

К.т.н., доцент Чертов О.Р., магістрант Переймибіда А.В.

**Національний технічний університету України
«Київський політехнічний інститут»**

СИСТЕМА ОПЕРАТИВНОГО ВИЯВЛЕННЯ ПРИХОВАНИХ ЗАЛЕЖНОСТЕЙ ПРИ АНАЛІЗІ РУХУ ТОВАРІВ

Abstract

Oleg Chertov, assoc. prof., PhD; Anton Pereimybida, student

System for online identification of hidden relations during goods movement analysis

This paper concerns different methods of data mining, namely association rules exploring. Common algorithms of finding association rules described. An improvement of chosen method is proposed to be implemented in the system for online identification of hidden relations.

Вступ

Наразі актуальною є проблема забезпечення ефективного управління випереджаючими логістичними потоками [1]. При цьому одним з найбільш важливих завдань є створення механізмів, що спрямовані на формування та поповнення запасів матеріальних ресурсів.

Існуючі методи [2] не враховують усю множину залежностей, які у неявному вигляді існують між сутностями логістичної системи та є суттєвими. Тому постає завдання отримання інформації щодо зазначених прихованих зв'язків. Виявлені залежності можуть бути використані при реорганізації розташування товарів у складських приміщеннях, зміни політики взаємодії із постачальниками тощо.

Останнім часом неухильно зростає інтерес до методів «виявлення знань в базах даних» (knowledge discovery in databases) [3]. Значні об'єми сучасних баз і сховищ даних та їх стрімке зростання викликали стійкий попит на нові алгоритми аналізу даних, одним з яких є пошук асоціативних правил, що дозволяє виявити приховані зв'язки у даних.

Постановка задачі

В роботі ставиться завдання розробки методу пошуку прихованих залежностей під час процесу оперативного аналізу руху товарів.

Правила виявлення прихованих залежностей

Одним із головних шляхів оптимізації діяльності логістичної системи є скорочення транспортних витрат. Задля цього потрібно визначати оптимальне розташування та групування складських товарів. Саме для виділення певних споріднених груп сутностей пропонується використання пошуку асоціативних правил.

Метою даного пошуку є встановлення наступних залежностей: якщо в транзакції зустрівся деякий набір елементів (товарів) X , то на підставі цього можна зробити висновок про те, що інший набір елементів Y також повинен з'явитися в цій транзакції.

Нехай $I = \{i_1, i_2, i_3, i_n\}$ — множина (набір) сутностей, що називаються елементами [4]. D — множина транзакцій, де кожна транзакція T — це набір елементів з I . Кожна транзакція є бінарним вектором, де $t[k]=1$, якщо елемент i_k присутній в транзакції, інакше — $t[k]=0$.

Підтримка правила (support, Supp) — показує, який відсоток транзакцій відповідає даному правилу. Оскільки правило будується на основі набору, то це означає, що правило $X \Rightarrow Y$ має підтримку, рівну підтримці набору I , що визначає дане правило.

$$Supp_{X \Rightarrow Y} = Supp_{X \cup Y} = \frac{|D_{X \cup Y}|}{|D|} \quad (1)$$

Достовірність (confidence, Conf) правила показує ймовірність того, що з наявності в транзакції набору X впливає наявність у ній набору Y . Достовірністю правила $X \Rightarrow Y$ є відношення кількості транзакцій, що містять набори X та Y , до кількості транзакцій, що містять набір X :

$$Conf_{X \Rightarrow Y} = \frac{|D_{X \cup Y}|}{|D_X|} = \frac{Supp_{X \cup Y}}{Supp_X} \quad (2)$$

Алгоритми пошуку асоціативних правил призначені для знаходження всіх правил $X \Rightarrow Y$, причому підтримка і достовірність цих правил повинні бути вищими за деякі наперед визначені пороги, що називаються відповідно мінімальною підтримкою і мінімальною достовірністю. Зазначені правила формуються з «частих наборів» — наборів (сукупностей) елементів, частота появи яких більша за деяке порогове значення.

Важливою властивістю, що скорочує область пошуку, є властивість антимонотонності. Згідно з нею, значення підтримки деякої множини елементів (деякого набору) не може перевищувати мінімальне значення підтримки довільної її підмножини.

