

## МЕТОД ПАРАЛЕЛЬНОГО ФОРМУВАННЯ КОРТЕЖІВ ОЗНАК ДАНИХ СЕГМЕНТУ СКЛАДНОКЛАСИФІКОВАНОЇ ІНФОРМАЦІЇ В УМОВАХ НЕВИЗНАЧЕНОСТІ ПРИЙНЯТТЯ РІШЕНЬ ТА АНАЛІЗУ ВПЛИВУ НА РОЗМЕЖОВАНІСТЬ ІНФОРМАЦІЇ

*В роботі наведено результати досліджень з визначення ознак даних, які мають визначальний вплив на розмежованість інформації в умовах невизначеності прийняття рішень інтелектуальними інформаційними системами. Розглядається загальна множина ознак з виділенням підмножини найбільшого впливу на розмежованість даних та формування кортежів ознак в зоні складнокласифікованої інформації при локальному зміщенні даних в гіперпросторі ознак.*

*Ключові слова: інтелектуальна інформаційна технологія, складнокласифікована інформація, прийняття рішень.*

## METHOD OF PARALLEL FORMATION OF TUPLES FEATURES OF DATA IN THE SEGMENT OF DIFFICULT TO CLASSIFY INFORMATION IN CONDITIONS OF UNCERTAINTY MAKING DECISION AND ANALYSIS OF THE IMPACT ON THE INFORMATION FRAGMENTATION

*Detection of atypical data and outliers is an important and difficult task in the classification of information. Information that is considered complex in turn is characterized by a set of features that determine its informativeness. Therefore, the search for signs, due to which the signs of information are considered atypical is an urgent and necessary task. A more difficult task is to find atypical features on a set of limited information, which can be considered data with the availability of alternative solutions, i.e. those data that are located at the class boundaries in the classification of information. Atypical data, zones between classes create difficulties in classifying information and constructing discriminant separation. A method for determining outliers of irrelevant traits based on grouping of class data using the minimum frame tree is proposed. Outliers are detected by minimizing the set of bipartite graph data of adjacent groups. This leads to spatial local demarcation and determination of the set of emission characteristics. The proposed method allows detecting outliers on different sets of features, both common and on the features of individual classes into which the information is divided.*

*Thus, the tree-like architecture of the computational process of trait research and formation of a set of traits, the change of which leads to compaction of atypical class data allows calculations in parallel and independently with parallelization at the stage of initial initialization class data of different classes.*

*The method of parallel detection of the set of atypical features of difficult to classify information is implemented. Atypical information is difficult to classify according to classification approaches. However, the analysis of these data is important from the point of view of data detection for studies with variable values of traits. The method allows determining the set of data that significantly affect the delimitation of atypical data.*

*Keywords: intelligent information technology, difficult to classify information, decision making system.*

### Постановка проблеми

Детектування нетипових даних та викидів є важливим та складним завданням у класифікації інформації. Інформація яка вважається складнокласифікованою в свою чергу характеризуються набором ознак, які визначають її інформативність. Тому пошук ознак, завдяки яким ознаки інформації вважаються нетиповими є актуальним та необхідним завданням. Складнішим завданням є пошук нетипових ознак на множині обмеженої інформації, якою можна вважати дані з наявністю альтернативних рішень, тобто ті дані, які розташовані на границях класів при класифікації інформації. Нетипові дані, зони між класами створюють складності при класифікації інформації та побудові дискримінантного розділення. Пропонується метод визначення викидів нерелевантних ознак базуючись на групуванні даних класу використовуючи мінімальне остове дерево. Викиди детектуються шляхом мінімізації множини дводольного графа даних суміжних груп. Це призводить до просторового локального розмежування та визначення множини викидів ознак. Запропонований метод дозволяє виявляти викиди на різних множинах ознак, як спільних, так і на ознаках окремих класів на які розмежовується інформація.

