DOI 10.31891/2307-5732-2022-311-4-176-179
УДК 004.031.6:621.3.07

ОСИПЕНКО В. В.
Київський національний університет технологій та дизайну
https://orcid.org/0000-0002-1077-1461
e-mail: vvo7@ukr.net
ЗЛОТЕНКО Б. М.
Київський національний університет технологій та дизайну
http://orcid.org/0000-0002-0870-8535
e-mail: zlotenco@ukr.net
КУЛІК Т. І.
Київський національний університет технологій та дизайну
http://orcid.org/0000-0002-1006-7853
e-mail: t-81@ukr.net
БІЛА Т. Я.
Київський національний університет технологій та дизайну
http://orcid.org/0000-0001-8937-5244
e-mail: bila.ty@knutd.edu.ua
ДЕМІШОНКОВА С. А.
Київський національний університет технологій та дизайну
https://orcid.org/0000-0001-5678-8114
e-mail: mashuk2007@ukr.net

ЗОВНІШНІЙ КРИТЕРІЙ СТАБІЛЬНОСТІ ВНУТРІШНЬОМНОЖИННИХ ВІДСТАНЕЙ В ЗАДАЧАХ ДІАГНОСТУВАННЯ СТАНІВ ТЕХНІЧНИХ ОБ'ЄКТІВ

Метою дослідження є адаптація критеріїв бікластеризації, які володіють властивостями зовнішнього доповнення для задач класифікації станів складних технічних об'єктів в сфері комп'ютерної інженерії. В роботі застосовані принципи індуктивного моделювання складних систем, зокрема, принципу зовнішнього доповнення, а також методологія теорії розпізнавання образів, методи індуктивного кластерного аналізу, математична статистика. Конструювання критерію стабільності внутрішньо множинних відстаней базується на схемах відомих критеріїв самоорганізації моделей, зокрема критерію несуперечностей моделей, які мають широке застосування в багатокрокових алгоритмах розпізнавання образів з інтелектуальним вибором оптимальних результатів. В роботі запропоновано застосування критерію стабільності внутрішньо множинних відстаней в задачах діагностування станів технічних об'єктів, зокрема, у сфері комп'ютерної інженерії. Оскільки для застосування такого критерію необхідна наявність цільової ознаки, адаптовано один із варіантів розбиття вихідної експериментальної бази даних на дві підмножини: підмножину цільових ознак і підмножину вхідних параметрів. Поняття внутрішньо множинних відстаней поширено на застосування в критеріях алгоритмів самоорганізації моделей оптимальної складності. Інтелектуальні алгоритми самоорганізації моделей оптимальної складності можуть бути застосовані для підвищення надійності експлуатації комп'ютерних систем.

Ключові слова: критерій, принцип зовнішнього доповнення, кластеризація, алгоритм, комп'ютерна система, самоорганізація моделей, інженерія.

Volodymyr OSYPENKO, Borys ZLOTENKO, Tetyana KULIK, Tatyana BILA., Svitlana DEMISHONKOVA
Kyiv National University of Technologies and Design

EXTERNAL CRITERION OF STABILITY OF INTRA-MULTIPLE DISTANCES IN TASKS OF DIAGNOSIS OF STATES OF TECHNICAL OBJECTS

The purpose of the research is adaptation of biclasticization criteria, which have the properties of external complement to the problems of classification of states of complex technical objects in the field of computer engineering. The principles of inductive modelling of complex systems, in particular, the principle of external complementarity, as well as the methodology of the theory of pattern recognition, methods of inductive cluster analysis, mathematical statistics are applied. The construction of the criterion of stability of intra-multiple distances is based on the schemes of known criteria of self-organization of models, in particular the criterion of model inconsistencies, which are widely used in multi-step image recognition algorithms with intelligent choice of optimal results.

The paper proposes the application of the criterion of stability of intra-multiple distances in the problems of diagnosing the state of technical objects, in particular, in the field of computer engineering. Since the application of such a criterion requires the presence of a target feature, one of the options for splitting the original experimental database into two subsets has been adapted: a subset of target features and a subset of input parameters. The concept of intra-multiple distances is extended to the application of algorithms of optimal complexity in the criteria of self-organization algorithms. Intelligent algorithms for self-organization of models of optimal complexity can be used to increase the reliability of computer systems.

Keywords: criterion, principle of external complement, clustering, algorithm, computer system, self-organization of models, engineering.

