УКРАЇНА

ПАТЕНТ

НА КОРИСНУ МОДЕЛЬ

№ 134493

СПОСІБ ВІДБОРУ ІНФОРМАТИВНИХ ОЗНАК ВЕЛИКИХ ДАНИХ ДЛЯ ПОБУДОВИ РОЗПІЗНАВАЛЬНИХ МОДЕЛЕЙ

Видано відповідно до Закону України "Про охорону прав на винаходи і корисні моделі".

Зареєстровано в Державному реєстрі патентів України на корисні моделі **27.05.2019.**

Заступник Міністра економічного розвитку і торгівлі України Ю.П. Бровченко

МІНІСТЕРСТВО ЕКОНОМІЧНОГО РОЗВИТКУ І ТОРГІВЛІ УКРАЇНИ ІДЕНТИФІКАЦІЙНИЙ КОД 37508596

(11) **134493**

(19) **UA**

(51) **МПК** (2019.01)
G06F 17/00

(21) Номер заявки: **u 2018 10874**

(22) Дата подання заявки: **02.11.2018**

(24) Дата, з якої є чинними права на корисну модель: **27.05.2019**

(46) Дата публікації відомостей про видачу патенту та номер бюлетеня: **27.05.2019, Бюл. № 10**

(72) Винахідники:
Олійник Андрій Олександрович, UA,
Льовкін Валерій Миколайович, UA

(73) Власник:
ЗАПОРІЗЬКИЙ НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ,
вул. Жуковського, 64, м. Запоріжжя, 69063, UA

(54) Назва корисної моделі:

СПОСІБ ВІДБОРУ ІНФОРМАТИВНИХ ОЗНАК ВЕЛИКИХ ДАНИХ ДЛЯ ПОБУДОВИ РОЗПІЗНАВАЛЬНИХ МОДЕЛЕЙ

(57) Формула корисної моделі:

Спосіб відбору інформативних ознак великих даних для побудови розпізнавальних моделей, який полягає в тому, що набір інформативних ознак формують на основі вибірки даних, будують модель, за якою оцінюють значення цільової функції помилки та як рішення приймають комбінацію ознак, для якої помилка є мінімальною, який **відрізняється** тим, що вибірку даних передають на вузли паралельної системи, де відбір інформативних ознак реалізують окремо від інших вузлів і використовують різні способи, що ґрунтуються на імовірнісному підході, інформацію про досліджені на окремих вузлах точки простору пошуку передають іншим вузлам для уникнення повторного оцінювання, збирають, розподіляють інформацію про процес відбору та узгоджують її, виконуючи оцінку концентрованості рішень навколо найкращих значень в обмежених областях пошуку на головному вузлі системи, за надмірної концентрації контрольних точок виконують введення в поточну множину рішень додаткових контрольних точок, рішення, які характеризуються найкращим значенням цільової функції, передають для подальшого оброблення, а всі інші рішення вибирають з наявного набору випадковим чином.

МІНІСТЕРСТВО ЕКОНОМІЧНОГО РОЗВИТКУ І ТОРГІВЛІ УКРАЇНИ

УКРАЇНА **(19) UA** **(11) 134493** **(13) U**

(51) МПК (2019.01)
G06F 17/00

(12) ОПИС ДО ПАТЕНТУ НА КОРИСНУ МОДЕЛЬ

(21) Номер заявки: **u 2018 10874**

(22) Дата подання заявки: **02.11.2018**

(24) Дата, з якої є чинними права на корисну модель: **27.05.2019**

(46) Публікація відомостей про видачу патенту: **27.05.2019, Бюл.№ 10**

(72) Винахідник(и):
Олійник Андрій Олександрович (UA),
Льовкін Валерій Миколайович (UA)

(73) Власник(и):
ЗАПОРІЗЬКИЙ НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ,
вул. Жуковського, 64, м. Запоріжжя, 69063 (UA)

(74) Представник:
Висоцька Наталя Іванівна, начальник патентно-інформаційного відділу НДЧ ЗНТУ

