

АСОЦІАТИВНІ МОДЕЛІ ДОСЛІДЖЕННЯ ЕКОНОМІЧНИХ ОБ'ЄКТІВ

ІВАНЧЕНКО Надія Олександрівна,

кандидат економічних наук, доцент,

Інститут міжнародних відносин Національного авіаційного університету

У статті розглядається асоціативні моделі DATA MINING (DM) один із варіантів кластеризації, що використовується бізнес-аналітикою для пошуку груп характеристик які переважно спостерігаються одночасно.

Ключові слова: інтелектуальний аналіз, технології, знання, моделі, трансакції, ринок, товари.

Постановка завдання

В інформаційній системі активно працюючого підприємства нагромаджуються великі обсяги різноманітних відомостей. Користувачі можуть аналізувати ці відомості за допомогою класичних аналітичних інструментів, вбудованих в прикладні рішення — звітів, зведених таблиць, діаграм. Ці інструменти дозволяють представити картину господарської діяльності підприємства в зручному для аналізу вигляді, визначити відхилення від плану, підготувати ухвалення управлінських рішень, спланувати нові показники на основі тенденцій минулих періодів [1].

Для успішного просування товарів завжди важливо знати, що і як продається, а також, хто є споживачем. Особливо це потрібно в умовах жорсткої конкуренції сьогодення. Вичерпну відповідь на перше питання дають такі засоби *Data Mining*, як аналіз ринкових кошиків і сиквенціальний аналіз. Знаючи зв'язки між покупками і тимчасові закономірності, можна оптимальним чином регулювати пропозицію. З іншого боку, маркетинг має можливість безпосередньо управляти попитом, але для цього необхідно знати якомога більше про поведінку споживачів. *Data Mining* дозволяє вирішувати завдання виділення груп споживачів зі схожими стереотипами поведінки, тобто сегментувати ринок. Для цього використовують кластеризацію і класифікацію.

Сиквенціальний аналіз допомагає торговим підприємствам приймати рішення про створення товарних запасів.

Асоціативні моделі використовується тоді, коли декілька подій пов'язані одна з одною. Вони характеризують близькість різних одночасно спостережуваних категоріальних характеристик і виражаються простими правилами. Такими характеристиками можуть бути товари, послуги та попит на них, що одночасно надаються споживачеві.

Основними перевагами асоціативних правил є легке сприйняття людиною, а саме:

- *корисні правила* — складаються з дійсної інформації, яка раніше була не відомою, але має логічне пояснення. Такі правила можуть бути використані для прийняття рішень, які приносять вигоду;

- *тривіальні правила* — складаються з дійсної та легко пояснюваної інформації яка є відомою. Такі правила не приносять ніякої користі, оскільки відображають відомі закони в досліджуваній галузі, або результати минулої діяльності. Іноді такі правила можуть використовуватись для перевірки виконання рішень, прийнятих на основі попереднього аналізу;

- *незрозумілі правила* — складаються з інформації, яка не може бути поясненою. Такі правила можуть бути отримані на основі аномальних значень або таємних знань. Напряму такі правила не можна використовувати для прийняття рішень. Для кращого розуміння потрібно проводити додатковий аналіз.

Аналіз останніх досліджень та публікацій

У повноцінну систему маркетингової аналітики входить: сховище даних (СД), засоби обробки, аналізу, та візуалізації результатів.

Концепція СД передбачає розділення структур збереження даних для оперативної обробки і обробки аналітичних запитів. Сховище даних накопичує трансакції, які надходять з систем оперативного обліку товарного ринку. Використання окремого СД обумовлено необхідністю об'єднання інформації з різнорідних джерел, даних і організації інформації в зручній для подальшого аналізу формі.

Для аналізу даних використовуються засоби багатовимірного зберігання і аналітичної обробки даних (OLAP — Online Analytical Processing), що представляють ринкові бізнес-операції у вигляді фактів (об'ємів продажу, числа одиниць на складі і т. д.) і вимірювань (сегмент-кластер, час, географія, постачальник, покупець, товар і т.д.). Засоби OLAP дозволяють здійснювати стратегічний огляд ситуації і в реальному часі одержувати відповіді на питання, що цікавлять маркетолога-аналітика [4].

