

БАГАТОКРИТЕРІАЛЬНИЙ ПОРІВНЯЛЬНИЙ АНАЛІЗ МЕТОДІВ DATA MINING ДЛЯ ВИКОРИСТАННЯ У СУБД

Анотація. Досліджено найпопулярніші методи Data Mining: класичні методи, нейронні мережі, методи візуалізації, дерева рішень, поліноміальні нейронні мережі, k-найближчого сусіда. Ці методи використовуються для інтелектуального аналізу даних, веденню статистики та пошуку закономірностей у великих об'ємах даних. Для побудови рейтингу розглянутих методів обрано кількісний багатокритеріальний метод – метод аналізу ієрархій «у абсолютних вимірюваннях» (нормативний MAI, MAI-H). Отриманий рейтинг може застосовуватись для вибору методу Data Mining під конкретну задачу.

Ключові слова: методи Data Mining, інтелектуальний аналіз даних, багатокритеріальний аналіз, метод аналізу ієрархій у абсолютних вимірюваннях.

Yevtushenko Halyna, Ilchenko Andrii
National Metallurgical Academy of Ukraine

MULTIPLE-CRITERIA COMPARATIVE ANALYSIS OF DATA MINING METHODS FOR USING IN DBMS

Summary. The most popular methods of Data Mining are investigated: classical methods, neural networks, visualization methods, decision trees, polynomial neural networks, k-neighboring neighbors. These methods are used to conduct statistics and search for regularities in large volumes of data. To construct the rating of the considered methods a quantitative multiple-criteria method the analytic hierarchy process in absolute measurements was chosen. The resulting rating can be used to select the Data Mining method for a specific task.

Keywords: data mining methods, multiple-criteria comparative analysis, analytic hierarchy process in absolute measurements.

Постановка проблеми. Корпоративні бази даних будь-якого підприємства включають в себе багато таблиць з великою кількістю зв'язків та даних. Як правило, кожен рядок у таблицях описує якийсь факт, дію, бізнес-процес і т.д. Наприклад, запис у таблиці продажів відображає факт продажу, ідентифікує який товар, коли і де було продано, якою формою оплати скористався покупець та, в залежності від примх підприємства, може відображати велику кількість інших статистичних даних.

Масиви даних настільки величезні, що люди ні просто не до снаги проаналізувати таку кількість інформації та отримати результат, на основі якого можна здійснювати прийняття рішень в майбутньому.

Саме вирішенню наведених проблем призначені методи інтелектуального аналізу даних (Data Mining). Ці методи дозволяють знаходити в сирих даних раніше невідомі нетривіальні практично корисні і доступні інтерпретації знань, необхідних для прийняття рішень у різних сферах людської діяльності.

Аналіз останніх досліджень і публікацій. Проблематика дослідження та класифікації методів Data Mining є актуальною [1-3]. При розробці систем для ведення бізнесу методи Data Mining займають дуже важливе місце, так як дають можливість приймати важливі для життя компанії рішення, опираючись на реальні дані.

У роботі [4] наведена класифікація цих методів, а також аналіз по окремим критеріям, проте загальний рейтинг не представлено.

Виділення невирішених раніше частин загальної проблеми. Багатокритеріальний порівняльний аналіз сучасних методів Data Mining, що використовуються для інтелектуаль-

ного аналізу даних, веденню статистики та пошуку закономірностей у великих об'ємах даних, є необхідним для обґрунтування вибору цих методів у конкретному проекті.

Мета статті. Дана робота присвячена багатокритеріальному аналізу та складанню рейтингу методів інтелектуального аналізу даних. Для побудови рейтингу розглянутих методів обрано кількісний багатокритеріальний метод – метод аналізу ієрархій «у абсолютних вимірюваннях» [5], який програмно реалізований у системі підтримки прийняття рішень NooTron [6].

MAI-H дозволяє порівнювати велику (у принципі – необмежену) кількість альтернатив; важливо також, що він надає можливість використання шкал інтенсивностей (лінгвістичних стандартів), які можна застосовувати для роздільної оцінки об'єктів (інструментів тестування у нашому випадку) [5].