### Аналіз останніх джерел

Детектування викидів даних є досить складним завданням в комп'ютерному зорі [1], нетиповому поводженню [2]. Складність полягає в тому, що важко визначити та дати оцінку нетиповості та викидам. Нетиповість визначається у порівнянні з іншими даними тієї ж групи. При класифікації визначення нетиповості ускладнюється тим, що виділяється окрема група даних, які вважаються нетиповими та не повинні належати до жодного класу. З цією метою розробляється методи групового визначення та класифікації із застосуванням ансамблів [3–5]. Використання ансамблю моделей дозволяє компенсувати неточності та більш об'єктивно підходити до отримання інформації про дані [6]. Вибір особливостей з використанням ансамблів моделей є перспективним напрямом [6, 7]. Поряд з тим розробляються методи використання проєкцій ментальних моделей на машинний рівень для доповнення ансамблів [4, 8].

Одноразова класифікація (ОКК) використовується для пошуку нових даних. Будується

класифікатор, який здатний розпізнавати, тобто детектувати нові дані серед множини об'єктів спостереження. Шляхом регулювання порога визначення детектор новизни можна перетворити в класифікатор об'єктів [9–11]. Також використовується ОКК для визначення викидів даних. Використовуються методи розділення типових даних від нетипових використовуючи техніку навчання без вчителя [12–14]. Також викиди можна детектувати методами візуальної аналітики [15, 16] та використовуючи методи групового детектування [17]. Однак ОКК можна вважати базовим підходом щодо застосування методів до розмічених та нерозмічених даних, який застосовується з пошуку нових даних та викидів.

Детектування викидів визначаються як для розмічених даних, так і для нерозмічених. Так можна визначити три категорії підходів [18]:

1. Контрольоване детектування викидів використовується для розмічених даних типового представлення та нетипового.

2. Напівконтрольоване детектування викидів використовуються тільки для нормальних даних. Використання для детектування викидів обумовлене тим, що розмічення типових даних є простішим завданням і порівнянні з нетиповими.

3. Неконтрольоване детектування викидів використовується для нерозмічених як типових, так і нетипових даних.

Таке розділення по категоріях є узагальнюючим та вказує на те, що детектування даних фактично загалом застосовується в усіх напрямках машинного навчання. Це вказує на важливість та необхідність розробки рішень, які б дозволили проводити ефективні методи детектування.

В сукупності деякі дослідження розділяють три основні напрямки, аномалії, викиди та новизну. Ці три елементи характеризуються низькою ймовірністю виникнення [20]. Таким чином розглядається нетиповість певних даних відносно основної сукупності. До певної міри існує відмінність у трактування цих понять та термінів в межах певних досліджень. Так новизна розглядається як нове надходження даних з позиції часу. Викид часто розглядають як шум даних, які необхідно видалити. Проте методи детектування таких даних фактично є одними і тими ж [20]. Тому можна використовувати методи виявлення зі спеціалізацією конкретних досліджень для таких даних.

**Метою роботи є:** розробка метода паралельного формування кортежів ознак даних сегменту складнокласифікованої інформації в умовах невизначеності прийняття рішень та аналізу впливу на розмежуваність інформації

#### Виклад основного матеріалу

Із загальної множини ознак  $X$  виділимо множину ознак, які позначимо  $\gamma$  - ознаки та формують множину  $X_\gamma \subset X$ .  $\gamma$  - ознаки не є характерними для загальної сукупності даних класу. Їхні вагові значення значною мірою впливають на просторове розташування даних в окремих проявах. Це призводить до віддалення даних, зниження щільності даних класу та зближення з даними та групами даних поза межами класу. Видалення цих даних сприяє зближенню даних класу та збільшенню їхньої щільності, а, відтак, подібності. Однією з головних характеристик є віддалення даних класу від даних, які не належать класу при видаленні або зміни вагових параметрів ознак.

