Recent Trends in Multi-Relational Data Mining

Prof. (Dr.) Vinay Kumar Singh

Director, Dr. Zakir Husain Institute, Muzaffarpur & Software Expert DDE, B.R.A. Bihar University, Muzaffarpur

ABSTRACT

Relational database theory has a long and rich history of ideas and developments concerning the efficient storage and processing of structured data, which should be exploited in successful data mining technology. Multi-Relational classification aims to build a classification model that utilizes information in different relations. These data mining technology enhanced the various fields of human life like business, education, agriculture, medical, scientific etc using Artificial Intelligence, Statistics, Pattern Recognition, Machine Learning and data visualization techniques.

Keywords: Data Mining, Multi-Relational Data Mining (MRDM), Knowledge Discovery in Databases (KDD), Business Intelligence (BI).

Introduction:

Data mining is the process of automatically discovering useful information in large data repositories. Data mining techniques are deployed to scour large databases to find novel and useful patterns that might otherwise remain unknown. They also provide capabilities to predict the outcome of a future observation. Data mining is an integral part of knowledge discovery in databases (KDD), which is the overall process of converting raw data into useful information. This process consists of a series of transformation steps, from data pre-processing to post processing of data mining results. Data mining refers to discover new patterns from a wealth of data in databases by focusing on the algorithms to extract useful knowledge.

Trends in Data Mining

Data mining is one of the most widely used methods to extract data from different sources and organize them for better usage. The development of efficient and effective data mining methods and systems, the construction of interactive and integrated data-mining environments, the design of data mining languages, and the application of data mining techniques to solve large application problems are important tasks for data mining researchers and data mining system and application developers. Some of the trends in data mining that reflect the pursuit of these challenges are:

Mining Complex Data Types

Scientific data sets are heterogeneous, typically involving semi-structured and unstructured data. Robust methods are needed for handling spatiotemporal data, related concept hierarchies, and complex geographic relationships.

Graph Mining

Graph Mining Graphs have become increasingly important in modelling complicated structures, such as circuits, images, chemical compounds, protein structures, biological networks, social networks, the Web, workflows, and XML documents. Many graph search algorithms have been developed in chemical informatics, computer vision, video indexing, and text retrieval. With the increasing demand for the analysis of large amounts of structured data, graph mining has become an active and important theme in data mining.

Data Mining For Intrusion Detection

An intrusion can be defined as any set of actions that threaten the integrity, confidentiality or availability of a network resource. An intrusion detection system for a largely complete network can typically generate thousands or millions of alarms per day representing an overwhelming task for security analysts. Anomaly detection builds models of normal network behaviour (called profiles), which it uses to detect a new pattern that significantly deviates from the profiles.

Data mining algorithms can be used for misuse detection and anomaly detection. Anomaly detection builds models of normal behaviour and automatically detects significant deviations from it. Supervised or unsupervised learning can be used. The techniques used must be efficient and scalable, and capable of handling network data of high volume, dimensionality, and heterogeneity.

Ubiquitous Data Mining

This type of data mining involves the mining of data from mobile devices to get information about individuals. Ubiquitous data mining having several challenges like complexity, privacy, cost, etc. This method has a lot of opportunities to be enormous in various industries especially in studying human-computer interactions.

Spatial And Geographic Data Mining

Spatial data mining technique includes extracting information from environmental, astronomical, and geographical data which also includes images taken from outer space. This type of data mining can reveal various aspects such as distance and topology which is mainly used in geographic information systems and other navigation applications.

Multimedia Data Mining

This is one of the latest methods which is catching up because of the growing ability to capture useful data accurately. It involves the extraction of data from different kinds of multimedia sources such as audio, text, hyper text, video, images, etc. and the data is converted into a numerical representation in different formats. This method can be used in clustering and classifications, performing similarity checks, and also identifying associations.

Distributed Data Mining

This type of data mining is gaining popularity as it involves the mining of a huge amount of information stored in different company locations or at different organizations. Highly sophisticated algorithms are used to extract data from different locations and provide proper insights and reports based upon them.

Time Series Data Mining

The primary application of this type of data mining is the study of cyclical and seasonal

trends. It also helps in identifying movements of components that exist within the data sets of stocks prices, currency exchange rates, customer's buying patterns and their behaviours, biomedical measurements, weather data etc. This practice is also helpful in analysing even random events which occur outside the normal series of events and is also used in trend analysis.

Business Data Mining

Nowadays business environment is more dynamic, so businesses must be able to react quicker, must be more profitable, and offer more high-quality services than ever before. Business data mining serves as a fundamental technology in enabling customer's transactions more accurately, faster and meaningfully. Data mining techniques of classification, regression, and cluster analysis are used in current business trends. Almost all of the current business data mining applications are based on the classification and prediction techniques for supporting business decisions, thus creating strong Business Intelligence (BI) system.

Although data mining is still in its infancy, companies in a wide range of industries - including retail, finance, health care, manufacturing transportation, and aerospace - are already using data mining tools and techniques to take advantage of historical data. By using pattern recognition technologies and statistical and mathematical techniques to sift through warehoused information, data mining helps analysts recognize significant facts, relationships, trends, patterns, exceptions and anomalies that might otherwise go unnoticed.

Conclusion:

I have described multi-relational data mining as the real approach of data mining paradigms related to the trends in the multi-relation data mining. Upcoming future trends in data mining will open wide opportunities for software companies to create such software with new methods and technique in data mining. I believe the future of data mining in finance, retail industry, telecommunications, etc will change data mining demand and technology. Consequently, I believe that relational data mining is one of the major research topics in the development of the next generation of data mining systems.

References:

1. Michael Steinbach, Pang-Ning Tan, and Vipin Kumar, Introduction to Data Mining, Pearson, 2005.

2. Foster Provost & Tom Fawcett, Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking, O'Reilly, 2013.

3. Bhanu Prakash Battula, B Kezia Rani, R Satya Prasad & T Sudha, Techniques in Computer Forensics: Recovery Perspective, International Journal of Security (IJS), Volume (3): Issue (2).

4. Guddattu, Vasudeva and Krishna (2011), Multilevel Analysis in Analyzing Speech Data, Indian Journal of Applied Linguistics, 37 (1). pp. 46-54.

5. S. Mitra Sitansu, Principles of Relational Database System, Prentice Hall, 1991.

✦ ✦ ✦