З метою автоматичного виявлення раніше невідомих знань в накопичених даних використовуємо технології інтелектуального аналізу даних DM, та Knowledge Discovery in Databases. На відміну від технології OLAP сам пошук закономірностей і шаблонів в даних здійснюється не користувачем системи, а самою технологією, що реалізує декілька алгоритмів DM [1].

Сучасні технології Data Mining (discovery-driven data mining) перелопачують інформацію з метою автоматичного пошуку шаблонів (патернів), характерних для яких-небудь фрагментів неоднорідних багатовимірних даних. На відміну від оперативної аналітичної обробки даних (online analytical processing, OLAP) в Data Mining тягар формулювання гіпотез і виявлення незвичайних (unexpected) шаблонів перекладено з людини на комп'ютер.

Таблиця 1

Приклади формулювань завдань при використанні методів OLAP і Data Mining

OLAP	Data Mining
Які середні показники травматизму для тих, що палять?	Які краще всього нещасні випадки?
Які середні розміри телефонних рахунків існуючих клієнтів порівняно з рахунками колишніх клієнтів (що відмовилися від послуг телефонної компанії)?	Які характеристики відрізняють клієнтів, які, ймовірно, збираються відмовитися від послуг телефонної компанії?
Яка середня величина щоденних покупок по вкраденій і не вкраденій кредитній картці?	Які схеми покупок характерні для шахрайства кредитними картками?

В основі більшості інструментів *Data Mining* лежать дві технології машинне навчання та візуалізація. Якість візуалізації визначається можливостями графічного відображення значень даних.

Ефективність методів машинного навчання в основному визначається їх здатністю досліджувати велику кількість взаємозалежних даних.

Обидві технології доповнюють одна одну в процесі *Data Mining* — аналізу. Візуалізація використовується для пошуку виключень, загальних тенденцій і залежностей та допомагає у вилученні даних на початковому етапі проекту. Машинне навчання використовується пізніше, для пошуку залежностей у проекті.

Основними властивостями знань виявленими за допомогою технології DM є [7]:

- *знання повинні бути новими*, раніше невідомими. Затрачені зусилля на відкриття знань, які вже відомі користувачу, на окупаються. Тому цінність представляють саме нові, раніше невідомі знання.

- *знання повинні бути нетривіальними*. Результати аналізу повинні відображати неочевидні, неочікувані закономірності в даних. Результати. Які могли бути отримані більш простими способами не оправдують залучення методів DM.

- *знання повинні бути практично корисними*. Найдені знання повинні бути застосовані на нових даних з достатньо високою вірогідністю. Корисність полягає в тому, щоб ці знання могли принести певну корисність при застосування.

- *знання повинні бути доступними для розуміння*. Знайдені закономірності повинні бути логічно пояснювальні та представлені в зрозумілому вигляді.

Результати та методи дослідження

У список основних задач, вирішуваних алгоритмами DM, входять: пошук асоціацій (зв'язків між різними характеристиками), кластеризація та сегментація (виявлення структури, груп, кластерів), дерева рішень, аналіз з виборчою дією, мережі впевненості, метод найближчих сусідів, нейронні мережі, нечітка логіка, генетичні алгоритми, регресійні методи (прогнозування, класифікація, відновлення функціональної залежності між характеристиками), еволюційне програмування. Візуалізація знайдених моделей та даних також виконує дуже важливу роль при аналізі.

З метою моделювання попиту були використані методи асоціативних правил для аналізу взаємозв'язків між товарами та товарними групами, що входять в одну транзакцію на товарному ринку.

Товари та набори об'єктів, що часто представлені в групах і становлять досліджувані набори, які позначимо множиною:

$$I = \{i_1, i_2 \dots i_n\},$$

де: i – об'єкти, що входять в аналізовані набори;

n – загальна кількість об'єктів.