Мінімізація дводольного графа нетипових даних  $\min_{\|X\|_0} \|E(B)\|_{\|X\|_0}$  s.t.  $M \geq \|X\|_0 \geq k, k \neq 0$  здійснюється з метою детектування  $\gamma$  ознак, тобто ознак, наявність яких негативно впливає на групування даних класу. Мінімізація здійснюється шляхом ущільнення даних класу в зоні нетипових даних  $\max(Dens_{V_{Class}})_{local}$  та отримання множини  $X_\gamma = \{\gamma_i\}_{i=1}^r$ . Оскільки в загальному випадку існує множина локальних мінімумів при мінімізації множини ребер дводольного графа  $(B)_{\|X\|_0}$  шляхів *direct* по досягненню кожного окремого мінімуму  $Local = \{local\}_{i=1}^n, n \in \square$  може існувати декілька  $Direct_{\min\|E(B)\|} = \{direct\}_{i=1}^n, n \in \square$ . В загальному випадку можна записати  $Direct \times Local$  s.t.  $Local = \min\|E(B)_{X_\gamma}\|$ . Певна множина шляхів досягнення локального мінімуму формується шляхом детектування  $\gamma$  ознак.

Зміна співвідношень взаємного положення даних в гіперпросторі ознак при зміні ознак  $a_i, i \in [1..m], m \in X$  визначається окремим станом при кожному шляху *direct* досягнення мінімуму  $\min\|E(B)\|$ . Відповідно шлях *direct* характеризується множиною змін ознак  $direct \rightarrow f\{a\}$ . Проте на множині *Direct* при досягненні локального мінімуму  $direct_i|_{\min\|E(B)\|} \neq direct_j|_{\min\|E(B)\|}, i \neq j \in |Direct|$

при збігові множин змін ознак  $\{a\}_{direct_i} \equiv \{a\}_{direct_j}$  за тих же умов. Таким чином зміна ознак  $\{a\}$  призводить до дискретизації станів простору даних  $(V \subset R^{CS}, dist)$  за характерним взаємним розташуванням даних. Відповідно множину шляхів представлено у вигляді множини векторів, а саме послідовних кортежів зміни ознак  $L \equiv \overline{Direct}$ . В загальному випадку траєкторії змін ознак було представлено у вигляді частково впорядкованої множини  $\langle L, \succ \rangle$ . Кожна траєкторія є унікальним вектором лінійно впорядкованим, тому в межах узагальнення траєкторії змін ознак набуде вигляду строго впорядкованої множини  $\langle L, \succ \rangle$ .

Для отримання траєкторій  $\langle L, \succ \rangle$  з пошуку множини мінімумів  $min \|E(B)\|$  необхідно виконати наступні кроки методу формування множини  $\gamma$  ознак  $X_\gamma \subset X$ .

- сформуванню початкову множину ознак, яка в загальному випадку визначається як  $XR_0 = X$ .

Необхідність формування початкової множини ознак зумовлена проекцією на множину дослідження.

Залежно від проекції в межах однокласової класифікації вибирається один з варіантів:

1. Область дослідження визначається усією множиною ознак  $X$  нетипових даних  $D_{Trans}$ .

$XR_0 = X$ .  $X = X|_{D_{Trans}}$  оскільки множина ознак визначають в межах множини нетипових даних.

2. Множина ознак формується на основі даних, які є підмножиною нетипових даних  $D_{Trans}$  і належать до даних класу  $D_{Class}$ ,  $D_{Class} \subset D_{Trans}$ . В такому випадку  $XR_0 = X|_{D_{Class}}$ .

3. Множина ознак формується на основі даних, які є підмножиною нетипових даних  $D_{Trans}$  та належать до даних поза межами класу  $D_{Non-class}$ ,  $D_{Non-class} \subset D_{Trans}$ . В такому випадку  $XR_0 = X|_{D_{Non-class}}$ .

Крім формування початкової множини ознак дослідження на основі даних за умовою відношення даних до класу, початкову множину можна сформуванню за відношенням до даних.

