

МОДЕЛЬ ПРОЦЕСУ ЧІТКОЇ КЛАСТЕРИЗАЦІЇ ПРОБЛЕМНИХ СИТУАЦІЙ В ОБЧИСЛЮВАЛЬНІЙ ТА ОРГАНІЗАЦІЙНІЙ ТЕХНІЦІ

Савчук Тамара, Петришин Сергій

Вінницький національний технічний університет

Анотація

Наведена модель процесу чіткої кластеризації проблемних ситуацій в обчислювальній та організаційній техніці, що являє собою сукупність відомостей про даний процес та всі чинники, які можуть вплинути на нього, зокрема інформаційні моделі проблемних ситуацій та їх станів, метрику, яка буде використовуватись при визначенні відстаней між такими ситуаціями та спосіб нормування значень параметрів.

Abstract

The following model of clustering problem situations in computer and organizational technique which is a collection of information about the process and all the factors that may affect it, including information model of problem situations and their conditions metrics to be used in determining the distances between such situations and the method of normalization values.

Вступ

У зв'язку з швидким науково-технічним прогресом значно збільшилась кількість обчислювальної та організаційної техніки (ООТ), яка застосовується і в побуті, і на виробництві. У зв'язку з цим зростає ймовірність виникнення ситуацій, які містять протиріччя та не мають однозначного вирішення відносно обставин і умов, в яких розгортається діяльність таких систем, тобто, проблемних ситуацій (ПС). У зв'язку з цим, постає необхідність у розробці методів та засобів, які б могли допомогти кластеризувати такі ситуації для підвищення ефективності їх усунення [1, 2].

Постановка задачі

Нехай

$X = \{x_1, \dots, x_n\}$ – вибірка ПС та/або їх станів, що мають місце в ООТ,

де x_1, \dots, x_n - ПС в ООТ,

$n = \overline{1, \infty}$ - кількість ПС в вибірці.

Тоді задача кластеризації ПС в ООТ зведеться до задачі розбиття означеної вибірки на k непересічних кластерів, так, щоб

- кожен кластер складався з ситуацій, які схожі між собою,
- ПС в ООТ, які знаходяться в різних кластерах значно відрізнялися [3].

Модель процесу чіткої кластеризації ПС в ООТ – сукупність відомостей про сам процес та всі чинники, які можуть вплинути на нього, організовані за певними правилами [4].

Виходячи з визначення можна зробити висновок про те, що для її розробки необхідно:

- 1) розробити інформаційну модель (ІМ) ПС та їх станів в ООТ;
- 2) визначити чітку кластеризацію ПС в ООТ як задачу кластеризації Data Mining;
- 3) визначити функції відстаней між ПС в ООТ;
- 4) провести нормування значень параметрів при кластерному аналізі ПС в ООТ.

Інформаційна модель проблемних ситуацій та їх станів в обчислювальній та організаційній техніці

ІМ ПС, яка має місце в ООТ визначається кортежем:

$$PS = \langle SP, DP, EP, A, CS, E \rangle$$

де SP – статичні параметри ООТ, в якій має місце ПС;

DP – динамічні параметри ООТ, в якій має місце ПС;

EP – параметри зовнішнього середовища;

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

CS – множина характеристик стану ПС в ООТ;

E – множина наслідків, до яких призвела ПС в ООТ.

Множину характеристик поточного стану ПС, яка має місце в ООТ, а отже, і ІМ стану ПС можна визначити кортежем:

$$CS = \langle RP, WP, WF, P, PD \rangle$$

де RP – множина коректних значень параметрів ООТ;

WP – множина некоректних значень параметрів ООТ;

WF – множина функцій, виконання яких порушені;

P – періодичність повторення неправильних станів;

PD – наявність явних фізичних пошкоджень ООТ.

Чітка кластеризація проблемних ситуацій в обчислювальній та організаційній техніці як задача кластеризації Data Mining

Серед проаналізованих ієрархічних та неієрархічних методів кластеризації Data Mining, які можуть бути використані в процесі чіткої кластеризації ПС в ООТ було з'ясовано, що прийнятним методом для розв'язання поставленої задачі є неієрархічний метод, що характеризується:

- використання мінімально можливої кількості характеристик та параметрів таких ситуацій при виконанні кластеризації і отримання при цьому максимально можливого якісного результату;

- можливість виконувати максимально якісний аналіз при використанні невеликої кількості попередньо кластеризованих ситуацій та їх станів, вся інформація про які містяться у сховищі даних, але і в свою чергу можливість працювати з потужними БД;

- простотою;

- наочністю;

- високою швидкістю.