Шість методів Data Mining проаналізовані та оцінені за наступними критеріями (табл. 1) [4]:

- 1) Точність роботи методу Data Mining.
- 2) Масштабованість – здатність методу Data Mining обробити більший обсяг роботи або бути легко розширеним.
- 3) Інтерпретованість – зрозумілість і прозорість реалізації методу Data Mining.
- 4) Придатність – наявність у методі Data Mining певних властивостей, необхідних для використання цього методу за призначенням.
- 5) Трудомісткість – складність при реалізації даного методу Data Mining розробниками.
- 6) Різносторонність – властивість методів Data Mining, яка відображає можливість застосування даних методів у різних сферах людської діяльності та їх адаптивність при зміні напрямку застосування.

Таблиця 1

Порівняльний аналіз методів Data Mining

Алгоритми	Точність	Масштабованість	Інтерпретованість	Придатність	Трудомісткість	Різносторонність	Швидкість	Популярність
	К1	К2	К3	К4	К5	К6	К7	К8
Класичні методи	Нейтрально	Високо	Нейтрально високо	Високо	Нейтрально	Нейтрально	Високо	Низько
Нейронні мережі	Високо	Низько	Низько	Низько	Нейтрально	Низько	Дуже низько	Низько
Методи візуалізації	Високо	Дуже низько	Високо	Високо	Дуже високо	Низько	Надзвичайно низько	Нейтрально високо
Дерева рішень	Низько	Високо	Високо	Нейтрально високо	Високо	Високо	Нейтрально високо	Нейтрально високо
Поліноміальні нейронні мережі	Високо	Нейтрально	Високо	Нейтрально високо	Нейтрально низько	Нейтрально	Нейтрально низько	Нейтрально
к-найближчого сусіда	Низько	Дуже низько	Нейтрально високо	Нейтрально	Нейтрально низько	Низько	Високо	Низько

Таблиця 2

Багатокритеріальний порівняльний аналіз методів Data Mining за допомогою MAI

Алгоритми	К1	К2	К3	К4	К5	К6	К7	К8	Глобальні пріоритети
	0,243	0,069	0,058	0,081	0,127	0,059	0,33	0,033	
Класичні методи	0,106	0,231	0,157	0,231	0,106	0,106	0,231	0,048	0,167
Нейронні мережі	0,231	0,048	0,048	0,048	0,106	0,048	0,033	0,048	0,095
Методи візуалізації	0,231	0,033	0,231	0,231	0,331	0,048	0,024	0,157	0,148
Дерева рішень	0,048	0,231	0,231	0,157	0,231	0,231	0,157	0,157	0,154
Поліноміальні нейронні мережі	0,231	0,106	0,231	0,157	0,07	0,106	0,07	0,106	0,131
к-найближчого сусіда	0,048	0,033	0,157	0,106	0,07	0,048	0,231	0,048	0,121

7) Швидкість – відображає власне швидкість отримання результату при застосуванні методів Data Mining.

8) Популярність – відображає частоту використання методів Data Mining при вирішенні задач.

На першому етапі у MAI-N необхідно визначити пріоритети критеріїв відносно мети задачі (складання рейтингу методів Data Mining для використання у СУБД) – результат наведено на рис. 1; на другому етапі – числову інтерпре-

тацію лінгвістичного опису кожного критерію (рис. 2).

Далі виконаємо багатокритеріальний порівняльний аналіз методів Data Mining за допомогою MAI, замінивши лінгвістичні описи оцінок методів на числові. Результат аналізу наведено у таблиці 2.

Візуалізація аналізу методів Data Mining відносно розглянутих критеріїв наведено на рисунку 3, а рейтинг розглянутих методів Data Mining наведено на рисунку 4.

