

Міністерство освіти і науки України
Вінницький національний технічний університет

ПЕТРИШИН СЕРГІЙ ІВАНОВИЧ



УДК 004.89

**ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ КЛАСТЕРИЗАЦІЇ СТАНІВ
КОМП'ЮТЕРНОЇ ТЕХНІКИ**

Спеціальність 05.13.06 – інформаційні технології

Автореферат
дисертації на здобуття наукового ступеня
кандидата технічних наук

Вінниця – 2016

Дисертацією є рукопис.

Роботу виконано у Вінницькому національному технічному університеті Міністерства освіти і науки України.

Науковий керівник: кандидат технічних наук, доцент
Савчук Тамара Олександрівна,
Вінницький національний технічний університет,
професор кафедри комп'ютерних наук.

Офіційні опоненти: доктор технічних наук, професор
Рудницький Володимир Миколайович,
Черкаський державний технологічний університет,
завідувач кафедри інформаційної безпеки та комп'ютерної інженерії

доктор технічних наук, старший науковий співробітник
Порєв Геннадій Володимирович,
Київський національний університет імені Тараса Шевченка,
доцент кафедри інтелектуальних та інформаційних систем.

Захист відбудеться «03» березня 2016 р. о 14³⁰ годині на засіданні спеціалізованої вченої ради Д 05.052.01 у Вінницькому національному технічному університеті за адресою: 21021, м. Вінниця, вул. Хмельницьке шосе, 95, ГНК, ауд. 210.

З дисертацією можна ознайомитися в бібліотеці Вінницького національного технічного університету за адресою: 21021, м. Вінниця, вул. Хмельницьке шосе, 95, ГНК.

Автореферат розісланий «02» лютого 2016 р.

Учений секретар
спеціалізованої вченої ради



С. М. Захарченко

ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

Актуальність теми. Швидкий науково-технічний прогрес сприяв значному збільшенню різновидів комп'ютерної техніки (КТ), яка на сьогоднішній день використовується і в побуті, і на виробництві. Це стало причиною зростання ймовірності виникнення ситуацій, що характеризуються порушенням виконання її основних функцій, або, взагалі, виходом із ладу такої техніки. У зв'язку з цим постає необхідність у розробці різних видів документацій для КТ, зокрема і інструкцій з її обслуговування, які будуть використовуватись в процесі ремонту та інших видів сервісного обслуговування такої техніки з метою підвищення його швидкості та якості. Означена супровідна документація може розроблятися як її виробником, так і сервісними підприємствами та компаніями. Особливістю етапу розробки інструкцій з обслуговування КТ є те, що розробнику наперед невідомо стани, а також їх класи, що визначає доцільність застосування технологій аналізу даних, зокрема кластеризації, для розв'язання задачі групування подібних станів КТ в таксони. До них будуть входити однотипні стани КТ, а тому дії, які необхідно виконувати для однокластерних станів, будуть однотипними. Отже, розробник інструкцій з обслуговування буде отримувати варіанти групувань станів КТ з рекомендаціями щодо подальших дій при кожному з них.

Сучасні методи аналізу даних та кластеризації розглянуті в роботах Григорія П'ятецького-Шапіро, Дж. Мак-Кіна, Г. Болла, Д. Холла, Г. Ланса, У. Уільямсона, Н. Джардайна, С. А. Айвазяна, І. Д. Манделя, А. А. Барсегяна, Д. С. Хайдукова, В. С. Берикова, Г. С. Лбова, Б. Г. Міркіна, Міхаеля Бері, питання моделювання та визначення якості декомпозиції представлено у роботах Р. Н. Кветного, В. М. Дубового, О. В. Бісікала, І. В. Кузьміна, Б. І. Мокіна, В. Б. Мокіна, С. В. Юхимчука, М. Д. Кацмана, С. Д. Штовби, А. І. Поворознюка, О. А. Павлова, В. Н. Вапніка, Б. Дюрана, М. Г. Загоруйка, О. Я. Червоненкіса, В. Є. Снитюка, В. О. Доровського, О. П. Ротштейна та інших провідних вчених.

Проте в процесі кластеризації заздалегідь невідомих станів складних об'єктів виникають ситуації, коли відхилення значень параметрів та характеристик станів знаходяться у допустимих межах, що не потребує віднесення їх до іншого кластеру. Існуючі методи кластеризації і аналізу даних недостатньо якісно працюють у таких умовах.

Отже, актуальним є підвищення якості розбиття множини станів комп'ютерної техніки за рахунок розробки та впровадження моделей, методів та відповідних інформаційних технологій, які будуть використовуватись для декомпозиції вибірки таких станів, що є основою для подальшого прийняття ефективних рішень щодо експлуатації такої техніки.

Оскільки в процесі опрацювання даних про стани КТ накопичуються потужні обсяги інформації, для її аналізу доцільно використовувати технології Data Mining, серед яких особливої уваги заслуговують методи кластерного аналізу, такі як K-MEANS і ФОРЕЛ, за умови їх удосконалення з урахуванням особливостей предметної області, яка досліджується.

Зв'язок роботи з науковими програмами, планами, темами.

Дисертаційна робота виконана відповідно до напряму наукових досліджень кафедри комп'ютерних наук Вінницького національного технічного університету «Моделі, методи, технології та пристрої інтелектуальних інформаційних систем управління, економіки, навчання та комунікацій».

Результати дисертаційного дослідження увійшли у звіти з держбюджетної науково-дослідної роботи на тему «Теоретичні методи аналізу швидкоплинних техногенних надзвичайних ситуацій для створення систем підтримки прийняття рішень керівниками ліквідації таких ситуацій» № 47-Д-307 (номер державної реєстрації 0108U000659), де здобувач був у складі авторів проміжних та заключного звітів.

Мета і задачі дослідження. Метою дослідження є підвищення якості кластеризації станів комп'ютерної техніки шляхом розробки і впровадження інформаційної технології для подальшого прийняття рішень щодо її експлуатації.

Для досягнення поставленої мети необхідно розв'язати такі задачі:

- дослідити стани комп'ютерної техніки, а також існуючі методи, технології та засоби аналізу даних про такі стани;
- розробити інформаційну модель процесу кластеризації станів комп'ютерної техніки;
- удосконалити метод кластеризації станів комп'ютерної техніки K-MEANS;
- удосконалити метод кластеризації станів комп'ютерної техніки ФОРЕЛ;
- розробити інформаційну технологію кластеризації станів комп'ютерної техніки;
- провести моделювання процесу кластеризації станів комп'ютерної техніки з використанням удосконалених методів K-MEANS і ФОРЕЛ з метою оцінювання якості розбиття;
- розробити програмне забезпечення для кластеризації станів комп'ютерної техніки.

Об'єктом дослідження є процес кластеризації станів комп'ютерної техніки.

Предметом дослідження є моделі, методи та інформаційні технології кластеризації станів комп'ютерної техніки.

Методи дослідження. Використано методи системного аналізу та синтезу для вивчення властивостей процесу кластеризації станів комп'ютерної техніки; методи інтелектуального аналізу даних для розробки інформаційної моделі кластеризації станів комп'ютерної техніки, та удосконалених методів кластеризації, які базуються на цій моделі; методи об'єктного моделювання процесу кластеризації станів комп'ютерної техніки; комп'ютерне моделювання для розробки програмного забезпечення та оцінювання якості кластеризації.