(54) СПОСІБ ВІДБОРУ ІНФОРМАТИВНИХ ОЗНАК ВЕЛИКИХ ДАНИХ ДЛЯ ПОБУДОВИ РОЗПІЗНАВАЛЬНИХ МОДЕЛЕЙ

(57) Реферат:

Спосіб відбору інформативних ознак великих даних для побудови розпізнавальних моделей полягає в тому, що набір інформативних ознак формують на основі вибірки даних, будують модель, за якою оцінюють значення цільової функції помилки та як рішення приймають комбінацію ознак, для якої помилка є мінімальною. Вибірку даних передають на вузли паралельної системи, де відбір інформативних ознак реалізують окремо від інших вузлів і використовують різні способи, що ґрунтуються на імовірнісному підході. Інформацію про досліджені на окремих вузлах точки простору пошуку передають іншим вузлам для уникнення повторного оцінювання. Збирають, розподіляють інформацію про процес відбору та узгоджують її, виконуючи оцінку концентрованості рішень навколо найкращих значень в обмежених областях пошуку на головному вузлі системи. За надмірної концентрації контрольних точок виконують введення в поточну множину рішень додаткових контрольних точок. Рішення, які характеризуються найкращим значенням цільової функції, передають для подальшого оброблення, а всі інші рішення вибирають з наявного набору випадковим чином.

UA 134493 U

Корисна модель належить до кібернетики й обчислювальної техніки і може бути використана для прискорення побудови, спрощення структури та підвищення ефективності розпізнавальних моделей шляхом відбору найбільш цінних ознак для побудови моделей.

Процесом як об'єктом технології корисної моделі є сукупність дій з вимірювання параметрів (ознак) об'єктів технічного діагностування (зразків), перетворення даних про об'єкти для оцінювання важливості ознак, а також керування процесом відбору інформативних ознак з метою автоматизації формування навчальних та тестових вибірок спостережень для побудови розпізнавальних моделей.

Продуктами, щодо яких у корисній моделі виконуються дії, є дані про значення параметрів зразків.

Продуктом, за допомогою якого у корисній моделі виконуються дії, є електронно-обчислювальна машина (ЕОМ).

Відомий спосіб відбору інформативних ознак [1], який полягає у тому, що в ньому спочатку виконують послідовне додавання інформативних ознак до заданої кількості шляхом визначення найменшої помилки зі всіх альтернативних варіантів додавання ознаки до наявного набору, після чого кількість ознак зменшують до заданої кількості шляхом послідовного вилучення з наявного набору ознак, що дають в комбінації з іншими ознаками найбільшу помилку. Це дозволяє при побудові розпізнавальних моделей використовувати менше даних, ніж у вихідній вибірці і тим самим прискорити процес побудови моделі та спростити її структуру.

Недоліками відомого способу є надто низька швидкість роботи для великих даних, де вибірки характеризуються великою кількістю ознак і екземплярів, оскільки для зменшення впливу можливої помилки ознаки поступово то додаються, то заново вилучаються, що призводить до додаткових витрат часу, а також призводять до неефективного виконання відбору інформативних ознак, оскільки не в повній мірі враховується вплив усіх ознак між собою, а тільки вплив наявного набору з додаванням або вилученням однієї ознаки.

Відомий спосіб відбору інформативних ознак для діагностики виробів [2], вибраний як найближчий аналог, який полягає в тому, що набір інформативних ознак формують на основі вибірки даних, генерують набір можливих комбінацій ознак та для кожної комбінації ознак будують модель, за якою оцінюють значення цільової функції помилки та як рішення приймають комбінацію ознак, для якої помилка є мінімальною, який відрізняється тим, що спочатку оцінюють індивідуальний вплив ознак на прогнозований параметр, після чого розраховують групове значення інформативності ознак, ознаки з оцінкою індивідуального впливу меншою за групове значення фіксують, зменшуючи розмірність комбінаторного простору пошуку, після чого ітеративно формують поточну підмножину рішень, для яких будують моделі й оцінюють помилки та формують нову підмножину рішень за певними правилами, які враховують попередньо отриману інформацію та вносять певне розмаїття.

