УДК 004.93

Т.А. Зайко, А.О. Олійник, С.О. Субботін

ІНДУКЦІЯ ЧИСЕЛЬНИХ АСОЦІАТИВНИХ ПРАВИЛ З ВРАХУВАННЯМ ІНДИВІДУАЛЬНОЇ ЗНАЧУЩОСТІ ОЗНАК

Анотація: Розглядається задача видобування чисельних асоціативних правил. Запропоновано метод індукції асоціативних правил з врахуванням значущості ознак, який дозволяє скоротити простір пошуку та час виявлення правил, збільшити рівні узагальнення та інтерпретабельності синтезованої бази асоціативних правил.

Ключові слова: асоціативне правило, база правил, індукція, значущість ознаки, нечітка логіка, транзакція.

Вступ

При розв'язанні задач діагностування, автоматичної класифікації, прогнозування та керування часто виникає потреба виявлення нових знань про досліджувані об'єкти або процеси [1,2]. Для обробки великих масивів даних та видобування з них нових знань широкого застосування набули методи пошуку асоціативних правил [3,4], що дозволяють виявляти нові закономірності вигляду "якщо умова, то дія" та синтезувати на їх основі бази правил, що є зрозумілими для експертів в прикладних областях [5].

Проте відомі методи індукції асоціативних правил у більшості випадків дозволяють виявляти лише бінарні правила [3,6], які не враховують чисельні значення ознак, що характеризують об'єкти або процеси, які підлягають аналізу. Методи виявлення чисельних асоціативних правил [7,8] можуть працювати с чисельними вибірками даних, проте їх робота пов'язана з проблемами вибору інтервалів дискретизації діапазонів значень змінних, визначення кількості таких інтервалів для кожної ознаки, що у деяких випадках призводить до суттєвого збільшення простору пошуку та вимог до обчислювальних ресурсів. Крім того, такі методи передбачають, що усі ознаки, які описують досліджувані об'єкти, мають однакову інформативність, що, як правило, на практиці не відповідає дійсності [2,9,10]. Включення таких ознак до синтезованої моделі або бази правил призводить до збільшення часу побудови моделі та ресурсів, необхідних для виконання цього процесу, а також до зменшення апроксимаційних та узагальнювальних властивостей побудованої моделі.

Тому актуальною є мета роботи – створення методу індукції чисельних асоціативних правил з врахуванням індивідуальної значущості ознак.

1. Постанова задачі

Нехай задана база транзакцій D (1):

$$D = \{T_1, T_2, \ldots, T_{N_D}\}, \tag{1}$$

у якій кожний елемент T_j, $j = 1, 2, \ldots, N_D$, містить інформацію про деякі взаємозалежні події, де $N_D = |D|$ – кількість елементів (транзакцій) у наборі даних D.

Елементи T_j можуть подаватися у вигляді (2):

$$T_j = (tid_j, item_j), \tag{2}$$

де tid_j – ідентифікатор j-ї транзакції T_j; $item_j = \left\{t_{1j}, t_{2j}, \ldots, t_{N_{item_j} j}\right\} \subseteq I$ – список елементів, що входять у транзакцію T_j; t_{ij} – i-й елемент списку $item_j$, $i = 1, 2, \ldots, N_{item_j}$; $N_{item_j} = |item_j|$ – кількість елементів множини $item_j$; $I = \{\tau_1, \tau_2, \ldots, \tau_{N_I}\}$ – множина можливих змінних (ознак), які можуть входити в список елементів $item_j$ кожної транзакції T_j, $j = 1, 2, \ldots, N_D$, набору даних D; τ_a – a-й елемент множини I, $a = 1, 2, \ldots, N_I$; $N_I = |I|$ – кількість елементів множини I.