Наукова новизна одержаних результатів полягає в тому, що:

– вперше запропоновано інформаційну модель процесу кластеризації станів комп'ютерної техніки, що, на відміну від існуючих, базується на моделі станів комп'ютерної техніки; враховує вагові коефіцієнти характеристик та параметрів таких станів в процесі визначення відстані між ними; використовує нормування значень цих параметрів та характеристик, що дозволило підвищити якість кластеризації;

– удосконалено метод кластеризації K-MEANS, який, на відміну від існуючих, передбачає врахування допустимих відхилень значень параметрів та характеристик комп'ютерної техніки при формуванні кластерів її станів, що дозволило підвищити якість кластеризації в умовах відомої кількості кластерів;

– удосконалено метод кластеризації ФОРЕЛ, який, на відміну від існуючих, передбачає розрахунок радіусів кластерів станів комп'ютерної техніки для зменшення кількості викидів у них, що дозволило підвищити якість розбиття множини таких станів в умовах відсутності відомостей про кількість кластерів;

– вперше розроблено інформаційну технологію кластеризації станів комп'ютерної техніки, в якій, на відміну від існуючих, використовуються інформаційна модель процесу кластеризації та удосконалені методи кластеризації K-MEANS і ФОРЕЛ, що дозволило підвищити якість розбиття множини таких станів.

Практичне значення одержаних результатів полягає в тому, що на основі запропонованого підходу до кластеризації станів комп'ютерної техніки отримано такі практичні результати:

– розроблено удосконалений алгоритм кластеризації станів комп'ютерної техніки K-MEANS;

– розроблено удосконалений алгоритм кластеризації станів комп'ютерної техніки ФОРЕЛ;

– розроблено програмне забезпечення для кластеризації станів комп'ютерної техніки.

Отримані на основі наукових досліджень практичні результати впроваджено в компанію LLC «ІКГок» у вигляді програмного забезпечення для кластеризації станів комп'ютерної техніки (акт впровадження №47/3 від 06.06.2014 р.), що дало можливість приймати ефективні управлінські рішення щодо її працездатності. Результати роботи впроваджено в департаменті енергетики, транспорту та зв'язку Вінницької міської ради як методіку аналізу даних про електроспоживання підприємств та установ міста, що базується на технологіях Data Mining, та інформаційну технологію аналізу даних про електроспоживання підприємств та установ міста, що базується на удосконаленому алгоритмі кластерного аналізу (акт впровадження від 23.06.2014 р.). Також отримані наукові результати впроваджено в навчальний процес при фаховій підготовці за освітньо-професійною програмою підготовки бакалаврів за напрямом 6.050101 «Комп'ютерні науки» у вигляді лекцій і лабораторного практикуму з дисципліни «Інтелектуальний аналіз даних» (акт

впровадження від 20.01.2015 р.). Впровадження результатів дослідження підтверджено відповідними актами.

Особистий внесок здобувача. Усі теоретичні та експериментальні дослідження виконані автором самостійно. У друкованих працях, опублікованих у співавторстві, автору дисертації належать: удосконалений метод ФОРЕЛ для кластеризації станів комп'ютерної техніки [1, 2], удосконалений метод кластеризації станів комп'ютерної техніки K-MEANS [3, 4], інформаційна модель процесу кластеризації станів комп'ютерної техніки [5], кластеризація станів складних об'єктів з використанням інформаційної технології [6 – 8], формалізація процесу оцінювання результатів кластеризації станів комп'ютерної техніки, як складних об'єктів [9], модель процесу чіткої кластеризації проблемних ситуацій в обчислювальній та організаційній техніці [10], комп'ютерне моделювання [11, 12].

Апробація результатів дисертації. Основні положення та результати виконаних в дисертаційній роботі досліджень були представлені на наукових семінарах кафедри комп'ютерних наук, на щорічних науково-технічних конференціях професорсько-викладацького складу, співробітників та студентів ВНТУ (в період з 2010 р. по 2015 р.), на міжнародних конференціях «ІНТЕРНЕТ-ОСВІТА-НАУКА» (2010, 2012, 2014 роки), Міжнародній науковій конференції, присвяченій 80-річчю з дня народження акад. І. В. Прангишвілі «Інформаційні та комп'ютерні технології, моделювання, управління», Тбілісі, Грузія (2010 рік); на 12-й Міжнародній науково-технічній конференції «Системний аналіз та інформаційні технології» (2010 рік), Міжнародній науково-практичній конференції «Актуальні проблеми інформаційних технологій, економіки та права» м. Чернівці (2011 рік), Міжнародній науково-технічній конференції «Обчислювальний інтелект (результати, проблеми, перспективи)» м. Черкаси (2011 рік), Міжнародній науковій конференції «Проблеми стійкості та оптимізації динамічних систем детермінованої та стохастичної структури» м. Чернівці (2010 рік), Всеукраїнській науково-практичній конференції «Проблеми інформатики та комп'ютерної техніки» м. Чернівці (2013 рік), міжнародній науково-технічній конференції «Електротехнічні і комп'ютерні системи» м. Одеса (2015 рік).

Публікації. За результатами досліджень опубліковано 12 праць, з них 1 стаття у науковому журналі, що входить до міжнародних наукометричних баз даних SCOPUS та INSPEC, 7 – у виданнях, що входить до переліку фахових видань, 1 публікація у вигляді тез доповідей, 1 – у збірниках матеріалів конференцій, 2 свідоцтва на реєстрацію авторського права на комп'ютерну програму.

Структура і обсяг дисертації. Дисертаційна робота складається зі вступу, 4-х розділів, висновків, переліку використаних джерел (139 бібліографічних посилань на 13 сторінках) та додатків на 29 сторінках. Основна частина дисертації становить 112 сторінок і містить 30 рисунків, 7 таблиць. Повний обсяг дисертації – 155 сторінок.

ОСНОВНИЙ ЗМІСТ РОБОТИ

У вступі обґрунтовано актуальність теми дисертаційної роботи, зазначено її зв'язок з науковими програмами, планами та темами, сформульовано мету та напрям досліджень, відзначено наукову новизну та практичне значення одержаних результатів, надано інформацію про їх впровадження, апробацію та публікації.

У першому розділі проведено аналіз сучасного рівня розвитку інформаційних технологій кластеризації. Досліджено поняття стану КТ, яким є сукупність значень параметрів та характеристик такої техніки в певний момент часу, а також наведено класифікацію станів КТ. Характерним для КТ є те, що для аналізу станів такої техніки необхідно опрацьовувати електричні величини, які можуть бути отримані шляхом вимірювання сили електричного струму, електричної напруги, електричного опору елементів, електричної ємності конденсаторів і, можливо, звукові сигнали та повідомлення BIOS. Проведено аналіз основних стадій життєвого циклу КТ та визначено етапи, на яких доцільно використовувати кластеризацію для аналізу її станів з метою формування в подальшому обґрунтованих рекомендацій щодо експлуатації такої техніки. У розділі сформульовано постановку задачі кластеризації станів КТ, яка буде розв'язуватись на етапах підготовки інструкції з технічного обслуговування такої техніки. Суть поставленої задачі полягає в розбитті множини станів КТ на непересічні підмножини, які називаються таксонами, щоб кожен з них складався із станів, при яких формуються подібні рекомендації щодо подальших дій при їх виникненні. У розділі наведено аналіз сучасних методів кластеризації станів КТ та визначено серед них як доцільні ФОРЕЛ та K-MEANS, що потребують удосконалення для усунення недоліків їх застосування в предметній області, яка аналізується. На основі проведеного аналізу визначено основні задачі дослідження.

