

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
СУМСЬКИЙ ДЕРЖАВНИЙ УНІВЕРСИТЕТ

ІНФОРМАТИКА, МАТЕМАТИКА,
АВТОМАТИКА

ІМА :: 2013

**МАТЕРІАЛИ
та програма**

НАУКОВО-ТЕХНІЧНОЇ КОНФЕРЕНЦІЇ

(Суми, 22-27 квітня 2013 року)

Суми
Сумський державний університет
2013

Пошук рідкісних цікавих асоціативних правил у великих масивах даних

Зайко Т.А., *асп.*; Олійник А.О., *доц.*;
Субботін С.О., *доц.*

Запорізький національний технічний університет, м. Запоріжжя

При виявленні асоціативних правил на основі заданих транзакційних баз даних, як правило, виконується пошук часто зустрічаючихся послідовностей, з яких на наступних етапах формуються правила [1, 2]. Використання такого підходу не дозволяє виявляти нові знання про досліджувані об'єкти та процеси у вигляді правил, які не є такими, що часто зустрічаються у заданих базах транзакцій, проте є цікавими для даної предметної області.

В запропонованому методі виявлення чисельних асоціативних правил після виконується етап пошуку цікавих правил, які рідко зустрічаються, вигляду $X \xrightarrow{Z} Y$. Для цього формується множина наборів, що рідко зустрічаються RI , і для кожного її елемента $A \in RI$ виконуються наступні дії: $X = A_{|A|}$ – останній елемент множини A ; $Y = A_{|A|-1}$ – передостанній елемент множини A ; $Z = A \setminus (X \cup Y)$. Тоді будемо витягати асоціативні правила вигляду $X \xrightarrow{Z} Y$ при виконанні наступних умов:

$$\begin{cases} \text{wsupp}(X \cup Y) < \beta_{\text{wsupp}(X \cup Y)}; \\ (\text{wsupp}(X \cup Z) \cap \text{wsupp}(Y \cup Z)) \geq \beta_{\text{wsupp}(Z)}; \\ w(X, Z) \cap w(Y, Z) \geq w_{\min}, \end{cases}$$

де $\beta_{\text{wsupp}(X \cup Y)}$ – зважена підтримка набору елементів, що входять у множину ω ; $\beta_{\text{wsupp}(\omega)}$ – граничне значення підтримки $\beta_{\text{wsupp}(X \cup Y)}$; $w(X, Z)$ – значення критерію оцінювання взаємозв'язку між множинами X та Z ; w_{\min} – граничне значення критерію w .

Після виявлення імплікацій вигляду $X \rightarrow Y$ та $X \xrightarrow{Z} Y$ на їх основі синтезується база асоціативних правил, що описує досліджувані об'єкти та процеси. Таким чином, запропонований метод виявлення чисельних асоціативних правил дозволяє витягати цікаві асоціативні правила не тільки на основі наборів, що часто зустрічаються, але і на основі множин елементів, які є рідкісними у заданих транзакційних базах даних, що дозволяє виявляти нові знання про досліджувані об'єкти та процеси.

1. A. Gkoulalas-Divanis, V. S. Verykios, *Association Rule Hiding for Data Mining* (New York : Springer-Verlag: 2010).