

ОГЛЯД МЕТОДІВ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ

Євсін Г.А.¹⁾, Львов С.Г.²⁾

¹⁾ НТУ «ХПІ», 61002, Харків, вул. Кирпичова, 2, hryhorii.yevsin@cit.khpi.edu.ua

²⁾ НТУ «ХПІ», 61002, Харків, вул. Кирпичова, 2, serhii.lvov@khpi.edu.ua

Для автоматичного аналізу даних використовується Data Mining –технологія інтелектуального аналізу даних з метою виявлення прихованих закономірностей у вигляді значущих особливостей, кореляцій, тенденцій і шаблонів.

Специфіка сучасних вимог до продуктивного оброблення інформації така:

- дані мають необмежений обсяг;
- дані є різномірними (кількісними, якісними, текстовими);
 - результати повинні бути конкретний і зрозумілий;
 - інструменти для обробки «сирих даних» повинні бути прості у використуванні.

Таблиця 1– Характеристика методів Data Mining

Назва	Характеристика
Класифікація(Classification)	Найпростіша і поширена задача Data Mining. Для вирішення задачі класифікації можуть використовуватися методи: найближчого сусіда (Nearest Neighbor); к-ближайшого сусіда (k-Nearest Neighbor); нейроні мережі (neural networks).
Кластеризація (Clustering)	Є логічним продовженням ідеї класифікації. Це задача складніша, особливість кластеризації полягає в тому, що класи об'єктів спочатку не визначені.
Асоціація (Associations)	В ході рішення задачі пошуку асоціативних правил відшукуються закономірності між зв'язаними подіями в наборі даних.
Послідовність (Sequence)	Послідовність дозволяє знайти тимчасові закономірності між транзакціями в часі.
Прогнозування (Forecasting).	В результаті рішення задачі прогнозування на основі особливостей існуючих даних оцінюються пропущені або ж майбутні значення цільових чисельних показників.
Візуалізація (Visualization, Graph Mining)	В результаті візуалізації створюється графічний образ аналізованих даних. Для вирішення задачі візуалізації використовуються графічні методи, що показують наявність закономірностей в даних.
Підведення підсумків (Summarization)	Задача, мета якої – опис конкретних груп об'єктів з аналізованого набору даних та інші.

В основу сучасної технології Data Mining встановлена концепція шаблонів (pattern), що відображають фрагменти багатоаспектних взаємостосунків в даних. Цими шаблонами є закономірності, властиві підвибіркам даних, які можуть бути компактно виражені у формі, зрозумілій людині. Пошук шаблонів проводиться

методами, не обмеженими рамками апріорних припущень про структуру вибірки і вид розподілів значень аналізованих показників.

Причини популярності Data Mining є: стрімке накопичення даних; загальна комп'ютеризація бізнес-процесів; проникнення Інтернет у всі сфери діяльності; прогрес в області інформаційних технологій: вдосконалення СУБД і сховищ даних; прогрес в області виробничих технологій: стрімке зростання продуктивності комп'ютерів, об'ємів накопичувачів, впровадження Grid систем.

Основна особливість Data Mining – це поєднання широкого математичного інструментарію (від класичного статистичного аналізу до нових кібернетичних методів) і останніх досягнень у сфері інформаційних технологій. В технології Data Mining гармонійно об'єдналися строго формалізовані методи і методи неформального аналізу, тобто кількісний і якісний аналіз даних.

Задачі Data Mining, залежно від моделей, що використовуються, можуть бути дескриптивними і прогнозуючими. В результаті вирішення описових (descriptive) задач аналітик одержує шаблони, що описують дані, які піддаються інтерпретації. Ці задачі описують загальну концепцію аналізованих даних, визначають інформативні, підсумкові, відмітні особливості даних.

Прогнозуючі (predictive) задачі ґрунтуються на аналізі даних, створенні моделі, прогнозі тенденцій або властивостей нових або невідомих даних.

Важливе положення Data Mining – нетривіальність розшукуваних шаблонів. Це означає, що знайдені шаблони повинні відображати неочевидні, несподівані (unexpected) регулярності в даних, складові так званих прихованих знань (hidden knowledge). Більшість аналітичних методів, що використовуються в технології Data Mining – це відомі математичні алгоритми і методи. Новою в їх застосуванні є можливість їх використання при рішенні тих або інших конкретних проблем, обумовлена новими можливостями технічних і програмних засобів, що з'явилися. Слід зазначити, що більшість методів Data Mining була розроблена в рамках теорії штучного інтелекту.

З появою і ускладненням інтелектуальних систем очевидну значущість мають підходи до побудови систем з використанням, так званих шарів (рівнів). Багаторівневий підхід – це модель взаємодії, в якій набір інтелектуальних систем або їх складових компонентів взаємодіє і обмінюється знаннями в деякому внутрішньому представленні.

Складність і різноманітність методів Data Mining вимагають створення спеціалізованих засобів для вирішення типових завдань аналізу інформації в конкретних галузях.

Список літератури:

1. Акіменко В.В. Прикладні задачі інтелектуального аналізу даних (DATA MINING). К.: КНУ ім. Тараса Шевченка, 2018. 152 с.
2. Черняк О. І. Інтелектуальний аналіз даних: підручник / О. І. Черняк, П. В. Захарченко. – К.: Знання, 2014. – 599 с.
3. Zumel N., Mount J. Practical Data Science with R. Manning Publications Co., 2014. 417 p.