У другому розділі розроблено методи кластеризації станів комп'ютерної техніки. Всі методи кластеризації розрізняються за метрикою та функціоналом якості розбиття, які використовуються в процесі їх роботи. Серед функціоналів якості розбиття було обрано середню відстань всередині та між кластерами станів КТ, суму середніх відстаней всередині та між кластерами її станів. Для одночасного врахування всіх із перерахованих функціоналів при оцінюванні результатів моделювання процесу кластеризації станів такої техніки в дисертаційній роботі запропоновано інтегральний функціонал якості розбиття множини станів на таксони, який враховує загальноприйняті критерії та обчислюється за формулою

$$K = 1 - \frac{\frac{F_0 + \Phi_0}{F_1 + \Phi_1}}{2}, \quad (1)$$

де F_0 – середня відстань всередині кластерів станів КТ;

F_1 – середня відстань між кластерами станів КТ;

Φ_0 – сума середніх відстаней всередині кластерів станів КТ;

Φ_1 – сума середніх відстаней між кластерами станів КТ.

Запропоновано інформаційну модель процесу кластеризації станів КТ. Інформаційна модель процесу кластеризації станів КТ – це модель, що описує істотні для даного процесу параметри та змінні величини, зв'язки між ними, його вхідні і вихідні значення (рис. 1). Виходячи із запропонованого визначення інформаційної моделі кластеризації станів КТ, її можна подати у вигляді кортежу

$$IMCS = \langle PS, a, q, m, Y \rangle, \quad (2)$$

де PS – множина станів КТ;

a – відстані між станами КТ;

q – параметри та характеристики станів КТ;

m – метод, за допомогою якого проводиться розбиття;

Y – результат кластеризації, множина кластерів, на які буде розбито множину станів КТ.

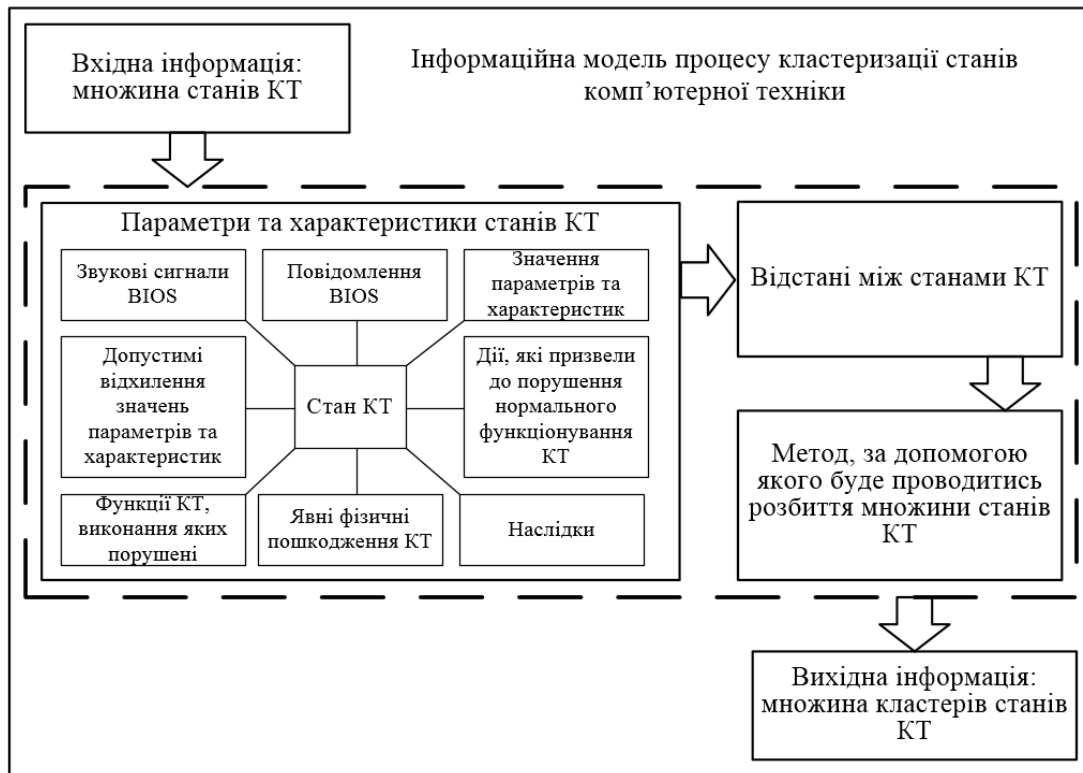


Рисунок 1 – Інформаційна модель процесу кластеризації станів комп'ютерної техніки

Модель стану КТ визначається кортежем

$$PS = \langle BIOS_s, BIOS_m, P, \Delta, WF, PD, A, E \rangle, \quad (3)$$

де $BIOS_s$ – звуковий сигнал BIOS, яким супроводжується завантаження КТ;

$BIOS_m$ – повідомлення BIOS, які виводяться на дисплей;

P – множина значень параметрів та характеристик КТ;

Δ – множина допустимих відхилень значень параметрів та характеристик КТ;

WF – множина функцій КТ, виконання яких порушені;

PD – наявність явних фізичних пошкоджень КТ;

A – дії, які призвели до порушення нормального функціонування КТ;

E – множина наслідків, до яких призвів стан КТ, який аналізується.

Серед проаналізованих метрик було обрано зважену евклідову відстань як таку, що враховує вагові коефіцієнти характеристик та параметрів станів КТ, що дозволило підвищити якість розбиття таких станів на кластери

$$a_{3E}(X_i, X_j) = \sqrt{\sum_{q=1}^m \lambda_q (x_{qi} - x_{qj})^2}, \quad (4)$$

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

$\lambda_q (0 \leq \lambda_q \leq 1 (q = \overline{1, m}))$ – вектор значень вагових коефіцієнтів, які відповідають характеристикам $x_{i1}, x_{i2}, \dots, x_{im}$ станів КТ;

$x_{qi} (q = \overline{1, m})$ – значення характеристик, що описують i -тий стан;

$x_{qj} (q = \overline{1, m})$ – значення характеристик, що описують j -тий стан.

Прийнято рішення використовувати нормування значень параметрів та характеристик станів КТ, що дозволило підвищити якість їх кластеризації. Для нормування значень параметрів використано співвідношення

$$q = \frac{x}{x'}, \quad (5)$$

де q – нормоване значення параметра x ;

x – поточне значення параметра стану КТ;

x' – нормативне (еталонне) значення параметра x стану КТ.