У випадку, якщо база транзакцій D містить крім бінарних, ще й дійсні змінні, елементи t_{ij} транзакції T_j подаються кортежем (3):

$$t_{ij} = \langle \tau_{ij}; v(\tau_{ij}) \rangle, \tag{3}$$

де τ_{ij} – ознака із множини I, що відповідає елементу t_{ij}; $v(\tau_{ij})$ – значення ознаки τ_{ij} в транзакції T_j, $v(\tau_{ij}) \in \Delta_{ij} = [\tau_{ij\,\min} \tau_{ij\,\max}]$; $\tau_{ij\,\min}$ і $\tau_{ij\,\max}$ – відповідно, мінімальне та максимальне значення з діапазону можливих значень Δ_{ij} ознаки τ_{ij}.

Тоді на основі заданої транзакційної бази даних D необхідно побудувати набір чисельних асоціативних правил у вигляді імплікацій $\langle X, v(X) \rangle \to \langle Y, v(Y) \rangle$, у яких набори X й Y не перетинаються [4, 9] (4):

$$\langle X, v(X) \rangle \to \langle Y, v(Y) \rangle : \; X \subset I, \; Y \subset I, \; X \cap Y = \emptyset \tag{4}$$

де $v(X)$ й $v(Y)$ – множини значень ознак, що належать множинам X і Y, відповідно.

Таким чином, у результаті синтезу асоціативних правил основі наявного набору даних D виконується пошук закономірностей між подіями $\tau_a \in I$, $a = 1, 2, \ldots, N_I$.

2. Метод синтезу чисельних асоціативних правил

Для можливості видобування асоціативних правил з транзакційних баз даних D, які містять чисельні атрибути, такі атрибути перетворюються до формату, доступного для застосування відомих методів пошуку асоціативних правил [3,4,6]. При цьому потрібно виконувати розбиття чисельних ознак на непересічні інтервали,

ISSN 1560-8956

кожний з яких розглядається потім як новий атрибут. Однак у таких випадках виникають проблеми вибору числа інтервалів і розбиття на інтервали, крім того суттєво зростає розмірність розв'язуваної задачі й вимоги до обчислювальних ресурсів ЕОМ.

Тому в розробленому методі синтезу чисельних асоціативних правил пропонується використовувати підхід на основі теорії нечітких множин [11], що дозволяє розбивати вихідні ознаки на нечіткі інтервали та працювати з кожною ознакою, а не з окремими інтервалами її розбиття. Крім того, у запропонованому методі при пошуку асоціативних правил використовуються розраховані оцінки індивідуальної інформативності ознак, що дозволяє враховувати їхню значимість у вихідній базі даних.

Пропонований метод може бути представлений наступними етапами:

– фаззифікація транзакційної бази даних D;
– визначення індивідуальної значущості ознак;
– обчислення граничних значень підтримки;
– побудова бази чисельних асоціативних правил.

На початковому етапі виконується фаззифікація бази транзакцій D, тобто приведення всіх її чисельних значень до нечіткого вигляду: $D \rightarrow FuzzyD$. Таке перетворення дозволить виділити нечіткі терми кожної ознаки для можливості виконання подальшого видобування асоціативних правил. Для фаззифікації бази транзакцій з метою наступного видобування чисельних асоціативних правил доцільно використовувати відомі функції належності [11,12], параметри яких пропонується вибирати виходячи з ідеї підтримки нечіткості знань, тобто таким чином, щоб забезпечувалося перетинання сусідніх інтервалів розбиття ознак.

У якості функцій належності доцільно використовувати такі функції, які дозволяють обмежувати інтервал значень ознак: трапецієподібну, П-подібну, трикутну функцію. Виходячи з особливостей розв'язуваної задачі й досліджуваних об'єктів або процесів, можна використовувати й інші функції належності [11,12]: сплайн-функцію, S-подібну й Z-подібну криві, сигмоідну функцію, функцію Гаусса, колоколоподібну функцію й інші. Для визначення значень параметрів функцій приналежності виконується розбиття кожної чисельної ознаки на деяку кількість інтервалів з наступним визначенням границь отриманих інтервалів.