Оперативне виявлення прихованих залежностей

Існуючі методи розраховані на так званий «статичний» аналіз. При цьому виконується багаторазове сканування усієї множини транзакцій, з метою визначення частих наборів та підрахунку підтримки і достовірності побудованих правил. Даний підхід не є прийнятним під час аналізу руху товарів у системі управління складом через процес постійного поповнення множини транзакцій, вилучення або додавання елементів. Звідси виникає необхідність у методах оперативного виявлення закономірностей та їх оновлення.

Пропонований метод спрямований на оперативне виявлення асоціативних правил та їх модифікацію з кожною новою транзакцією. Його робота здійснюється за два етапи.

На першому етапі проводиться аналіз зрізу усієї бази транзакцій традиційним статичним методом пошуку асоціативних правил. При цьому формуються дві множини: 1) правила, що задовольнили деяке визначене користувачем значення підтримки і достовірності, та 2) так звані «кандидати» — тобто правила, підтримка яких знаходиться в деякому заданому околі.

Для зберігання інформації про набори та транзакції використовується деревовидна структура *FP-tree* [5, 6]. Спочатку створюється порожній кореневий вузол, який відповідає пустій множині та помічається як *null*. Далі для кожної наступної транзакції відбувається додавання наборів у існуюче дерево. При цьому дотримуються такого правила: якщо для наступного елемента в дереві зустрічається вузол, ім'я якого співпадає з іменем елемента, то новий вузол не створюється, а відбувається інкрементація значення відповідного вузла. Інакше для даного товару створюється новий вузол і йому присвоюється значення 1. Між вузлами, що відповідають одному предмету, також встановлюються посилання. Для прикладу, маємо множину з п'яти елементів — $\{a, b, c, d, e\}$ та три транзакції $\{a, b\}$, $\{b, c, d\}$ та $\{a, b, c, d, e\}$. Побудоване дерево зображене на рис.1.

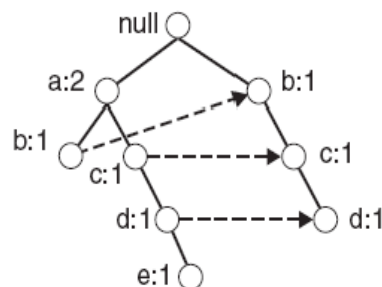


Рис. 1. FP-дерево.

Використання даної деревовидної структури дозволяє зменшити витрати пам'яті для зберігання транзакцій, надає зручний механізм їх модифікації та поповнення.

На другому етапі методу в динаміці з кожною новою транзакцією проводиться аналіз лише двох отриманих на першому етапі множин. Це дозволяє зменшити витрати ресурсів і часу та забезпечує можливість оперативного аналізу руху товарів. При цьому коригуються отримані правила, обчислюються нові значення підтримки, а також відсіюються або додаються нові правила на основі множини кандидатів. Саме поєднання двох зазначених етапів дозволяє підприємству отримувати актуальні дані про приховані зв'язки товарів.

Висновки

В даній роботі була розглянута проблема виявлення прихованих залежностей при аналізі руху товарів. Для її вирішення був досліджений апарат пошуку асоціативних правил та запропонований новий метод, котрий послідовно об'єднує класичний пошук залежностей на базі підрахунку підтримки і достовірності правил та їх перерахунку на відносно невеликій множині раніше виявлених правил і кандидатів до них.

Література

1. Логистика / Под ред. Б.А. Аникина, Т.А. Родкиной.. — 2-е. — Москва : Проспект, 2010. — 406 с.
2. Сток Д. Р. Стратегическое управление логистикой. — М : Инфра-М, 2005. — 797 с.
3. Барсегян А.А., Куприянов М.С., Степаненко В.В., Холод И.И. Технологии анализа данных: Data Mining, Visual Mining, Text Mining, OLAP. — СПб. : БХВ-Петербург — 2007. — 384 с.
4. Agrawal R. Mining Association Rules Between Sets of Items in Large Databases / Agrawal R., Imielinski T., Swami A // SIGMOD Conference, 1993. — 563 с.
5. Agarwal, R., Aggarwal, C., V. Prasad. A tree projection algorithm for generation of frequent itemsets // Journal of Parallel and Distributed Computing, 2001. — 1841 с.
6. Han, Jiawei; Pei, Jian; Yin, Yiwen; Mao, Runying. Mining Frequent Patterns Without Candidate Generation: A Frequent-pattern Tree Approach // Data Mining and Knowledge Discovery, 2004. — 295 с.