Існуючі різновиди методу K-MENS передбачають необхідність задавати на його вхід кількість кластерів k ; вибір випадковим чином станів КТ, які будуть початковими центрами кластерів; чутливість до викидів, які можуть бути в множині вхідних даних, що негативно впливає на якість розбиття, оскільки початкові центри можуть бути випадково обраними із одного кластера. Крім того, збільшується ймовірність появи викидів у таксонах через те, що не враховуються особливості предметної області, яка аналізується. Тому в дисертаційній роботі було запропоновано удосконалення зазначеного методу, що дозволяє підвищити якість розбиття множини станів КТ за рахунок визначення початкових центрів кластерів на основі значень потенціалів, а також виділення в окремі таксон станів, які могли бути помилково віднесені до кластера за рахунок допустимих відхилень значень параметрів та характеристик такої техніки, що враховує особливості предметної області. Для цього обчислюється для кожного стану значення потенціалу, який показує можливість формування кластера в його околі

$$P_1(X_i) = \sum_{j=1, j \neq i}^n \exp(-\varepsilon \cdot a_{3E}(X_i, X_j)), i = 1, \dots, n, \quad (6)$$

де $P_1(X_i)$ – значення першого потенціалу станів КТ;

n – кількість станів КТ, що підлягають кластеризації;

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

ε – додатна константа, яка характеризує масштаб відстаней між станами

КТ

$$\varepsilon = \frac{1}{\overline{a_{\varepsilon E}(X_i, X_j)}}, \quad (7)$$

де $\overline{a_{\varepsilon E}(X_i, X_j)}$ – середнє значення відстаней між станами КТ.

Чим щільніше розміщені стани в околі потенційного центра кластера, тим вище значення його потенціалу. Центром першого кластера μ_{y_1} обирається стан з найбільшим потенціалом. Оскільки декілька станів КТ з найбільшими значеннями потенціалів, як правило, розміщені поруч, то для знаходження інших центрів кластерів необхідно виключити вплив знайденого. Для цього значення потенціалів перераховується таким чином: від поточного значення віднімається значення потенціалу знайденого центра

$$P_j(X_i) = P_{j-1}(X_i) - P_{j-1}(\mu_{y_{(j-1)}}) \cdot \exp(-\beta \cdot a_{\varepsilon E}(X_i, \mu_{y_{(j-1)}})), i = 1, \dots, n, \quad (8)$$

де $P_j(X_i)$ – значення j -го потенціалу станів КТ;

n – кількість станів КТ, що підлягають кластеризації; $a_{\varepsilon E}(X_i, \mu_{y_{(j-1)}})$ – відстань між станом X_i та центром кластера $\mu_{y_{(j-1)}}$;

β – додатна константа, яка характеризує масштаб розміру одного кластера ($\beta \approx \varepsilon$).

Центрами кластерів μ_{y_j} на кожному кроці обираються ситуації з найбільшим значенням потенціалу. Таку операцію необхідно проводити до тих пір, поки не буде знайдено k попередніх центрів кластерів.

Після формування множини початкових центрів кластерів виконується розбиття вибірки станів КТ з використанням відомого підходу, що передбачає віднесення всіх станів до одного з k кластерів – того, відстань до центра якого є мінімальною. Далі виконується уточнення місця розташування центрів кожного кластера, якими стають стани КТ, ознаки яких розраховуються як середнє арифметичне ознак станів, що входять до цього кластера

$$\mu_{y_j} = \frac{1}{|y_j|} \sum_{X_i \in y_j} X_i, y \in Y, j = 1, \dots, k, \quad (9)$$

де y_j – j -тий кластер станів КТ.

Виконується така кількість ітерацій розбиття, поки центри кластерів стануть стійкими (тобто при кожній ітерації центрами кластерів будуть одні й ті самі стани). Тоді дисперсія всередині кластера буде мінімізована, а між кластерами – максимізована.

З метою підвищення якості розбиття множини станів КТ відбувається ініціалізація додаткового кластера та віднесення до нього станів, які можуть бути зараховані до кластерів з урахуванням допустимих відхилень значень параметрів та характеристик таких станів.

В формалізованому вигляді алгоритм K-MEANS, в основу якого покладено запропонований удосконалений метод, містить такі основні кроки.

1. Задати кількість кластерів k , на які буде розбито множини станів КТ.
2. Знайти значення потенціалів всіх станів КТ з використанням формули (6).
3. Визначити стан КТ, який може бути центром першого кластера за формулою

$$\mu_{y_1} = \arg \max_{x_i \in X} P_1(X_i), i = 1, \dots, n. \quad (10)$$

4. ($j=1$) поки $j \leq k$ виконувати:

- 4.1. Порядковий номер j поточного потенціалу збільшити на 1.
- 4.2. Обчислити значення поточного потенціалу для всіх станів, за виключенням тих, які були обрані як початкові центри кластерів за формулою (8).
- 4.3. Визначити точку, яка може бути центром j -го кластера

$$\mu_{y_j} = \arg \max_{x_i \in X} P_j(X_i), i = 1, \dots, n. \quad (11)$$

5. Віднести кожен стан до найближчого центра кластера

$$y_j = \arg \min_{y \in Y} a_{zE}(X_i, \mu_y), i = 1, \dots, n, \quad (12)$$

де y_j – кластер станів КТ.

6. Обчислити нове положення центрів з використанням виразу (4).

7. Виконувати пункти 7-8, поки центри кластерів y_i не перестануть змінюватись.

8. Ініціалізувати додатковий кластер станів КТ $y_0 = \emptyset$.

9. Обчислити значення допустимого відхилення відстані Δ , яка визначається з використанням зваженої евклідової відстані.

10. Для всіх кластерів $y_i (i = 1, \dots, n)$ виконувати:

- 10.1. Поки $|y_0| \neq const$ для всіх центрів кластерів $\mu_{y_j} (j \neq i)$ виконувати:

- 10.1.1. Знайти

$$X' = \arg \min_{X_l \in y_i} a_{zE}(X_l, \mu_{y_j}), l = 1, \dots, |y_i|, \quad (13)$$

де X' – найближчий стан до центра кластера $y_j (j \neq i)$.

- 10.1.2. Якщо $a_{zE}(X', \mu_{y_j}) - \Delta < a_{zE}(X', \mu_{y_i}) - \Delta$ то

$$y_i = y_i \setminus X', y_0 = y_0 \cup \{X'\}. \quad (14)$$

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 якість декомпозиції множини станів КТ, оскільки користувач не володіє інформацією про розміри кластерів. У зв'язку з цим в дисертаційній роботі було 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

$$R = \frac{(MAX - MIN) \times (100 - \eta)}{100} + MIN - \Delta, \quad (15)$$

де MAX, MIN – максимальна та мінімальна відстань 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 виключення можливості 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, виключаються з некластеризованих, а робота методу повторюється, поки множина некластеризованих станів не буде пустою.

В формалізованому вигляді алгоритм ФОРЕЛ, в основу якого покладено запропонований удосконалений метод, містить такі основні кроки.

1. Формування множини некластеризованих станів КТ

$$U = \{X_1, X_2, \dots, X_n\}, \quad (16)$$

де U – множина некластеризованих станів КТ;

n – кількість станів КТ, що підлягають кластеризації.

2. Знаходження значень відстаней між некластеризованими точками:

а) мінімальної

$$MIN = \min_{i,j=1,\dots,n,i \neq j} a_{3E}(X_i, X_j), \quad (17)$$

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

б) максимальної

$$MAX = \max_{i,j=1,\dots,n,i \neq j} a_{3E}(X_i, X_j) \quad (18)$$

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

4. Знаходження значення радіуса кластера з використанням формули (15).

