

УДК 004.93

[0000-0003-2655-6667] **І. І. Коваленко**, *д.т.н., професор*,[0000-0002-0547-3689] **Є. О. Давиденко**, *к.т.н.*,

e-mail: davydenko@chmnu.edu.ua

[0000-0003-4372-7472] **А. В. Швед**, *к.т.н.*Чорноморський національний університет імені Петра Могили  
вул. 68 Десантників, 10, Миколаїв, 54003, Україна

## МЕТОДИКА ПОШУКУ АСОЦІАТИВНИХ ПРАВИЛ

З появою великих об'ємів інформації, що зберігається, актуальними стали задачі, пов'язані з необхідністю їх обробки. Дані, що накопичуються, характеризуються невпорядкованістю та неструктурованістю, коли кожна одиниця зберігання не може бути представлена скінченною кількістю ознак. Для обробки великих масивів неструктурованих даних в останні роки широко використовуються методи пошуку асоціативних правил. Проблема полягає в тому, що кількість можливих асоціацій зі збільшенням кількості предметів у кожній з транзакцій збільшується експоненціально та потребує значних обчислювальних затрат. Тому в процесі формування асоціативних правил широко використовуються методики, що дають змогу зменшити кількість асоціацій, які потрібно проаналізувати.

У роботі запропоновано підхід до формування вибірки предметів, найбільш характерних для спрощеного масиву транзакцій, на основі якої формуються асоціативні бінарні відношення та розраховуються їх характеристики для визначення того, чи є такі відношення правилами. Для цього спочатку масив стискається за допомогою пошуку транзакцій з однаковими предметними наборами з використанням відношення еквівалентності. Потім, для виявлення транзакцій з предметами, що часто повторюються, виконується попарний перетин предметних наборів.

**Ключові слова:** транзакція, асоціації, асоціативні правила, масив даних.

**Вступ.** Нині у зв'язку зі швидким зростанням об'ємів інформації, що зберігається, виникають задачі, пов'язані з необхідністю їх обробки з метою пошуку нових закономірностей та встановлення і виявлення нових знань. Такі задачі з'являються на промислових підприємствах, а також в організаціях, що займаються роздрібною торгівлею, фінансовим аналізом, логістикою та комунікаціями [1]. Для аналізу даних широко використовуються методи штучного інтелекту (нечіткі моделі, нейронні мережі, дерева рішень, байєсівські мережі та ін.), однак такі методи, як правило, використовуються для обробки структурованих даних, що представляються у вигляді масивів, які містять значення ознак та вихідних даних параметрів елементів вибірки даних [2–4].

У той же час дані, що накопичуються, характеризуються невпорядкованістю та неструктурованістю, коли кожна одиниця зберігання не може бути представлена скінченною кількістю ознак (атрибутів) [1]. Такі дані представляють, як правило, у вигляді транзакцій (множина подій, що відбуваються одночасно). Для обробки великих масивів неструк-

турованих даних в останні роки широко використовуються методи пошуку асоціативних правил, що дозволяють виявляти нові закономірності типу «якщо умова, то наслідок». Найпростіший алгоритм пошуку асоціативних правил розглядає всі можливі комбінації умов та наслідків, оцінює для них підтримку та достовірність, а потім виключає всі асоціації, що не задовольняють заданим обмеженням.

Проблема тут полягає в тому, що кількість можливих асоціацій зі збільшенням кількості предметів у кожній з транзакцій збільшується експоненціально та потребує значних обчислювальних затрат. Тому в процесі формування асоціативних правил широко використовуються методики, що дають змогу зменшити кількість асоціацій, які потрібно проаналізувати.

**Аналіз публікацій та останніх досягнень.** У низці публікацій [5–10 та ін.] розглядаються питання розробки методів, методик та алгоритмів пошуку асоціативних правил. Так, у роботі [1] виконаний достатньо глибокий огляд існуючих підходів для пошуку числових, узагальнених, часових та нечітких асоціативних правил. Робота [5] присвячена порів-

няльному аналізу інструментарію пошуку асоціативних правил (алгоритми Apriori, DHP, Partition, DIC та ін.). В роботі [8] розглядається застосування алгоритмів асоціативних правил для прикладної задачі, пов'язаної з продажем товарів будівельної галузі.