Недоліками відомого способу є те, що він характеризується низькою швидкістю роботи для вибірок великого обсягу та може призводити до отримання не найціннішої множини інформативних ознак за умов невдалого початкового виділення комбінацій ознак, оскільки на це впливає тільки результат обчислення індивідуального впливу ознак і відповідне усереднене значення за всіма ознаками.

В основу корисної моделі поставлено задачу створення способу відбору інформативних ознак великих даних для побудови розпізнавальних моделей з метою підвищення швидкості процесу відбору інформативних ознак та покращення ефективності розпізнавальних моделей за рахунок виділення найбільш цінних ознак.

Поставлена задача вирішується тим, що у способі відбору інформативних ознак великих даних для побудови розпізнавальних моделей набір інформативних ознак формують на основі вибірки даних, будують модель, за якою оцінюють значення цільової функції помилки та як рішення приймають комбінацію ознак, для якої помилка є мінімальною. Причому вибірку даних передають на вузли паралельної системи, де відбір інформативних ознак реалізують окремо від інших вузлів і використовують різні способи, що ґрунтуються на імовірнісному підході, інформацію про досліджені на окремих вузлах точки простору пошуку передають іншим вузлам для уникнення повторного оцінювання, збирають, розподіляють інформацію про процес відбору та узгоджують її, виконуючи оцінку концентрованості рішень навколо найкращих значень в обмежених областях пошуку на головному вузлі системи, за надмірної концентрації контрольних точок виконують введення в поточну множину рішень додаткових контрольних точок, рішення, які характеризуються найкращим значенням цільової функції, передають для подальшого оброблення, а всі інші рішення вибирають з наявного набору випадковим чином.

У порівнянні з найближчим аналогом відмінними ознаками є одночасне виконання відбору інформативних ознак на наборі вузлів паралельної системи з використанням різних способів,

оцінювання концентрованості контрольних точок навколо найкращих значень в обмежених областях пошуку та внесення в поточну множину рішень додаткових контрольних точок. Це дозволяє підвищити швидкість відбору інформативних ознак та отримати вибірки даних, які призводять до отримання більш точних результатів розпізнавання образів.

У технічному рішенні, що заявляється, нові ознаки при взаємодії з відомими дають новий технічний результат, що дозволяє вирішити поставлену задачу.

Ознаки, що відрізняють технічне рішення, яке заявляється, від найближчого аналога, не виявлені в інших технічних рішеннях при вивченні цієї галузі техніки.

Спосіб відбору інформативних ознак великих даних для побудови розпізнавальних моделей виконують таким чином.

Спочатку до пам'яті ЕОМ, яка є головним вузлом Pr_0, уводять вихідну вибірку даних $S =< P, T >$, що містить Q екземплярів, кожен з яких характеризується значеннями ознак p_{q1}, p_{q2},..., p_{qM}, і вихідного параметра t_q, де p_{qm} - значення m-ої ознаки q-го екземпляру, $q = 1, 2, ..., Q$, $m = 1, 2, ..., M$;

M - загальне число ознак у множині S.

Далі визначають наступні параметри роботи:

- набір способів відбору ознак MS, що може включати еволюційний пошук з групуванням ознак, еволюційний метод з кластеризацією ознак, мультиагентні методи з прямим і непрямим зв'язком між агентами, метод відбору ознак на основі дерев рішень, метод відбору ознак на основі асоціативних правил;

- число елітних рішень $x_{elit,j}$, що автоматично переносяться на наступну ітерацію пошуку.

Окрім того обчислюють значення індивідуальних інформативностей $V(p_m)$ ознак p_m, що характеризують взаємозв'язок ознаки з вихідним параметром T, використовуючи коефіцієнт парної кореляції.

