

Вінницький національний технічний університет

УДОСКОНАЛЕНИЙ МЕТОД ФОРЕЛ ДЛЯ КЛАСТЕРИЗАЦІЇ СТАНІВ КОМП'ЮТЕРНОЇ ТЕХНІКИ

Вступ. Сучасний стан розвитку суспільства характеризується різким зростанням обсягів інформації у різних галузях, що зумовило необхідність широкого застосування комп'ютерної техніки (КТ) та, відповідно, спричинило зростання попиту на неї.

У зв'язку із зростанням кількості та різноманіття комп'ютерної техніки зростає і складність задач прийняття рішень щодо експлуатації такої техніки при виникненні несправностей в процесі її роботи, оскільки кожен новий її вид має свої особливості функціонування, що відображається в її станах. Тому доцільною є розробка і удосконалення методів та засобів кластеризації станів комп'ютерної техніки з метою формування в подальшому обґрунтованих рекомендацій щодо забезпечення коректного функціонування такої техніки [1].

Постановка задачі. Множину станів КТ X необхідно розбити на k непересічних підмножини, які називаються таксонами (кластерами), таким чином, щоб кожен з них складався із станів, при яких формуються подібні рекомендації щодо подальших дій при їх виникненні.

Результати роботи. Розробка удосконаленого методу ФОРЕЛ для кластеризації станів комп'ютерної техніки. Серед множини методів кластеризації для аналізу станів КТ доцільно застосувати метод ФОРЕЛ, який потрібно вдосконалити для усунення недоліків існуючих його модифікацій та адаптації до відповідної предметної області.

До особливостей відомого методу кластеризації станів КТ ФОРЕЛ [2-5] відносять такі:

- його продуктивність є невисокою;
- метод є збіжним за скінченне число кроків;
- в лінійному просторі центром кластера може виступати як будь-яка точка, так і стан КТ;
- на першому етапі обирається один із станів КТ як початковий об'єкт, від якого проводитиметься кластеризація, що, в свою чергу, впливатиме на якість розбиття;
- наявність апріорних знань про діаметри кластерів;
- відомий метод не враховує особливостей станів КТ, що негативно впливає на якість їхньої кластеризації.

Метод ФОРЕЛ та його відомі модифікації передбачають: необхідність задання радіуса кластера R ; можливість віднесення до кластера об'єктів з інших таксонів через неправильний вибір радіуса, що негативно впливає на достовірність декомпозиції множини станів КТ, оскільки користувач не має інформації про розміри кластерів. У зв'язку з цим запропоновано удосконалення методу, що дозволяє підвищити якість розбиття множини станів КТ за рахунок аналізу отриманих кластерів і виявлення станів, які є викидами, а також визначення радіуса кластера R , що враховує допустимі відхилення значень параметрів та характеристик станів КТ і показника якості кластеризації η , яка б задовольнила користувача

$$R = \frac{(MAX - MIN) \times (100 - \eta)}{100} + MIN - \Delta, \quad (1)$$

де MAX , MIN – максимальна та мінімальна відстані між двома станами КТ у множині;

η – показник якості кластеризації $0 \leq \eta \leq 100$;

Δ – значення допустимого відхилення відстані між станами КТ.

Після визначення радіуса центр кластера поміщається в будь-який з некластеризованих станів КТ та відносяться до таксона стани, відстань до яких від центра менша за R . Після чого визначається новий центр (новим центром кластера стає центр мас знайденого таксона) та відбувається повторне формування кластера, до якого відносяться стани КТ, відстань від знайденого центра до яких менша за радіус R . Такі дії виконуються поки центр мас не стабілізується.

Для унеможливлення появи викидів серед станів всередині кластерів після формування кожного з них проводиться відповідний аналіз та віднесення їх до множини некластеризованих.

Стани КТ, які належать новому таксону, вилучаються з некластеризованих, а робота методу повторюється, поки множина некластеризованих станів не буде пустою.

У формалізованому вигляді алгоритм кластеризації станів КТ ФОРЕЛ (рис.1), в основу якого покладено запропонований удосконалений метод, містить такі основні кроки.

1. Формування множини некластеризованих станів КТ

$$U = \{X_1, X_2, \dots, X_n\}, \quad (2)$$

де U – множина некластеризованих станів КТ;

n – кількість станів КТ, що підлягають кластеризації.