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

$$K_0 = \{X_i \in U \mid a_{3E}(X_i, X_0) \leq R\}, \quad (19)$$

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

$$X_0 = \arg \min_{X_i \in K_0} \sum_{X_j \in K_0} a_{3E}(X_i, X_j), \quad (20)$$

де X_0 – центр мас кластера.

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

5.5. Знайти

$$\bar{a}_0 = \frac{\sum_{X_0, X_i \in K_0} a_{3E}(X_i, X_0)}{|K_0|}, i = 1, \dots, |K_0|, \quad (21)$$

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

5.6. Знайти

$$X_{\max} = \arg \max_{X_i \in K_0} a_{3E}(X_i, X_0), i = 1, \dots, |K_0|, \quad (22)$$

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

5.7. Якщо $(a_{3E}(X_{\max}, X_0) \geq 2 \cdot \bar{a}_0)$, то

X_{\max} вилучити з K_0

$$K_0 = K_0 \setminus \{X_{\max}\} \quad (23)$$

та X_{\max} додати до U

$$U = U \cup \{X_{\max}\}. \quad (24)$$

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

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

$$U = U \setminus K_0. \quad (25)$$

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

Запропонований метод кластеризації станів КТ ФОРЕЛ дозволяє підвищити якість кластеризації за рахунок визначення радіуси кластерів і визначення станів, які є викидами.

У третьому розділі розроблено технологію кластеризації станів комп'ютерної техніки. Проведено моделювання процесу кластеризації станів КТ з використанням удосконаленого методу K-MEANS. Результати показали, що за рахунок запропонованих удосконалень відомого методу K-MEANS щодо визначення початкових центрів кластерів при кластеризації станів КТ, точність віднесення таких станів до певного кластера підвищується від 3% до 7%, що доводить доцільність використання удосконаленого методу K-MEANS для розв'язання поставленої задачі. Проведено моделювання процесу кластеризації станів КТ з використанням удосконаленого методу ФОРЕЛ. Результати моделювання показали, що запропонований удосконалений метод кластеризації станів КТ, який базується на алгоритмі ФОРЕЛ, дозволяє провести їх кластеризацію з покращенням значення інтегрального функціоналу якості розбиття K на 2-6%.

Розроблено інформаційну технологію кластеризації станів комп'ютерної техніки як сукупність методів, програмно-технологічних засобів кластеризації таких станів, об'єднаних у технологічний ланцюжок, що забезпечує збір, зберігання та обробку інформації про стани КТ для отримання інформації нової якості про такі стани – програмного забезпечення (рис. 2).

Основні кроки автоматизованої інтелектуальної обробки даних про стани КТ з метою їх кластеризації такі.

1. Завантаження вхідних даних про стан/множину станів для проведення кластеризації.

2. Підготовка вхідних даних про завантажені стани, що передбачає проведення нормування значень всіх параметрів та характеристик таких станів, яке проводиться з використанням способів, що описані в пункті 2.2 дослідження.

3. Класифікація станів КТ (стани з бази даних (БД) та нові стани) за класифікаційними ознаками BIOSs і BIOSm.

4. Для всіх класів станів КТ, до яких було віднесено нові стани виконувати кроки 5–15.

5. Залежно від наявності або відсутності відомостей про кількість кластерів буде виконуватись розбиття множини станів КТ з використанням удосконаленого методу K-MEANS (перейти до 7 кроку) або удосконаленого методу ФОРЕЛ (перейти до 6 кроку).

6. Проведення кластерного аналізу станів КТ за допомогою удосконаленого методу кластеризації ФОРЕЛ. В процесі аналізу

використовуються всі відомі параметри та характеристики станів. Для визначення відстані між станами застосовується зважена евклідова відстань.

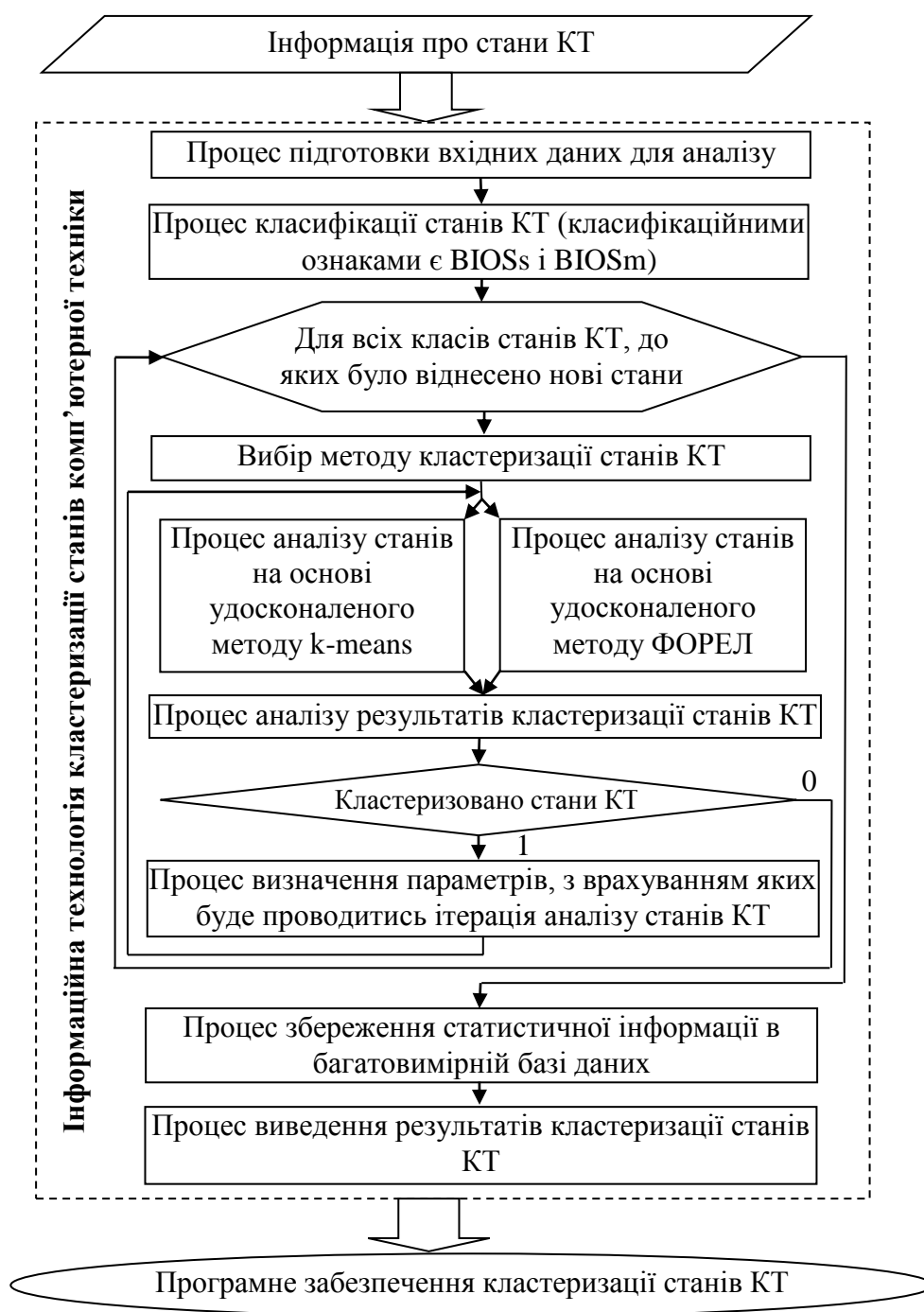


Рисунок 2 – Структурна схема інформаційної технології кластеризації станів комп'ютерної техніки

7. Проведення кластерного аналізу станів КТ за допомогою удосконаленого методу кластеризації K-MEANS. В процесі аналізу використовуються всі відомі параметри та характеристики станів. Для визначення відстані між станами застосовується зважена евклідова відстань.