Далі по вузлах $Ps = \{Pr_1, Pr_2, ..., Pr_{NPr-1}\}$ обчислювальної системи розподіляють способи відбору ознак MS та передають доступ до вхідної вибірки S. При цьому на низькоітеративні способи виділяють по одному вузлу, а серед інших вузлів рівномірно розподіляють більш складні способи відбору інформативних ознак, що використовують як базис еволюційний і мультиагентний підходи.

Після цього на кожному вузлі Ps виконують процес скорочення ознак вибірки S за допомогою відповідного способу.

Вузли Ps надсилають головному вузлу сигнал Sgn_{inj} про завершення відбору ознак на j-му вузлі Pr_j за досягнення деякого з критеріїв зупинки:

$Crit_1$ - успішне знаходження комбінації ознак P_{red}, що задовольняє заданим мінімально прийнятним умовам пошуку (наприклад, мінімально прийнятному значенню критерію оптимальності набору ознак);

$Crit_2$ - максимально прийнятна кількість ітерацій пошуку;

$Crit_3$ - максимально прийнятна кількість обчислень значень цільової функції.

З головного вузла Pr_0 на вузли Ps відправляють сигнали Sgn_{outj} про необхідність завершення відбору ознак на вузлі Pr_j за умов:

- отримання від будь-якого вузла Ps сигналу Sgn_{inj} про успішне завершення пошуку при задоволенні критерію $Crit_1$;

отримання від деякої множини вузлів Ps (наприклад, не менше ніж від половини вузлів Ps) сигналу Sgn_{inj} про завершення пошуку при задоволенні критеріїв $Crit_2$ або $Crit_3$,

досягнення максимально допустимого часу пошуку $Crit_4$, після якого на кожному вузлі Ps виконують завершення поточної ітерації пошуку і передачу інформації на головний вузол про множину досліджених контрольних точок $Xe \in XS$ і відповідні їм значення цільової функції $V(Xe)$.

У процесі пошуку зберігають інформацію $Inf_k =< Xe_k, V(Xe_k) >$ про досліджені на кожному вузлі Ps точки $Xe \in XS$ простору пошуку XS.

По завершенні відбору ознак на вузлах Ps виконують збір і розподіл поточної інформації про процес відбору. Для цього на головний вузол Pr_0 з вузлів Ps надсилають інформацію $InformPr_j$, про множини досліджених контрольних точок:

$$InformPr_j = \{Inform_1, Inform_2, ..., Inform_{Nxj}\},$$

$$Inform_k = < x_k, V(x_k), InformL(x_k), InformM(x_k) >,$$

де $Inform_k$ - інформація про к -е рішення;

N_{xj} - - кількість контрольних точок, досліджених на j-му вузлі;

$x_k - k$ - є рішення, що оцінене у процесі відбору ознак і відповідає k-й дослідженій контрольній точці Xe_k у просторі пошуку XS: $x_k \to Xe_k$;

$V(x_k)$ - значення цільової функції k -го рішення;

$InformL(x_k)$ - прапор, що відбиває присутність рішення x_k у множині рішень $R(iter) = [x_1, x_2, ..., x_{N_x}]$ на останній ітерації пошуку $iter$;

$InformM(x_k)$ - перелік способів, на яких було оцінено рішення x_k.

Після отримання N_{pr} інформації $InformPr_j$ з усіх вузлів Ps виконують її об'єднання на головному вузлі Pr_0: $Inform = \bigcup_{j=1} InformPr_j$. Якщо при об'єднанні множин $InformPr_j$ виникає ситуація, коли одне і те ж рішення x_k присутнє у різних множинах, то в змінну $Inform(x_k)$ заносять перелік всіх способів, в яких приймало участь рішення x_k. Значення цільової функції $V(x_k)$ вибирають найкращим серед оцінок, отриманих на різних вузлах.

