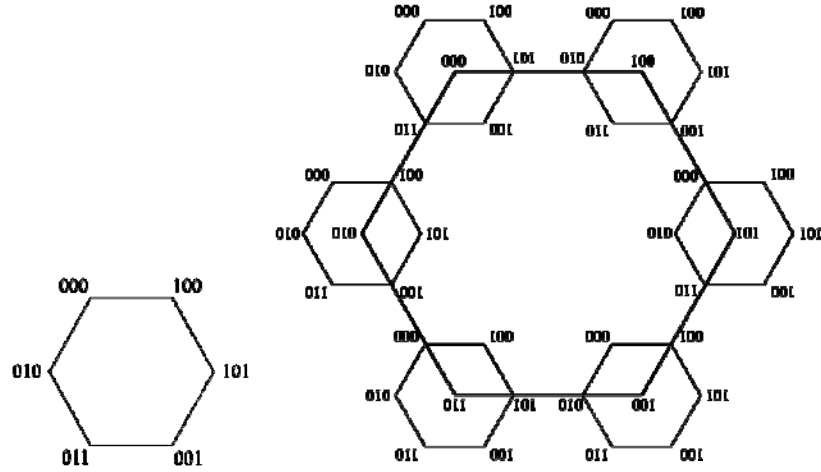


**Figure 7:** Labeling of HHM(0) and HHM(1).

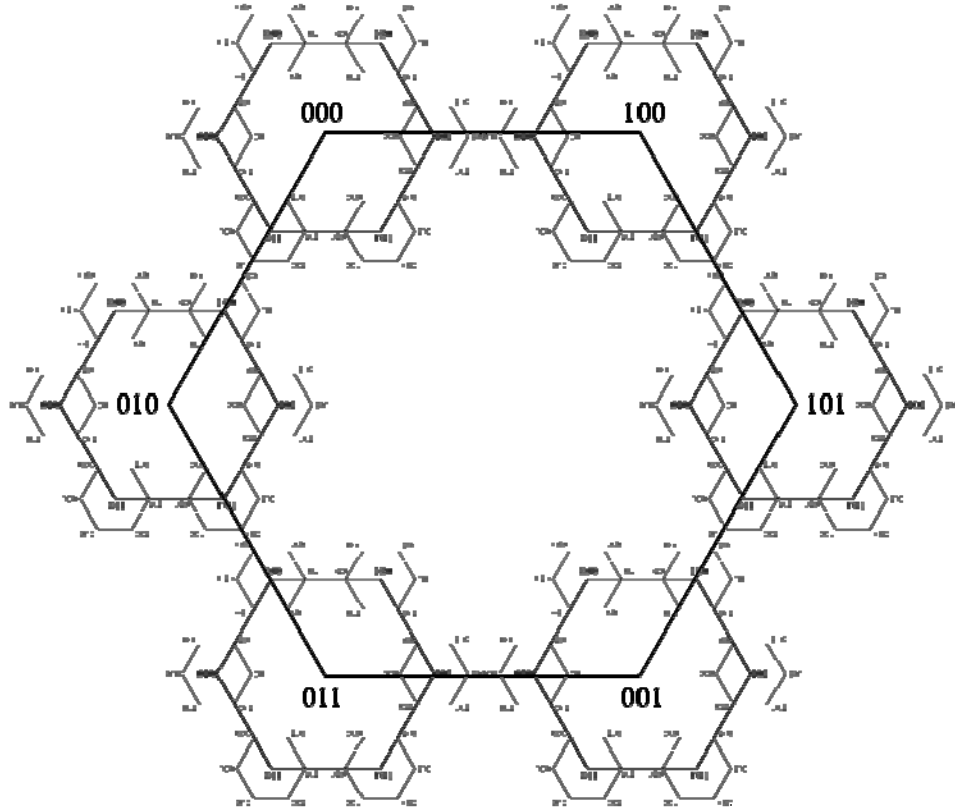


**Figure 8:** Labeling of HHM(2).

**Algorithm 1.** The recurrence relation of the following labelling algorithm on HHM( $n$ ) can be written as  $T(n)=T(n-1)+\theta(1)$ . Solving this recurrence relation, obtained running time of is  $T(n)=\theta(n)$ .

```

function V(n)
  if n = 0
    V(n) ← {000,010,011,001,101,100}    // initial value
    return V(n);
  else
    V(n) ← V(n-1)||V(0);
    return V(n);
  end
end

```

#### 4. Topology properties of HHM

The Hamilton graph is the graph that contains Hamilton cycle which has a path visits each vertex only once and returned to the starting point. In this section of the study, The Hamiltonian properties of HHM(n) is analyzed.

**Theorem 1.** HHM(n) graphs are one type of a Hamiltonian graphs.

**Proof.** By using mathematical induction, the proof for this theorem can be done.

*Base Step:*  $k = 0, k = 1$  or  $k = 2$

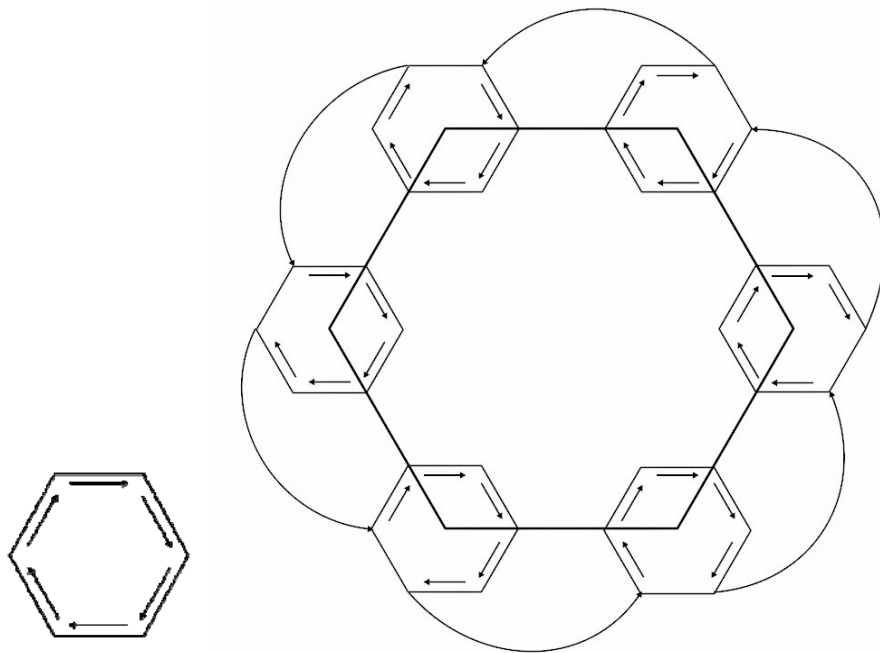


**Figure 9:** Base step for the mathematical induction.

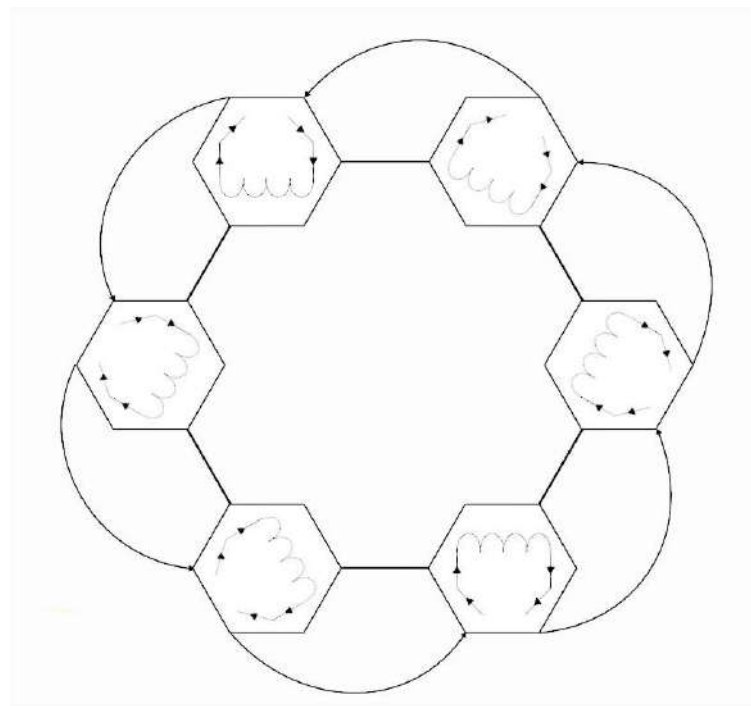


**Figure 10:** Hamiltonian property proof for HHM(k).

*Hypothesis Step:* We are assuming that HHM(k-1) is a Hamiltonian graph.