1. Множина ознак формується на основі ознак, які належать тільки нетиповим даним

$XR_0 = \sigma_{x:\forall x \in X|_D} (X|_{D_{Trans}})$ ,  $D_{Trans} \subset D$ ,  $X|_{D_{Trans}} \setminus X|_D \neq \emptyset$ .

2. Множина ознак формується на основі ознак, які належать тільки нетиповим даним та відносяться до даних класу  $D_{Class}$   $XR_0 = \sigma_{x:\forall x \in X|_D \wedge \forall x \in X|_{D_{Class}}} (X|_{D_{Trans}})$ ,  $D_{Trans} \subset D$ ,  $X|_{D_{Trans}} \setminus X|_D \neq \emptyset$ .

3. Множина ознак формується на основі ознак, які належать тільки нетиповим даним та відносяться до даних поза межами класу  $D_{Non-class}$   $XR_0 = \sigma_{x:\forall x \in X|_D \wedge \forall x \in X|_{D_{Non-class}}} (X|_{D_{Trans}})$ ,  $D_{Trans} \subset D$ ,  $X|_{D_{Trans}} \setminus X|_D \neq \emptyset$ .

Подальшим розвитком функції відбору множини ознак дослідження є формування початкової множини на основі відношення до даних.

1. Множина ознак формується на основі ознак, які належать тільки нетиповим даним та відносяться до даних класу  $D_{Class}$  та не відносяться до даних поза межами класу  $D_{NonClass}$

$XR_0 = \sigma_{x:\forall x \in X|_D \wedge \forall x \in X|_{D_{Class}} \wedge \forall x \in X|_{D_{Non-class}}} (X|_{D_{Trans}})$ ,  $D_{Trans} \subset D$ ,  $X|_{D_{Trans}} \setminus X|_D \neq \emptyset$ .

2. Множина ознак формується на основі ознак, які належать тільки нетиповим даним та відносяться до даних поза межами класу  $D_{Non-class}$  та не відносяться до даних класу  $D_{Class}$

$XR_0 = \sigma_{x:\forall x \in X|_D \wedge \forall x \in X|_{D_{Class}} \wedge \forall x \in X|_{D_{Non-class}}} (X|_{D_{Trans}})$ ,  $D_{Trans} \subset D$ ,  $X|_{D_{Trans}} \setminus X|_D \neq \emptyset$ .

3. Множина ознак формується на основі ознак, які належать тільки нетиповим даним та відносяться до даних поза межами класу  $D_{Non-class}$  та відносяться до даних класу  $D_{Class}$ , тобто виключно на спільних

ознаках  $XR_0 = \sigma_{x:\forall x \in X|_D \wedge \forall x \in X|_{D_{Class}} \wedge \forall x \in X|_{D_{Non-class}}} (X|_{D_{Trans}})$ ,  $D_{Trans} \subset D$ ,  $X|_{D_{Trans}} \setminus X|_D \neq \emptyset$ .

- на основі вибраної початкової множини ознак, що визначається областю дослідження, обчислюється матриця евклідових відстаней  $Dist_0$  нетипових даних  $D_{Trans}$ .

- побудова мінімального остового дерева  $MST_0$  на обчислені матриці відстаней  $Dist_0$ .

- виділення множини ребер  $E(B)$  дводольного графа  $B_0$  з вершинами підмножини вершин мінімального остового дерева  $V_B \subset V_{MST}$ , де ребра дводольного графа визначаються інцидентними вершинами, як належать до множини даних класу та множини даних поза межами класу  $\{e_{u,v} \mid \forall u \in V_{Class}, \forall v \in V_{Non-class}\}$ .

Послідовні чотири етапи обчислень є базовим блоком перетворень вхідних даних

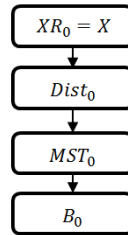


Рис. 1. Послідовна структура перетворення вхідних даних ознак області дослідження

Перетворення вхідних даних, які визначають область дослідження ознак виділяє необхідну інформативність даних, що проявляється у побудові дводольного графа в межах концепції однокласової класифікації.