Після отримання інформації $InformPr_j$ про множини досліджених контрольних точок x_k на головному вузлі Pr_0 виконують оцінювання їх концентрованості в поточній множині рішень $P(iter) = \{x_1, x_2, ..., x_{N_x}\}$ навколо найкращих значень в обмежених областях пошуку $v_{cons}(iter)$ для визначення рівномірності покриття простору пошуку XS у процесі відбору ознак, задля чого виконують розбиття поточної множини рішень $R(iter)$ на кластери $Cl(iter) = \{Cl_1, Cl_2, ..., Cl_{NCl}\}$ на основі застосування методів кластерного аналізу.

Потім для визначення концентрованості рішень навколо найкращих значень в обмежених областях пошуку обчислюють такі критерії:

1) середня відстань $dC(Cl_c)$ між рішеннями в конкретному кластері:

$$dC(Cl_c) = \frac{2}{|Cl_c|(|Cl_c|-1)} \sum_{k=1}^{|Cl_c|} \sum_{u=k+1}^{|Cl_c|} d(x_k, x_u), x_k, x_u \in Cl_c, \quad (1)$$

де $d(x_k, x_u)$ - відстань між точками x_k і x_u простору пошуку XS, що належать кластеру Cl_c. Відстань $d(x_k, x_u)$ обчислюється за формулою:

$$d(x_k, x_u) = \frac{1}{M} \sum_{m=1}^{M} |g_{mk} - g_{mu}|$$

де g_{mk} і $g_{mu} - m$ -і координати k-то і u-го рішення, відповідно;

2) дисперсія $DC(Cl_c)$ рішень x_k у межах кластеру Cl_c відображає середню відстань від центру $\overline{x_c}$ до рішень x_k, що відносяться до кластеру Cl_c:

$$DC(Cl_c) = \frac{1}{|Cl_c|} \sum_{x_k \in Cl_c} d(x_k, \overline{x_c}), \quad (3)$$

$$d(x_k, x_c) = \sqrt{\sum_{m=1}^{M} (g_{mk} - \overline{g_{mClc}})^2}, \quad (4)$$

$$\overline{g_{mClc}} = \frac{1}{|Cl_c|} \sum_{x_k \in Cl_c} g_{mk}, \quad (5)$$

де $d(x_k, \overline{x_c})$ - відстань між рішеннями x_k і центром с-го кластеру $\overline{x_c} = \{\overline{g_{1Clc}}, \overline{g_{2Clc}}, \overline{g_{MClc}}$;

$\overline{g_{mClc}} - m$ -а координата центру с -го кластеру;

3) середньокластерна відстань $dC(iter)$ між рішеннями на поточній ітерації $iter$ пошуку:

$$dC(iter) = \frac{\sum_{c=1}^{N_{Cl}} |Cl_c| dC(Cl_c)}{\sum_{c=1}^{N_{Cl}} |Cl_c|} = \frac{1}{N_x} \sum_{c=1}^{N_{Cl}} |Cl_c| dC(Cl_c) \quad ; \quad (6)$$

4) середньокластерна дисперсія $DC(iter)$ рішень на поточній ітерації $iret$:

$$dC(iter) = \frac{1}{N_x} \sum_{c=1}^{N_{Cl}} |Cl_c| dC(Cl_c) \quad ; \quad (7)$$

5) коефіцієнт концентрованості рішень на поточній ітерації:

$$v_{conc}(iter) = \frac{dC(iter)}{d(iter)}, \quad (8)$$

$$d(iter) = \frac{2}{N_x(N_x - 1)} \sum_{k=1}^{N_x} \sum_{u=k+1}^{N_x} d(x_k, x_u), x_k, x_u \in R(iter), \quad (9)$$

де $d(iter)$ - середня відстань між усіма рішеннями на поточній ітерації. Використовуючи оцінки дисперсії рішень $DC(iter)$, коефіцієнт концентрованості рішень на поточній ітерації обчислюють наступним чином:

$$v_{conc}(iter) = \frac{DC(iter)}{D(iter)}, \quad (10)$$

$$D(iter) = \frac{1}{N_x} \sum_{k=1}^{N_x} d(x_k, \bar{x}), \quad (11)$$

$$d(x_k, \bar{x}) = \sqrt{\sum_{m=1}^{M} (g_{mk} - \overline{g_m})^2}, \quad (12)$$

$$\overline{g_m} = \frac{1}{N_x} \sum_{k=1}^{N_x} g_{mk}, \quad (13)$$

де $D(iter)$ - дисперсія рішень на поточній ітерації;