*Final Step:* In Figure 10, HHM(k) is being a Hamiltonian graph can be seen. The HHM(k) includes six HHM(k-1) and there is a Hamiltonian path in each of these HHM(k-1) in Figure 10. HHM(k) that includes six Hamiltonian HHM(k-1) graphs and in HHM(k) there is Hamiltonian path showed in Figure 10.

**Theorem 2. (i)** Number of nodes in HHM(n) is  $6^n$ .

**(ii)** The recurrence relation  $E(n) = 6E(n-1) + 6$  to calculate number of edges of HHM(n) with  $E(0)=0$  initial value.

**Proof.** It can be easily proved that the theorem using mathematical induction. (see [27])

**Corollary 1.** Using gray code in the labeling of HHM(n) requires us to use excess bits, although it makes it easier to label. We need to use 3-bit Gray code for 6 nodes when  $n=1$ , 6-bit Gray code for 36 nodes when  $n=2$ , 9-bit Gray code for 216 nodes when  $n=3$ . Normally, it is enough to use 8 bits for labeling 216 nodes.

## 5. Routing of HHM

There are several routing techniques in honeycomb structure. Rais et al. assume each cell is a node in the network and developed its routing strategy accordingly [28]. Zhang et al. gives 3 different algorithms on honeycomb mesh routing, a unicast routing algorithm XYZ-Route nodes in the network send messages to neighbor of its neighbor on anticlockwise direction with shortest path, a fault-tolerant routing algorithm HFT-Rout provides communication between nodes including the fact that there are convex faults, and a multicast algorithm, named HMTREE-Route is used for routing a message in the presence of more than one destination along the same path as possible, also several copies can be distributed to different paths at the tree's crossroads during the routing process [29]. Also in the study of Stojmenovic XYZ-Route algorithm used [24].

In this section, the routing algorithms in HHM(n) will be investigated using the routing algorithms of hypercube [7]. Firstly, two examples are given. It is seen that a path cannot always be found as in hypercube in example 3.

**Example 2.** Let source node  $s = 010101$  and destination node  $d = 101100$ .

$$r = s \oplus d = 010101 \oplus 101100 = 111001$$

$$s_k = s \oplus 2^i, k = 1, 2, \dots, i = 0, 1, 2, \dots$$

$$s_1 = s \oplus 2^0 = 010101 \oplus 000001 = 010100$$

$$s_2 = s_1 \oplus 2^3 = 010100 \oplus 001000 = 011100$$

$$s_3 = s_2 \oplus 2^4 = 011100 \oplus 010000 = 001100$$

$$s_4 = s_3 \oplus 2^5 = 001100 \oplus 100000 = 101100 = d$$

The route is  $s_1 \rightarrow s_2 \rightarrow s_3 \rightarrow s_4$ .

**Example 3.** Let source node  $s = 010101$  and destination node  $d = 101010$ .

$$r = s \oplus d = 010101 \oplus 101010 = 111111$$

$$s_k = s \oplus 2^i, k = 1, 2, \dots, i = 0, 1, 2, \dots$$

$$s_1 = s \oplus 2^0 = 010101 \oplus 000001 = 010100$$

$$s_2 = s_1 \oplus 2^1 = 010100 \oplus 000010 = 010110 \text{ (this node is not defined)}$$

**Algorithm 2.** The following algorithm is calculated unicast routing in HHM(n) as example 2 and 3. The running time of the Algorithm2 is obtained as  $T(n)=\theta(3n)$ .

```

Unicast_Routing(G,Source,Destination)

r=Source⊕Destination;
L=length_r;
P=[];           % define empty array
P[0]=Source;
k=1;
for i=0 to L-1
    if r[i]!=0 then
        if P[k]=P[k-1]⊕2i ∈ V then
            k=k+1;
        else
            print('deadlock: this node isn't defined');
            break;
        end
    end
end
return P;
end

```

## 6. Conclusions

In this paper, we firstly define HHM(n) and investigate topology properties of HHM(n). The results obtained are given below:

1. HHM(n) has a fractal structure, that is self-repeating shapes that grow or shrink forever, forming the body in a similar object,
2. HHM(n) graph is a Hamiltonian graph, starting at any node of a graph, it crosses all nodes only once and can come to the starting node,
3. HHM(n) is scalable,
4. Number of nodes of HHM(n) is  $6^n$ ,
5. The recurrence relation  $E(n) = 6E(n-1) + 6$  to calculate number of edges of HHM(n) with  $E(0) = 0$  initial value,
6. Deadlocks occur when the unicast routing algorithm on the HHM is similar to the unicast routing algorithm on the hypercubes. In future studies, this problem will be tried to be eliminated,
7. HHM is basically defined in a similar structure to HEFN. Both are built over the hypercube. Also, HHM and HEFN are isomorphic graphs because they have same adjacency spectrum. These similarities can be shown in detail in future studies.

## 7. References

- [1] T. Thomadsen, Hierarchical Network Design, PHD tesis Informatics Math. Model. - Tech. Univ. Denmark, pp. 1–34, 2014.
- [2] A. Liu, J. Yuan, S. Wang, and J. Li, On g-good-neighbor conditional connectivity and diagnosability of hierarchical star networks, Discret. Appl. Math., vol. 293, pp. 95–113, 2021.
- [3] J. R. Kim, J. U. Lee, and J. B. Jo, Hierarchical spanning tree network design with Nash genetic algorithm, Comput. Ind. Eng., vol. 56, no. 3, pp. 1040–1052, 2009.
- [4] V. H. Rabarijaona, F. Kojima, and H. Harada, Hierarchical mesh tree protocol for efficient multi-hop data collection, IEEE Wirel. Commun. Netw. Conf. WCNC, vol. 2, pp. 2008–2013, 2014.
- [5] A. Dehghani and K. RahimiZadeh, Design and performance evaluation of Mesh-of-Tree-based