- наступним кроком є виділення ознаки дослідження  $X$  з ініціалізованої множини ознак  $XR_0$  та формування нового простору ознак  $XR_{i,j} = XR_{i-1} \setminus \{x_j\}$ ,  $i = 0..|L|$ ,  $\forall l \in L$ ,  $j = 1..|XR_{i-1}|$ . На цьому етапі кількість досліджуваних ознак визначається множиною  $XR_{i-1}$ , тобто множиною ознак попереднього етапу дослідження формування ознак. Кожна ознака досліджується окремим етапом незалежно від інших ознак, що може виконуватись у паралельному режимі. Кількість паралельних потоків  $|XR_{i-1}| - 1$ .

- обрахунок блока послідовної структури, який в загальному представлений таким чином:  $Dist(XR) \rightarrow MST \rightarrow B \rightarrow E(B)$ .

- рішення  $TF$  про належність ознаки  $x_{i,j}$  до множини  $\gamma$  ознак визначається співвідношенням

$$TF : |E(B_{i-1})| - |E(B_i)| > 0 \quad (1)$$

- наступним кроком є формування кортежу ознак послідовного збільшення компактності класу  $\langle TR_{i,j} \rangle = \langle TR_{i-1,j} \rangle \cup x_j$ .

- умовою завершення формування кортежу ознак є умова обернена до рішення  $TF$ , тобто до наявності ознаки формування кортежу

$$|E(B_{i-1})| - |E(B_i)| \leq 0 \quad (2)$$

Сформований кортеж відповідає траєкторії  $l_k \in L, k = 1..|L|$ ,  $l_k = \langle TR_{i,j} \rangle$  пошуку локального мінімуму  $\min \|E(B)\|$  наявності міжкласових зв'язків  $link = \{Class, Non-class\}$  на основі МОД.

- об'єднуються траєкторії за елементами кортежів у множину  $X_\gamma$  за унікальними ознаками:

$$X_\gamma = \sigma_{\exists l_k, \forall x \in XR_0} \left( \bigcup \langle TR \rangle \right) \quad (3)$$

Таким чином формується множина ознак  $X_\gamma$ , тобто ознак зміна яких призводить до просторового ущільнення нетипових даних шляхом зменшення міжкласових ребер мінімального остового дерева побудованого на цих даних. Загальний граф процесу синтезу траєкторій зміни ознак представлений на рисунку 2.