$d(x_k, \bar{x})$ - відстань між рішенням x_k і центральним рішенням $\bar{x}$ на ітерації $iter$;

$\overline{g_m} - m$ -а координата центрального рішення $\bar{x}$.

Величина критерію $v_{conc}(iter)$ знаходиться в інтервалі (0;1). Чим ближче цей критерій до одиниці, тим менш згрупованими є рішення (відповідно, простір пошуку покрито контрольними точками більш рівномірно). Значення критерію $v_{conc}(iter)$, близькі до нуля, свідчать про суттєву концентрацію рішень навколо найкращих значень в обмежених областях пошуку;

6) максимальна кількість контрольних точок $x_k \in Cl_c$, згрупованих у межах однієї обмеженої області пошуку найкращих значень (в області кластеру Cl_c):

$$N_{max\,Ncl}(iter) = \max_{c=1,2,...,N_{cl}} (|Cl_c|). \quad (14)$$

Чим більше величина критерію $N_{max\,Ncl}(iter)$, тим більша кількість рішень згрупована в межах однієї обмеженої області пошуку найкращих значень, і, відповідно, тим менш рівномірно розподілені рішення в просторі пошуку XS на ітерації $iter$.

Далі розглядають можливість прийняття рішення про надмірну концентрацію контрольних точок навколо найкращих значень в обмежених областях пошуку на основі значень інтегрального критерію $v_{conc}(iter)$ та критерію $N_{max\,Ncl}(iter)$. У разі, якщо значення хоча б одного з даних критеріїв вище заданого порогового значення, приймають рішення про надмірну концентрацію контрольних точок навколо найкращих значень в обмежених областях пошуку. Для підвищення рівномірності покриття простору пошуку вводять в поточну множину рішень $R(iter+1)$ додаткові контрольні точки x_a У кількості, рівній уже наявним рішенням у множині $R(iter)$. Обчислюють середні значення $\overline{g_m}$ $m - x$ координат центрального рішення $\bar{x}$ використовуючи формулу (13).

Потім на основі обчислених значень $\overline{g_m}$ і випадково згенерованого числа rand[0;1] створюють нові рішення $x_a = \{g_{1a}, g_{2a}, \ldots, g_{Ma}\}$, m-у координату g_{ma} яких визначають за формулою

$$g_{ma} = \begin{cases} 1, & rand[-1;1] > (\overline{g_m} - V(p_m)); \\ 0, & rand[-1;1] \le (\overline{g_m} - V(p_m)). \end{cases} \quad (15)$$

Після цього на вузлах Ps перезапускають процедуру відбору. При цьому нові початкові множини рішень $R_j(iter+1)$ для відповідних способів відбору ознак формують на основі множини $R(iter+1)$. Для цього спочатку відбирають рішення $x_{elit,j}$, які на попередній ітерації характеризувалися найкращим значенням цільової функції: $V(x_{elit,j}) = \max_{x_k \in R_j(iter)} (V(x_k))$. Потім з множини $R(iter+1)$ випадковим чином вибирають рішення x_k, загальне число яких вибирається з потреб способу відбору ознак, що використовується на j-му вузлі Pr_j обчислювальної системи.

Потім на основі початкових множин контрольних точок $R_j(iter+1)$ на вузлах Pr_1, $Pr_2, \ldots, Pr_{NPr-1}$ виконують відбір ознак.

Роботу закінчують тоді, коли буде досягнутий один з наступних критеріїв зупинки: $Crit_1$, $Crit_5$ - перевищення загального максимально допустимого часу пошуку на паралельній системі або $Crit_6$ - досягнення максимально допустимого числа перезапусків відбору ознак на вузлах Ps.