Наприклад, вони відповідають наступному переліку об'єктів:

$$I = \{Товар_1, Твар_2, Товар_3, Товар_4, Товар_5, Товар_6, Товар_7\}$$

Набори об'єктів з множини I є транзакціями, знаходяться у експериментальній БД та піддаються аналізу є підмножиною множини I :

$$T = \{i_j \mid i_j \in I\}$$

Такі транзакції на ринку товарів відповідають наборам, що купуються споживачем і зберігаються в БД у вигляді транзакції - товарного чека або накладної. В них перераховуються товари, що придбані покупцем, їх ціна, кількість та ін. Набір транзакцій, інформація про яких доступна для аналізу, визначається множиною:

$$D = \{T_1, T_2, \dots, T_r, \dots, T_m\},$$

де: m – кількість доступних для аналізу транзакцій.

Множину транзакцій позначаємо:

$$D_{ij} = \{T_r \mid i_j \in T_r; j = 1..n; r = 1..m\} \subseteq D$$

Довільний набір об'єктів, що визначає попит на ринку позначимо:

$$F = \{i_j \mid i_j \in I; j = 1..n\}$$

Множина транзакцій, в які входить набір F , позначимо:

$$D_F = \{T_r \mid F \subseteq T_r; r = 1..m\} \subseteq D$$

Для визначення підтримки набору F до загальної кількості транзакцій використаємо функцію $Supp(F)$:

$$Supp(F) = \frac{|D_F|}{|D|}$$

При пошуку аналітик може вказати мінімальне значення підтримки його наборів, що задовольняє попит $Supp_{min}$. Набір товарів може бути багаторазовим, якщо значення його підтримки більше мінімального значення підтримки, заданого користувачем:

$$Supp(F) > Supp_{min}$$

Таким чином, при пошуку асоціативних правил попиту товарів на ринку треба знайти множину всіх частих наборів:

$$L = \{F \mid Supp(F) > Supp_{min}\}$$

Оскільки правило будується на підставі набору, то, значить, правило $X \geq Y$ має підтримку, рівну підтримці набору F , який складають X і Y :

$$Supp_{x \Rightarrow y} = Supp_F = \frac{|D_{F=X \cup Y}|}{|D|}$$

Очевидно, що правила, побудовані на підставі одного і того ж набору, мають однакову підтримку.

Для визначення достовірності правила $X \geq Y$ є відношення числа транзакцій, що містять набори X і Y , до числа транзакцій, що містять набір X . Це відношення показує вірогідність того, що з наявності в транзакції набору X слідує наявність в ній набору Y

$$Conf_{X \Rightarrow Y} = \frac{|D_{F=X \cup Y}|}{|D_X|} = \frac{Supp_{X \cup Y}}{Supp_X}$$

З експерименту видно, що чим більше достовірність, тим правило краще, причому у правил, побудованих на підставі одного і того ж набору, достовірність буде різна.

Але достовірність не дозволяє оцінити корисність правила. Якщо відсоток наявності в транзакціях набору Y за умови наявності в них набору X менше, ніж відсоток безумовної наявності набору Y , т. е. це значить, що вірогідність випадково вгадати наявність в транзакції набору Y більше, ніж передбачено за допомогою правила $X \geq Y$. Для виправлення такої ситуації вводимо міру покращення:

$$Conf_{X \Rightarrow Y} = \frac{|D_{F=X \cup Y}|}{|D_X|} < Supp_Y,$$

Покращення правила показує, чи корисним є правило випадкового вгадування та відповідає відношенню числа транзакцій, що містять набори X і Y , до кількості транзакцій, що містять набір X , і кількості транзакцій, що містять набір Y :

$$impr_{X \Rightarrow Y} = \frac{|D_{F=X \cup Y}|}{|D_X \cup D_Y|} = \frac{Supp_{X \cup Y}}{Supp_X Supp_Y}.$$

Якщо покращення більше одиниці, то це значить, що за допомогою правила передбачити наявність набору Y вірогідніше, ніж випадкове вгадування, якщо менше одиниці, то навпаки.