Робота [6] присвячена комбінованому застосуванню асоціативного аналізу та методу дерев рішень для вирішення задач економічної галузі. В цілому слід відзначити, що сучасні інструментальні засоби пошуку асоціативних правил – такі, наприклад, як Apriori, Apriori Scale та ін. – базуються на побудові вербально-числових діапазонів (норма, вище норми, нижче норми) значень предметних наборів, що становлять транзакції. Проте повністю проблему пошуку асоціативних правил існуючі методи та алгоритми не знімають, що потребує розробки нових підходів і рішень. Один із можливих таких підходів розглянуто в цій роботі.

**Метою роботи** є розробка методики пошуку асоціативних правил на основі попереднього аналізу і стиснення масиву транзакцій, що представляють різні предметні набори.

**Виклад основного матеріалу.** Нехай маємо масив транзакцій  $T_r = T_{r_1}, T_{r_2}, \dots, T_{r_i}, \dots, T_{r_n}$ , що представлений відповідними предметними наборами  $X_1, X_2, \dots, X_i, \dots, X_n$ . На першому етапі такий масив може бути стиснений за допомогою пошуку транзакцій з однаковими предметними наборами з використанням відношення еквівалентності ( $\sim$ ), наприклад,  $(X_2 \sim X_3 \sim X_6) = X_2^*$ . Тоді вихідний масив матиме вигляд  $X_1, X_2^*, \dots, X_i, \dots, X_n$ . Далі для виявлення транзакцій із предметами, що часто повторюються, виконується попарний перетин предметних наборів  $X_1 \cap X_2^*, X_1 \cap X_i, \dots, X_1 \cap X_n; X_2^* \cap X_i, X_2^* \cap X_n; X_i \cap X_n$ . За результатами отриманих перетинів може бути сформована вибірка предметів найбільш характерних для спрощеного масиву транзакцій, на основі якої формуються асоціативні бінарні відношення та розраховуються їх характеристики для визначення того, чи є такі відношення правилами.

Розглянемо запропонований підхід на наступному прикладі. Нехай виконано  $n=10$  транзакцій ( $T_r$ ) за вихідним предметним набором  $X_0$  обсягом  $m=12$  предметів:

$$X_0 = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}, x_{12}\}.$$

Отримані при цьому транзакції мають наступний вигляд:

$$T_{r_1}: X_1 = \{x_2, x_5, x_6\};$$

$$T_{r_2}: X_2 = \{x_1, x_3, x_5, x_8\};$$

$$T_{r_3}: X_3 = \{x_4, x_5, x_6, x_7, x_8, x_9\};$$

$$T_{r_4}: X_4 = \{x_1, x_3, x_5, x_8\};$$

$$T_{r_5}: X_5 = \{x_2, x_5, x_6\};$$

$$T_{r_6}: X_6 = \{x_3, x_4, x_5, x_6, x_9, x_{11}\};$$

$$T_{r_7}: X_7 = \{x_2, x_5, x_6\};$$

$$T_{r_8}: X_8 = \{x_7, x_8, x_9, x_{10}, x_{12}\};$$

$$T_{r_9}: X_9 = \{x_1, x_3, x_5, x_8\};$$

$$T_{r_{10}}: X_{10} = \{x_7, x_8, x_9, x_{10}, x_{11}, x_{12}\}.$$

Припустимо, що в загальну кількість предметних наборів  $X_i$  можуть входити набори як з однаковими предметами, так і з різним асортиментом предметів. Тоді перша підгрупа транзакцій може бути об'єднана відношенням еквівалентності. В нашому прикладі це наступні транзакції:  $X_1^* = (X_1 \sim X_5 \sim X_7)$  та  $X_2^* = (X_2 \sim X_4 \sim X_9)$ . Це дає можливість зменшити загальну кількість транзакцій з 10 до 6:  $X_1^*, X_2^*, X_3, X_6, X_8, X_{10}$ . З метою виявлення предметів, що найчастіше повторюються в перерахованих предметних наборах, виконаємо їх попарний перетин:

$$X_1^* \cap X_2^* = \{x_2, x_5, x_6\} \cap \{x_1, x_3, x_5, x_8\} = \{x_5\};$$

$$X_1^* \cap X_3 =$$

$$\{x_2, x_5, x_6\} \cap \{x_4, x_5, x_6, x_7, x_8, x_9\} = \{x_5, x_6\};$$

$$X_1^* \cap X_6 =$$

$$\{x_2, x_5, x_6\} \cap \{x_3, x_4, x_5, x_6, x_9, x_{11}\} = \{x_5, x_6\};$$

$$X_1^* \cap X_8 =$$

$$\{x_2, x_5, x_6\} \cap \{x_7, x_8, x_9, x_{10}, x_{12}\} = \emptyset;$$

$$X_1^* \cap X_{10} =$$

$$\{x_2, x_5, x_6\} \cap \{x_7, x_8, x_9, x_{10}, x_{11}, x_{12}\} = \emptyset;$$

$$X_2^* \cap X_3 =$$

$$\{x_1, x_3, x_5, x_8\} \cap \{x_4, x_5, x_6, x_7, x_8, x_9\} = \{x_5, x_8\};$$

$$X_2^* \cap X_6 =$$

$$\{x_1, x_3, x_5, x_8\} \cap \{x_3, x_4, x_5, x_6, x_9, x_{11}\} = \{x_3, x_5\};$$

$$X_2^* \cap X_8 =$$

$$\{x_1, x_3, x_5, x_8\} \cap \{x_7, x_8, x_9, x_{10}, x_{12}\} = \{x_8\};$$

$$X_2^* \cap X_{10} = \{x_1, x_3, x_5, x_8\} \cap$$

$$\{x_7, x_8, x_9, x_{10}, x_{11}, x_{12}\} = \{x_8\};$$

$$X_3 \cap X_6 = \{x_4, x_5, x_6, x_7, x_8, x_9\} \cap \{x_3, x_4, x_5, x_6, x_9, x_{11}\} = \{x_4, x_5, x_6, x_9\};$$

$$X_3 \cap X_8 = \{x_4, x_5, x_6, x_7, x_8, x_9\} \cap$$

$$\{x_7, x_8, x_9, x_{10}, x_{12}\} = \{x_7, x_8, x_9\};$$

$$X_3 \cap X_{10} = \{x_4, x_5, x_6, x_7, x_8, x_9\} \cap$$

$$\{x_7, x_8, x_9, x_{10}, x_{11}, x_{12}\} = \{x_7, x_8, x_9\};$$

$$\begin{aligned} X_6 \cap X_8 &= \{x_3, x_4, x_5, x_6, x_9, x_{11}\} \cap \\ \{x_7, x_8, x_9, x_{10}, x_{12}\} &= \{x_9\}; \\ X_6 \cap X_{10} &= \{x_3, x_4, x_5, x_6, x_9, x_{11}\} \cap \\ \{x_7, x_8, x_9, x_{10}, x_{11}, x_{12}\} &= \{x_9, x_{11}\}; \\ X_8 \cap X_{10} &= \{x_7, x_8, x_9, x_{10}, x_{12}\} \cap \\ \{x_7, x_8, x_9, x_{10}, x_{11}, x_{12}\} &= \{x_7, x_8, x_9, x_{10}, x_{12}\}; \end{aligned}$$

Об'єднавши результати отриманих перетинів, можна отримати вибірку предметів, що найчастіше зустрічаються в транзакціях.

$$\begin{aligned} X &= \{x_5\} \cup \{x_5, x_6\} \cup \{x_5, x_6\} \cup \\ \{x_5, x_8\} \cup \{x_3, x_5\} \cup \{x_8\} \cup \{x_8\} \cup \\ \{x_4, x_5, x_6, x_9\} \cup \{x_7, x_8, x_9\} \cup \{x_7, x_8, x_9\} \cup \\ \{x_9\} \cup \{x_9, x_{11}\} \cup \{x_7, x_8, x_9, x_{10}, x_{12}\} &= \\ \{x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}, x_{12}\}. \end{aligned}$$