2. Знаходження значень відстаней між некластеризованими точками:

а) мінімальної

$$MIN = \min_{i,j=1,\dots,n,i \neq j} a_{3E}(X_i, X_j), \quad (3)$$

де $a_{3E}(X_i, X_j)$ – відстань між станами X_i та X_j ;

б) максимальної

$$MAX = \max_{i,j=1,\dots,n,i \neq j} a_{3E}(X_i, X_j) \quad (4)$$

3. Визначення допустимого відхилення відстані Δ , з використанням зваженої евклідової відстані.

4. Знаходження значення радіуса кластера з використанням формули (1).

5. За умови $U \neq \emptyset$ (у вибірці є некластеризовані стани КТ):

5.1. Обрати довільний стан $X_0 \in U$ випадковим чином.

5.2. Сформувати кластер станів КТ – сферу з центром в X_0 і радіусом R

$$K_0 = \{X_i \in U \mid a_{3E}(X_i, X_0) \leq R\}, \quad (5)$$

де K_0 – сформований кластер станів КТ;

$a_{3E}(X_i, X_0)$ – відстань від стану КТ X_i до центра кластера X_0 .

5.3. Помістити центр кластера в його центр мас

$$X_0 = \operatorname{argmin}_{X_i \in K_0} \sum_{X_j \in K_0} a_{3E}(X_i, X_j) \quad (6)$$

де X_0 – центр мас кластера.

5.4. Виконувати пункти 5.1–5.3, поки центр X_0 не стабілізується.