Як правило, ознаки, що описують досліджувані об'єкти або процеси, мають різну інформативність [13], тому з метою видобування цікавих асоціативних правил, що адекватно описують досліджувані залежності, доцільно враховувати індивідуальну значимість ознак. Оскільки вихідний параметр у транзакційних базах даних, як правило, не заданий, пропонується оцінювати індивідуальну значимість ознак за допомогою параметрів, що характеризують границі областей групування екземплярів (транзакцій) у

просторі ознак. Для обчислення індивідуальної інформативності ознак пропонується використовувати підхід, що враховує границі інтервалів розбиття ознак у кластерах. У даному методі пропонується сортувати масив значень кожної ознаки τ_a за зростанням. Ліва l_{ak} й права r_{ak} границі k-го інтервалу Δ_{ak} a-ї ознаки τ_a обираються таким чином, щоб екземпляри (транзакції) зі значенням ознаки $\tau_a \in \Delta_{ak} = [l_{ak}; r_{ak})$ відносилися до одного кластеру K_b, а екземпляри із сусідніх інтервалів – до інших кластерів $K_c \neq K_b$.

У якості міри інформативності a-ї ознаки в транзакційній базі даних D доцільно використовувати кількість інтервалів $N_{\text{інт. }a}$, на які розбивається діапазон її значень $\Delta_a = [\tau_{a\min}; \tau_{a\max}]$: чим менше кількість таких інтервалів, тим більше інформативність ознаки.

Тому значущість ознаки τ_a будемо обчислювати по одній з формул:

– відношення мінімальної кількості інтервалів серед усіх ознак до величини $N_{\text{інт. }a}$ a-ї ознаки (5):

$$w_a = \frac{\min\limits_{A=1,2,\ldots,|I|} N_{\text{інт. }A}}{N_{\text{інт. }a}}; \tag{5}$$

– нормоване значення величини $N_{\text{інт. }a}$ (6):

$$w_a = 1 - \frac{N_{\text{інт. }a} - \min\limits_{A=1,2,\ldots,|I|} N_{\text{інт. }A}}{\max\limits_{A=1,2,\ldots,|I|} N_{\text{інт. }A} - \min\limits_{A=1,2,\ldots,|I|} N_{\text{інт. }A}} = \frac{\max\limits_{A=1,2,\ldots,|I|} N_{\text{інт. }A} - N_{\text{інт. }a}}{\max\limits_{A=1,2,\ldots,|I|} N_{\text{інт. }A} - \min\limits_{A=1,2,\ldots,|I|} N_{\text{інт. }A}}. \tag{6}$$

Важливим етапом є визначення граничних значень підтримки наборів елементів, яке в запропонованому методі відбувається з використанням інформації про індивідуальну значущість ознак, розраховану раніше.

У розробленому методі видобування чисельних асоціативних правил підтримку транзакції T_j будемо розраховувати як перетинання функцій належності ознак, що входять у транзакцію T_j (7):

$$\operatorname{supp}(T_j) = \bigcap_{\tau_a \in T_j}^{\mu_a(T_j)}, \tag{7}$$

де $\mu_a(T_j)$ – значення функція приналежності a-ї ознаки, обчислене для її значення в транзакції T_j.

Тоді підтримка набору X визначається як сума підтримок усіх транзакцій, які містять цю множину (8):

$$\operatorname{supp}(X) = \sum_{X \subseteq T_j} \operatorname{supp}(T_j) = \sum_{X \subseteq T_j} \bigcap_{\tau_a \in T_j}^{\mu_a(T_j)}. \tag{8}$$

ISSN 1560-8956

Крім того, передбачається можливість видобування наборів, які не часто зустрічаються, однак є цікавими та дозволяють виявляти нові знання про досліджувані об'єкти або процеси [14].

При побудові бази асоціативних правил у процесі їх видобування використовуються значення індивідуальної інформативності ознак, розраховані раніше, що дозволяє враховувати значущість кожного атрибута при пошуку правил. При генерації нових наборів-кандидатів у процесі синтезу асоціативних правил враховується властивість антимонотонності підтримки [4], застосування якої дозволяє суттєво скоротити простір пошуку.

Запропонований метод видобування чисельних асоціативних правил передбачає фаззифікацію заданої бази транзакцій і автоматичне розбиття діапазонів значень ознак на інтервали, враховує індивідуальну значущість ознак, використовує критерії для оцінювання непрямих асоціацій, що знижує ступінь участі користувача в процесі пошуку асоціативних правил, зменшує ймовірність видобування правил, що некоректно описують досліджувані об'єкти та процеси.