Функція відстані, яка буде використовуватись при чіткій кластеризації проблемних ситуацій в обчислювальній та організаційній техніці

Серед проаналізованих метрик: Евклідова відстань, квадрат Евклідової відстані, зважену Евклідову відстань, відстань за Хемінгом, пікова відстань, прийнятною для чіткої кластеризації ПС в ООТ є зважена Евклідова відстань, оскільки вона враховує «важливість» кожної характеристики такої ситуації, що підвищує якість аналізу, що призведе до підвищення якості кластеризації [5]:

$$a_{3E}(X_i, X_j) = \sqrt{\lambda_1 \cdot (x_{i1} - x_{j1})^2 + \lambda_2 \cdot (x_{i2} - x_{j2})^2 + \dots + \lambda_m \cdot (x_{im} - x_{jm})^2}$$

де $a_{3E}(X_i, X_j)$ – зважена Евклідова відстань між двома ПС в ООТ X_i та X_j ;

$\lambda_1, \lambda_2, \dots, \lambda_m$ ($0 \leq \lambda_l \leq 1$ ($l = \overline{1, m}$))) – вектор значень вагових коефіцієнтів, які відповідають характеристикам $x_{i1}, x_{i2}, \dots, x_{im}$ ПС в ООТ;

$x_{i1}, x_{i2}, \dots, x_{im}$ – вектор значень характеристик, який описує i -ту ПС в ООТ;

$x_{j1}, x_{j2}, \dots, x_{jm}$ – вектор значень характеристик, який описує j -ту ПС в ООТ.

Для визначення вектора значень вагових коефіцієнтів $\lambda_1, \lambda_2, \dots, \lambda_m$ використовують

навчальні вибірки ПС в ООТ або досвід експертів.

Нормування значень параметрів при кластерному аналізі проблемних ситуацій в обчислювальній та організаційній техніці

Нормування – це коригування вектора значень характеристик ПС в ООТ відповідно до деяких функцій перетворення для спрощення їх порівняння. Нормування даних: потрібне, коли несумісність одиниць вимірювання змінних може відбитися на результатах кластерного аналізу (обчислення, засновані на змішаних добутках), та рекомендується в тих випадках, коли результати аналізу можуть бути поліпшені, якщо виразити результати в певних сумісних одиницях [6].

Серед проаналізованих способів нормування прийнятним для чіткої кластеризації ПС в ООТ є спосіб:

$$q = \frac{x}{x'}$$

де q – нормоване значення параметра u ;

x – поточне значення параметра ПС в ООТ;

x' – нормативне (еталонне) значення параметра x ПС в ООТ.

Наведений спосіб доцільно використовувати оскільки для його застосування достатньо знати тільки поточні та еталонні значення параметрів та воно не призводить до зміни результатів розв'язку задачі кластеризації, що є важливим при кластеризації ПС в ООТ.

Висновки

Таким чином, наведена модель, що являє собою сукупність відомостей про даний процес та всі чинники, які можуть вплинути на нього, зокрема ІМ ПС та їх станів в ООТ, зважену Евклідову відстань, як метрику, яка буде використовуватись при визначенні відстаней між такими ситуаціями та спосіб нормування значень параметрів може бути використаною при процесу чіткої кластеризації ПС в ООТ.

Список використаних джерел:

3. В. Е. Афанасьевская Применение методов нечеткой кластеризации при решении задач диагностики авиационных двигателей/ В.Е. Афанасьевская, О. С. Радивоненко, А. Ю. Соколов, М. Л. Угрюмов // Авиационно-космическая техника и технология - Харьков, 2010 - № 8 (75), С. 128-132.

4. Т. О. Савчук 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” - Львів, 2014 - №1, С. 193-198.

5. 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.

6. Юхимчук С.В. Моделі автоматизації вироблення рекомендацій керівнику гасіння пожежі на залізничному транспорті / Юхимчук С.В., Кацман М.Д. : Монографія. – Вінниця: УНІВЕРСУМ-Вінниця, 2008. -144 с. - ISBN 978-966-641-235-8.

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. Савчук Т.А. Нормирование значений параметров при кластер-анализе чрезвычайных ситуаций на железнодорожном транспорте / Т.А. Савчук, С. И. Петришин // «Информационные и компьютерные технологии, моделирование, управление», Международная научная конференция посвящённая 80-летию со дня рождения академика И.В. Прангишвили. – Грузия, Тбилиси, 1 - 4 ноября 2010 – С. 165.