Постановка проблеми у загальному вигляді та її зв'язок із важливими науковими чи практичними завданнями

Відомо, що в алгоритмах самоорганізації моделей, а за сучасною термінологією – індуктивного моделювання складних систем (ІМСС), який базується на добре відомому в сфері інтелектуальних технологій і системного аналізу Методу групового урахування аргументів (МГУА) повинні застосовуватися критерії, які володіють властивостями зовнішнього доповнення [1–3]. Їх застосування з необхідністю

вимагають розділення множини вхідних об'єктів на дві підмножини з метою незалежного конструювання моделей на виділених підмножинах. Робота так званого зовнішнього критерію в цьому випадку полягає у співставленні за певними правилами найважливіших характеристик індуктивних моделей для подальшого вибору кращих (або єдиної оптимальної) з них.

Аналіз останніх джерел

В роботі мова йде про один із найважливіших атрибутів будь-якого математичного моделювання – критеріїв відбору (селекції) кращих результатів і самоорганізація моделей в алгоритмах кластер-аналізу не є виняток. Для вирішення завдання кластеризації за методологією ІМСС тут розглядається певний аналог критерію мінімуму зміщення виходів моделей, отриманих незалежно на виділених попередньо підмножинах, який ще називають критерієм несуперечності [1, 4]. Знаходження мінімуму критерію селекції дозволяє конструювати оптимальну підмножину ознак ("ансамбль" інформативних ознак) [5], що дає можливість установити компактні утворення – кластери – на вибірковому масиві об'єктів.

Формулювання цілей статті

Метою роботи є адаптація критеріїв бікластеризації, які володіють властивостями зовнішнього доповнення для задач класифікації станів складних технічних об'єктів в сфері комп'ютерної інженерії.

Виклад основного матеріалу

Вище сказано, що для застосування індуктивних технологій індуктивного моделювання складних систем вхідну множину необхідно розділити на дві рівні частини з однаковими статистичними характеристиками. Найпростіше, це можна зробити, розділяючи вхідну вибірку на частини за правилом "парний-непарний" елемент, хоча можна застосовувати й інші, більш складні алгоритми.

Критерій несуперечності кластеризацій. Нехай основною характеристикою кластера буде середньоквадратична похибка [6]. Вираз для критерію мінімуму зміщення запишемо так:

$$n_{cons} = \frac{\left| \sum_{i=1}^{K} \sum_{j \in O_2} \left\| \omega_j - m_i^{o_1} \right\| - \sum_{i=1}^{K} \sum_{j \in O_1} \left\| \omega_j - m_i^{o_2} \right\| \right|}{\sum_{i=1}^{K} \left(\sum_{j \in O_2} \left\| \omega_j - m_i^{o_1} \right\| + \sum_{j \in O_1} \left\| \omega_j - m_i^{o_2} \right\| \right)} \Rightarrow \min, \quad i = 1, 2, ..., K, \tag{1}$$

де K – число виділених кластерів, $m_i^{o_1}$ і $m_i^{o_2}$ – центри кластерів, виділених на підмножинах O_1 і O_2 відповідно, ω_j – j-те зображення вибіркової множини.

Для того, щоб суми модулів відхилень

$$\sum_{i=1}^{K} \sum_{j \in O_2} \left\| \omega_j - m_i^{o_2} \right\| \qquad \text{і} \qquad \sum_{i=1}^{K} \sum_{j \in O_2} \left\| \omega_j - m_i^{o_2} \right\| \tag{2}$$

монотонно зменшувалися і, внаслідок чого, критерій n_{cons} (1) сягав би мінімуму, необхідно вводити в розгляд відповідний порядок встановлення кластерів. Крім того, для застосування критерію (1) необхідно накласти також і умову виділення однакової кількості кластерів на підвибірках O_1 і O_2. Очевидно, що така умова вимагає достатньо великої вибірки вихідних даних для задоволення вимоги стійкості статистичних характеристик. Такі жорсткі умови можна обійти, якщо застосувати Проте це окрема тема, як відноситься більше до побудови алгоритму кластеризації (в даному випадку – бікластеризації) і тут не розглядається.

Внутрішньомножинна відстань у вибіркових множинах. Відомо, що відстані всередині множини K точок в евклідовому просторі $\mathbb{R}^n$ визначається як [7]:

$$\sum_{i=1}^{K} \sum_{j \in O_2} \left\| \omega_j - m_i^{o_2} \right\| \qquad \text{і} \qquad \sum_{i=1}^{K} \sum_{j \in O_2} \left\| \omega_j - m_i^{o_2} \right\|, \tag{3}$$

де ω^i і ω^j $(i \neq j)$ – вектори заданої множини Ω, x_k^i – k-та компонента (ознака) цього вектора.