Визначимо частоти  $f$  появи елементів вибірки  $X$  в усіх отриманих перетинах:

$$\begin{aligned} f(x_3) &= 0,06; \\ f(x_4) &= 0,06; \\ f(x_5) &= 0,4; \\ f(x_6) &= 0,2; \\ f(x_7) &= 0,2; \\ f(x_8) &= 0,4; \\ f(x_9) &= 0,4; \\ f(x_{10}) &= 0,06; \\ f(x_{11}) &= 0,06; \\ f(x_{12}) &= 0,06. \end{aligned}$$

Виходячи з цього, можна із множини  $X$  виділити підмножину  $Y \subseteq X$  предметів, які найчастіше зустрічаються:  $\{x_5, x_8, x_9\}$ , що дозволяють сформувати наступні бінарні асоціації:  $x_5 \rightarrow x_8$ ;  $x_8 \rightarrow x_5$ ;  $x_5 \rightarrow x_9$ ;  $x_9 \rightarrow x_5$ ;  $x_9 \rightarrow x_8$ ;  $x_8 \rightarrow x_9$ .

Для аналізу отриманих бінарних асоціацій скористаємося низкою характеристик, що вказують на наявність у них можливих зв'язків [11, 12]:

1. Підтримка асоціативного правила – це кількість транзакцій, що містять як умову, так і наслідок. Наприклад, для асоціації  $x_i \rightarrow x_j$  можна записати:

$$S(x_i \rightarrow x_j) = P(x_i \cap x_j) = \frac{\text{кількість транзакцій, що містять } x_i \text{ та } x_j}{\text{загальна кількість транзакцій}}.$$

2. Достовірність асоціативного правила  $x_i \rightarrow x_j$  являє собою міру точності правила і визначається як відношення кількості транзакцій, що містять і умову, і наслідок, до кількості транзакцій, що містять тільки умову:

$$C(x_i \rightarrow x_j) = \frac{P(x_i \cap x_j)}{P(x_i)} = \frac{\text{кількість транзакцій, що містять } x_i \text{ та } x_j}{\text{кількість транзакцій, що містять } x_i}.$$

Якщо підтримка й достовірність достатньо великі, можна з великою ймовірністю стверджувати, що будь-яка майбутня транзакція, що включає умову, буде також містити і наслідок.

3. Ліфт (оригінальна назва – «інтерес», також зустрічається термін «покращення») – це відношення частоти появи умови в транзакціях, що також містять і наслідок, до частоти появи наслідку в цілому:

$$L(x_i \rightarrow x_j) = \frac{C(x_i \rightarrow x_j)}{S(x_j)}.$$

Вважається, що ліфт є узагальненою мірою зв'язку двох предметних наборів: при  $L > 1$  зв'язок позитивний, при  $L = 1$  зв'язок відсутній, а при  $L < 1$  – зв'язок негативний.

4. Левередж – це різниця між частотою, з якою умова та наслідок з'являються спільно (тобто підтримка асоціації), та добутком частот появи підтримок умови та наслідку окремо:

$$T(x_i \rightarrow x_j) = S(x_i \rightarrow x_j) - S(x_i) * S(x_j).$$

Ліфт та левередж можуть використовуватися для подальшого обмеження набору асоціацій, що розглядаються, шляхом встановлення порогу значимості, нижче якого асоціації відкидаються.

Проведемо розрахунок перерахованих показників для отриманих асоціативних правил ( $x_5 \rightarrow x_8$ ):

$$\begin{aligned} S(x_5 \rightarrow x_8) &= 0,4; \\ C(x_5 \rightarrow x_8) &= 0,5; \\ L(x_5 \rightarrow x_8) &= \frac{C(x_5 \rightarrow x_8)}{S(x_8)} = \frac{0,5}{0,6} = 0,83; \end{aligned}$$

$$T(x_5 \rightarrow x_8) = S(x_5 \rightarrow x_8) - S(x_5) * S(x_8) = 0,4 - 0,48 = -0,08.$$