8. Перевірка результату на наявність таких станів, які виокремлені в окремий кластер, тобто таких, які знаходяться на великій відстані від всіх

сформованих кластерів або на однаковій відстані до двох чи більше кластерів, і за допомогою методу кластеризації були віднесені до окремого кластера.

9. БД аналізованих станів КТ доповнюється станами, які були проаналізовані, окрім тих, які знайдені на 8 кроці.

10. Якщо є знайдені кластери з одного стану, то перевіряється чи цей стан є новим, чи він є збереженим в БД. Якщо він не є новим, то відбувається перехід до 17 кроку, інакше – до 11.

11. Всі параметри та характеристики, за якими проводилась кластеризація станів КТ, класифікуються користувачем за їх ваговими коефіцієнтами, кількість класів також визначається користувачем. Для зручності класи параметрів та характеристик нумеруються від 1 до h . До класу з номером 1 входять параметри та характеристики, ваговий коефіцієнт яких наближається до максимального, а до класу з номером h – до мінімального.

12. Виконання кроків 4-5 з використанням параметрів з класів від 1 до $(h-1)$ та значенням h , що дорівнює $(h-1)$.

13. Перевірка, чи знайдений на 8 кроці стан віднесений до кластера, потужність якого більша за 1.

14. Якщо стан віднесений до кластера, потужність якого більша за 1, то БД аналізованих станів КТ доповнюється станом, який знайдений на 8 кроці та відбувається перехід до кроку 16.

15. Якщо $h > 1$ виконується крок 13, інакше знайдений на 8 кроці стан відноситься до нового кластера та виконується крок 16.

16. БД аналізованих станів КТ доповнюється кластеризованими станами.

17. Подання результатів кластеризації станів КТ користувачу для формування висновку щодо подальшої експлуатації КТ та формування супровідної документації.

Отже, розроблено ІТ кластеризації станів КТ, що дозволило підвищити якість кластеризації за рахунок використання запропонованої ІМ процесу кластеризації та удосконалених методів кластеризації станів КТ K-MEANS і ФОРЕЛ.

У четвертому розділі наведено практичну реалізацію інформаційної технології кластеризації станів комп'ютерної техніки. Розроблено трирівневу клієнт-серверну архітектуру програмного забезпечення (ПЗ) для кластеризації станів КТ, що пропонується для використання при розробці клієнтського додатка та сервера додатків.

Для розробки клієнтського додатка ПЗ для кластеризації станів КТ було обрано бібліотеки Underscore.js, AngularJS, та JQuery. Для розробки серверного додатка системи кластеризації станів КТ було обрано мову програмування C# та технології WebAPI. Для маніпулювання даними про станів КТ використано СУБД MS SQL SERVER.

Для аналізу результатів застосування інформаційної технології (ІТ) кластеризації станів КТ було проаналізовано стани КТ LLC «ІКрок» протягом 2 років.

Для оцінювання доцільності застосування запропонованої ІТ було проведено експериментальні дослідження щодо декомпозиції множини станів

за допомогою удосконаленого та відомих методів ФОРЕЛ та К-MEANS. Моделювання з використанням існуючих методів було проведено з використанням ППП MATLAB.

На основі отриманих результатів було розраховано значення інтегрального функціонала якості кластеризації множини станів КТ з використанням відношення (1), що становив для:

– відомого методу ФОРЕЛ – $K_{\text{ФОРЕЛ}} = 0,64$, для удосконаленого – $K_{\text{мФОРЕЛ}} = 0,71$;

– існуючого методу К-MEANS – $K_{\text{кК-MEANS}} = 0,68$, для удосконаленого – $K_{\text{мК-MEANS}} = 0,76$,

при цьому час, який був витрачений на аналіз даних, був практично однаковий для відомих і відповідних удосконалених методів.

Таким чином, підвищення якості кластеризації станів КТ з використанням удосконаленого методу ФОРЕЛ становить $\frac{0,71}{0,64} \approx 1,11$, а з використанням

методу К-MEANS – $\frac{0,76}{0,68} \approx 1,12$.

Було проведено експериментальні дослідження із розбиття вибірок станів КТ різної потужності з використанням як відомих, так і запропонованих удосконалених методів кластеризації К-MEANS і ФОРЕЛ, що показали доцільність використання останніх в процесі аналізу станів КТ. При цьому для отримання достовірних результатів кількість кластерів в процесі дослідження змінювалась. Результати експериментальних досліджень щодо середнього значення якості розбиття вибірки станів КТ з урахуванням кількості кластерів та відповідної якості кластеризації наведено в таблиці 1.

Таблиця 1 – Результати експериментальних досліджень щодо середнього значення якості розбиття вибірки станів КТ з урахуванням кількості кластерів та відповідної якості кластеризації

Потужність вибірки	Метод						
	CURE	BIRCH	PAM	К-MEANS	К-MEANS удосконалений	ФОРЕЛ	ФОРЕЛ удосконалений
50	0,87	0,87	0,8	0,8	0,85	0,82	0,86
100	0,85	0,84	0,79	0,79	0,85	0,81	0,85
200	0,82	0,8	0,76	0,77	0,82	0,78	0,83
500	0,79	0,76	0,72	0,75	0,8	0,75	0,79
1000	0,73	0,69	0,69	0,7	0,76	0,72	0,76
2000	0,69	0,63	0,66	0,67	0,73	0,7	0,74
5000	0,61	0,6	0,63	0,64	0,69	0,66	0,7
10000	0,55	0,54	0,6	0,6	0,65	0,61	0,66

Отже, експериментальні дослідження та результати впровадження розробленої ІТ показали, що якість кластеризації станів при використанні

запропонованої ІТ кластеризації станів КТ підвищилась в 1,115 рази, тобто на 11,5%.

У додатках наведено акти впровадження результатів дисертації, лістинг програмної реалізації основних модулів ПЗ для кластеризації станів КТ, інструкцію користувача ПЗ для кластеризації станів КТ, процедуру вимірювання значень електричних показників КТ, вибірку станів КТ.

ЗАГАЛЬНІ ВИСНОВКИ

У дисертаційній роботі наведено теоретичне узагальнення і нове вирішення актуальної наукової задачі, яка полягає в підвищенні якості кластеризації станів комп'ютерної техніки шляхом розробки і впровадження відповідної інформаційної технології. Означена технологія буде використовуватись при підготовці інструкцій з обслуговування такої техніки, що є необхідними в процесі її ремонту та інших видів сервісного обслуговування при прийнятті рішень щодо експлуатації.

