

Метод пошуку асоціативних правил



Сфери застосування асоціативних правил:

- роздрібна торгівля: визначення товарів, які варто просувати спільно; вибір місцезнаходження товару у крамниці; аналіз споживчого кошика; прогнозування попиту
- перехресні продажі
- маркетинг: пошук ринкових сегментів; тенденції споживчої поведінки
- сегментація клієнтів: виявлення спільних характеристик клієнтів компанії; виявлення груп покупців
- оформлення каталогів, аналіз збутових кампаній фірми; визначення послідовностей покупок клієнтів
- аналіз Web-логів

Основні поняття про асоціативні правила:

Нехай $I = \{i_1, i_2, \dots, i_m\}$ – множина елементів, D – множина транзакцій, де кожна транзакція T – це набір елементів з I , $T \subseteq I$.

Кожна транзакція – це множина елементів (подій), що відбулися одночасно, і являє собою бінарний вектор, де $t[k] = 1$, якщо i_k елемент присутній у транзакції, інакше $t[k] = 0$. Транзакція T містить X , певний набір елементів з I , якщо $X \subseteq T$.

Розширеною транзакцією називається транзакція, розширена предками всіх елементів, що входять у цю транзакцію.

Асоціативним правилом (association rule) називається імплікація $X \rightarrow Y$, де $X \subset I$, $Y \subset I$ та $X \cap Y = \emptyset$.

Підтримкою (support) правила $X \rightarrow Y$ називається частка s , якщо s % транзакцій з D , містять $X \cup Y$, $\text{supp}(X \rightarrow Y) = \text{supp}(X \cup Y)$.

Вірогідність правила показує яка імовірність того, що з X впливає Y . Правило $X \rightarrow Y$ є справедливим з вірогідністю (confidence) c , якщо c % транзакцій з D , що містять X , також містять Y ,
 $\text{conf}(X \rightarrow Y) = \text{supp}(X \rightarrow Y) / \text{supp}(X)$.

Методи пошуку асоціативних правил
призначені для знаходження всіх
правил $X \rightarrow Y$, причому підтримка і
вірогідність цих правил повинні бути
вище деяких наперед визначених
порогів, називаних відповідно
мінімальною підтримкою
(minsupport) і *мінімальною*
вірогідністю (minconfidence).

Задача знаходження асоціативних правил розбивається на дві підзадачі:

- 1) знаходження всіх наборів елементів, що задовольняють порогові minsupport . Такі набори елементів називаються такими, що часто зустрічаються;
- 2) генерація правил зі знайдених наборів елементів з вірогідністю, що задовольняє порогові minconfidence .

Методи вирішення подібного класу задач – *AIS (1993)*, *SETM*, *Apriori*, *DHP*, *Partition*, *DIC* тощо

Узагальненим асоціативним правилом (generalized association rule) називається імплікація $X \rightarrow Y$, де $X \subset I$, $Y \subset I$ та $X \cap Y = \emptyset$ і де жоден з елементів, що входять у набір Y , не є предком жодного елемента, що входить у X . Підтримка і вірогідність підраховуються так само, як і у випадку асоціативних правил.

Правило $X \rightarrow Y$ називається узагальненим, тому що елементи, які входять у нього можуть знаходитися на будь-якому рівні таксономії. Також \underline{x} називається *предком* x та \bar{x} - *нащадком* x .

Основною *відмінністю* узагальнених асоціативних правил від асоціативних правил є те, що одержувані правила включають елементи, що є предками елементів, які входять у множину транзакцій.

Ієрархією (таксономією) елементів називається ліс спрямованих ациклічних дерев, листами яких є елементи транзакцій, а внутрішніми вузлами – групи елементів.

Уведення додаткової інформації про групування елементів у виді ієрархії допомагає установити асоціативні правила не тільки між окремими елементами, але і між різними рівнями ієрархії (групами).

Виявлення узагальнених асоціативних правил:


Нехай I – ліс спрямованих дерев. Дуги в I – це залежності між елементами. Нехай елементи, що належать I , розташовані в деякій ієрархії. Якщо є дуга від a до b , то говорять, що a – предок b та b – нащадок a (a – це узагальнення b).

Необхідно знайти закономірності, що є узагальненими асоціативними правилами виду $X \rightarrow Y$, причому $\text{supp}(X \rightarrow Y) \geq \text{minsupport}$ та $\text{conf}(X \rightarrow Y) \geq \text{minconfidence}$.

Визначення “цікавих” правил

Правило $X \rightarrow Y$ називається *R-цікавим* щодо правила-предка, якщо підтримка правила $X \rightarrow Y$ у R разів більше очікуваної підтримки правила $X \rightarrow Y$ щодо предка або якщо вірогідність правила $X \rightarrow Y$ у R разів більше очікуваної вірогідності правила $X \rightarrow Y$ щодо правила-предка.

Цікавим називається правило, якщо в нього немає предків або воно є *R-цікавим* щодо усіх своїх найближчих предків.



Частково цікавим називається правило, якщо в нього немає предків або воно є R -цікавим щодо будь-якого свого найближчого предка.

Необхідно знайти закономірності, що є узагальненими асоціативними правилами виду $X \rightarrow Y$, причому підтримка правила $X \rightarrow Y$ більше або дорівнює деякому наперед заданому значенню мінімальної підтримки і вірогідність більше або дорівнює значенню мінімальної вірогідності і правила $X \rightarrow Y$ є цікавими або частково цікавими.

Метод обчислення узагальнених асоціативних правил

Етап 1. Пошук множин елементів, що часто зустрічаються, підтримка яких більше ніж заданий поріг підтримки (мінімальна підтримка).

Етап 2. Обчислення правил на основі знайдених на попередньому етапі множин елементів, що часто зустрічаються. Основна ідея обчислення правил на основі множин, що часто зустрічаються, полягає в такому: якщо $ABCD$ – це множина елементів, що часто зустрічається, то на основі цієї множини можна побудувати правила $X \rightarrow Y$ (наприклад, $AB \rightarrow CD$), причому $X \cup Y = ABCD$. Підтримка правила дорівнює підтримці множини, що часто зустрічається. Вірогідність правила обчислюється за формулою $conf(X \rightarrow Y) = supp(X \rightarrow Y) / supp(X)$. Правило додається до результуючого списку правил, якщо вірогідність цього правила більше порога $minconf$.

Етап 3. З результуючого списку правил видаляються всі «нецікаві» правила.

Метод пошуку асоціативних правил Apriori:

- *передобробка даних*: приведення всіх даних до бінарного вигляду, зміна структури даних (подання даних у вигляді таблиці), упорядкування елементів
- крок 1: знаходження наборів елементів, що часто зустрічаються
- крок 2: витягнення правил з набору елементів, що часто зустрічаються



Програмні засоби реалізації методу пошуку асоціативних правил:

- Deductor – метод “Асоціативні правила”, візуалізатор “Дерево правил”, візуалізатор “Що - якщо”
- R – бібліотеки arules, arulesViz