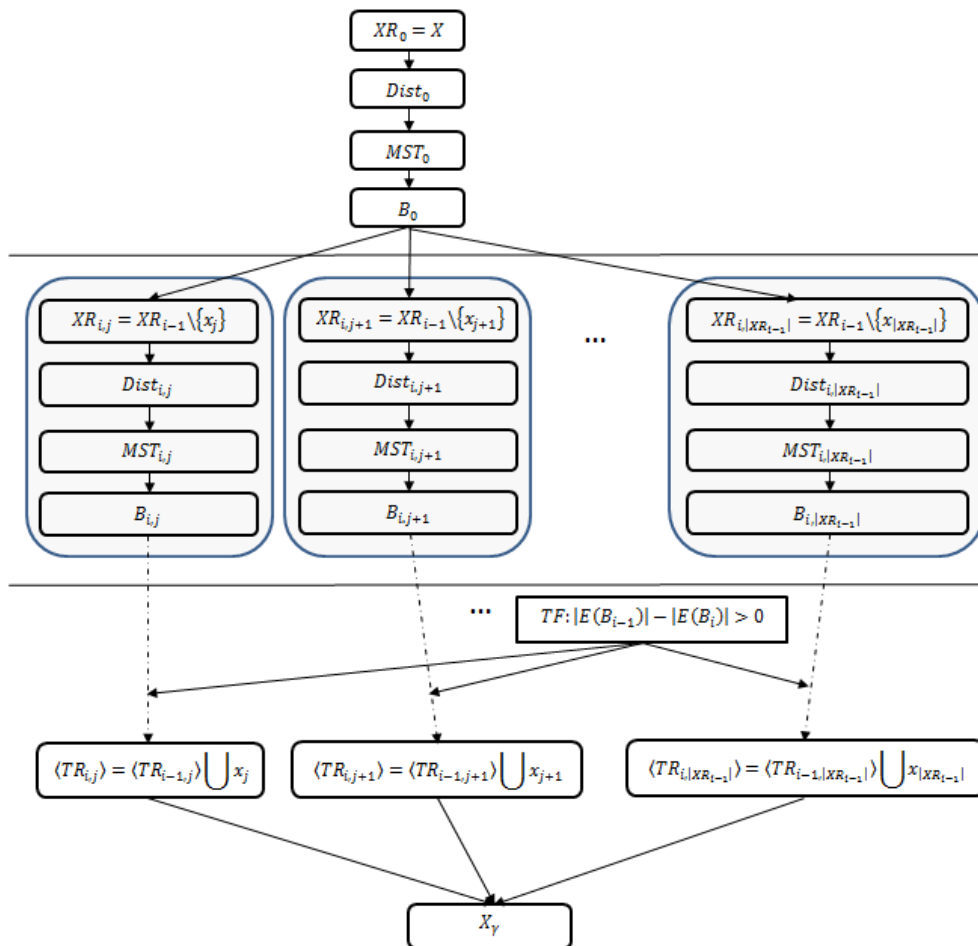


Рис. 2. Граф процесу синтезу траєкторій зміни ознак у паралельній формі

Обчислювальний блок послідовної структури з обробки даних, а саме з множини вихідних ознак  $XR_0$  та отримання множини ребер дводольного графа  $E(B)$  є однаковим при пошуку ознаки при послідовному застосуванні для побудови кортежу, так і при паралельному формуванні кортежів. Обчислювальний процес побудови траєкторій пошуку мінімального локального мінімуму множини міжкласових ребер даних є незалежним та може бути успішно розпаралеленим.

Кожна траєкторія є унікальним набором ознак у вигляді кортежу та отримується з деревовидної структури, яка будується на основі зменшення кількості міжкласових ребер  $\Delta E(B)$ . Вершини розгалуження траєкторії є точкам розпаралелювання процесу незалежного обчислення. В точці розгалуження визначається множина ознак дослідження на основі ініціалізованої множини ознак  $XR_0$  та формується за відрахуванням множини ознак з кортежу частини траєкторії до точки розгалуження  $\langle TR_{i-1,j} \rangle$

$$XR_{i,j} = XR_0 \setminus \langle TR_{i-1,j} \rangle \tag{4}$$

Необхідно зазначити, що зменшення множини ознак початкової ініціалізації  $XR_0$  в процесі пошуку локальних мінімумів є незначним в наслідок формування множини ознак при побудові траєкторій з невеликої кількості ознак при досягненні умови завершення траєкторії  $|E(B_{i-1})| - |E(B_i)| \leq 0$ ,

$$|XR_0| \square |\langle TR \rangle|, \tag{5}$$

тому зменшення обчислювальної складності з рахунок зменшення множини досліджуваних ознак в загальному об'ємі обчислень є несуттєвим.

Оціночні дослідження з довжин траєкторій показали, що загалом траєкторії є порівняно недовгими ланцюгами даних у порівнянні з множиною початкової ініціалізації, які складають незначну кількість від загальних ознак нетипових даних. Відповідно в такі мірі і зменшуються обчислення на останньому етапі

порівняно з першим при досягненні локального мінімуму, що є несуттєвим. Таким чином кількість паралельних процесів на початковому етапі визначається множиною початкової ініціалізації  $XR_0$ . Загальна кількість обчислень послідовної структури можна визначити таким чином:

$$|XR_0|(|X_y|+1) \quad (6)$$

Без врахування зменшення початкової множини ознак за рахунок вершин траєкторій, початкова множина ознак  $XR_0$  досліджується на кожній вершині траєкторій, яка є вершиною розгалуження дерева.