З метою аналізу та дослідження ефективності запропонованого методу індукції асоціативних правил оцінимо його обчислювальну складність O – кількість елементарних операцій, необхідних для розв'язання конкретної задачі.

При оцінюванні обчислювальної складності розробленого методу будемо враховувати, що він складається з чотирьох етапів, описаних раніше.

Обчислювальна складність етапу, пов'язаного з фаззифікацією транзакційної бази даних D, може бути оцінена виходячи з того, що для кожної ознаки τ_a ($\tau_a \in I, \quad a = 1, 2, \ldots, |I|$) у кожній транзакції T_j буде визначатися її нечітке значення. Отже, обчислювальна складність першого етапу складе: $O_1 = O_1(|I| \cdot |N_D|)$.

Визначення індивідуальної значущості w_a ознак τ_a у базі транзакцій D пов'язане з необхідністю їх угруповування в множини компактно розташованих транзакцій. Розглянемо підхід, що враховує границі інтервалів розбиття ознак у кластерах. Вище описано, що такий підхід пов'язаний з необхідністю сортування кожної ознаки τ_a. Тому обчислювальна складність даного етапу безпосередньо пов'язана з обчислювальною складністю використовуваного методу сортування й може бути визначена як $O_2(|I| \cdot O_{сорт.})$. Ефективними методами сортування є методи, що використовують дерева (турнірне сортування, сортування за допомогою пошукового дерева), метод пірамідального сортування, метод швидкого сортування К. Хоара, обчислювальна складність яких становить $O_{сорт.} = O_{сорт.}(N_D \log_2(N_D))$ [15]. Отже, обчислювальна складність етапу визначення значущості ознак може бути оцінена в такий спосіб: $O_2 = O_2(|I| \cdot N_D \log_2(N_D))$.

Етап визначення граничних значень, як правило, припускає участь користувача при виборі значень відповідних порогів. У випадку обчислення значення коефіцієнта α, що враховує значущість самої довгої транзакції в базі даних D, буде потрібно виконати не більш $|I|$ раз операцію додавання відповідних величин w_a ($a : \tau_a \in T_j, |T_j| = \max\limits_{T_J \in D} |T_J|$). Тому обчислювальна складність цього етапу визначається так: $O_3 = O_3(|I|)$.

Безпосереднє видобування асоціативних правил пов'язане з побудовою множини наборів, які часто зустрічаються FI, що у свою чергу вимагає визначення значень підтримок кожного з кандидатів, максимальна кількість яких не перевищує $|I|^2$. Складність цього процесу складе $O_{FI}\left(|I|^2\right)$ операцій. Процес видобування асоціативних правил із множини FI припускає обробку кожної підмножини $A \in FI$, на що буде потрібно $O_{извл.}\left(|I|^2\right)$ операцій. Тому обчислювальна складність четвертого етапу складе: $O_4 = O_{FI}\left(|I|^2\right) + O_{извл.}\left(|I|^2\right) = O_4\left(|I|^2\right)$.

Таким чином, загальна обчислювальна складність запропонованого методу може бути визначена в такий спосіб (9):

$$O = O_1(|I| \cdot |N_D|) + O_2(|I| \cdot N_D \log_2(N_D)) + O_3(|I|) + O_4\left(|I|^2\right) = \\ = O\left(|I| \cdot N_D \log_2(N_D) + |I|^2\right). \tag{9}$$

Оцінка обчислювальної складності показує, що кількість елементарних операцій, необхідних для видобування асоціативних правил за допомогою запропонованого методу, квадратично залежить від кількості $|I|$ ознак $\tau_a \in I$ у транзакційній базі даних D, а також пропорційна величині $N_D \log_2(N_D)$, де N_D – кількість транзакцій в D. Така оцінка дозволяє зробити висновок про те, що запропонований метод є ефективним, оскільки залежність його елементарних операцій від розміру вхідних даних є поліноміальною.