- hierarchical wireless network-on-chip for multicore systems, *J. Parallel Distrib. Comput.*, vol. 123, no. July, pp. 100–117, 2019.
- [6] S. L. Zhao, R. X. Hao, and J. Wu, The generalized 4-connectivity of hierarchical cubic networks, *Discret. Appl. Math.*, vol. 289, pp. 194–206, 2021.
  - [7] S. K. Yun and K. H. Park, Hierarchical Hypercube Networks (HHN) for massively parallel computers, *J. Parallel Distrib. Comput.*, vol. 37, no. 2, pp. 194–199, 1996.
  - [8] J. Fang, G. Sun, N. Qiu, T. Pang, S. Li, and Q. Li, On hierarchical honeycombs under out-of-plane crushing, *Int. J. Solids Struct.*, vol. 135, no. March, pp. 1–13, 2018.
  - [9] A. Ajdari, B. H. Jahromi, J. Papadopoulos, H. Nayeb-Hashemi, and A. Vaziri, Hierarchical honeycombs with tailorable properties, *Int. J. Solids Struct.*, vol. 49, no. 11–12, pp. 1413–1419, 2012.
  - [10] R. Oftadeh, B. Haghpanah, D. Vella, A. Boudaoud, and A. Vaziri, Optimal Fractal-Like Hierarchical Honeycombs, *Phys. Rev. Lett.*, vol. 104301, no. September, pp. 1–5, 2014.
  - [11] X. Xu, Y. Zhang, J. Wang, F. Jiang, and C. H. Wang, Crashworthiness design of novel hierarchical hexagonal columns, *Compos. Struct.*, vol. 194, no. March, pp. 36–48, 2018.
  - [12] E. George C, M. Cheng, C. J. Bettinger, J. T. Borenstein, R. Langer, and L. E. Freed, Accordion-like Honeycombs for Tissue Engineering of Cardiac Anisotropy, *Nat. Mater.*, vol. 7, no. May 2014, 2008.
  - [13] B. Rajan, A. William, C. Grigorious, and S. Stephen, On Certain Topological Indices of Silicate , Honeycomb and Hexagonal Networks, vol. 3, no. 5, pp. 530–535, 2012.
  - [14] A. Boudjemai, R. Amri, A. Mankour, H. Salem, M. H. Bouanane, and D. Boutchicha, Modal Analysis and Testing of Hexagonal Honeycomb Plates Used for Satellite Structural Design, *Mater. Des.*, vol. 35, pp. 266–275, 2012.
  - [15] L. N. Lester and J. Sandor, Computer Graphics on a Hexagonal Grid, *Comput. Graph.*, vol. 8, no. 4, 1985.
  - [16] F. G. Nocetti, I. Stojmenovic, and J. Zhang, Addressing and Routing in Hexagonal Networks with Applications for Tracking Mobile Users and Connection Rerouting in Cellular Networks, *IEEE Trans. PARALLEL Distrib. Syst.*, vol. 13, no. 9, pp. 963–971, 2002.
  - [17] J. Carle, J. F. Myoupo, and D. Seme, All-to-all Broadcasting Algorithms on Honeycomb Networks and Applications, *Parallel Process. Lett.*, vol. 9, no. 4, pp. 539–550, 1999.
  - [18] P. Manuel, B. Rajan, I. Rajasingh, and C. M. M, On minimum metric dimension of honeycomb networks, vol. 6, pp. 20–27, 2008.
  - [19] A. N. Altıntaş Tankül and B. Selçuk, On Hamiltonian Properties of Honeycomb Meshes, *Anatol. Sci.*, vol. 44, no. 1, pp. 184–190, 2019.
  - [20] B. Selçuk and A. N. Altıntaş Tankül, Perfect Matching of Fractal Honeycomb Meshes, pp. 38–46, 2019.
  - [21] T. Zhang and K. Ding, A new Proof of Honeycomb Conjecture by Fractal Geometry Methods, *Front. Mech. Eng. China*, no. April, 2013.
  - [22] C. Science, Higher dimensional honeycomb networks, vol. 2, no. 4, pp. 391–420, 2001.
  - [23] J. Carle, J. F. Myoupo, and I. Stojmenovic, Higher Dimensional Honeycomb Networks, *J. Interconnect. Networks*, vol. 2, no. 4, pp. 391–420, 2001.
  - [24] I. Stojmenovic, Honeycomb Networks : Topological Properties and Communication Algorithms, vol. 8, no. 10, pp. 1036–1042, 1997.
  - [25] D. Xu, J. Fan, X. Jia, S. Zhang, and X. Wang, Hamiltonian properties of honeycomb meshes, *Inf. Sci. (Ny)*, vol. 240, pp. 184–190, 2013.
  - [26] A. Karci, New Interconnection Networks: Fibonacci Cube and Extended Fibonacci Cubes Based Hierarchic Networks, pp. 869–874, 2001.
  - [27] A. Karci and B. Selçuk, A new hypercube variant : Fractal Cubic Network Graph Engineering Science and Technology , an International Journal A new hypercube variant : Fractal Cubic Network Graph, *Eng. Sci. Technol. an Int. J.*, no. February 2015, 2014.
  - [28] A. Rais, K. Bouragba, and M. Ouzzif, Routing and clustering of sensor nodes in the honeycomb architecture, *J. Comput. Networks Commun.*, vol. 2019, 2019.
  - [29] W. Zhang, X. Yang, Y. Liang, and Z. Huang, Routing algorithms in honeycomb meshes, *Int. J. Parallel, Emergent Distrib. Syst.*, vol. 24, no. 5, pp. 367–382, 2009.

# Handwritten Digits Recognition Using SVM, KNN, RF and Deep Learning Neural Networks

Yevhen Chychkarov<sup>a</sup>, Anastasiia Serhienko<sup>b</sup>, Iryna Syrmamiikh<sup>a</sup>, Anatolii Kargin<sup>c</sup>

<sup>a</sup> Donetsk State University of Management, Mariupol, Ukraine

<sup>b</sup> Pryazovskyi State Technical University, Mariupol, Ukraine

<sup>c</sup> Ukrainian state university of railway transport, Kharkiv, Ukraine

## Abstract

This article discusses several classification algorithms of recognizing numbers from photographic images or with manual input, namely: support vector machine (SVM), K-nearest neighbors (KNN), random forest (RF) and several variants of neural networks. The success rates of the algorithms in the field of handwriting recognition were compared.

Six variants of recognition technology were analyzed and tested: using classifier from Scikit-learn package and using deep learning neural networks. To construct and train neural networks or train classifiers, 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 one-dimensional representation). Preliminary processing of images (filtering, scaling, etc.) was carried out using the OpenCV library. For recognition, each image of a 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 devised. For handwritten digit recognition, the best recognition accuracy is provided by a convolutional neural network, as 97.6% of car ladle digits were recognized correctly with it.

To improve the recognition accuracy for handwritten digits, it is necessary to perform two additional stages of image preprocessing and dataset transformation.

After building recognition models using all the algorithms mentioned above, the recognition accuracy of all handwritten digits on the test program turned out to be within 98-100%. For industrial images regardless of the used neural network version, the recognition accuracy was 96-98%.

## Keywords

Scikit-learn; Classifier; Keras; TensorFlow; MNIST; Python; Deep learning; Neural networks; Digit recognition.

## 1. Introduction

Handwriting recognition is the ability of a computer or device to accept handwritten text as input from sources such as printed documents, photographs, images, or input stream from other devices. Direct input to the touch screen (this is also a handwriting input), which is interpreted as text, is widely used. An example is smartphones or tablets with a touchscreen that can use handwriting (by finger or stylus) input as text or number input.

There are many applications and technologies currently available for handwriting recognition and optical character recognition (OCR).

---

CMIS-2021: The Fourth International Workshop on Computer Modeling and Intelligent Systems, April 27, 2021, Zaporizhzhia, Ukraine  
EMAIL: [influence@meta.ua](mailto:influence@meta.ua) (Ye. Chychkarov); [invitory@i.ua](mailto:invitory@i.ua) (A. Serhienko); [siv.2017@ukr.net](mailto:siv.2017@ukr.net) (I. Syrmamiikh); [kargin@kart.edu.ua](mailto:kargin@kart.edu.ua) (A. Kargin)

ORCID: 0000-0002-4362-5129 (Ye. Chychkarov); 0000-0003-1328-2572 (A. Serhienko); 0000-0003-0524-6731 (I. Syrmamiikh); 0000-0003-2885-9071 (A. Kargin)



© 2021 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

There are a few examples of technical application of optical number or digit recognition from photographs, i.e. recognition of railway carriage numbers, car license plates, product marking, readings of recording devices, etc.

This article discusses several classification algorithms for recognizing numbers in photographic images or with manual input, namely: support vector machine (SVM), K-nearest neighbors (KNN), random forest (RF) and several variants of neural networks. The algorithms success rates in the field of handwriting recognition were compared. In the methods section of this article, brief information is given about handwriting recognition and compared machine learning methods. In the experimental section, the values obtained as a result of the study were compared. Evaluations were made on the compared machine learning algorithms.

## **2. Analysis of literature data and formulation of the problem**

Optical Character Recognition (OCR) is an important area of research in artificial intelligence and character recognition [1]. For optical character recognition, many applications have been developed that solve the problems of text information extraction, automatic recognition of car license plates and railway carriage numbers [2].

By now a wide range of studies have been carried out, including a comprehensive study and implementation of various popular algorithms for recognizing handwritten text, including handwritten numbers. The task of segmentation and recognition of image areas containing handwritten or printed characters is relevant due to the presence of numerous technical applications.

For example, license plate recognition in work [3] or registration and recognition of wagons and tank numbers in work [4-5] are performed using neural network technologies. The functionality of all these systems is approximately the same – they automate the process of reading numbers and store the received information.

Character recognition is a classification task that includes recognizing a set of characters in an image, dividing them into 10 classes in case of numbers or 26 classes in case of the Latin alphabet letters.

Many systems are currently available for identifying printed text. However, according to [2], the identification of handwritten characters is still a problem in the field of pattern recognition. A large number of research and development dedicated to the OCR system, consider extensive handwritten digits recognition capabilities.

There are some problems that hinder the implementation of character recognition, namely:

- low quality of photo or scanned copy, but this problem is partially solved by image preliminary processing.
- presence of distorted characters, especially when working with handwritten documents or numbers, due to the peculiarities of character style.
- similarity between outlines of some characters.

For example, in work [6], the effect of preliminary processing and segmentation of license plate number images on the results of their recognition was noted. For any recognition methods, incorrect character segmentation does not allow to achieve accurate results.

