***Т. П. Макаренко,*** *магістр*

*Вищий навчальний заклад Укоопспілки*

*«Полтавський університет економіки і торгівлі»*

**ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ РИНКУ ПРАЦІ В УКРАЇНІ**

Інтелектуальний аналіз даних (ІАД,Data Mining), або розвідка даних – термін, що застосовується для опису здобуття знань у бази даних, дослідження даних, обробки зразків даних, очищен-ня і збору даних. Це процес виявлення кореляції, тенденцій, шаблонів, зв’язків і категорій.

Термін Data Mining дістав назву від двох понять: дані – data і переробка сирого матеріалу (гірської руди) – mining.

Data Mining – предметна область що виникла і розвивається на базі таких наук, як прикладна статистика, розпізнавання образів, штучний інтелект, теорія баз даних тощо.

Виникнення і розвиток Data Mining зумовлені різними факто-рами, серед яких вирізняємо основні: вдосконалення програмно-апаратного забезпечення; вдосконалення технологій зберігання і запису даних; накопичення великої кількості ретроспективних даних; вдосконалення алгоритмів обробки інформації.

Сутність і мету технології Data Mining можна описати так: це технологія, призначена для пошуку у великих інформаційних масивах даних неочевидних, об’єктивних, корисних на практиці закономірностей. ІАД здійснюється за допомогою використання технологій розпізнавання шаблонів, а також статистичних і ма-тематичних методів.

При розвідці даних багаторазово виконуються операції і пе-ретворення над «сирими» даними (відбір ознак, стратифікація, кластеризація, візуалізація і регресія), що призначені для знахо-дження:

– структур, які інтуїтивно зрозумілі для людей і краще роз-кривають суть бізнес-процесів, що лежать в основі їх протікання;

**- © ПУЕТ -** 47

– моделей, які можуть передбачити результат або значення певних ситуацій, використовуючи історичні або суб’єктивні дані.

Інтелектуальний аналіз даних – процес автоматичного по-шуку прихованих закономірностей або взаємозв’язків між змін-ними у великих масивах необроблених даних, що поділяється на задачі класифікації, моделювання і прогнозування. Класичне визначення цього терміна дав у 1996 р. один із засновників цього напряму Г. П’ятецький-Шапіро.

Data Mining – це процес виявлення у необроблених даних раніше невідомих нетривіальних, практично корисних і доступ-них інтерпретацій знань, необхідних для прийняття рішень у різних сферах діяльності.

За визначенням SAS Institute, Data Mining – це процес ви-

ділення, дослідження і моделювання великих обсягів даних для виявлення невідомих до цього структур (patterns) з метою до-сягнення переваг у бізнесі.

За визначенням Gartner Group, Data Mining – це процес, мета якого – виявляти нові кореляції, зразки і тенденції у результаті просіювання великого обсягу даних з використанням методик розпізнавання зразків і статистичних та математичних методів.

В основу технології Data Mining покладено концепцію шаб-лонів (patterns), що є закономірностями, які властиві вибіркам даних і можуть бути подані у формі, зрозумілій людині.

Задачі Data Mining:

1. Класифікація (Classification) – виявляються ознаки, які ха-рактеризують групи об’єктів досліджуваного набору даних – класи; за цими ознаками новий об’єкт можна віднести до того або іншого класу. Для вирішення задач класифікації можуть ви-користовуватися методи: найближчий сусід (Nearest Neighbor); к-найближий сусід (k-Nearest Neighbor); байє-совські мережі (Bayesian Networks); індукція дерев рішень; нейронні мережі

(neural networks).

2. Кластеризація (Clustering) – результатом її є поділ об’єктів на групи.

3. Асоціація (Associations) – знаходять закономірності між пов’язаними подіями у наборі даних. Найбільш відомий алго-

48 **- © ПУЕТ -**

ритм рішення задачі пошуку асоціативних правил – алгоритм Аргіогі.

1. Послідовність (Sequence), або послідовна асоціація (sequential association), – дає можливість знайти часові законо-мірності між транзакціями. Завдання послідовності подібне до асоціації, але її метою є встановлення закономірностей між по-діями, пов’язаними за часом, тобто послідовність визначається високою ймовірністю ланцюжка пов’язаних за часом подій.
2. Прогнозування (Forecasting) – на основі особливостей іс-торичних даних оцінюються майбутні значення показників. За-стосовуються методи математичної статистики, нейронні мережі тощо.

Визначення відхилень (Deviation Detection), аналіз відхилень або викидів – виявлення й аналіз даних, що найбільше відрізня-ються від загальної чисельності даних, виявлення нехарактерних шаблонів.

Оцінювання (Estimation) – зводиться до прогнозу безперервних значень ознак.

Аналіз зв’язків (Link Analysis) – задача знаходження залеж-ностей у наборі даних.

Візуалізація (Visualization, Graph Mining) – створюється гра-

фічний образ аналізованих даних. Для вирішення задач візуа-лізації використовуються графічні методи, що показують наяв-ність закономірностей в даних.

Підбивання підсумків (Summarization) – опис конкретних груп об’єктів за допомогою аналізованого набору даних.

***Список використаних інформаційних джерел***

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 1. | [Електронний | ресурс]. | – | Режим | доступу: |
|  | http://pidruchniki.com/1623021247786/informatika/intelektualni\_ |
|  | tehnologiyi\_data\_mining\_text\_mining. |  |  |
| 2. | [Електронний | ресурс]. | – | Режим | доступу: |

http://stud.com.ua/62442/menedzhment/intelektualniy\_analiz\_ danih.