Основні наукові і практичні результати роботи:

1. Виходячи з аналізу сучасного рівня розвитку інформаційних технологій кластеризації встановлено, що доцільними методами для кластеризації станів комп'ютерної техніки є методи ФОРЕЛ та К-MEANS, які потрібно удосконалити для усунення недоліків відомих методів та адаптації їх до відповідної предметної області. З врахуванням недоліків методів, моделей кластеризації та переваг запровадження інформаційних технологій зроблено висновок про доцільність застосування таких технологій при розв'язанні задачі кластеризації станів комп'ютерної техніки для подальшого прийняття рішень щодо її експлуатації. Таким чином, розробка інформаційної технології кластеризації станів комп'ютерної техніки та вдосконалення методів кластеризації таких станів К-MEANS і ФОРЕЛ є актуальною теоретичною і прикладною задачею.

2. Запропоновано інформаційну модель процесу кластеризації станів комп'ютерної техніки, що містить в собі модель станів комп'ютерної техніки, враховує вагові коефіцієнти характеристик та параметрів таких станів в процесі визначення відстані між ними та використовує нормування значень цих параметрів та характеристик, що дозволило підвищити якість розбиття. При цьому, вхідною інформацією є множина станів комп'ютерної техніки, вихідною – множина кластерів станів комп'ютерної техніки. Серед проаналізованих метрик було обрано зважену евклідову відстань як таку, яка враховує вагові коефіцієнти характеристик та параметрів таких станів та дозволить підвищити якість розбиття множини станів комп'ютерної техніки.

3. Удосконалено метод кластеризації станів комп'ютерної техніки К-MEANS, що, на відміну від існуючих, передбачає визначення початкових центрів кластерів на основі потенціалів, а також враховує особливості станів такої техніки та виділяє в окремий таксон стани, які могли бути помилково віднесені до кластера за рахунок допустимих відхилень значень параметрів та

характеристик. Це дозволило підвищити якість кластеризації станів комп'ютерної техніки на 7%.

4. Удосконалено метод кластеризації станів комп'ютерної техніки ФОРЕЛ, який, на відміну від відомих, 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%.

5. Розроблено інформаційну технологію кластеризації станів комп'ютерної техніки, що дозволило підвищити якість їх кластеризації на 11,5% за рахунок використання запропонованої інформаційної моделі та удосконалених методів кластеризації K-MEANS і ФОРЕЛ, а також підтримки можливості аналізу сигналів та повідомлень BIOS.

7. На основі запропонованої інформаційної технології розроблено програмне забезпечення для кластеризації станів комп'ютерної техніки.

8. Результати дисертаційних досліджень впроваджені в компанії LLC «ІКРОК» у вигляді програмного забезпечення для кластеризації станів комп'ютерної техніки, що дало можливість приймати ефективні управлінські рішення; в департаменті енергетики, транспорту та зв'язку Вінницької міської ради як методика аналізу даних про електроспоживання підприємств та установ міста, що базується на технологіях Data Mining, та інформаційна технологія аналізу даних про електроспоживання підприємств та установ міста, що базується на удосконаленому алгоритмі кластерного аналізу; в навчальний процес при фаховій підготовці за освітньо-професійною програмою підготовки бакалаврів за напрямом 6.050101 «Комп'ютерні науки». Результати досліджень використані в НДР «Теоретичні методи аналізу швидкоплинних техногенних надзвичайних ситуацій для створення систем підтримки прийняття рішень керівниками ліквідації таких ситуацій» № 47-Д-307 (номер державної реєстрації 0108U000659). Впровадження результатів дослідження підтверджено відповідними актами.

СПИСОК ОПУБЛІКОВАНИХ ПРАЦЬ ЗА ТЕМОЮ ДИСЕРТАЦІЇ

1. Савчук Т. О. Удосконалений метод ФОРЕЛ для кластеризації станів комп'ютерної техніки / Т. О. Савчук, С. І. Петришин // Збірник наукових праць Дніпродзержинського державного технічного університету (технічні науки). – 2015. – № 1(26). – С. 245-252

2. Савчук Т. О. Ідентифікація проблемних ситуацій та їх станів в складних технічних системах з використанням модифікованого алгоритму ФОРЕЛ / Т. О. Савчук, С. І. Петришин // Вісник Національного університету „Львівська політехніка”. – 2014. – № 783. – С. 187-193. – ISSN 0321-0499

3. Савчук Т. О. Удосконалений метод кластеризації станів комп'ютерної техніки K-MEANS / Т. О. Савчук, С. І. Петришин // Вісник Тернопільського національного технічного університету. – 2015. – № 2(78). – С. 198-206. – ISSN 1727-7108

4. Савчук Т. О. Використання модифікованого алгоритму k-means для ідентифікації проблемних ситуацій та їх станів в складних технічних системах / Т. О. Савчук, С. І. Петришин // Вимірювальна та обчислювальна техніка в технологічних процесах. – 2014. – № 1. – С. 117-121. – ISSN 2219-9365

5. Савчук Т. О. Інформаційна модель процесу кластеризації станів комп'ютерної техніки / Т. О. Савчук, С. І. Петришин // Електротехнічні та комп'ютерні системи. – 2015. – № 19(95). – С. 182-186. – ISSN 2221-3805

6. Савчук Т. О. Кластеризація станів комп'ютерної техніки з використанням інформаційної технології / Т. О. Савчук, С. І. Петришин // Вісник Хмельницького національного університету (серія: технічні науки). – 2015. – № 4(227). – С. 149-152. – ISSN 2307-5732

7. Савчук Т. О. Інформаційна технологія ідентифікації проблемних ситуацій та їх станів в складних технічних системах / Т. О. Савчук, С. І. Петришин // Вісник Національного університету „Львівська політехніка”. – 2014. – № 805. – С. 193-198.

8. T. O. Savchuk. Identification of technogenic emergency situations in railway transport using cluster analysis/ T. O. Savchuk, S. I. Petrishyn, Laura Sugurova, Andrzej Smolarz // Przegląd Elektrotechniczny – Warsaw, 2014. – № 11/2014. – P. 177–184. – Access mode: <http://pe.org.pl/articles/2014/11/46.pdf>.

9. Савчук Т. О. Формалізація процесу оцінювання результатів кластеризації складних об'єктів / Т. О. Савчук, С. І. Петришин // Всеукраїнська науково-практична конференція «Проблеми інформатики та комп'ютерної техніки» (ПКТ – 2013), 27 – 31 травня, 2013 : Тези доповідей. – Чернівці, 2013 – С. 150–152.

10. Савчук Т. О. Модель процесу чіткої кластеризації проблемних ситуацій в обчислювальній та організаційній техніці / Т. О. Савчук, С. І. Петришин // «ІНТЕРНЕТ-ОСВІТА-НАУКА-2014», дев'ята міжнародна конференція ІОН-2014, 14 – 17 жовтня, 2014 : Збірник праць. – Вінниця, 2014 – С. 79–81.

11. Савчук Т. О. Комп'ютерна програма «Кластерний аналіз станів комп'ютерної техніки з використанням модифікованого методу K-MEANS» / Т. О. Савчук, С. І. Петришин // Свідоцтво про реєстрацію авторського права на твір № 61714. – Рішення від 17.09.2015.