Частинне середнє між фіксованим об'єктом ω^j і усіма іншими об'єктами $\left\{\omega^i, i = 1, 2, ..., k\right\}$, $i \neq j$ визначається як:

$$d^2(\omega^i, \left\{\omega^j\right\}) = \frac{1}{K-1} \sum_{i=1}^{K} \sum_{k=1}^{n} \left(x_k^i - x_k^j \right)^2. \tag{4}$$

Звідси середня відстань по всіх об'єктах множини Ω, що визначить внутрішньо множинну відстань, можна подати у вигляді:

$$d^2(\left\{\omega^i\right\}, \left\{\omega^j\right\}) = \frac{1}{K} \sum_{j=1}^{K} \left[\frac{1}{K-1} \sum_{i=1}^{K} \sum_{k=1}^{n} \left(x_k^i - x_k^j \right)^2 \right] = \frac{1}{K(K-1)} \sum_{k=1}^{K} \sum_{k=1}^{K} \sum_{k=1}^{n} \left(x_k^i - x_k^j \right)^2. \tag{5}$$

Для використання поняття внутрішньомножинної відстані в критеріях селекції, вираз (5) подамо через вибіркові дисперсії значень компонент векторів образів. Оскільки розглядається одна і та ж вибіркова множина, то $\overline{\left(x_k^i\right)^2} = \overline{\left(x_k^j\right)^2}$. Тепер внутрішньо множинну відстань можна отримати шляхом наступних перетворень:

$$\overline{d^2} = \frac{1}{K-1}\sum_{k=1}^{n}\left[\frac{1}{K^2}\sum_{i=1}^{K}\sum_{j=1}^{K}\left(x_k^i - x_k^j\right)^2\right] = \frac{1}{K-1}\sum_{k=1}^{n}\left[\frac{1}{K^2}\sum_{i=1}^{K}\sum_{j=1}^{K}\left(x_k^i\right)^2 - \frac{2}{K^2}\sum_{i=1}^{K}\sum_{j=1}^{K}\left(x_k^i x_k^j\right)^2 + \frac{1}{K^2}\sum_{i=1}^{K}\sum_{j=1}^{K}\left(x_k^j\right)^2\right] =$$
$$= \frac{1}{K-1}\sum_{k=1}^{n}\left[\frac{1}{K}\sum_{i=1}^{K}\overline{\left(x_k^i\right)^2} - 2\overline{\left(x_k^i\right)\left(x_k^j\right)} + \frac{1}{K}\sum_{i=1}^{K}\overline{\left(x_k^j\right)^2}\right] = \frac{2}{K-1}\sum_{k=1}^{n}\left[\overline{\left(x_k^i\right)^2} - \overline{x_k^i}^2\right], \quad i = 1,2,...,K. \tag{6}$$

Так як

$$\overline{\left(x_k^i\right)^2} - \overline{x_k^i}^2 = \tilde{\sigma}_k^2 \tag{7}$$

є зміщена вибіркова дисперсія *k*-ї компоненти для *K* об'єктів множини $\left\{\omega^i, i = 1,2,...,K\right\}$, то

$$\overline{d^2} = \frac{2K}{K-1}\sum_{k=1}^{n}\tilde{\sigma}_k^2 . \tag{8}$$

Враховуючи, що незміщена вибіркова дисперсія задається співвідношенням

$$\sigma_k^2 = \frac{K}{K-1}\tilde{\sigma}_k^2, \ \kappa = 1,...,n . \tag{9}$$

то

$$\overline{d^2} = 2\sum_{k=1}^{n}\sigma_k^2 . \tag{10}$$

Приймаючи до уваги те, що в моделюванні приймають участь об'єкти виключно із однієї статистично однорідної вибірки, вираз (10) перепишемо у вигляді:

$$d_o = 2\sigma_0^2 , \tag{11}$$

де

$$\sigma_0^2 = \frac{1}{K-1}\sum_{i=1}^{K}\left(x_i^0 - \overline{x}^0\right)^2 - \tag{12}$$

незміщена вибіркова дисперсія множини точок.