In works [7-9], [12] the study was focused on the comparison of CNN different models with the fundamental algorithms of machine learning to different attributes, such as performance, runtime, complexity, etc., for their explicit evaluation.

In work [7], the author concluded that the Multilayer Perceptron classifier gave the most accurate results with minimum error rate followed by Support Vector Machine, Random Forest Algorithm, Bayes Net, Naive Bayes, j48, and Random Tree respectively. Authors [8] reported a comparison between SVM, CNN, KNN, RFC and were able to achieve the highest accuracy of 98.72% using CNN (which took maximum execution time) and lowest accuracy using RFC. In work [9] the authors made a detailed comparison of SVM, KNN & MLP models to classify the handwritten text and concluded that KNN and SVM predict all the classes of dataset correctly with 99.26% accuracy, but the process was a little complicated with MLP when it failed classifying number 9, for which the authors suggested to use CNN with Keras to improve the classification.

Improving the accuracy of handwritten digit recognition is achieved by increasing the complexity of the used deep learning neural networks. For example, in [6-7] a convolution neural network for handwritten digit recognition using MNIST datasets was used. The authors of work [10] used a convolutional neural network with 7 layers including 5 hidden layers along with gradient descent and back propagation model to find and compare the accuracy on different epochs, thereby getting maximum accuracy of 99.2%. In work [11] the same authors briefly discussed different components of CNN, its advancement from LeNet-5 to SENet and comparisons between different model like AlexNet, DenseNet and ResNet. The research outputs: the LeNet-5 and LeNet-5 (with distortion) achieved test error rate of 0.95% and 0.8% respectively on MNIST data set, the architecture and accuracy rate of AlexNet is the same as LeNet-5, but much bigger with around 4096000 parameters and “Squeeze-and-Excitation network” (SENet) became a winner of ILSVRC-2017 since they had reduced the top-5 error rate to 2.25% and by far the most sophisticated model of CNN that exists.

### **3. Goal and objectives of the research**

This paper provides a reasonable understanding of machine learning and deep learning algorithms like SVM, KNN, RF, CNN, and MLP for handwritten digit recognition. Furthermore, it provides information about the algorithm which is efficient in performing the task of digit recognition. The related work that has been done in this field followed by the methodology and implementation of all the algorithms for their better understanding will be discussed in the following sections of this paper. Next, it presents the conclusion and result. The last section of this paper contains used citations and references.

Goals of this research:

1. Evaluation of recognition accuracy for machine learning and deep learning algorithms such as SVM, KNN, RF, CNN and MLP in relation to real sets of handwritten numbers.
2. Analysis of algorithms influence of an image preliminary processing on recognition accuracy.
3. Evaluation of the possibilities to use the considered algorithms for solving technical problems associated with the processing of handwritten digits noisy images (on the example of recognizing cast iron ladle numbers).

## **4. Methodology**

### **4.1. Equipment and dataset**

The comparison of the algorithms (Support vector machines, KNN, Random Forest, Multi-layered perceptron & Convolutional neural network) is based on the characteristic chart of each algorithm on common grounds like dataset, the number of epochs, complexity of the algorithm, accuracy of each algorithm, specification of the device (Ubuntu 20.10, i5 9th gen processor, 8GB memory) used to execute the program and runtime of the algorithm under ideal condition.

To build and train the model, a well-known and fairly complete base of handwritten digits MNIST was selected [13-15]. This database (MNIST - Modified NIST) contains a total of 70,000 handwritten images of numbers and is part of the larger NIST database [16], which contains handwritten images segmented with images of specially prepared templates populated by respondents from the Census Bureau and students of the US educational institutions. In the MNIST database, images from different authors have been placed in different parts to enhance uniqueness.

### **4.2. Algorithms and methods used**

The Support Vector Machine (SVM) was first proposed by Vapnik and since then has attracted a high degree of interest in the machine learning research community [17]. SVM is a supervised machine learning algorithm. During the training, SVM learns the relationship of each data and tag in the existing training set. Typically, data items are placed in n-dimensional space, where n is the number of features. A particular coordinate represents the value of the feature. Points are classified by finding a hyperplane that distinguishes between two classes. The algorithm chooses a hyperplane that



separates the classes correctly. SVM chooses extreme vectors that help in creating the hyperplane. These extreme cases are called support vectors, and hence the algorithm is termed as Support Vector Machine. There are mainly two types of SVMs, linear and non-linear SVM [18]. Kernel function selection is an important step in the process of SVM to solve a problem [19]. SVM is successful in solving classification problems compared to many other techniques.

Decision trees form a classification model as a tree structure for the solution of a problem. The tree structure and rules are easy to understand. This simplifies the implementation of the algorithm. Decision trees method consists of simple sequential decision making operations. One of the most important steps in creating the tree structure is choosing the attribute value for which the branching in the tree will be determined [20,21].

A multilayer perceptron is a class of feedforward artificial neural networks that consists of at least three layers: input, hidden, and output. Except for the input neurons, all neurons use a non-linear activation function.

The number of hidden layers can be increased to any number according to the problem, without limitation on the number of nodes. The specific number of hidden layers or the number of nodes in a hidden layer is difficult to determine due to the unstable nature of the model and is therefore chosen experimentally. Each hidden layer in the model can have different activation functions for processing. For teaching purposes, it uses a supervised learning method called backpropagation. In MLP, a node connection consists of a weight that is tuned to synchronize with each connection as the model is trained. [22]

KNN is one of the classification methods. Nearest neighbor searching is the following problem: We are given a set  $S$  of  $n$  data points in a metric space,  $X$ , and the task is to preprocess these points so that, given any query point  $q \in X$ , the data point nearest to  $q$  can be reported quickly. This is also called the closest-point problem and the post office problem. Nearest neighbor searching is an important problem in a variety of applications [23]. The KNN algorithm is used to assign a new observation to the one of the classes that is most common among the  $k$  neighbors of a given element, the classes of which are already known. Classification is made according to the threshold value determined with the average of the  $k$  data that appears to be the closest. The performance of the method is influenced by the closest neighbor number, threshold value, similarity measurement and sufficient number of normal behaviors in the learning cluster [24-25].

Random forest algorithm can be used in both classification and regression problems like decision trees. The logic of work is to create more than one decision tree and produce average results with the help of these trees. The reason why this algorithm is called random is that it offers extra randomness during the creation of the tree structure. When splitting a node, instead of looking for the best attribute directly, it looks for the best attribute in a subset of random attributes. This situation creates more diverse trees [21, 26-27].

Deep convolutional neural networks (CNNs) are a specialized kind of ANNs that use convolution in place of general matrix multiplication in at least one of their layers [28]. CNN is a deep learning algorithm that is widely used for image recognition and classification. Unlike neural networks with a simpler architecture, which have one or more hidden layers, CNNs are composed of many layers. Such a feature allows them to compactly represent highly nonlinear and varying functions [29]. CNNs involve many connections, and the architecture is typically comprised of different types of layers, including convolution, pooling and fully connected layers, and realize form of regularization [30]. In order to learn complicated features and functions that can represent high-level abstractions (e.g., in vision, language, and other AI-level tasks), CNNs would need deep architectures. Deep architectures, and CNNs, consist of a large number of neurons and multiple levels of latent calculations of non-linearity. According to [31], each level of CNN architecture represents features at a different level of abstraction defined as a composition of lower-level features.

CNN uses a filter (kernel) which is an array of weights to extract features from the input image. CNN employs different activation functions at each layer to add some non-linearity [32].

Many works have noted the importance of image preprocessing for recognizing handwritten numbers or letters.

In particular, [33] applied various preprocessing methods to improve the performance of CNN models. Translations, rotations and elastic deformations have been studied for a variety of in the area of machine learning goals [34-36].

According to the authors [33], preprocessing with the combination of elastic and rotation improves the accuracy of the three analyzed networks up to 0.71%.

According to [37], many of today's OCR systems are built following traditional approaches to image processing and work great with printed text but if use them for handwritten text recognition in images it can get unexpected results with poor recognition quality.

The quality of a learned system is primarily dependent of the size and quality of the training set. In work [36] was proposed a simple technique for vastly expanding the training set on base of elastic distortions. In [38] the MNIST dataset size was increased by four times by using affine transformations and the prior knowledge of transform invariant properties. In the raised method, the elastic distortions are applied to each sample of the training set to extend nine new samples.