12. Савчук Т. О. Комп'ютерна програма «Кластеризація станів комп'ютерної техніки з використанням модифікованого методу ФОРЕЛ» / Т. О. Савчук, С. І. Петришин // Свідоцтво про реєстрацію авторського права на твір № 61714. – Рішення від 19.09.2015.

АНОТАЦІЯ

Петришин С. І. Інформаційна технологія кластеризації станів комп'ютерної техніки. – На правах рукопису.

Дисертація на здобуття наукового ступеня кандидата технічних наук зі спеціальності 05.13.06 – інформаційні технології. Вінницький національний технічний університет, Вінниця, 2016.

Дисертаційну роботу присвячено підвищенню якості кластеризації станів комп'ютерної техніки шляхом розробки і впровадження інформаційної технології для подальшого прийняття рішень щодо експлуатації такої техніки. Запропоновано інформаційну модель процесу кластеризації станів комп'ютерної техніки, удосконалено методи кластеризації станів комп'ютерної техніки: K-MEANS, який враховує допустимі відхилення значень параметрів та характеристик комп'ютерної техніки при формуванні кластерів її станів; ФОРЭЛ, що передбачає розрахунок радіусів кластерів для зменшення кількості викидів у таких кластерах. Розроблено інформаційну технологію кластеризації станів комп'ютерної техніки, яка базується на запропонованій інформаційній моделі та удосконалених методах кластеризації K-MEANS і ФОРЭЛ. Це дозволило підвищити якість кластеризації станів комп'ютерної техніки. Розроблено програмне забезпечення для кластеризації станів комп'ютерної техніки.

Ключові слова: інформаційна технологія, стан комп'ютерної техніки, програмне забезпечення, кластеризація, метод K-MEANS, метод ФОРЭЛ.

АННОТАЦИЯ

Петришин С. И. Информационная технология кластеризации состояний компьютерной техники. – На правах рукописи.

Диссертация на соискание ученой степени кандидата технических наук по специальности 05.13.06 – информационные технологии. Винницкий национальный технический университет, Винница, 2016.

Диссертационная работа посвящена повышению качества кластеризации состояний компьютерной техники путем разработки и внедрения информационной технологии для дальнейшего принятия решений по дальнейшей эксплуатации такой техники. Объектом исследования является процесс кластеризации состояний компьютерной техники. Общий научный результат работы – решение актуальной научно-практической задачи повышения качества кластеризации состояний компьютерной техники с помощью новой информационной технологии кластеризации таких состояний.

На основе анализа состояний компьютерной техники и современных методов кластеризации таких состояний установлено, что целесообразными методами для кластеризации состояний КТ являются методы ФОРЭЛ и K-MEANS, которые нужно усовершенствовать для устранения недостатков классических методов и адаптации их к анализируемой предметной области, что приведет к повышению качества кластеризации таких состояний.

Предложена информационная модель процесса кластеризации состояний КТ, которая есть моделью, представленной в виде информации, описывающей существенные для данного процесса параметры и переменные величины, связи между ними, его входные и выходные значения. Входной информацией для данного процесса является множество состояний компьютерной техники,

выходной – множество кластеров состояний компьютерной техники. Кроме входящей и исходящей информации, информационная модель процесса кластеризации состояний компьютерной техники включает функции расстояний между состояниями КТ, функции нормирования значений параметров при кластерном анализе состояний КТ, методы, с помощью которых будет проводиться разбиение. Среди проанализированных метрик было избрано взвешенное евклидово расстояние как учитывающее весовые коэффициенты характеристик и параметров таких состояний, что позволит повысить качество разбиения множества состояний КТ. Предложено использовать нормирование значений параметров и характеристик состояний КТ, что позволило повысить качество их кластеризации.

Усовершенствован метод кластеризации состояний компьютерной техники K-MEANS, что, в отличие от классического, предусматривает определение начальных центров кластеров на основе потенциалов, а также учитывает особенности состояний компьютерной техники и выделяет в отдельный таксон состояния, которые могли быть ошибочно включены в кластер за счет допустимых отклонений значений параметров и характеристик такой техники, что позволило повысить качество кластеризации таких состояний. Усовершенствован метод кластеризации состояний компьютерной техники ФОРЭЛ, что, в отличие от классического, предусматривает расчет радиусов кластеров с учетом особенностей предметной области и определяет состояния, которые являются выбросами и извлекает их из кластеров, что позволило повысить качество кластеризации таких состояний.

Разработана информационная технология кластеризации состояний компьютерной техники, что позволило повысить качество их кластеризации на 11,5% за счет использования предложенной информационной модели и усовершенствованных методов кластеризации K-MEANS и ФОРЭЛ, а также поддержки возможности анализа сигналов и сообщений BIOS.

На основе предложенной информационной технологии разработано программное обеспечение кластеризации состояний компьютерной техники.

Результаты диссертационного исследования внедрены в компании LLC «ИКРОК» в виде программного обеспечения кластеризации состояний компьютерной техники; в департаменте энергетики, транспорта и связи Винницкого городского совета как методика анализа данных по электропотреблению предприятий и учреждений города, основанная на технологиях Data Mining и информационная технология анализа данных по электропотреблению предприятий и учреждений города; в учебный процесс при профессиональной подготовке по образовательно-профессиональной программе подготовки бакалавров по направлению 6.050101 «Компьютерные науки». Результаты исследований использованы в НИР «Теоретические методы анализа быстротекущих техногенных чрезвычайных ситуаций для создания

систем поддержки принятия решений руководителями ликвидации таких ситуаций» № 47-Д-307 (номер государственной регистрации 0108U000659). Внедрение результатов исследования подтверждено соответствующими актами.

Ключевые слова: информационная технология, состояние компьютерной техники, программное обеспечение, кластеризация, метод K-MEANS, метод ФОРЭЛ.

ABSTRACT

Petrishyn S. I. Information technology of clustering states of computer equipment. - The manuscript.

The thesis for the degree of candidate of technical sciences, specialty 05.13.06 - information technologies. Vinnytsia National Technical University, Vinnytsia, 2016.

The thesis is dedicated to improving the quality of clustering states of computer equipment through the development and implementation of information technology for further decisions on the operation of such equipment. An information model of clustering state of computer equipment has been suggested, clustering methods classes of computer equipment K-MEANS have been improved, taking into account the allowable deviation values of parameters and characteristics of computer equipment in the formation of clusters of states and FOREL, which involves the calculation of the radius of clusters to reduce emissions in these clusters. The information technology of clustering states of computer equipment based on the proposed information model and improved methods of clustering K-MEANS and FOREL are developed. This enhanced the quality of clustering states of computer equipment. The software for clustering states of computer technology has been developed.

Keywords: information technology, the state of computer equipment, software, clustering, K-MEANS method, FOREL method.

Підписано до друку 01.02.2016 р. Формат 29,7×42 ¼

Наклад 100 прим. Зам. № 2016-025.

Віддруковано в комп'ютерному інформаційно-видавничому центрі

Вінницького національного технічного університету

м. Вінниця, Хмельницьке шосе, 95. Тел.: 59-87-38

Свідоцтво суб'єкта видавничої справи

серія ДК №3516 від 01.07.2009 р.