3. Експерименти та результати

З метою проведення експериментів по дослідженню властивостей і характеристик запропонованого методу видобування чисельних асоціативних правил він був програмно реалізований мовою програмування C#.

Для дослідження властивостей і характеристик розробленого методу видобування чисельних асоціативних правил використовувалися тестові дані, представлені у вигляді бази транзакцій D, чисельні характеристики якої наведені нижче: $N_D = 100000$ – кількість транзакцій T_j $(j = 1, 2, \ldots, 100000)$ у базі D; $|I| = 10000$ – кількість елементів (ознак), $\tau_a \in I, a = 1, 2, \ldots, 10000$, з яких могли формуватися транзакції; $|T_j| = 10$ – середня кількість ознак у транзакціях бази D.

Залежність часу видобування асоціативних правил від значення мінімальної підтримки wminsupport наведена на рис. 1

ISSN 1560-8956

(при цьому значення інших параметрів установлювалися такими: wminconfidence = 75%, $\beta_{\text{wsupp}(X \cup Y)} = 5\%$, $\beta_{\text{wsupp}(Z)} = 75\%$, $w_{\min} = 75\%$). Запропонований метод порівнювався з методами FARM і FWARM, запропонованими в [7] і [8], відповідно.

З рис. 1 видно, що час видобування асоціативних правил суттєво залежить від значення мінімальної підтримки, оскільки незначні значення цього порога приводять до генерування занадто великої кількості наборів, які вважаються такими, що часто зустрічаються, що у свою чергу приводить до необхідності обробки істотних обсягів інформації.

Крім того, з рис. 1 видно, що запропонований метод є більш ефективним у порівнянні з існуючими [7,8], оскільки в ньому формується масив *FI* наборів, що часто зустрічаються, який залежить не тільки від кількості поява конкретних наборів у базі даних *D*, але й від оцінок значущості ознак, що не дозволяє генерувати мало значущі набори й, відповідно, зменшує розмір *FI*, а значить і час, затрачуваний на його обробку.

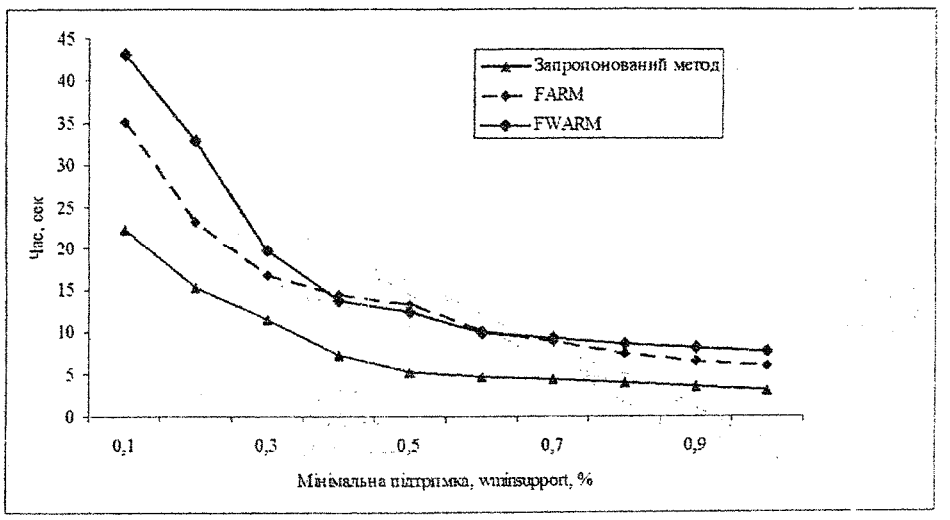

Рис. 1 – Графік залежності часу видобування асоціативних правил від значення wminsupport

Важливо відзначити, що результати експериментів показали, що час роботи методу практично не залежить від значення мінімальної вірогідності wminconfidence. Це пов'язане з тим, що даний параметр використовується на завершальному етапі методу при перевірці на вірогідність виявлених правил і не впливає на кількість оброблюваних наборів даних.

На рис. 2 зображено графік залежності часу функціонування методу від кількості транзакцій N_D у базі *D* (при цьому значення wminsupport становило 0,5 %, інші параметри методу та бази *D* не змінювалися).