These distortions improve the results on MNIST substantially.

## 5. Implementation and Computer experiment results

To compare the algorithms based on working accuracy, execution time, complexity, and the number of epochs (in deep learning algorithms) this paper used three different classifiers: Support Vector Machine Classifier, KNN Classifier, Random Forest Classifier, Multilayer Perceptron Classifier.

This set of algorithms was implemented using the Scikit-Learn package in the Python programming language [38].

More complicated Multilayer Perceptron Classifier and Convolutional Neural Network Classifier were implemented using the TensorFlow with Keras frontend package in the Python programming language.

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.

### 5.1. Pre-Processing of Images

To select areas of images containing recognizable digits, we used OpenCV library tools. The findContours function or the Maximally stable extremal region extractor (mser) algorithm were used to highlight the contours of the digits.

To recognize numbers in photographs of real cast iron ladle cars, an algorithm based on boundary detection was used [39]. The algorithm for preprocessing the image and highlighting the area containing the digits of the number included the following stages:

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).

Recognition images to be scaled to a size of 28x28 (784 cells in a one-dimensional representation). Each pixel value of the images lies between 0 to 255 followed by Normalizing these pixel values by converting the dataset into 'float32' and then dividing by 255.0 so that the input features will range

between 0.0 to 1.0. Next, one-hot is performed encoding to convert the y values into zeros and ones, making each number categorical.

## 5.2. Implementation of the considered algorithms

Classical classification algorithms were implemented on the basis of the package Scikit-learn [40]. It is an open source machine learning library that supports supervised and unsupervised learning. It also provides various tools for model fitting, data pre-processing, model selection and evaluation, and many other utilities.

To implement the Support Vector Machine algorithms, the module `sklearn.svm` was used. The support vector machines in scikit-learn support both dense (`numpy.ndarray` and convertible to that by `numpy.asarray`) and sparse (any `scipy.sparse`) sample vectors as input. However, to use an SVM to make predictions for sparse data, it must fit for such data. For optimal performance, use C-ordered `numpy.ndarray` (dense) or `scipy.sparse.csr_matrix` (sparse) with `dtype=float64`.

For binary and multiclass dataset classification, scikit-learn implements the SVC, NuSVC, and LinearSVC classes. SVC and NuSVC are similar methods, except for slightly different sets of parameters and they have different mathematical formulations (see section Mathematical formulation). On the other hand, LinearSVC is another (faster) implementation of Support Vector Classification for the case of a linear kernel.

The implementation of C-Support Vector Classification is based on `libsvm`. The fit time scales at least quadratically with the number of samples and may be impractical beyond tens of thousands of samples.

The tuning of the classifier was carried out by choosing the type of kernel, the regularization parameter and Kernel coefficient for 'rbf', 'poly' and 'sigmoid' kernel types.

A random forest classifier is a meta estimator that fits a number of decision tree classifiers on various sub-samples of the dataset and uses averaging to improve the predictive accuracy and control over-fitting. The sub-sample size is controlled by the `max_samples` parameter if `bootstrap=True` (default), otherwise the whole dataset is used to build each tree. The adjustment of this classifier was carried out by selecting the optimal value of `max_samples` parameter (it is a number of trees in the forest).

Scikit-learn package implements two different nearest neighbors classifiers: `KNeighborsClassifier` implements learning based on the k-nearest neighbors of each query point, where k is an integer value specified by the user. `RadiusNeighborsClassifier` implements learning based on the number of neighbors within a fixed radius r of each training point, where r is a floating-point value specified by the user.

The k-neighbors classification in `KNeighborsClassifier` is the most commonly used technique. The optimal choice of the value k is highly data-dependent: in general, a larger one suppresses the effects of noise, but makes the classification boundaries less distinct.

In this work, `KNeighborsClassifier` was used with the choice number of neighbors required for each sample.

The original Multilayer perceptron was also implemented using scikit-learn package. As the experience of setting up `MLPClassifier` showed, the following parameters have the main influence on the accuracy of character recognition:

- `hidden_layer_sizes` represents the number of neurons in the i-th hidden layer;
- `solver` represents the solver for weight optimization;
- `learning_rate_init` represents the initial used learning rate. It controls the step-size in updating the weights (only used when `solver='sgd'` or `'adam'`).

It was found that a result acceptable in terms of accuracy is achieved when choosing a sufficiently large number of neurons in the hidden layers, `sgd` or `adam` optimizers, and a moderate initial training speed.

All variants of the classifier models after tuning on the MNIST dataset were saved using the `joblib` module.

The other variant of Multilayer perceptron was implemented using TensorFlow package with Keras interface.

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 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.

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.

The implementation of handwritten digit recognition by Convolutional Neural Network was done using TensorFlow and Keras.

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. 1.

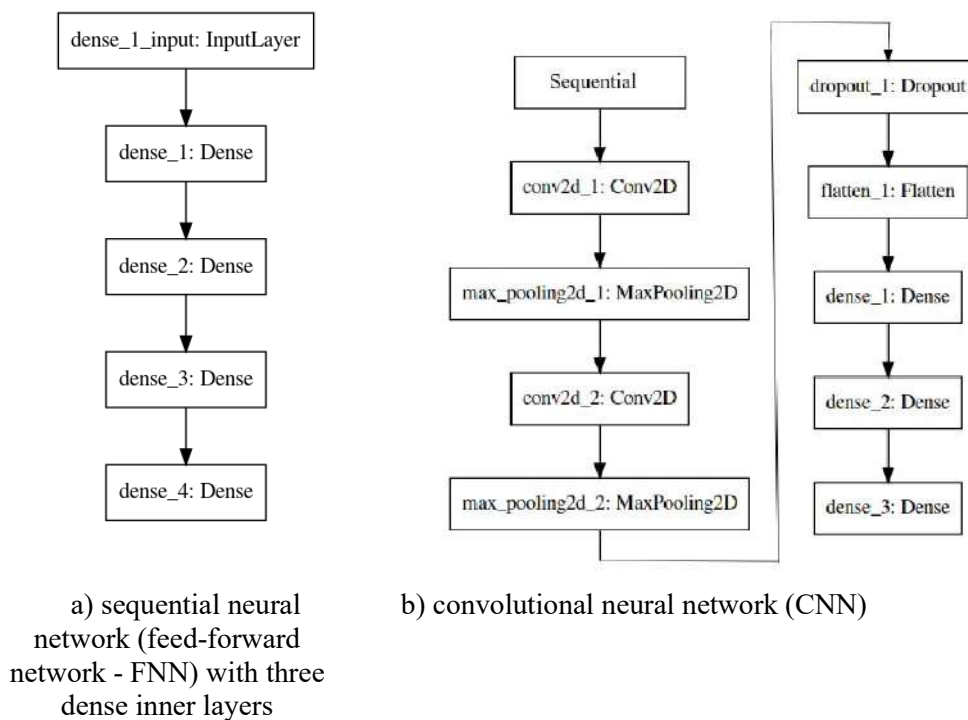


**Figure 1:** Structures variants of the deep learning neural network based on Keras framework, which were used to recognize car number elements

A simpler version of a fully connected neural network (Fig. 1, a) 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% [33].

A more accurate digit recognition of areas of interest was achieved using a convolutional neural network (CNN, see Fig. 1b). The disadvantage of this type of networks is a significantly longer duration of training.

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.

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:

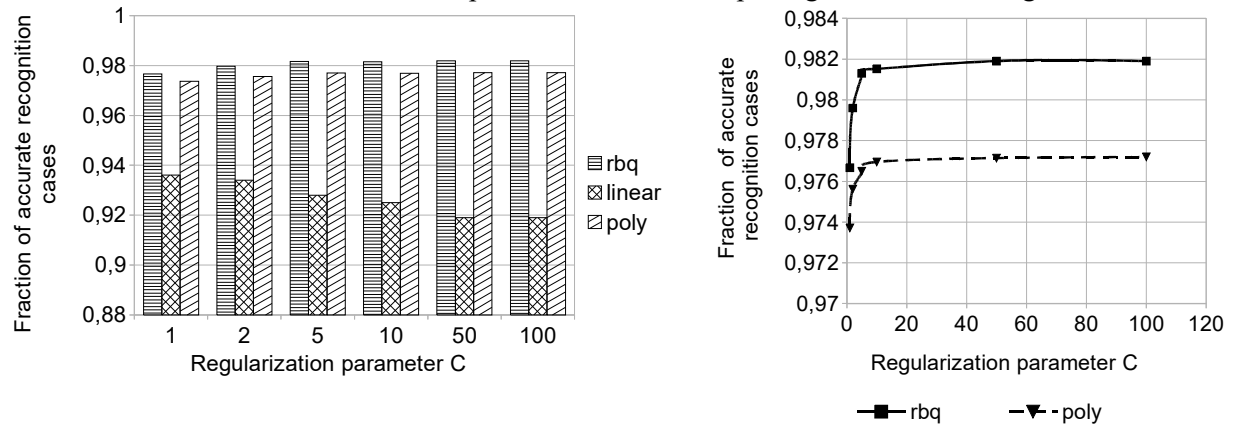
- architecture of the model, allowing to restore the model;
- model weight;
- training configuration (loss calculation function, optimizer)
- state of the optimizer, which allows to resume training exactly where it was stopped.