Кардинальне число множини ознак визначається загальною кількістю ознак, які формують множину  $X_\gamma$  зміна яких призводить до покращення розмежування нетипових даних класів.

$$|X_\gamma| \quad (7)$$

Характеристика визначається відносно множини ознак дослідження. Тобто на множині ознак нетипових даних, які належать до класу  $|X_\gamma|_{Class}$ , знаходяться поза межами класу  $|X_\gamma|_{NonClass}$  та ознак, які є спільними для даних класу та даних поза межами класу  $|X_\gamma|_{Class-NonClass}$ .

### Висновки

Таким чином, деревоподібна архітектура обчислювального процесу дослідження ознак та формування множини ознак, зміна яких призводить до ущільнення групування нетипових даних класу дозволяє проводити обчислення паралельно та незалежно з розпаралелюванням на етапі початкової ініціалізації множини ознак дослідження та на вершинах траєкторій пошуку локальних мінімумів множини ребер з'єднання класових даних різних класів.

Реалізований метод паралельного детектування множини нетипових ознак приграничних складнокласифікованих даних. Нетипові дані є складнокласифіковані за підходами класифікації. Однак аналіз цих даних є важливим з точки зору детектування даних для досліджень із змінними значеннями ознак. Метод дозволяє визначити множину даних, які значною мірою впливають на розмежування нетипових даних.

### Література

1. Zaheer M. Z. Cleaning Label Noise with Clusters for Minimally Supervised Anomaly Detection / M. Z. Zaheer, J. Lee, M. Astrid, A. Mahmood, S.-I. Lee // arXiv:2104.14770 [cs]. — 2021.
2. Sultani W. Real-world Anomaly Detection in Surveillance Videos / W. Sultani, C. Chen, M. Shah // arXiv:1801.04264 [cs]. — 2019.
3. Barmak O. V. Characteristics for choice of models in the ansables classification / O. V. Barmak, Y. V. Krak, E. A. Manziuk // Problems in programming, ISSN 1727-4907. — 2018. — № 2–3. — Pp. 171–179. doi: 10.15407/pp2018.02.171.
4. Krak I. Data Classification Based on the Features Reduction and Piecewise Linear Separation / I. Krak, O. Barmak, E. Manziuk, A. Kulas // Intelligent Computing and Optimization. ICO'2019: Intelligent Computing and Optimization. Advances in Intelligent Systems and Computing. Cham. Intelligent Computing and Optimization. — 2020. — Vol. 1072. — Pp. 282–289. doi: 10.1007/978-3-030-33585-4\_28.
5. Manziuk E. A. Approach to creating an ensemble on a hierarchy of clusters using model decisions correlation / E. A. Manziuk, W. Wójcik, O. V. Barmak, I. V. Krak, A. I. Kulas, V. A. Drabovska, V. M. Puhach, S. Sundetov, A. Mussabekova // Przegląd Elektrotechniczny, ISSN 0033-2097. — 2020. — Vol. 96, № 9. — Pp. 108–113. doi: 10.15199/48.2020.09.23.
6. Zvarevashe K. Ensemble Learning of Hybrid Acoustic Features for Speech Emotion Recognition / K. Zvarevashe, O. Olugbara // Algorithms. — 2020. — Vol. 13, № 3. — Pp. 70. doi: 10.3390/a13030070.
7. Alexandropoulos S.-A. N. A new ensemble method for outlier identification / S.-A. N. Alexandropoulos, S. B. Kotsiantis, V. E. Piperigou, M. N. Vrahatis // 2020 10th International Conference on Cloud Computing, Data Science Engineering (Confluence). — 2020. — Pp. 769–774. doi: 10.1109/Confluence47617.2020.9058219.
8. Krak I. Approach to Piecewise-Linear Classification in a Multi-dimensional Space of Features Based on Plane Visualization / I. Krak, O. Barmak, E. Manziuk, H. Kudin // Lecture Notes in Computational Intelligence and Decision Making. ISDMCI'2019: Advances in Intelligent Systems and Computing. Advances in Intelligent Systems and Computing. Cham. Lecture Notes in Computational Intelligence and Decision Making. — 2020. — Vol. 1020. — Pp. 35–47. doi: 10.1007/978-3-030-26474-1\_3.
9. Oosterlinck D. From one-class to two-class classification by incorporating expert knowledge: Novelty detection in human behaviour / D. Oosterlinck, D. F. Benoit, P. Baecke // European Journal of Operational Research. — 2020. — Vol. 282, № 3. — Pp. 1011–1024. doi: 10.1016/j.ejor.2019.10.015.
10. Abati D. Latent Space Autoregression for Novelty Detection / D. Abati, A. Porrello, S. Calderara, R. Cucchiara // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019. — Pp. 481–490.
11. Rodríguez-Ruiz J. A one-class classification approach for bot detection on Twitter / J. Rodríguez-Ruiz, J. I. Mata-Sánchez, R. Monroy, O. Loyola-González, A. López-Cuevas // Computers & Security. — 2020. — Vol. 91. — Pp. 101715. doi: 10.1016/j.cose.2020.101715.
12. Gong D. Memorizing Normality to Detect Anomaly: Memory-Augmented Deep Autoencoder for Unsupervised Anomaly Detection / D. Gong, L. Liu, V. Le, B. Saha, M. R. Mansour, S. Venkatesh, A. van den Hengel // Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019. — Pp. 1705–1714.
13. You C. Provable Self-Representation Based Outlier Detection in a Union of Subspaces / C. You, D. P. Robinson, R. Vidal // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017. — Pp. 3395–3404.