Крива, зображена на рис. 2 і побудована за результатами застосування запропонованого методу, підтверджує оцінку обчислювальної складності *O*, наведену вище, згідно з якою час видобува-

ння асоціативних правил за допомогою запропонованого методу є пропорційним величині $N_D \log_2 (N_D)$.

Графік залежності кількості витягнутих асоціативних правил $N_{АП}$ від кількості транзакцій N_D у базі D наведено на рис. 3.

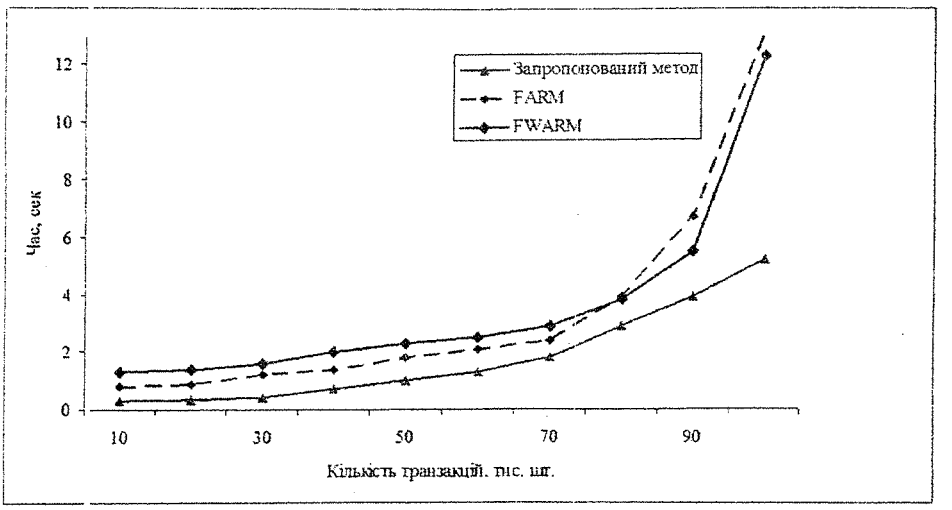

Рис. 2 – Графік залежності часу видобування асоціативних правил від кількості транзакцій N_D у базі D

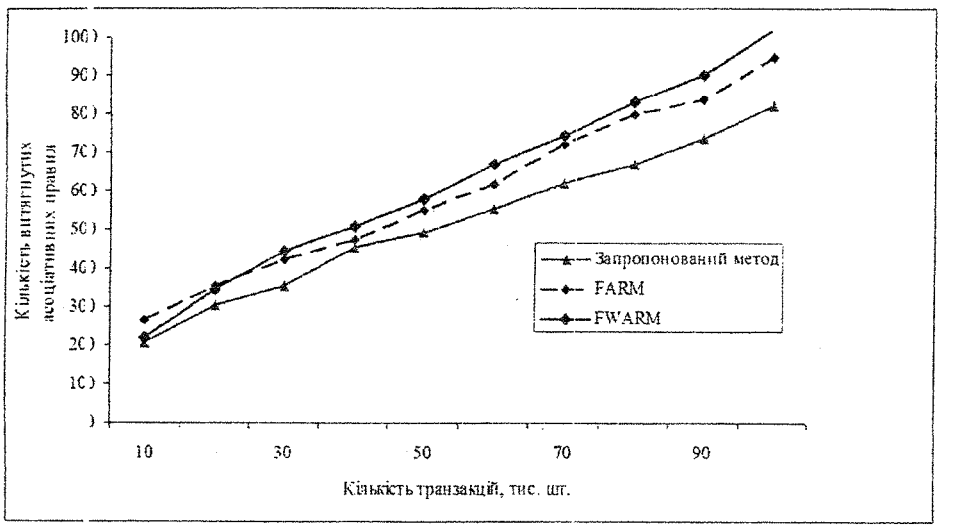

Рис. 3 – Графік залежності кількості витягнутих асоціативних правил $N_{АП}$ від кількості транзакцій N_D у базі D

За результатами експериментів видно, що кількість згенерованих асоціативних правил $N_{АП}$ пропорційна кількості записів N_D у базі D. Запропонований метод витягав у середньому на 16–23 % правил менше в порівнянні з методами FARM і FWARM [7, 8], що пояснюється використанням апріорної інформації про значущість w_a ознак $\tau_a \in I$, що дозволяло не розглядати деякі набори (з низькими оцінками індивідуальної значущості w_a ознак) як ті, що часто зустрічаються, й, відповідно, не тільки скорочувало час пошуку, але й зменшувало кількість витягнутих правил.