Назва	Кр1	Кр2	Кр3	Кр4	Кр5	Кр6	Кр7	Кр8	ЛПр.
Кр1 Точність	1	7	5	2	1	4	1	5	0.243
Кр2 Масштабованість	1/7	1	1	1	1	1	1/4	2	0.069
Кр3 Інтерпретованість	1/5	1	1	1	1/3	1	1/5	2	0.058
Кр4 Придатність	1/2	1	1	1	1	1	1/5	3	0.081
Кр5 Трудомісткість	1	1	3	1	1	3	1/5	4	0.127
Кр6 Різносторонність	1/4	1	1	1	1/3	1	1/6	2	0.059
Кр7 Швидкість	1	4	5	5	5	6	1	7	0.330
Кр8 Популярність	1/5	1/2	1/2	1/3	1/4	1/2	1/7	1	0.033

Dim	Lam	CI	CR
8.000	8.400	0.057	0.041

Рис. 1. Матриця парних порівнянь критеріїв оцінки методів Data Mining

	Название	Кр1	Кр2	Кр3	Кр4	Кр5	Кр6	Кр7	Кр8	ЛПр.
Кр1	Надзвичайно низько	1	1/2	1/3	1/4	1/5	1/6	1/7	1/8	0.024
Кр2	Дуже низько	2	1	1/2	1/3	1/4	1/5	1/6	1/7	0.033
Кр3	Низько	3	2	1	1/2	1/3	1/4	1/5	1/6	0.048
Кр4	Нейтрально низько	4	3	2	1	1/2	1/3	1/4	1/5	0.070
Кр5	Нейтрально	5	4	3	2	1	1/2	1/3	1/4	0.106
Кр6	Нейтрально високо	6	5	4	3	2	1	1/2	1/3	0.157
Кр7	Високо	7	6	5	4	3	2	1	1/2	0.231
Кр8	Дуже високо	8	7	6	5	4	3	2	1	0.331

Dim	Lam	CI	CR
8.000	8.288	0.041	0.029

Рис. 2. Матриця парних порівнянь лінгвістичного опису кожного критерію для порівняння методів Data Mining

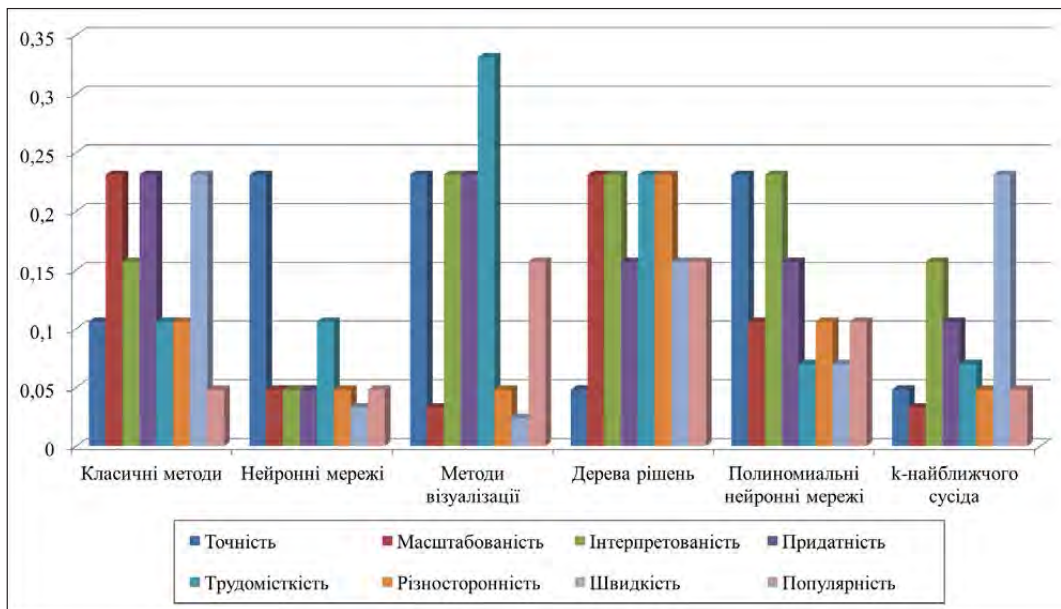


Рис. 3. Багатокритеріальний порівняльний аналіз методів Data Mining за допомогою MAI