У результаті з заданої вихідної вибірки даних S отримують вибірку $S^* = < P^*, T >$, що містить найважливіші ознаки основних екземплярів даних, тобто $P^* = \{p_{qr} | q = \overline{1,Q}, r \in P_{red}, |P_{red}| < M\}$.

На основі запропонованого способу здійснювався відбір інформативних ознак для задачі розпізнавання автомобільних засобів, яка характеризувалася вибіркою даних з 10000 екземплярів, а кожен екземпляр даної вибірки представляв зображення транспортного засобу і був сформований значеннями 26 ознак і 1 цільового параметра, який визначав, чи належить екземпляр класу, що розглядається (мотоцикліст, легковий автомобіль, вантажний автомобіль, автобус, мінівен або нерозпізнаний об'єкт). Це дозволило скоротити час виконання відбору інформативних ознак у 3,29 разу у порівнянні з прототипом та в середньому на 7,93 % підвищити точність розпізнавання за рахунок відсутності концентрації навколо найкращих значень в обмежених областях пошуку у окремих випадках. Це свідчить про збільшення швидкості виконання відбору інформативних ознак та збільшення точності розпізнавання за допомогою отриманої в результаті вибірки даних і відповідно про прийнятність якості сформованої вибірки зразків для побудови розпізнавальної моделі.

Технічним результатом внаслідок використання запропонованого способу є:

підвищення швидкості процесу відбору інформативних ознак за рахунок розпаралелювання даного процесу на вузлах системи та збереження інформації про досліджені на кожному вузлі точки простору пошуку;

збільшення точності розпізнавання за рахунок використання одразу декількох різних способів, а також введення додаткових точок у простір пошуку за випадків надмірної концентрації контрольних точок навколо найкращих значень в обмежених областях пошуку, що дозволяє підвищити різноманітність множини рішень на поточній ітерації та більш рівномірно покрити простір пошуку контрольними точками.

Джерела інформації:

1. Загоруйко, Н.Г. Прикладные методы анализа данных и знаний /Н.Г. Загоруйко. - Новосибирск: ИМ СО РАН, 1999. - 270 с.

2. Пат. № 18294 Україна, MnKG06F 19/00. Спосіб відбору інформативних ознак для діагностики виробів /С.О. Субботін, А.О. Олійник; заявник Запорізький національний технічний університет. - № u200603087; Заявл. 22.03.06; Опубл. 15.11.06; Бюл. № 11. - 4 с.

ФОРМУЛА КОРИСНОЇ МОДЕЛІ

Спосіб відбору інформативних ознак великих даних для побудови розпізнавальних моделей, який полягає в тому, що набір інформативних ознак формують на основі вибірки даних, будують модель, за якою оцінюють значення цільової функції помилки та як рішення приймають комбінацію ознак, для якої помилка є мінімальною, який **відрізняється** тим, що вибірку даних передають на вузли паралельної системи, де відбір інформативних ознак реалізують окремо від інших вузлів і використовують різні способи, що ґрунтуються на імовірнісному підході, інформацію про досліджені на окремих вузлах точки простору пошуку передають іншим вузлам для уникнення повторного оцінювання, збирають, розподіляють інформацію про процес відбору та узгоджують її, виконуючи оцінку концентрованості рішень навколо найкращих значень в обмежених областях пошуку на головному вузлі системи, за надмірної концентрації контрольних точок виконують введення в поточну множину рішень додаткових контрольних точок, рішення, які характеризуються найкращим значенням цільової функції, передають для подальшого оброблення, а всі інші рішення вибирають з наявного набору випадковим чином.

Комп'ютерна верстка Л. Литвиненко

Міністерство економічного розвитку і торгівлі України, вул. М. Грушевського, 12/2, м. Київ, 01008, Україна

ДП "Український інститут інтелектуальної власності", вул. Глазунова, 1, м. Київ – 42, 01601