Критерій стабільності внутрішньомножинних відстаней. Нехай вихідні експериментальні дані представлені у вигляді матриці:

$$\left\{X\right\} = \begin{pmatrix} x_{o1} & x_{11}...x_{i1}...x_{N1} \\ ... & ... \quad ... \quad ... \\ x_{oj} & x_{1j}...x_{ij}...x_{Nj} \\ ... & ... \quad ... \quad ... \\ x_{om} & x_{1m}...x_{im}...x_{Nm} \end{pmatrix} \tag{13}$$

Під вибірками O_1 і O_2 будемо розуміти наступне. Як на вибірці O_1 так і на вибірці O_2 кластеризації підлягають усі об'єкти вихідної множини $\left\{X\right\}$. Різниця тут має місце в тому, що кластеризація на O_1 здійснюється на основі всіх можливих ансамблів ознак [1] без участі цільової ознаки, а на O_2 – тільки по цільовій ознаці x_o. При цьому для можливості співставлення кластеризацій на O_1 центри кластерів обчислюються за значеннями цільових ознак образів, попавши в окремі кластери. Таким чином, на вибірці O_1 число кластеризацій дорівнює кількості усіх можливих комбінацій ознак в ансамблях, а на O_2 тільки одна, яка відповідає кластеризації по осі x_o.

Дамо наступне *означення*: функціонал $\rho(\sigma)$, що відображає рівність внутрішньомножинних відстаней по *K* кластерах на вибірках O_1 і O_2 відповідно, називається критерієм внутрішньомножинних відстаней.

Використовуючи формальний запис для критерію несуперечності, а також (11), критерієм внутрішньомножинних відстаней запишемо наступним чином:

$$\rho(\sigma) = \frac{\left|\sum_{i=1}^{K}\left(\delta_{0i}^2\right)_{O_1} - \sum_{i=1}^{K}\left(\delta_{0i}^2\right)_{O_2}\right|}{\sum_{i=1}^{K}\left(\delta_{0i}^2\right)_{O_1} + \sum_{i=1}^{K}\left(\delta_{0i}^2\right)_{O_2}} \Rightarrow \min . \tag{14}$$

Нехай на O_1 і O_2 виділено однакова кількість кластерів. Тоді, якщо на основі деякого ансамблю ознак $\left\{x_1,...,x_{n^*}\right\}$ множина об'єктів класифікуються таким чином, щоб в кожний з *K* кластерів на O_1 попали ті ж об'єкти, що і у випадку кластеризації по цільовій ознаці x_o, то із (11) і попереднього означення маємо:

$$\sum_{i=1}^{K}\left(\delta_{0i}^2\right)_{O_1} - \sum_{i=1}^{K}\left(\delta_{0i}^2\right)_{O_2} = 0 . \tag{15}$$

Ця важлива властивість характерна і для критерію мінімуму зміщення, який має широке застосування в МГУА. Очевидно, що на добитися ідеального виконання рівності нулю в (15) і, тим більше за наявності шумів і недостатніх об'ємів інформації, на практиці не реально. Тому, нехай $\mathfrak{I}$ – підмножина кращих значень критерію (16). Тоді вибору підлягає ансамбль (підпростір ознак $n^* \le N$), для якого:

$$\rho^*(\sigma) = \min_{\mathfrak{I}} \{\rho(\sigma)\}. \quad (16)$$

Поданий критерій внутрішньомножинних відстаней створений спеціально для застосувань в індуктивних алгоритмах кластеризації. Такі алгоритми призначені не лише для отримання розбиття вхідної множини векторів-образів на кластери, але й для конструювання підпросторів інформативних ознак [8–10].

Висновки з даного дослідження і перспективи подальших розвідок у даному напрямі

В роботі розглянуто критерій внутрішньомножинних відстаней для застосувань в інтелектуальних алгоритмах кластеризації ІММС. Індуктивні алгоритми є важливими інструментами в проблемах діагностування складних технічних об'єктів і технологічних процесів. Такими об'єктами досліджень не в останню чергу є комп'ютерні системи і мережі в процесах експлуатації в складних умовах навколишнього середовища, за коротких вибірок вхідної зашумленої інформації. Описаний критерій застосовується зазвичай системно з іншими критеріями індуктивної самоорганізації, наприклад критерієм стабільності міжмножинних відстаней [8], критерієм балансу [1, 2] в різних варіаціях та ін.

Література

1. Ивахненко А. Г. Индуктивный метод самоорганизации моделей сложных систем / А. Г. Ивахненко. – К. : Наукова думка, 1981. – 296 с.

2. Ivakhnenko A. G. Inductive learning algorithms for complex systems modeling / Ivakhnenko A. G., Madala H. R. – New York : Boca Raton, CRC Press, 1994. – 384 p.

