

ВИКОРИСТАННЯ МОДИФІКОВАНОГО АЛГОРИТМУ K-MEANS ДЛЯ ІДЕНТИФІКАЦІЇ ПРОБЛЕМНИХ СИТУАЦІЙ ТА ЇХ СТАНІВ В СКЛАДНИХ ТЕХНІЧНИХ СИСТЕМАХ

В статті запропоновано використання модифікованого алгоритму K-MEANS для ідентифікації проблемних ситуацій та їх станів в складних технічних системах. За рахунок запропонованих модифікацій класичного алгоритму K-MEANS стосовно визначення початкових центрів кластерів при ідентифікації проблемних ситуацій та їх станів в складних технічних системах, підвищується точність віднесення таких ситуацій до певного кластеру від 5,3% до 5,4%.

Ключові слова: кластеризація, алгоритм K-MEANS, проблемні ситуації, ідентифікація, складні технічні системи.

T.O. SAVCHUK, S.I. PETRISHYN
Vinnytsia National Technical University

USING A MODIFIED K-MEANS ALGORITHM TO IDENTIFY PROBLEMATIC SITUATIONS AND CONDITIONS IN COMPLEX ENGINEERING SYSTEMS

In the article proposed to use a modified K-MEANS algorithm to identify problem situations and states in complex technical systems. Due to proposed modifications of the classical K-MEANS algorithm on defining the initial cluster centers in the identification of problematic situations and states in complex technical systems, increasing accuracy of classifying such situations to a certain cluster from 5.3% to 5.4%.

Keywords: clustering algorithm K-MEANS, problem situations, identification, complex technical systems.

Вступ

Сталих тенденцій розвитку складних технічних систем, що будуть використовуватись як в побуті так і на виробництві, можливо досягти при поєднанні їх технічного вдосконалення з розвитком методів та засобів ідентифікації проблемних ситуацій, що можуть виникнути в процесі їх функціонування. Методи та засоби ідентифікації повинні базуватися на досконалих алгоритмах аналізу інформації про такі ситуації та/або їх стани, оскільки саме так можливо досягти максимальної якості роботи таких засобів.

Постановка задачі

Нехай

$$X = \{x_1, \mathbf{K}, x_n\}$$

– вибірка проблемних ситуацій та/або їх станів, що мають місце в складних технічних системах,

де x_1, \mathbf{K}, x_n – проблемні ситуації в складних технічних системах,

$n = 1, \infty$ – кількість проблемних ситуацій в вибірці.

Тоді задача ідентифікації проблемних ситуацій та/або їх станів в складних технічних системах зведеться до задачі розбиття означеної вибірки на k непересічних кластерів, так, щоб кожен кластер складався з ситуацій та/або їх станів, які схожі між собою, проблемні ситуації в складних технічних системах та/або їх стани, які знаходяться в різних кластерах значно відрізнялися [1].

Аналіз сучасних засобів розв'язання задачі ідентифікації проблемних ситуацій та їх станів в складних технічних системах

Серед множини систем ідентифікації стану складних технічних систем можна виділити такі:

1) ТОРОС - це система методик, інструкцій, регламентів, програмних і технічних засобів, які орієнтовані на задачі планування, ресурсного забезпечення, виконання і контролю діяльності по технічному обслуговуванню, ремонту, аналізу і оцінці технічного стану гірських машин і обладнання. Її призначення полягає в інформаційній підтримці прийняття рішень, направлених на зниження ризиків відказів гірських машин і обладнання. Ця система будується на базі програмного комплексу TRIM - спеціалізованого програмного продукту класу ЕАМ, який орієнтований на задачі управління експлуатацією, технічним обслуговуванням і ремонтом обладнання [2].

2) Hitachi Diagnostic Kit (EDL) - система для ідентифікації стану і налаштування роботи конструктивних блоків будівельної та гірничодобувної техніки Hitachi. Виконує ідентифікацію та налаштування обладнання не лише механічних, гідравлічних систем, а також і електрообладнання. Hitachi Diagnostic Kit (EDL) включає системи ідентифікації стану та моніторингу, які за інформацією, що надходить від датчиків і вимикачів отримує дані для техобслуговування гідравлічних систем та іншого обладнання. При виявленні проблем система ранжує несправності за ступенем важливості і повідомляє про необхідні коригувальні дії [3].