$$\begin{aligned} 1. \quad (x_8 \rightarrow x_5): \\ S(x_8 \rightarrow x_5) &= 0,4; \\ C(x_8 \rightarrow x_5) &= 0,66; \\ L(x_8 \rightarrow x_5) &= \frac{C(x_8 \rightarrow x_5)}{S(x_5)} = \frac{0,66}{0,8} = 0,825; \end{aligned}$$

$$T(x_8 \rightarrow x_5) = S(x_8 \rightarrow x_5) - S(x_8) * S(x_5) = 0,4 - 0,48 = -0,08.$$

$$\begin{aligned} 2. \quad (x_5 \rightarrow x_9): \\ S(x_5 \rightarrow x_9) &= 0,2; \\ C(x_5 \rightarrow x_9) &= 0,25; \\ L(x_5 \rightarrow x_9) &= \frac{C(x_5 \rightarrow x_9)}{S(x_9)} = \frac{0,25}{0,4} = 0,625; \end{aligned}$$

$$T(x_5 \rightarrow x_9) = S(x_5 \rightarrow x_9) - S(x_5) * S(x_9) = 0,2 - 0,32 = -0,12.$$

$$3. \quad (x_9 \rightarrow x_5):$$

$$S(x_9 \rightarrow x_5) = 0,2;$$

$$C(x_9 \rightarrow x_5) = 0,5;$$

$$L(x_9 \rightarrow x_5) = \frac{C(x_9 \rightarrow x_5)}{S(x_5)} = \frac{0,5}{0,8} = 0,625;$$

$$T(x_9 \rightarrow x_5) = S(x_9 \rightarrow x_5) - S(x_9) * S(x_5) = 0,2 - 0,32 = -0,12.$$

4.  $(x_8 \rightarrow x_9)$ :

$$S(x_8 \rightarrow x_9) = 0,3;$$

$$C(x_8 \rightarrow x_9) = 0,5;$$

$$L(x_8 \rightarrow x_9) = \frac{C(x_8 \rightarrow x_9)}{S(x_9)} = \frac{0,5}{0,4} = 1,25;$$

$$T(x_8 \rightarrow x_9) = S(x_8 \rightarrow x_9) - S(x_8) * S(x_9) = 0,3 - 0,24 = 0,06.$$

5.  $(x_9 \rightarrow x_8)$ :

$$S(x_9 \rightarrow x_8) = 0,3;$$

$$C(x_9 \rightarrow x_8) = 0,75;$$

$$L(x_9 \rightarrow x_8) = \frac{C(x_9 \rightarrow x_8)}{S(x_8)} = \frac{0,75}{0,6} = 1,25;$$

$$T(x_9 \rightarrow x_8) = S(x_9 \rightarrow x_8) - S(x_9) * S(x_8) = 0,3 - 0,24 = 0,06.$$

Отримані результати зведемо в таблицю 1.

Таблиця 1 – Показники асоціативних правил

Асоціації	S	C	L	T
$x_5 \rightarrow x_8$	0,4	0,5	0,83	-0,08
$x_8 \rightarrow x_5$	0,4	0,66	0,825	-0,08
$x_5 \rightarrow x_9$	0,2	0,25	0,625	-0,12
$x_9 \rightarrow x_5$	0,2	0,5	0,625	-0,12
$x_8 \rightarrow x_9$	0,3	0,5	1,25	0,06
$x_9 \rightarrow x_8$	0,3	0,75	1,25	0,06

На основі отриманих результатів можна говорити про те, що тільки дві бінарні асоціації  $(x_8 \rightarrow x_9)$  та  $(x_9 \rightarrow x_8)$  можуть бути віднесені до асоціативних правил.

**Висновок.** База транзакцій на початковому етапі її формування являє собою масив неупорядкованих даних, що потребують своєї структуризації і класифікації по кількості предметних наборів. У зв'язку з цим у роботі запропонована методика пошуку асоціативних правил, в основі якої лежать процедури «стиснення» вихідного масиву транзакцій до їх запису в базу даних.