ISSN 1560-8956

На рис. 4 відображено результати експериментів по дослідженню кількості витягнутих асоціативних правил $N_{АП}$ від значення мінімальної підтримки wminsupport.

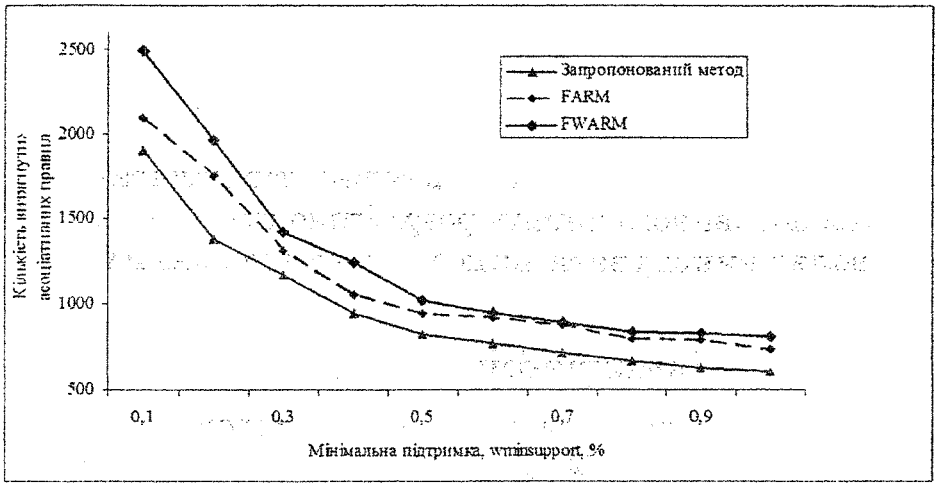

Рис. 4 – Графік залежності кількості витягнутих асоціативних правил $N_{АП}$ від значення мінімальної підтримки wminsupport

Як видно з рис. 4, розроблений метод генерує менше асоціативних правил (особливо при незначних значеннях wminsupport до 0,5 %), що також пояснюється використанням у процесі видобування правил оцінок індивідуальної значущості w_a, що дозволяє не витягати правила, які не представляють інтерес при аналізі досліджуваних об'єктів і процесів.

Таким чином, результати експериментів показали, що розроблений метод дозволяє витягати з баз транзакцій чисельні асоціативні правила, використовуючи при цьому апріорну інформацію про значущість ознак, що скорочує простір пошуку та час видобування правил, зменшує кількість витягнутих правил, і, відповідно, підвищує рівні узагальнення й інтерпретабельності синтезованої бази асоціативних правил.

Висновки

У роботі вирішено актуальну задачу автоматизації видобування чисельних асоціативних правил.

Наукова новизна роботи полягає в тому, що запропоновано метод видобування чисельних асоціативних правил, основними етапами якого є: фаззифікація транзакційної бази даних, визначення індивідуальної значущості ознак, обчислення граничних значень підтримки й побудова бази чисельних асоціативних правил. Запропонований метод передбачає фаззифікацію заданої бази транзакцій і автоматичне розбиття діапазонів значень ознак на інтервали, враховує індивідуальну значущість ознак, використовує критерії для оцінювання непрямих асоціацій, що знижує ступінь участі користувача в процесі пошуку асоціативних правил, зменшує ймовірність виявлення правил, які некоректно описують досліджувані

об'єкти й процеси, а також дозволяє витягати набори, що не тільки часто зустрічаються, але й рідко виникаючі цікаві асоціативні правила. Використання апріорної інформації про значущість ознак у розробленому методі дозволяє скоротити простір пошуку та час видобування правил, зменшити кількість витягнутих правил, і, відповідно, підвищити рівні узагальнення й інтерпретабельності синтезованої бази асоціативних правил.