3) Комплекс програмних, апаратних та програмно-апаратних засобів фірми ULTRA-X, які

використовуються для ідентифікації стану та ремонту персональних комп'ютерів та ноутбуків різних виробників [4]. За допомогою окремих компонентів можна оцінювати стан окремих частин комп'ютера і отримувати рекомендації щодо дій при виникненні певних проблемних ситуацій. Серед засобів компанії ULTRA-X варто виділити такі:

- комбіновані індикатори POST кодів BIOS mini PCI, mini PCI E, LPC, ELPC, I2C;
- індикатор для перевірки цілісності контактів шини PCI;
- індикатор для перевірки цілісності контактів роз'ємів модулів DDR2 і DDR3;
- індикаторні тестери для перевірки процесорних гнізд;
- PC POWER PCI-2.22;
- комплект P.H.D. PCI-2;
- QuickTech PRO Premium та інші.

4) Launch X431 IV - система, яка служить для ідентифікації стану електронних систем управління автомобілів європейського, американського і азійського виробництва [5]. Оснащується відкритою діагностичною платформою на базі операційної системи Linux і успадковує переваги попередніх модифікацій. Комплектація Launch X431 IV також включає інтегрований міні-принтер, адаптери для підключення широкого спектру комунікаційних роз'ємів і набір програмного забезпечення структурування за вживаності для конкретних марок автомобілів. Забезпечує тестування несправностей більшості систем і конструктивних блоків автомобілів: анти-блокувальної системи гальм, аудіо-системи, газорозрядної лампи, гідро підсилювача керма, датчика кута повороту рульового колеса, дверей, двигуна, дзеркал, клімат-контролю, коліс, кондиціонера, круїз-контролю, кузова, GPS-навігації, пневматичної ідвіски, подушки безпеки, приладової панелі, радіо, ручного гальма, сидінь, телевізора, трансмісії, тяги, центрального замка.

5) Nanascan 10 - система для ідентифікації несправностей автомобілів різних марок і виконання сервісних функцій [6]. Дружній інтерфейс сканера підтримує російську мову. Підтримує роботу по протоколах бортової діагностики:

- OBD-I.
- OBD-II: ISO 9141-2, ISO14230 (KWP).
- J1850 (PWM & VPW) /
- CAN 2.0A / 2.0B.
- Оригінальні діагностичні протоколи виробників автомобілів.

Передбачена можливість оновлення програмного забезпечення і баз кодів несправностей, виправлення збою в роботі сканера на основі прямого обміну даними з виробником через Інтернет. Система оснащується операційною системою Windows XP. Потужна апаратна платформа включає процесор Intel® Atom 1,6 ГГц, адаптерами Wi-Fi, Bluetooth і роз'єм Ethernet RJ45 для взаємодії з локальними мережами.

В роботі наведено невелика кількість систем, які використовуються для ідентифікації проблемних ситуацій та/або їх станів в складних технічних системах. На сьогоднішній день швидкими темпами розвиваються напрями по розробці систем ідентифікації станів проблемних ситуацій в автомобілях різних марок, комп'ютерах та ноутбуках різних марок, медичному обладнанні і т.д., проте не розробляються такі системи, які могли б працювати з різними системами, тобто не розробляються масштабовані системи ідентифікації стану проблемних ситуацій в складних технічних системах.

До переваг існуючих рішень можна віднести такі:

- висока швидкість ідентифікації проблемних ситуацій та/або їх станів в складних технічних системах, обробки заданих ситуацій;
- відсутність необхідності втручання експертів для аналізу результатів, тому що більшість існуючих систем надає рекомендації, щодо усунення проблемних ситуацій, або ж їх можна дізнатись на основі вихідної інформації від системи та довідників.

До недоліків існуючих рішень можна віднести:

- неможливість обробки нечітких даних;
- направленість на ідентифікацію стану лише при конкретно визначених проблемних ситуаціях;
- неможливість отримання проміжних чи узагальнених результатів при недостатній повноті вхідних параметрів;
- такі системи є практично немасштабованими, тобто орієнтованими на конкретний вид складних технічних систем;
- такі системи мають високу вартість.

Отже, актуальним є удосконалення існуючих алгоритмів кластеризації, які будуть використовуватись для ідентифікації проблемних ситуацій та/або їх станів в складних технічних системах з урахуванням їх особливостей.

Розробка модифікованого алгоритму K-MEANS кластеризації проблемних ситуацій та їх станів в складних технічних системах

Використання класичного алгоритму K-MEANS для кластеризації проблемних ситуацій та/або їх станів в складних технічних системах при здатності працювати з потужними об'ємами даних про такі ситуації та при наперед визначеній кількості кластерів k , має такі недоліки [7, 8]:

- вибір випадковим чином k проблемних ситуацій в складних технічних системах, які і будуть

вважатись початковими центрами кластерів, на першому кроці алгоритму, негативно впливає на якість кластеризації, тобто точність віднесення проблемної ситуації або її стану до коректного кластеру;

- чутливість до викидів, які можуть бути в множині проблемних ситуацій, негативно впливає та якість їх ідентифікації.