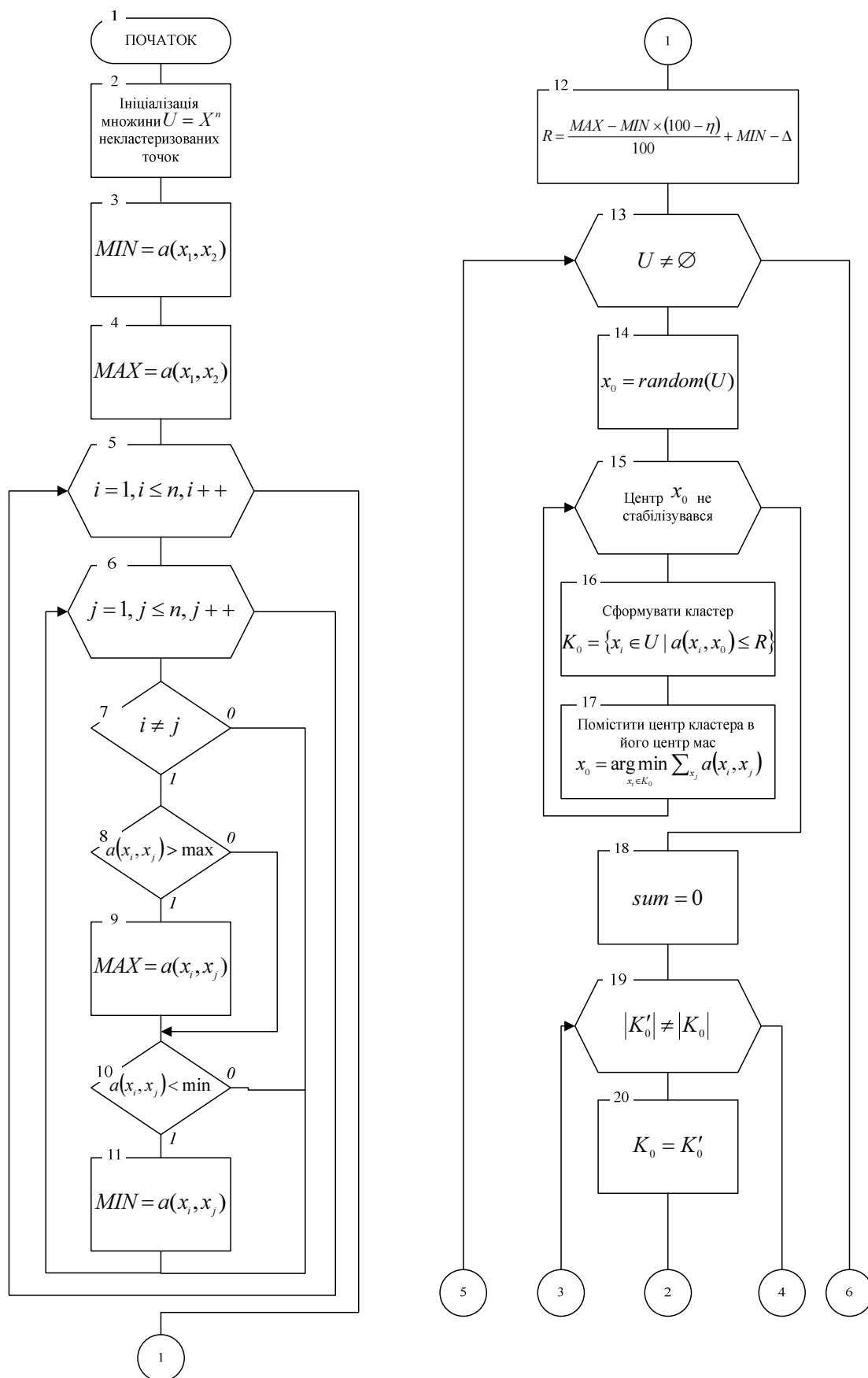


Рисунок 1 – Схема удосконаленого алгоритму кластеризації станів КТ ФОРЕЛ

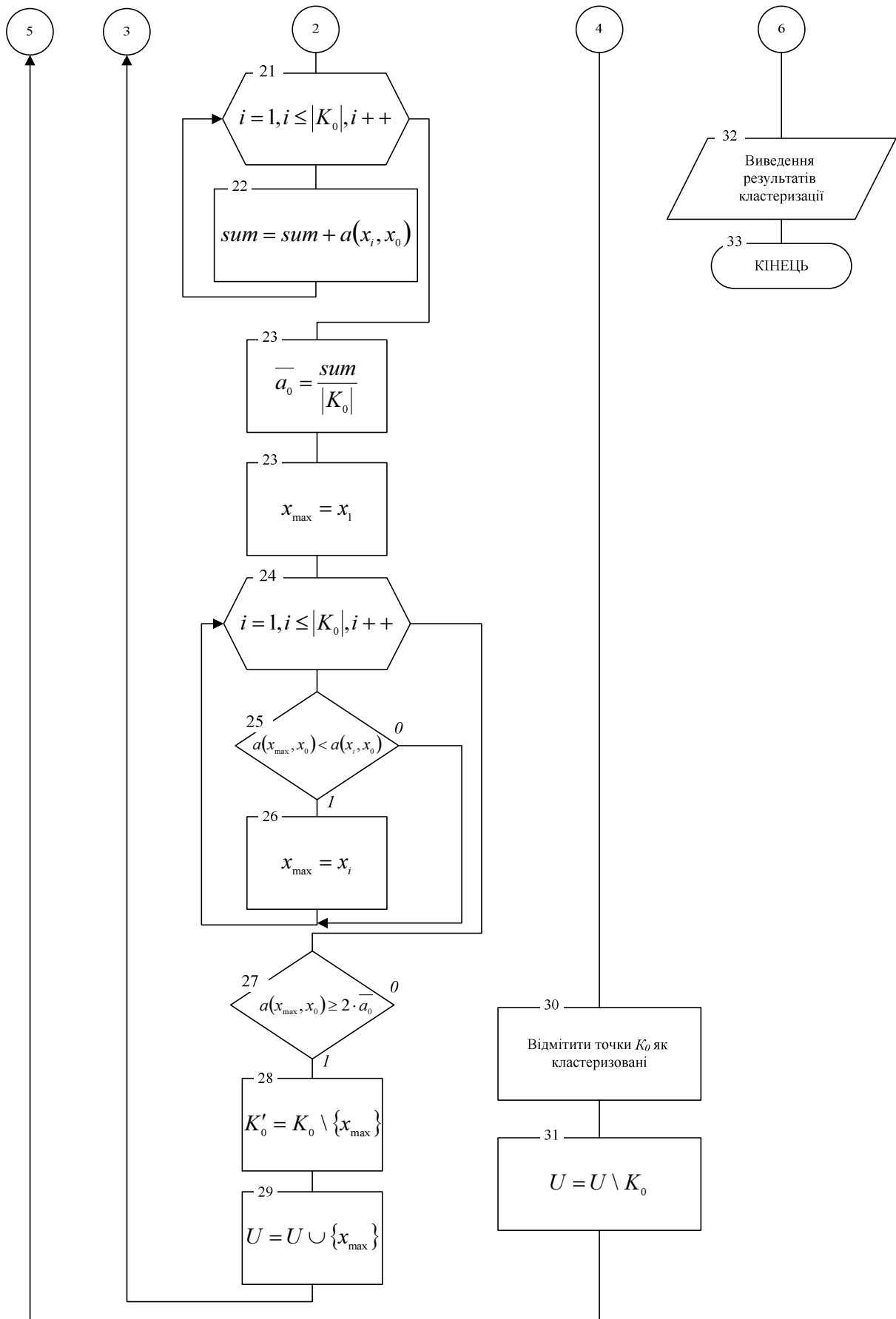


Рисунок 1, аркуш 2

5.5. Знайти

$$\bar{a}_0 = \frac{\sum_{X_0, X_i \in K_0} a_{zE}(X_i, X_0)}{|K_0|}, i = 1, \dots, |K_0|, \quad (7)$$

де \bar{a}_0 – середнє значення відстані між центром кластера та станами КТ, які йому належать.

5.6. Знайти

$$X_{\max} = \arg \max_{X_i \in K_0} a_{zE}(X_i, X_0), i = 1, \dots, |K_0|, \quad (8)$$

де X_{\max} – максимально віддалений від центра кластера стан.

5.7. Якщо $(a_{zE}(X_{\max}, X_0) \geq 2 \cdot \bar{a}_0)$, то

X_{\max} вилучити з K_0

$$K_0 = K_0 \setminus \{X_{\max}\} \quad (9)$$

та X_{\max} додати до U

$$U = U \cup \{X_{\max}\}. \quad (10)$$

5.8. Виконувати пункти 5.6 та 5.7, поки не буде вилучень X_{\max} з K_0 .

5.9. Вилучити стани, що віднесені до кластера K_0 (як кластеризовані)

$$U = U \setminus K_0. \quad (11)$$

6. Виконувати п. 5, поки $U = \emptyset$ (всі стани КТ кластеризовані).

Отже, удосконалений метод кластеризації станів КТ ФОРЕЛ розраховує радіуси кластерів і визначає стани, які є викидами, та вилучає їх з кластерів до множини некластеризованих станів, що дозволило зменшити кількість викидів в кластерах та підвищити якість кластеризації. Також на основі удосконаленого методу розроблено удосконалений алгоритм ФОРЕЛ.