### Список літератури

- [1] Т. А. Зайко, А. А. Олейник, и С. А. Субботин, "Ассоциативные правила в интеллектуальном анализе данных", *Вісник Національного технічного університету «ХП»*. *Інформатика та моделювання*, № 39, с. 82-96, 2013. [Електронний ресурс]. Режим доступу: [http://nbuv.gov.ua/UJRN/Vsriim\\_2013\\_39\\_14](http://nbuv.gov.ua/UJRN/Vsriim_2013_39_14)
- [2] A. Shved, and Ye. Davydenko, "The analysis of uncertainty measures with various types of evidence", in *2016 IEEE First Internat. Conf. on Data Stream Mining & Processing (DSMP)*, Lviv, 2016, pp. 61-64. doi: 10.1109/DSMP.2016.7583508.
- [3] I. Kovalenko, Ye. Davydenko, and A. Shved, "Development of the procedure for integrated application of scenario prediction methods", *Eastern-European Journal of Enterprise Technologies*, vol. 2, iss. 4 (98), pp. 31-38, 2019. doi: 10.15587/1729-4061.2019.163871
- [4] К. О. Antipova, Ye. O. Davydenko, I. I. Kovalenko, and A. V. Shved, "Modelling of group expert judgments under conditions of complex uncertainty", *East European Scientific Journal*, no. 5 (45), pp. 4-10, 2019.
- [5] В. А. Биллиг, Е. И. Корнеева, и Н. А. Сябро, "Ассоциативные правила. Сравнительный анализ инструментария", *Программные продукты, системы и алгоритмы*, № 2, 2016.
- [6] Е. В. Галкина, "Комбинированное применение метода дерева решений и ассоциативного анализа в управлении", *Международный научно-исследовательский журнал*, № 9 (51), с. 29-32, 2016. doi: 10.18454/IRJ.2016.51.095.
- [7] В. И. Городецкий, В. В. Самойлов, "Ассоциативный и причинный анализ и ассо-

- циативные байесовские сети", *Труды СПИИРАН*, вып. 9, с. 13-65, 2009.
- [8] Е. К. Джуматов, А. С. Вишня, и С. А. Филиппов, "Применение алгоритмов ассоциативных правил для выявления рекомендуемых к продаже товаров строительной отрасли", *ТЕОРИЯ. ПРАКТИКА. ИННОВАЦИИ: Междунар. науч.-техн. журн.*, № 1, с. 1-15, 2018.
- [9] J.-M. Adamo, *Data mining for association rules and sequential patterns: sequential and parallel algorithms*. New York: Springer-Verlag, 2001.  
doi: 10.1007/978-1-4613-0085-4.
- [10] Y. S. Koh, and N. Rountree, *Rare association rule mining and knowledge discovery: technologies for infrequent and critical event detection*. New York: Information Science Reference, 2009.  
doi: 10.4018/978-1-60566-754-6.
- [11] А. В. Молдавская, "Метод формирования многоуровневых последовательных паттернов", *Проблеми програмування*, № 2-3, с. 158-163, 2016. Режим доступа: <http://dspace.nbuv.gov.ua/handle/123456789/126401>
- [12] С. А. Субботин, А. А. Олейник, Е. А. Гофман, С. А. Зайцев, и А. А. Олейник, *Интеллектуальные информационные технологии проектирования автоматизированных систем диагностирования и распознавания образов: монография / под ред. С. А. Субботина*. Харьков: Компания Смит, 2012.
- References**
- [1] Т. А. Zaiko, А. А. Oleinik, and S. A. Subbotin, "Associative rules in data mining", *Visnyk Natsionalnoho tekhnichnoho universytetu «KhPI». Informatyka ta modeliuвання*, no. 39, pp. 82-96, 2013 [in Russian].
- [2] A. Shved, and Ye. Davydenko, "The analysis of uncertainty measures with various types of evidence", in *2016 IEEE First Internat. Conf. on Data Stream Mining & Processing (DSMP)*, Lviv, 2016, pp. 61-64.  
doi: 10.1109/DSMP.2016.7583508
- [3] I. Kovalenko, Ye. Davydenko, and A. Shved, "Development of the procedure for integrated application of scenario prediction methods", *Eastern-European Journal of Enterprise Technologies*, vol. 2, iss. 4 (98), pp. 31-38, 2019.  
doi: 10.15587/1729-4061.2019.163871
- [4] К. О. Antipova, Ye. O. Davydenko, I. I. Kovalenko, and A. V. Shved, "Modelling of group expert judgments under conditions of complex uncertainty", *East European Scientific Journal*, no. 5 (45), pp. 4-10, 2019.
- [5] V. A. Billig, E. I. Korneyeva, and N. A. Siabro, "Associative rules. Benchmarking toolkit", *Programmnye produkty, sistemy i alhoritmy*, no. 2, 2016 [in Russian].
- [6] E. V. Galkina, "The combined application of the decision tree method and associative analysis in management", *Mezhdunarodnyiy nauchno-issledovatel'skiy zhurnal*, no. 9 (51), pp. 29-32, 2016 [in Russian].  
doi: 10.18454/IRJ.2016.51.095
- [7] V. I. Gorodetskiy, and V. V. Samoylov, "Associative and causal analysis and associative Bayesian networks", *Trudy SPIIRAN*, no. 9, pp. 13-65, 2009 [in Russian].
- [8] Е. К. Dzhumatov, А. S. Vishnya, and S. A. Filippov, "The use of associative rule algorithms for identifying construction products recommended for sale", *ТЕОРИЯ. ПРАКТИКА. INNOVATSII: Internat. sci-tech. journ.*, no. 1, pp. 1-15, 2018 [in Russian].
- [9] J.-M. Adamo, *Data mining for association rules and sequential patterns: sequential and parallel algorithms*. New York: Springer-Verlag, 2001.  
doi: 10.1007/978-1-4613-0085-4.
- [10] Y. S. Koh, and N. Rountree, *Rare association rule mining and knowledge discovery: technologies for infrequent and critical event detection*. New York: Information Science Reference, 2009.  
doi: 10.4018/978-1-60566-754-6.
- [11] А. В. Moldavskaya, "The method of forming multi-level sequential patterns", *Problemy prohramuvannia*, no. 2-3, pp. 158-163, 2016 [in Russian].
- [12] S. A. Subbotin, А. А. Oleynik, Е. А. Gofman, S. А. Zaytsev, and А. А. Oleynik, *Intelligent information technology for the design of automated systems for diagnosing and recognizing images: monograph*. Khar'kov: Kompaniya Smit, 2012 [in Russian].