Практична цінність отриманих результатів полягає в тому, що на основі запропонованого методу розроблено програмне забезпечення, що дозволяє виконувати видобування чисельних асоціативних правил.

Бібліографічний список

1. Рассел С. Искусственный интеллект: современный подход / С. Рассел, П. Норвиг. – М.: Вильямс, 2006. – 1408 с.
2. Encyclopedia of artificial intelligence / Eds.: J. R. Dopico, J. D. de la Calle, A. P. Sierra. – New York : Information Science Reference, 2009. – Vol. 1–3. – 1677 p.
3. Koh Y. S. Rare Association Rule Mining and Knowledge Discovery / Y. S. Koh, N. Rountree. – New York : Information Science Reference. – 2009. – 320 p.
4. Adamo J.-M. Data mining for association rules and sequential patterns: sequential and parallel algorithms / Adamo J.-M. – New York : Springer-Verlag. – 2001. – 259 p.
5. Zhao Y. Post-mining of association rules: techniques for effective knowledge extraction / Y. Zhao, C. Zhang, L. Cao. – New York : Information Science Reference. – 2009. – 372 p.
6. Субботін С. О. Подання й обробка знань у системах штучного інтелекту та підтримки прийняття рішень : навч. посібник / С. О. Субботін. – Запоріжжя: ЗНТУ, 2008. – 341 с.
7. Dubois D. A Systematic Approach to the Assessment of Fuzzy Association Rules / D. Dubois, E. Hullermeier, H. Prade // Data Mining and Knowledge Discovery. – 2006. – Vol. 13. – P. 167-192.
8. Khan M. S. Weighted Association Rule Mining from Binary and Fuzzy Data / M. S. Khan, M. Muyeba, F. Coenen // Lecture Notes in Computer Science. – 2008. – Vol. 5077. – P. 200-212.
9. Интеллектуальные информационные технологии проектирования автоматизированных систем диагностирования и распознавания образов : монография / [С. А. Субботин, Ан. А. Олейник, Е. А. Гофман, С. А. Зайцев, Ал. А. Олейник ; под ред. С. А. Субботина]. – Харьков : ООО "Компания Смит", 2012. – 317 с.

ISSN 1560-8956

10. Субботін С. О. Неітеративні, еволюційні та мультиагентні методи синтезу нечіткологічних і нейромережних моделей: монографія / С. О. Субботін, А. О. Олійник, О. О. Олійник. – Запоріжжя: ЗНТУ, 2009. – 375 с.

11. Zadeh L. Fuzzy sets / L. Zadeh // Information and Control. – 1965. – № 8. – P. 338–353.

12. Гибридные нейро-фаззи модели и мультиагентные технологии в сложных системах : монография / [В. А. Филатов, Е. В. Бодянский, В. Е. Кучеренко и др. ; под общ. ред. Е. В. Бодянского]. – Дніпропетровськ : Системні технології, 2008. – 403 с.

13. Зайко Т. А. Определение индивидуальной значимости признаков для извлечения численных ассоциативных правил / Т. А. Зайко , А. А. Олейник, С. А. Субботин // Искусственный интеллект и его приложения : III-й Межвузовский научно-исследовательский семинар, Магнитогорск, 25 декабря 2012 г. : материалы семинара. – Магнитогорск : МаГУ, 2012. – С. 105–108.

14. Зайко Т. А. Пошук рідкісних цікавих асоціативних правил в великих масивах даних / Т. А. Зайко, А. О. Олійник, С. О. Субботін // Інформатика, математика, автоматика : науково-технічна конференція, Суми, 22–27 квітня 2013 р. : матеріали та програма конференції. – Суми : СумДУ, 2013. – С. 29.

15. Кнут Д. Искусство программирования. В 3-х томах. Т. 2 Сортировка и поиск / Д. Кнут. – М. : Вильямс, 2007. – 824 с.

Отримано 19.02.2013