Висновки. Таким чином, удосконалено метод кластеризації станів комп'ютерної техніки ФОРЕЛ, який передбачає розрахунок радіусів кластерів з урахуванням особливостей предметної області, визначає стани, що є викидами та вилучає їх з кластерів. Це дозволило підвищити якість кластеризації таких станів на 6%.

ЛІТЕРАТУРА

1. Перспективы развития компьютерных систем [Электронный ресурс] – режим доступа: <http://www.itstan.ru/komp-tehnika/perspektivy-razvitija-kompjuternyh-sistem.html>.
2. Загоруйко Н.Г. Прикладные методы анализа данных и знаний / Н.Г.Загоруйко. – Новосибирск: ИМ СО РАН, 1999. – 270с. – ISBN 5-86134-060-9.
3. Загоруйко Н.Г. Алгоритмы обнаружения эмпирических закономерностей / Н.Г.Загоруйко, В.Н.Ёлкина, Г.С.Лбов. – Новосибирск: Наука, 1985. – 112с. – ISBN 5-1601212-А.
4. Interpretation of FOREL clusters [Electronic resource]. – Access mode: <http://www.biomedcentral.com/content/supplementary/1471-2164-7-318-s1.pdf>.
5. Крюков А.В. Кластерный анализ режимов систем тягового электроснабжения для целей ситуационного управления [Электронный ресурс] / А.В.Крюков, В.П.Закарюкин, Н.А.Абрамов // Винеровские чтения: 4-й Всерос. конф.: материалы конф. – Режим доступа: [http://berestneva.am.tpu.ru/Papers/KONF2009/%F7%C9%CE%C5%D2%CF%D7%D3%CB%C9%C5%20%DE%D4%C5%CE%C9%D1/2009%20\(F\)/fscommand/pdf/042.pdf](http://berestneva.am.tpu.ru/Papers/KONF2009/%F7%C9%CE%C5%D2%CF%D7%D3%CB%C9%C5%20%DE%D4%C5%CE%C9%D1/2009%20(F)/fscommand/pdf/042.pdf).

Надійшла до редколегії 10.03.2015.