Означені недоліки класичного алгоритму K-MEANS можна усунути за рахунок введення механізму обрахунку потенціалів проблемних ситуацій та/або їх станів, як обраних початковими центрами кластерів.

Тоді, модифікований алгоритм K-MEANS кластеризації проблемних ситуацій та/або їх станів в складних технічних системах матиме такий вигляд:

1. Знайти значення потенціалів всіх проблемних ситуацій в складних технічних системах:

$$P_1(x_i) = \sum_{j=1, j \neq i}^n \exp(-a \cdot a(x_i, x_j)), i = 1, \mathbf{K}, n, \quad (1)$$

де $P_1(x_i)$ - значення першого потенціалу проблемних ситуацій в складних технічних системах;

n - кількість проблемних ситуацій в складних технічних системах, що підлягають кластеризації;

$a(x_i, x_j)$ - відстань між i -ю та j -ю проблемними ситуаціями;

a - позитивна константа, яка характеризує масштаб відстаней між проблемними ситуаціями та/або їх станами в складних технічних системах (2.13)

$$a = \frac{1}{a(X)}, \quad (2)$$

де $\overline{a(X)}$ - середнє значення відстаней між проблемними ситуаціями та/або їх станами в складних технічних системах;

X - множина проблемних ситуацій в складних технічних системах, що підлягають кластеризації;

2. Визначити проблемну ситуацію, що має місце в складній технічній системі, яка може бути центром першого кластера:

$$m_{y1} = \arg \max_{x_i \in X} P_1(x_i), i = 1, \mathbf{K}, n, \quad (3)$$

де m_{y1} - центр першого кластера проблемних ситуацій та/або їх станів в складних технічних системах;

3. Порядковий номер j поточного потенціалу збільшити на 1.

4. Обрахувати значення поточного потенціалу для всіх проблемних ситуацій, за виключенням тих, які були обрані в якості початкових центрів кластерів.

$$P_j(x_i) = P_{j-1}(x_i) - P_{j-1}(m_{y(j-1)}) \cdot \exp(-b \cdot a(x_i, m_{y(j-1)})), i = 1, \mathbf{K}, n, \quad (4)$$

де $P_j(x_i)$ - значення j -го потенціалу проблемних ситуацій в складних технічних системах;

n - кількість проблемних ситуацій в складних технічних системах, що підлягають кластеризації;

$a(x_i, m_{y(j-1)})$ - відстань між i -ю проблемною ситуацією та центром кластера $m_{y(j-1)}$;

b - позитивна константа, яка характеризує масштаб розміру одного кластера ($b \approx a$);

5. Визначити точку, яка може бути центром j -го кластера:

$$m_{yj} = \arg \max_{x_i \in X} P_j(x_i), i = 1, \mathbf{K}, n, \quad (5)$$

де m_{yj} - центр j -го кластера проблемних ситуацій та/або їх станів в складних технічних системах;

6. Виконувати пункти 3-5 ПОКИ $j \leq k$;

7. Віднести кожен ситуацію до найближчого центра кластера:

$$y_j = \arg \min_{y \in Y} a(x_i, m_y), i = 1, \mathbf{K}, n; \quad (6)$$

де y_j - кластер проблемних ситуацій та/або їх станів в складних технічних системах;

8. Обчислити нове положення центрів:

$$m_{yj} = \frac{\sum_{i=1}^n [y_i = y] f_i(x_i)}{\sum_{i=1}^n [y_i = y]}, y \in Y, j = 1, \mathbf{K}, k; \quad (7)$$

9. Виконувати пункти 7-8 ПОКИ центри кластерів Y_j не перестануть змінюватись.

Таким чином, основними етапами ідентифікації проблемних ситуацій та/або їх станів в складних технічних системах з використанням модифікованого алгоритму K-MEANS є:

Етап 1. Завдання кількості кластерів k , на які буде розбито множину проблемних ситуацій в

складних технічних системах

Етап 2. Визначення точок (проблемних ситуацій в складних технічних системах), які можуть бути початковими центрами кластерів проблемних ситуацій та/або їх станів в складних технічних системах. Для цього обчислюється для кожної проблемної ситуації значення потенціалу, який показує можливість формування кластера в її окрузі. Чим щільніше розміщені об'єкти в окрузі потенційного центра кластера, тим вище значення його потенціалу.