14. Lai C.-H. Robust Subspace Recovery Layer for Unsupervised Anomaly Detection / C.-H. Lai, D. Zou, G. Lerman // arXiv:1904.00152 [cs, stat]. — 2019.
15. Barmak O. Using piecewise hyper linear classification in multidimensional feature space for text content / O. Barmak, E. Manziuk, I. Krak // 2019 IEEE 14th International Conference on Computer Sciences and Information Technologies (CSIT). — 2019. — Vol. 2. — Pp. 119–123. doi: 10.1109/STC-CSIT.2019.8929798.
16. Barmak O. Classification Based Hierarchical Clustering Prediction Variability in the Ensembles of Models Using a Statistical Approach / O. Barmak, E. Manziuk, I. Krak // 2020 IEEE 15th International Conference on Computer Sciences and Information Technologies (CSIT). — 2020. — Vol. 1. — Pp. 11–14. doi: 10.1109/CSIT49958.2020.9322019.
17. Barmak O. Diversity as The Basis for Effective Clustering-Based Classification / O. Barmak, I. Krak, E. Manziuk // CEUR Workshop Proceedings. ICST'2020: 9th International Conference "Information Control Systems & Technologies. Odessa, Ukraine. CEUR Workshop Proceedings. — 2020. — Vol. 2711. — Pp. 53–67.
18. Cheng Z. Unsupervised Outlier Detection via Transformation Invariant Autoencoder / Z. Cheng, E. Zhu, S. Wang, P. Zhang, W. Li // IEEE Access. — 2021. — Vol. 9. — Pp. 43991–44002. doi: 10.1109/ACCESS.2021.3065838.

E. A. МАНЗІОК

ORCID ID: 0000-0002-7310-2126

[eduard.em.km@gmail.com](mailto:eduard.em.km@gmail.com)

Рецензія/Peer review : 24.05.2021 р.

Надрукована/Printed :30.06.2021 р.