**I. I. Kovalenko**, *D. Tech. Sc., professor*,

**Ye. O. Davydenko**, *Ph. D. (Eng.)*,

e-mail: davydenko@chmnu.edu.ua

**A. V. Shved**, *Ph. D. (Eng.)*

Petro Mohyla Black Sea National University  
68 Desantnykiv str., 10, Mykolaiv, 54003, Ukraine

## METHOD FOR ASSOCIATIVE RULES SEARCH

*With the appearance of large volumes of stored information, the tasks related to the need of their processing become urgent. Artificial intelligence methods are used to analyze various kinds of data. Accumulating data is characterized by disordered and unstructured data when each storage unit cannot be represented by a finite number of features. In recent years, associative rule search methods have been widely used to process large arrays of unstructured data. The problem is that the number of possible associations with the increase of the number of items in each of the transactions increases exponentially and requires considerable computational cost. Therefore, in the process of associative rules formation, the techniques, which allow to reduce the number of associations that need to be analyzed, are widely used.*

*The purpose of this work is to develop a methodology for finding associative rules based on preliminary analysis and compression of an array of transactions representing different subject sets.*

*The paper proposes an approach to the formation of a sample of items most characteristic for a simplified array of transactions, on the basis of which associative binary relations are formed and their characteristics are calculated to determine whether such relations are rules. To do this, the array is first compressed by searching for transactions with the same subject sets using an equivalence relation. Then, to detect transactions with recurring items, a pairwise intersection of the subject sets is performed.*

*At the initial stage of formation, the transaction database is an array of disordered data that needs to be structured and classified by the number of subject sets. As a result of research, the paper offers a method of finding associative rules, which are based on the procedures of "compressing" the original array of transactions to their entry in the database.*

**Keywords:** *transaction, associations, associative rules, data array.*