3. Лур'є І. А. Гібридизація алгоритму індуктивного кластер-аналізу з використанням оцінки щільності розподілу даних [Електронний ресурс] / Лур'є І. А., Осипенко В. В., Литвиненко В. І., Таиф М. А., Корніловська Н. В. – Lviv Polytechnic National University Institutional Repository, 2015. – URL : http://ena.lp.edu.ua.

4. Осипенко В. В. Два підходи до розв'язання задачі кластеризації у широкому сенсі з позицій індуктивного моделювання / В. В. Осипенко // Вісник НУБіП України. Сер. Енергетика і автоматика. – 2014. – № 1. – С. 83–97. – URL : http://nbuv.gov.ua/j-pdf/eia_20141_11.pdf.

5. Duda R. O. Pattern Classification, 2nd Edition / Duda R. O., Hart P. E., Stork D. G., John. – New York : Wiley & Sons, 2001. – 738 p.

6. Сеньо П. С. Теорія ймовірностей та математична статистика : підручник / Сеньо П. С. – 1-е вид. – К. : Центр навчальної літератури, 2004. – 448 с.

7. Ту Дж. Принципи розпізнавання образів / Ту Дж., Гонсалес Р. – М. : Мир, 1978. – 414 с.

8. Васильев В. И. Распознающие системы : справочник / Васильев В. И. – 2-е изд. перераб. и доп. – К. : Наукова думка, 1983. – 422 с.

9. Babichev S. Implementation of the objective clustering inductive technology based on DBSCAN clustering algorithm / S. Babichev, V. Lytvynenko, V. Osypenko // 2017 IEEE 12th Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT): Proceedings, Sept. 05-08, 2017. – Lviv, 2017. – P. 479–484. DOI: 10.1109/STC-CSIT.2017.8098832.

10. Орешков В. EM – масштабируемый алгоритм кластеризации [Електронний ресурс] / Орешков В. – URL : https://loginom.ru/blog/em.

References

1. Ivakhnenko A. H. Ynduktyvnyi metod samoorhanyzatsiyi modelei slozhnykh system / A. H. Yvakhnenko // Kyev: Naukova dumka, 1981. – 296 p.

2. Ivakhnenko A. G. Inductive learning algorithms for complex systems modeling / Ivakhnenko A. G., Madala H. R. – New York : Boca Raton, CRC Press, 1994. – 384 p.

3. Lur'ie I. A. Hibrydyzatsiya alhorytmu induktyvnoho klaster-analizu z vykorystanniam otsinky shchilnosti rozpodilu danykh / Lur'ie I. A., Osypenko V. V., Lytvynenko V. I., Tayf M. A., Kornilovska N. V. – Lviv Polytechnic National University Institutional Repository. – 2015. – URL: http://ena.lp.edu.ua.

4. Osypenko V. V. Dva pidkhody do rozv'iazannia zadachi klasteryzatsii u shyrokomu sensi z pozytsii induktyvnoho modeliuvannia / V. V. Osypenko // Visnyk NUBiP Ukrainy. Ser. Enerhetyka i avtomatyka. – 2014. – № 1. – P. 83-97.

5. Duda R. O. Pattern Classification, 2nd Edition / Duda R. O., Hart P. E., Stork D. G., John. – New York : Wiley & Sons, 2001. – 738 p.

6. Seno P. S. Teoriya ymovirnostei ta matematychna statystyka. Pidruchnyk, 1-e vyd. / Seno P. S. – K. : Tsentr navchalnoi literatury, 2004. – 448 p.

7. Tu Dzh. Pryntsypy rozpiznavannia obraziv / Tu Dzh., Honsales R. – M.: Myr, 1978. – 414 p.

8. Vasylev V. Y. Raspoznaiushchye system: spravochnyk / Vasylev V. Y. – 2-e izd., pererab. y dop. – K. : Naukova dumka, 1983. – 422 p.

9. Babichev S. Implementation of the objective clustering inductive technology based on DBSCAN clustering algorithm / S. Babichev, V. Lytvynenko, V. Osypenko // 2017 IEEE 12th Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT): Proceedings, Sept. 05-08, 2017. – Lviv, 2017. – P. 479–484. DOI: 10.1109/STC-CSIT.2017.8098832.

10. Oreshkov V. EM – masshtabiruemyi algoritm klasterizatsii. URL: https://loginom.ru/blog/em.

Рецензія/Peer review : 13.06.2022 р. Надрукована/Printed :02.08.2022 р.