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).

However, when using customized models for recognizing handwritten numbers for other technical objects – carriage numbers, photographs of meter readings, etc., for all model variants (including those based on CNN), the number of errors increased sharply – up to 40-50%, depending on the dataset.

With proper tuning of all the models mentioned above (from KNN to CNN – as the complexity grows), the accuracy of the MNIST test sample estimate was 97.5-98.5%, which is slightly inferior to the best results achieved using preconfigured or convolutional neural networks.

The results of the SVC classifier as part of the scikit-learn package are shown in Fig. 2.



a) comparison of recognition results for different kernels (kernels «rbf», «linear», «poly» were used )

b) comparison of recognition results for different values of the regularization parameter

**Figure 2:** The results of the recognition accuracy evaluation for SVC classifier with different calibration parameters

As can be seen from the results shown in Fig. 2, the best accuracy of image recognition from the MNIST sample using the SVC classifier is achieved for the «rbf» kernel and a regularization parameter of at least 50.

Similar studies were carried out for other classification options - KNN Classifier, Random Forest Classifier, Multilayer Perceptron Classifier. The results of the analysis of the influence of the tuning parameters of the KNN Classifier and Random Forest (RF) Classifier classifiers are shown in Fig. 3.

As can be seen from fig. 3, a decrease in the number of nearest neighbors for the KNN Classifier or an increase in the number of trees in the forest for RF Classifier leads to an increase in recognition accuracy. However, both of the considered classification algorithms do not have any reserves for improving the recognition accuracy by varying the settings.

The settings of the multilayer perceptron had a great influence on the recognition accuracy. The results of the computational experiment are shown in Fig. 4. As can be seen from the graphs presented, the recognition accuracy of digit samples from the test sample increases with an increase in the number of neurons in the hidden layers (it was assumed to be the same for all layers) and with an increase in the number of hidden layers.

The graphs in Fig. 4 are built using the MLP Classifier method from the scikit-learn package. For a similar classifier built using the Keras package, similar results were obtained for the effect on the recognition accuracy of the number of hidden layers and the number of neurons in a layer.

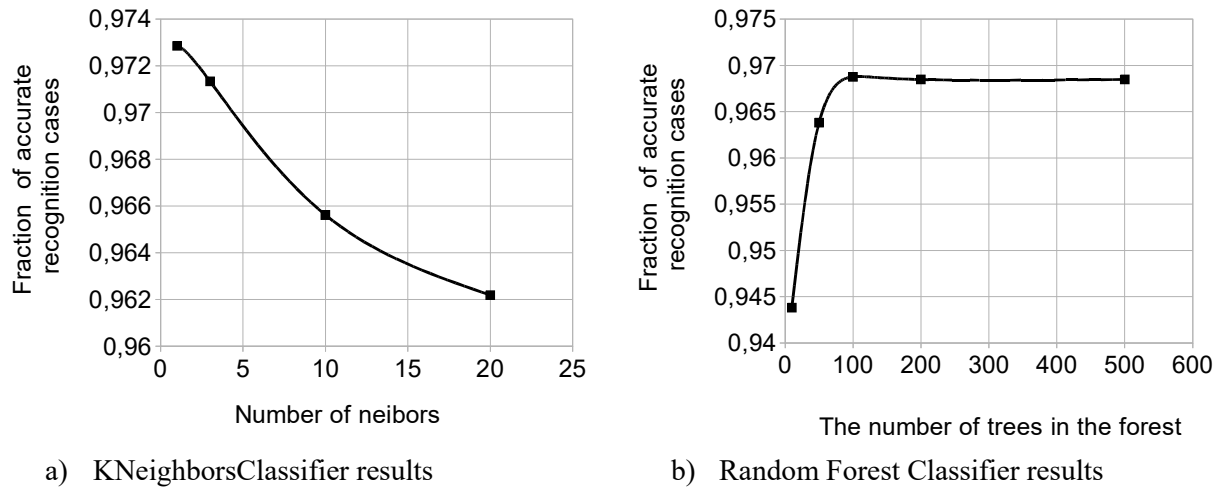


**Figure 3:** The results of the recognition accuracy evaluation for KNeighborsClassifier and Random Forest Classifier with different calibration parameters

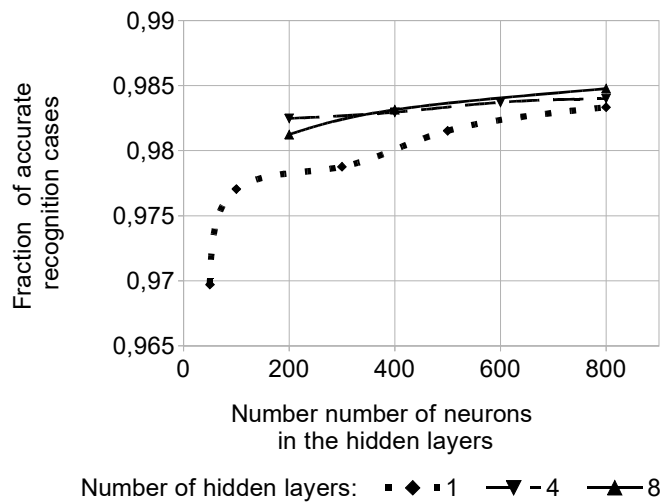


**Figure 4:** The results of the recognition accuracy evaluation for MLP Classifier (scikit-learning package) with different parameters

When testing various recognition models, some features of the MNIST dataset were established (fig. 5):

- all images of numbers have a clearly defined border area;
- images are grayscale, the brightness of pixels varies across the width of the digit image;
- the images of the numbers are centered in relation to the 28x28 area.

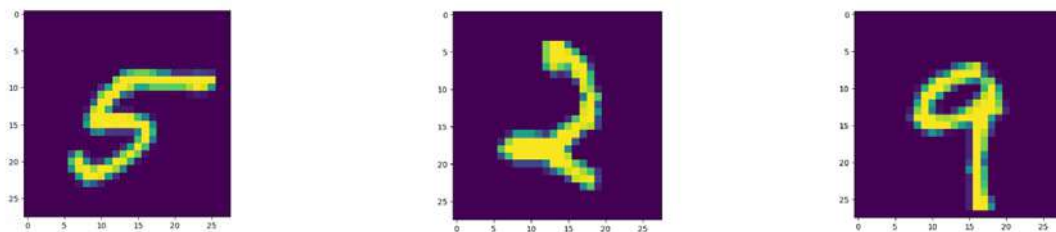


**Figure 5:** Sample images from the MNIST database

However, the area of interest images for recognition are actually black and white, either due to the binarization of the image at a preprocessing stage or are initially black and white.

To improve the recognition accuracy, it is necessary to perform two additional stages of image preprocessing:

- - after highlighting the area of interest contour exactly along the boundaries of the digit, this part of the image is centered in the square area;
- - the border of the image is added with a width of 15-25% of the size of the square area.

The width of the border area significantly affects the reliability of digit recognition (both printed and handwritten).

Results of the study of influence of the added width of the boundary region in the recognition accuracy set of handwritten characters 0-9 are shown at Fig. 6. A computational experiment for constructing this curve was performed using Scikit-learn implementation of the SVC algorithm.

An analysis of the effect of the width of the added boundary region was carried out for variants of digit recognition using neural networks. The shape of the curve in Fig. 6 for variants with a multilevel perceptron or a CNN network remained exactly the same as for the classification algorithms. An example of recognizing a set of handwritten digits 0-9 is shown in Fig. 7.