У останньому випадку можна використовувати заперечливе правило, тобто правило, яке передбачає відсутність набору Y :

$$X \Rightarrow \text{не} Y.$$

У такого правила покращення буде більше одиниці, оскільки

$$Supp_{\text{не} Y} = 1 - Supp_Y.$$

Таким чином, можна одержати правило, яке передбачає результат краще, ніж випадковим чином.

Дані оцінки використовується при генерації правил оцінювання попиту на ринку товарів. При пошуку асоціативних правил задаємо мінімальні значення перерахованих змінних. В результаті ті правила, які не задовольняють цим умовам, відкидаються і не включаються до рішення задачі. З цієї точки зору не можна об'єднувати різні правила, хоча і що мають загальне смислове навантаження. Наприклад, наступні правила:

$$X = \{i_1, i_2\} \Rightarrow Y = \{i_3\},$$

$$X = \{i_1, i_2\} \Rightarrow Y = \{i_4\}$$

не можна об'єднати в одне:

$$X = \{i_1, i_2\} \Rightarrow Y = \{i_3, i_4\},$$

оскільки достовірності їх будуть різні, отже, деякі з них можуть бути виключені, а деякі – ні.

Якщо об'єкти мають додаткові атрибути, які впливають на склад об'єктів в транзакціях, а отже, і в наборах, то вони повинні враховуватися в правилах, що генеруються. В цьому випадку умовна частина правил міститиме не тільки перевірку наявності об'єкту в транзакції, але і складніші операції порівняння: більше, менше, включення. Результуюча частина правил також може містити твердження щодо значень атрибутів. Наприклад, якщо у товарів розглядається ціна, то правила можуть мати наступний вигляд:

якщо товар_1. ціна < 3 то товар_2. ціна < 6.

Дане правило говорить про те, що якщо попит є на товар_1 за ціною менше 3, то, ймовірно, буде попит на товар_2 за ціною менше 6.

Отримані висновки

Засоби інтелектуального аналізу даних *DM* надають можливість моделювати попит товарів на ринку, прогнозувати збут продукції, складати маркетингові плани, прогнозувати вірогідність покупки клієнтом певного товару на підставі його віку, демографічних характеристик і інших ознак; визначити групи клієнтів, що характеризуються схожою поведінкою.

ДЖЕРЕЛА ТА ЛІТЕРАТУРА

1. Дюк В., Самойленко А. Data Mining: учебный курс. – СПб: Питер, 2001. – 368 с.
2. Корнеев В. В., Гареев А. Ф., Васютин С. В., Райх В. В. Базы данных. Интеллектуальная обработка информации. – М.: Нолидж, 2000. – 352 с., ил.
3. Ситник В. Ф. Засоби дейтамайнінгу для аналізу бізнесових рішень // Науково-технічна інформація. – 2002. – № 3. – С. 60-64.
4. Ситник В. Ф., Краснюк М. Т. Интеллектуальный анализ данных (дейтамайнинг): Навч. посіб. – К.: КНЕУ, 2007. – 376 с.
5. Berry, Michael J. A. and Gordon Linoff. Data Mining Techniques for Marketing, Sales, and Customer Support. – New York: Wiley Computer Publishing, 1997.
6. Филиппов В. А. Интеллектуальный анализ данных: методы и средства. – М.: Эдиториал УРСС, 2001.
7. Чубукова И. А. Data Mining: учебное пособие. – М.: Интернет-университет информационных технологий: БИНОМ: Лаборатория знаний, 2006. – 382 с.

Иванченко Н. О. Ассоциативные модели исследования экономических объектов / Институт международных отношений Национального авиационного университета.

В статье рассматриваются ассоциативные модели DATA MINING (DM) один из вариантов кластеризации, что используется аналитикой бизнеса для поиска групп характеристик, наблюдаемый преимущественно одновременно.

Ключевые слова: интеллектуальный анализ, технологии, знания, модели, транзакции, рынок, товары.

Ivanchenko N. O. Associative models of research of economic objects / Institute of International Relations National Aviation University.

In the article is associative models of DATA MINING (DM) that is used by analytic business geometry for the search of groups of descriptions which are mainly observed simultaneously.

Keywords: *intellectual analysis, technologies, knowledge's, models, transactions, market, commodities.*