ТЕХНІЧНІ НАУКИ

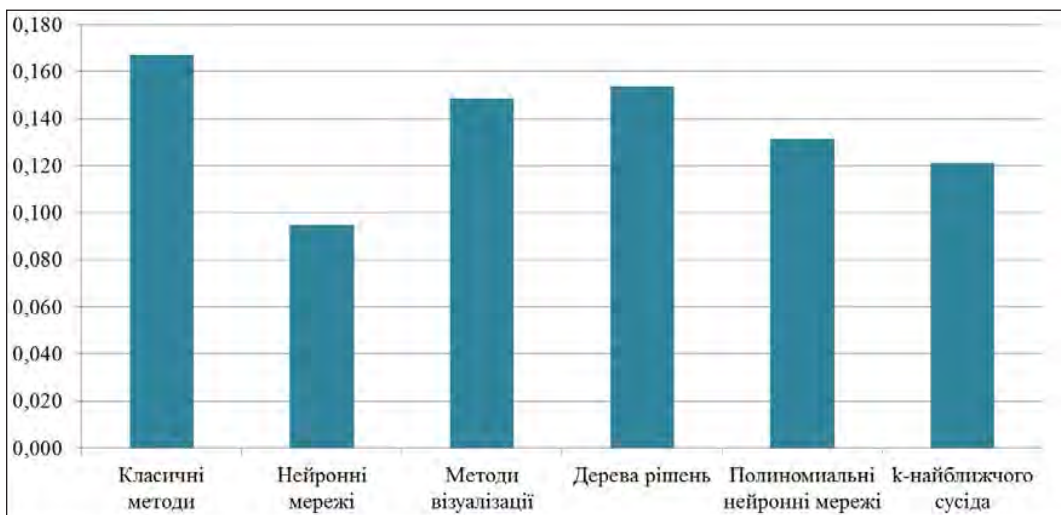


Рис. 4. Результат багатокритеріального порівняльного аналізу методів Data Mining за допомогою MAI

Висновки.

1. У процесі виконання роботи проаналізовані найпопулярніші методи Data Mining, що використовуються для інтелектуального аналізу даних, веденню статистики та пошуку закономірностей у великих об'ємах даних. Також на основі описаних критеріїв було складено рейтинг на основі методу аналізу ієрархій у абсолютних вимірюваннях із застосуванням СППР NooTron.

2. Згідно отриманого вище рейтингу (рис. 4) перше місце зайняли класичні методи

Data Mining – тобто саме ці методи буде доцільно використовувати при інтелектуальному аналізі даних у більшості випадків.

3. Якщо є пріоритетність певного критерію, то згідно діаграми на рис. 3 можна вибрати метод Data Mining, який потрібен для певної задачі, але знову ж оптимальними за всіма критеріями є класичні методи.

4. Результати проведеного аналізу можуть бути використані для обґрунтування вибору методів Data Mining для проведення інтелектуального аналізу даних у конкретному проекті.

Список літератури:

1. Михалев А.И. Компьютерные методы интеллектуальной обработки данных: [учебное пособие] / А.И. Михалев, Е.А. Винокурова, С.Л. Сотник. – Днепропетровск: НМетАУ, ИК «Системные технологии», 2014. – 209 с.
2. Барсегян А.А. Методы и модели анализа данных: OLAP И Data Mining / А.А. Барсегян, М.С. Куприянов, В.В. Степаненко, И.И. Холод. – СПб.: BHV, 2004. – 331 с.
3. Дюк В.А. Data Mining – интеллектуальный анализ данных // Информационные технологии [Електронний ресурс]. – Режим доступу: <http://www.inftech.webservis.ru/it/database/datamining/ar2.html>, вільний. – Загол. з екрану.
4. Інтелектуальний аналіз даних (Data mining) – Огляд методів DataMining [Електронний ресурс]. – Режим доступу: <http://intellect-tver.ru/?p=165>, вільний. – Загол. з екрану.
5. Саати Т.Л. Принятие решений при зависимостях и обратных связях: Аналитические сети / Т.Л. Саати. – М.: Изд-во ЛКИ, 2008. – 360 с.
6. Система поддержки принятия решений NooTron [Електронний ресурс]. – Режим доступу: <http://nootron.net.ua>, вільний. – Загол. з екрану.