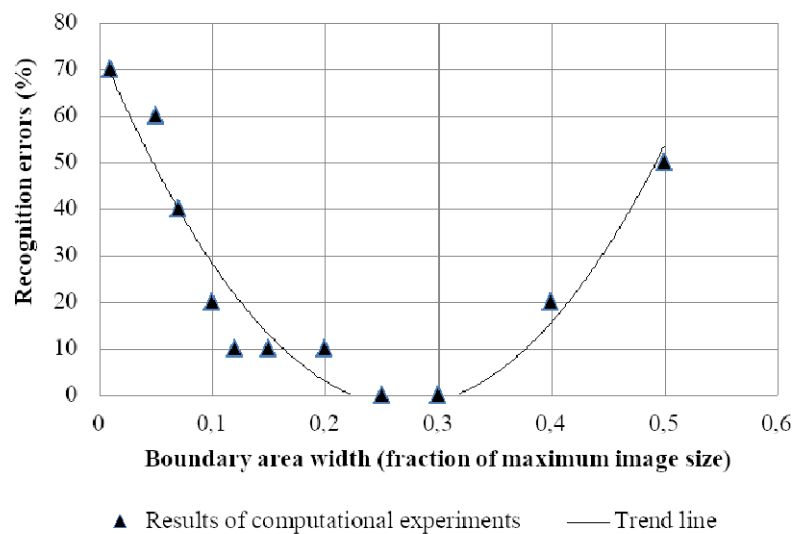
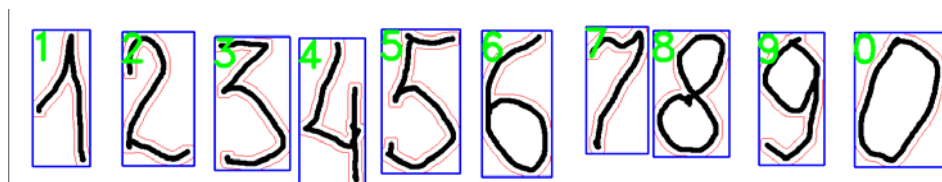
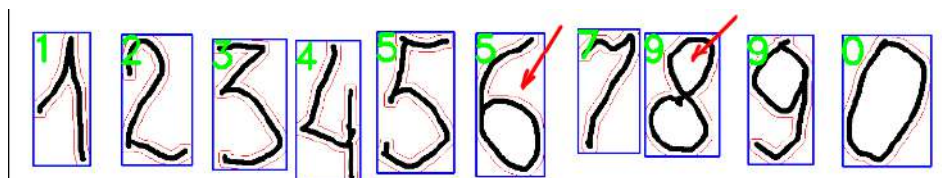


**Figure 6:** Dependence of the recognition error for a group of handwritten numbers on the width of the added border area



a) Border width 20% of the maximum image size. Recognition accuracy is 100%.



b) Border width 50% of the maximum image size. Recognition accuracy 80% (recognition errors is indicated by arrows).

**Figure 7:** Examples of recognition results with CNN deep learning network for a group of handwritten digits with different widths of the added border area

To assess the reasons of the recognition failure, a test program in Python was developed that allows to visually draw numbers and monitor the state of the recognition area and the result (Fig. 8).

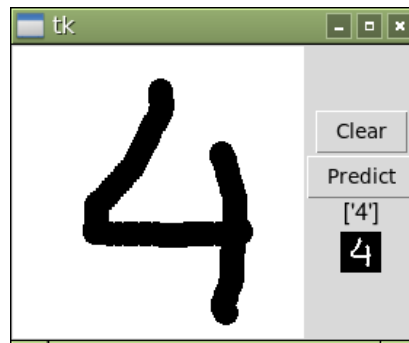


**Figure 8:** Appearance of the program window for testing handwritten character recognition

After building recognition models using all the algorithms mentioned above, the recognition accuracy of all handwritten digits on the test program turned out to be in the range of 98-100% (one or no errors per 50 drawn digits).

Another reserve for improving the recognition accuracy is setting up the base for building models to work with black and white images. For this, images from the MNIST dataset were converted to black and white.

The results obtained are largely related to the peculiarities of highlighting areas of interest containing digits in the image. The selection of the contour (the `findContours` function in OpenCV) is carried out exactly along the border of the founded contour. The selected area in the form of a rectangle described around the contour of the digit, without region of interest additional processing, noticeably differs from the base for recognition (MNIST).

A similar result has been achieved in industrial images. Regardless of the used neural network version, the recognition accuracy was 96-98%. For elements of the ladle car number binarization and morphological “closing” operations (`MORPH_CLOSE` in terms of the OpenCV library) were performed before recognition.

## 6. Conclusion

This research paper has implemented some models namely Support Vector Machine Classifier, KNN Classifier, Random Forest Classifier, Multilayer Perceptron Classifier, Multi-Layer Perceptron and Convolutional Neural Network for handwritten digit recognition using MNIST datasets. It compared them based on their working accuracy.

1. After a simple setup, all these algorithms demonstrated almost the same accuracy of handwritten digit recognition, differing within +1%.
2. It was found that CNN gave the most accurate results for handwritten digit recognition, but the only drawback is that it took an exponential amount of computing time.
3. To improve the recognition accuracy, it is necessary to perform two additional stages of image preprocessing and dataset transformation:
  - after highlighting the area of interest contour exactly along the boundaries of the digit, this part of the image is centered in the square area;
  - the border of the image is added with a width of 15-25% of the size of the square area;
  - converting images from the MNIST dataset to black and white form.

After building recognition models using all the algorithms mentioned above, the recognition accuracy of all handwritten digits on the test program turned out to be in the range of 98-100% (one or no errors per 50 drawn digits). A similar result was obtained when recognizing the generated sets of digits with different shapes – the recognition accuracy reaches the recognition accuracy on the MNIST test sample. For industrial images regardless of the used neural network version, the recognition accuracy was 96-98%.



## 7. References

- [1] Pramanik, R., Bag, S., (2018). Shape decomposition-based handwritten compound character recognition for Bangla OCR. *Journal of Visual Communication and Image Representation*, 50, 123-134. doi:10.1016/j.jvcir.2017.11.016
- [2] Sarkhel, R., Das, N., Saha, A. K., Nasipuri, M., (2016). A multi-objective approach towards cost effective isolated handwritten Bangla character and digit recognition. *Pattern Recognition*, 58, 172-189. doi:10.1016/j.patcog.2016.04.010
- [3] González, A., Bergasa L. M., (2013). A text reading algorithm for natural images. *Image and Vision Computing*. 31, 3, 255–274. doi: 10.1016/j.imavis.2013.01.003
- [4] Sanver, U., (2014). Identification of train wagon numbers. In: *Proceedings of the 2014 IEEE NW Russia Young Researchers in Electrical and Electronic Engineering Conference*, St. Petersburg, 2014, pp. 63-68. doi: 10.1109/ElConRusNW.2014.6839203.
- [5] Lisanti, G., Karaman, S., Pezzatini, D., Del Bimbo, A., (2015). A Multi-Camera Image Processing and Visualization System for Train Safety Assessment. <https://arxiv.org/pdf/1507.07815.pdf>
- [6] Surekha, P., Gurudath, P., Prithvi, R., Ritesh Ananth, V.G., (2018). Automatic license plate recognition using image processing and neural network. *ICTACT journal on image and video processing*, 08, 04, 1786-1792. doi: 10.21917/ijivp.2018.0251.
- [7] Shamim, S. M., Miah, Md Badrul, Sarker, Angona, Rana, Masud, Jobair, Abdullah. (2018). Handwritten Digit Recognition Using Machine Learning Algorithms. *Indonesian Journal of Science and Technology*. 18. 10.17509/ijost.v3i1.10795.
- [8] Dutt, Anuj, Dutt, Aashi, (2017). Handwritten Digit Recognition Using Deep Learning, *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*. 6, 7, 990-997.
- [9] Norhidayu binti Abdul Hamid, Nilam Nur Binti Amir Sharif, (2017). Handwritten recognition using SVM, KNN, and Neural networks, <https://arxiv.org/pdf/1702.00723.pdf>
- [10] Siddique, Fathma, Sakib, Shadman, Siddique, Abu Bakr. (2019). Recognition of Handwritten Digit using Convolutional Neural Network in Python with Tensorflow and Comparison of Performance for Various Hidden Layers. [https://www.researchgate.net/publication/335908757\\_Recognition\\_of\\_Handwritten\\_Digit\\_using\\_Convolutional\\_Neural\\_Network\\_in\\_Python\\_with\\_Tensorflow\\_and\\_Comparison\\_of\\_Performance\\_for\\_Various\\_Hidden\\_Layers](https://www.researchgate.net/publication/335908757_Recognition_of_Handwritten_Digit_using_Convolutional_Neural_Network_in_Python_with_Tensorflow_and_Comparison_of_Performance_for_Various_Hidden_Layers)
- [11] Farhana, S., Abu Sufian, Dutta, P., Advancements in Image Classification Using Convolutional Neural Network. <https://arxiv.org/pdf/1905.03288.pdf>
- [12] Wang, H., Zhou, Z., Li, Y., Chen, Z., Lu, P., Wang, W., Liu, W., Yu, L., (2017). Comparison of machine learning methods for classifying mediastinal lymph node metastasis of non-small cell lung cancer from 18F-FDG PET/CT images. *EJNMMI Research* 7:11 DOI 10.1186/s13550-017-0260-9. <https://arxiv.org/pdf/1702.02223.pdf>
- [13] Arif, R. B., Siddique, M. A. B., Khan, M. M. R., Oishe, M. R., (2018). 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)*, Dhaka, Bangladesh, 2018. pp. 112-117.
- [14] Siddique, F., Sakib, S., Siddique, M.A.B., (2019). Handwritten Digit Recognition using Convolutional Neural Network in Python with Tensorflow and Observe the Variation of Accuracies for Various Hidden Layers. *Preprints* 2019, 2019030039. doi: 10.20944/preprints201903.0039.v1.
- [15] LeCun, Y., The MNIST database of handwritten digits. <http://yann.lecun.com/exdb/mnist>.
- [16] 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>.
- [17] Vapnik, V., *The Nature of Statistical Learning Theory*. NY: Springer-Verlag (1995).