Етап 3. Обрання центра першого кластера, яким стає проблемна ситуація з найбільшим потенціалом. Оскільки декілька проблемних ситуацій в складних технічних системах з найбільшими значеннями потенціалів, як правило розміщені одна біля одної, тому для знаходження інших центрів кластерів таких ситуацій необхідно виключити вплив знайденого центра кластера, для цього значення потенціалів перераховується наступним чином: від поточних значень потенціалу віднімають значення щойно знайденого центра. Таку операцію необхідно проводити до тих пір поки не буде знайдено k попередніх центрів кластерів проблемних ситуацій в складних технічних системах.

Етап 4. Віднесення всіх проблемних ситуацій в складних технічних системах до одного з k кластерів – того, відстань до чийого центра є мінімальною.

Етап 5. Розрахунок нових центрів кожного кластера, якими стають проблемні ситуації, що мають місце в складних технічних системах, ознаки яких розраховуються як середнє арифметичне ознак ситуацій, що входять у цей кластер.

Етап 6. Відбувається така кількість ітерацій (повторення виконання кроків 4-5), поки центри кластерів стануть стійкими (при кожній ітерації центрами кластерів будуть одні й ті самі проблемні ситуації), дисперсія всередині кластера буде мінімізована, а між кластерами – максимізована.

Експериментальні дослідження ідентифікації вибірки проблемних ситуацій потужністю 100 з використанням модифікованого та класичного алгоритмів K-MEANS показав, що середнє значення відносного показника якості розбиття [9] в залежності від заданої кількості кластерів склав для класичного алгоритму від 83,2% до 83,9%, а для модифікованого алгоритму від 88,6% до 89,2% (таблиця 1), що свідчить про доцільність використання запропонованого модифікованого алгоритму K-MEANS для ідентифікації проблемних ситуацій та/або їх станів.

Таблиця 1

Результати експериментальних досліджень класичного та модифікованого алгоритмів K-MEANS

Кількість кластерів	20	25	30
Назва Алгоритму			
K-MEANS	83,2%	83,5%	83,9%
Модифікований K-MEANS	88,6%	88,9%	89,2%

Висновки

Таким чином, за рахунок запропонованих модифікацій класичного алгоритму K-MEANS стосовно визначення початкових центрів кластерів при ідентифікації проблемних ситуацій та/або їх станів в складних технічних системах, підвищується точність віднесення таких ситуацій до певного кластера від 5,3% до 5,4%, що доводить доцільність використання модифікованого алгоритму K-MEANS для розв'язання поставленої задачі.

Література

1. Савчук Т.О. Оцінювання надзвичайних ситуацій на залізничному транспорті, що базується на кластерному аналізі./Савчук Т.О., Петришин С.І.// Тези XXXIX науково-технічної конференції професорсько-викладацького складу, співробітників та студентів університету з участю працівників науково-дослідницьких організацій та інженерно-технічних працівників підприємств м.Вінниці та області. ВНТУ – м. Вінниця – 2010р.
2. ТОРОС http://www.trim.ru/docs/toros_mining.pdf
3. Hitachi Diagnostic Kit (EDL) http://www.autoscaners.ru/catalogue/?catalogue_id=751
4. ULTRA-X, Inc. <http://www.uxd.com/>
5. Launch X431 IV http://www.autointhebox.com/launch-x431-iv-master-scanner_c30
6. Hanascan 10 http://www.autoscaners.ru/catalogue/?catalogue_id=hanascan_10_multimarochnyy_skaner
7. Барсегян А.А. Технология анализа данных: Data Mining, Visual Mining, Text Mining, OLAP / А. А. Барсегян, М. С. Куприянов, В.В. Степаненко, И. И. Холод. – 2-е изд., перераб. и доп. – Спб.: БХВ, Петербург, 2007. – 384с.: ил. – ISBN 5-94157-991-8.
8. Айвазян С.А., Бухштабер В.М., Енюков И.С. Прикладная статистика: Классификация и снижение размерности. – М.:Финанси и статистика, 1989. – 607 с.
9. Савчук Т.О. Оцінювання результатів моделювання процесу кластерного аналізу надзвичайних ситуацій на залізничному транспорті./Савчук Т.О., Петришин С.І.// Інформаційні технології та комп'ютерна