- [18] Srivastava, Durgesh, Bhambhu, Lekha. (2010). Data classification using support vector machine. Journal of Theoretical and Applied Information Technology. 12. 1-7. <http://www.jatit.org/volumes/research-papers/Vol12No1/1Vol12No1.pdf>
- [19] Gunn, S.R., Support vector machines for classification and regression, (1998). <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.10.7171&rep=rep1&type=pdf>
- [20] Safavian, S. R., Landgrebe, D., (1991). A survey of decision tree classifier methodology. In: IEEE Transactions on Systems, Man, and Cybernetics, vol. 21, no. 3, pp. 660-674, May-June 1991, doi: 10.1109/21.97458.
- [21] Géron, A., "Hands-On machine learning with Scikit-Learn and TensorFlow concepts," O'Reilly Publishing, 2017.
- [22] Luca Parisi. M-arcsinh: An Efficient and Reliable Function for SVM and MLP in scikit-learn. <https://arxiv.org/pdf/2009.07530.pdf>
- [23] Arya, S., Mount, D.M., Netanyahu, N.S., Silverman, R. Wu, A.Y., (1998). An optimal algorithm for approximate nearest neighbor searching in fixed dimension. Journal of the ACM, 45, 6, 891-923. <https://graphics.stanford.edu/courses/cs468-06-fall/Papers/03%20AMNSW%20-%20JACM.pdf>
- [24] Babu, U., Chintla, Aneel, Venkateswarlu, Y. (2014). Handwritten Digit Recognition Using Structural, Statistical Features and K-nearest Neighbor Classifier. International Journal of Information Engineering and Electronic Business. 6. 62-68. 10.5815/ijieeb.2014.01.07.
- [25] Toghi, Behrad, Grover, Divas. (2018). MNIST Dataset Classification Utilizing k-NN Classifier with Modified Sliding Window Metric. [https://www.researchgate.net/publication/327742790\\_MNIST\\_Dataset\\_Classification\\_Utilizing\\_k-NN\\_Classifier\\_with\\_Modified\\_Sliding\\_Window\\_Metric](https://www.researchgate.net/publication/327742790_MNIST_Dataset_Classification_Utilizing_k-NN_Classifier_with_Modified_Sliding_Window_Metric)
- [26] Bernard, S., Heutte, L., Adam, S., (2007). Using Random Forests for Handwritten Digit Recognition. Proceedings of the International Conference on Document Analysis and Recognition, ICDAR. 2. 1043-1047. 10.1109/ICDAR.2007.4377074.
- [27] Gilles Louppe. UNDERSTANDING RANDOM FORESTS: from theory to practice <https://arxiv.org/pdf/1407.7502.pdf>, arXiv:1407.7502v3 [stat.ML] 3 Jun 2015
- [28] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning (Adaptive Competition and Machine Learning). The MIT Press, p. 779, 2016.
- [29] Dalto, M., (2015). Deep Neural Networks for Time Series Prediction with Application in Ultra-Short-Term Wind Forecasting, IEEE, 1657–1663. [https://www.researchgate.net/publication/312928742\\_Deep\\_neural\\_networks\\_for\\_time\\_series\\_prediction\\_with\\_applications\\_in\\_ultra-short-term\\_wind\\_forecasting](https://www.researchgate.net/publication/312928742_Deep_neural_networks_for_time_series_prediction_with_applications_in_ultra-short-term_wind_forecasting).
- [30] Ferreira, A., Giraldi, G., (2017). Convolutional Neural Network Approaches to Granite Tiles Classification, Expert Systems with Applications, 84, 1–11, <https://doi.org/10.1016/j.eswa.2017.04.053>
- [31] Bengio, Y., (2009). Learning Deep Architectures for AI. [http://www.cs.cmu.edu/~10701/slides/deep\\_learning\\_paper.pdf](http://www.cs.cmu.edu/~10701/slides/deep_learning_paper.pdf).
- [32] Nwankpa, C.E., Ijomah, W., Gachagan, A., (2018). Marshall, S., Activation Functions: Comparison of Trends in Practice and Research for Deep Learning, <https://arxiv.org/pdf/1811.03378.pdf>
- [33] Tabik, Siham & Peralta, Daniel & Herrera-Poyatos, Andrés & Herrera, Francisco. (2017). A snapshot of image Pre-Processing for convolutional neural networks: Case study of MNIST. International Journal of Computational Intelligence Systems. 10. 555. 10.2991/ijcis.2017.10.1.38.
- [34] Yann LeCun, Leon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. Proc. IEEE, 86(11):2278–2324, 1998.
- [35] Geoffrey Hinton, Li Deng, Dong Yu, George Dahl, Abdel-rahman Mohamed, Navdeep Jaitly, Andrew Senior, Vincent Vanhoucke, Patrick Nguyen, Tara Sainath, and Brian Kingsbury. Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups. IEEE Signal Process. Mag., 29(6):82–97, 2012.
- [36] P. Y. Simard, D. Steinkraus and J. C. Platt, "Best practices for convolutional neural networks applied to visual document analysis," Seventh International Conference on Document Analysis

- and Recognition, 2003. Proceedings., Edinburgh, UK, 2003, pp. 958-963, doi: 10.1109/ICDAR.2003.1227801.
- [37] Stark, F., Hazirbas, C., Triebel, R., Cremers, D. CAPTCHA Recognition with Active Deep Learning, In: GCPR Workshop on New Challenges in Neural Computation, 2015. <https://vision.in.tum.de/media/spezial/bib/stark-gcpr15.pdf>
  - [38] Lauer, F., Suen, C. Y., Bloch, G. (2007). A trainable feature extractor for handwritten digit recognition. Pattern Recognition, 40(6), 1816-1824. doi:10.1016/j.patcog.2006.10.011
  - [39] Scikit-learn. User guide. [https://scikit-learn.org/stable/user\\_guide.html](https://scikit-learn.org/stable/user_guide.html)
  - [40] Serhiienko, A., Chychkarov, Y., Syrmamiikh, I., (2020). Information Technology of Hot-metal Ladle Car Handwritten Numbers Recognition from Photo Image. In: Computer Modeling and Intelligent Systems (CMIS-2020): Proceedings of the Third International Workshop (Zaporizhzhia, April 27 – May 1, 2020 y.). – Zaporizhzhia, 2020, Vol. I-2608, P. 900–912. <http://ceur-ws.org/Vol-2608/paper67.pdf>