References

1. Savchuk T.O. Evaluation of emergency situations in rail transport, based on cluster analysis./ Savchuk T.O., Petrishin S.I. // Proceedings XXXIX scientific conference of the faculty, staff and students with the participation of research organizations and engineering and technical personnel of enterprises Vinnitsa region. VNYU - Vinnitsa - 2010.
2. TOROS http://www.trim.ru/docs/toros_mining.pdf
3. Hitachi Diagnostic Kit (EDL) http://www.autoscaners.ru/catalogue/?catalogue_id=751
4. ULTRA-X, Inc. <http://www.uxd.com/>
5. Launch X431 IV http://www.autointhebox.com/launch-x431-iv-master-scanner_c30
6. Hanascan 10 http://www.autoscaners.ru/catalogue/?catalogue_id=hanascan_10_multimarochnyy_skaner
7. Barseghyan A.A. Data analysis Technology: Data Mining, Visual Mining, Text Mining, OLAP / A. A. Barseghyan, M.S. Kupryyanov, V. Stepanenko, I. cold. - 2nd ed. - St. Petersburg.: BHV, St. Petersburg, 2007. - 384p.: Ref. - ISBN 5-94157-991-8.
8. Ayvazian S.A., Buchstaber V., Enyukov I.S. Applied statistics: Classification and reduction of dimension. - Moscow: Finance and Statistics, 1989. - 607 p.
9. Savchuk T.O. Evaluation results modeling process, cluster analysis of emergency situations in rail transport. / Savchuk T.O., Petrishin S.I.// Information technology and computer engineering. - 2012. - № 1, p. 18-24.

Рецензія/Peer review : 18.5.2014 р.

Надрукована/Printed :25.6.2014 р.

УДК 007:621.391:681.3

В.А. ВЫШИНСКИЙ

Институт кибернетики им. В.М. Глушкова НАН Украины

ТРУДНОСТИ РАЗВИТИЯ ВЫЧИСЛИТЕЛЬНОЙ ТЕХНИКИ В УСЛОВИЯХ КРИЗИСА ФИЗИКИ

Развитие Computer science сегодня затребовало знаний существования материи на нано уровне. Исследования показали, что современная физика не в состоянии удовлетворить это «не праздное» любопытство. В статье рассматриваются причины тому. Особое внимание обращено на нерешенность шестой проблемы Д. Гильберта, которая позволила бы выйти из такой сложной ситуации в естественных науках. В статье приводится разрешение этой шестой проблемы.

Ключевые слова: вычислительная техника, физика.

V.A. VYSHINSKIY

V.M. Glushkov Institute of cybernetics of National academy of Science of Ukraine

DIFFICULTIES IN THE DEVELOPMENT OF IN THE COMPUTER SCIENCE WITHIN THE CRISIS OF PHYSICS

Abstract — The contemporary progress of Computer science requires knowledge of existence of matter on the nano level. The investigation have shown that the modern physics is unable to satisfy this not "idle" curiosity. The reasons of this are examined in the paper. The special attention is paid to the unsolved D. Gilbert sixth problem that would allow to go out from such difficult situation in natural sciences. To the article permission of this sixth problem is driven.

Keywords: computer science, physics.

Будущее вычислительной техники (ВТ) тесно связано с развитием технологий манипулирования веществом на нано уровне. Это обусловлено потребностями научно-технического прогресса в наращивании объемов и качества обработки информации, которые могут быть достигнуты только при увеличении интеграции вычислительных схем в аппаратуре. Ведь, указанная интеграция тесно связана с дальнейшим уменьшением в объеме и затратах на потребление энергии элементарными вычислительными схемами по сравнению с современной микро схемной их реализацией, и которую можно достичь уже на нано уровне существования материи. В то же время, по-видимому, повторение технологии обработки информации, на которой зиждиться сегодня ВТ, на таком измельченном уровне, является, по крайней мере, не серьезной. Настало время, когда разработки должны опираться на другие информационные процессы, в основе которых положены операции, взятые у принципиально новых универсальных алгоритмических систем, и в которых информационное содержимое операнда кардинально отличается от бита, байта, действительного числа. Ведь эти незыблемые в обработке единицы информации настолько малы, что при их отображении в материальные сгустки на нано уровне возникают проблемы, разрешение которых на современном этапе развития технологий не возможно.

Одна из таких проблем уже сегодня стала проявляться в попытках реализации новых идей, связанных с созданием квантового компьютера. Дело в том, что энергетическая величина сигнала, с помощью которой представляется такая малая единица информации как q-бит, аналог бита в квантовом компьютере, в нано аппаратуре: для ее хранения, транспортировки и обработки, становится соизмеримой с тепловым шумом. Тогда в процессе квантовых вычислений возникает проблема обнаружения в этом шуме рабочего сигнала, разрешение которой в ближайшие 100 лет не предвидится. Таким образом, ставится под большое сомнение создание работоспособного квантового компьютера.

Чтобы выработать конструктивные предложения по информационному содержимому операнда машины нано уровня следует, как бы, еще раз обратиться к термину информация, определений которой