e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:03/March-2023 Impact Factor- 7.868 www.irjmets.com

HYPERTEXT AND HYPERMEDIA IN DATA MINING

Safa Ismail Nakade[*1]

[*1]Department Of Information Technology, Mumbai University, Mumbai, India.

DOI : https://www.doi.org/10.56726/IRJMETS34113

ABSTRACT

Since its development, knowledge has played an important role in human activities. Data mining is the process of knowledge discovery, which acquires knowledge by analyzing data stored in very large repositories, analyzes it from different perspectives, and summarizes the results into useful information. Due to the importance of extracting knowledge/information from huge data repositories, data mining has become a very important and guaranteed branch of engineering that directly or indirectly affects human life in various fields. The purpose of this article is to examine a number of future trends in the field of data mining, focusing on those considered to be the most promising and applicable to future data mining applications.

Keywords: Present And Future Of Data Mining, Data Mining, Data Mining Trends, Data Mining Applications.

I. INTRODUCTION

Data Mining (DM) originated from Database Knowledge Discovery (KDD), also known as Data Mining (DM), which extracts new knowledge from large databases. A number of techniques are currently being used in this burgeoning field, including statistical analysis and machine learning-based methods. With the rapid development of the World Wide Web and the rapid increase in unstructured databases, new technologies and applications are constantly emerging in its field. The main challenges in data mining are:

- Data mining that processes large amounts of data across different sites Data volumes can easily exceed the terabyte limit.

- Data mining is a very computationally intensive process involving very large sets of data. Often data must be partitioned and distributed for parallel processing to achieve acceptable time and space performance

- Input data changes rapidly. In many fields of application, the data to be exploited is either produced at a high rate, or actually provided in the form of a stream. In these cases, the knowledge must be used quickly and efficiently to make it useful and update the

II. SCOPE OF DATA MINING

Data mining derives its name from the similarity between searching for valuable business information in large databases (for example, searching for related products in gigabytes of scanner data from store) and finding valuable business information in a Similarities between exploring precious veins in the mountains.

Both processes require sifting through large volumes of material or intelligently examining them to find exactly where the value lies. With a database of sufficient size and quality, data mining techniques can generate new business opportunities by providing:

- **Automatic prediction of trends and behaviours**. Data mining automates the process of finding predictive information in large databases. Questions that previously required in-depth hands-on analysis can now be answered quickly and directly from the data. A classic example of a prediction problem is target marketing. Data mining uses data from past promotional mailings to identify the targets most likely to maximize the ROI of future mailings. Other forecasting issues include predicting bankruptcies and other forms of default, and identifying population groups that are likely to react similarly to a given event.
- **Automatically discover previously unknown patterns**. Data mining tools analyze databases and identify previously hidden patterns in a single step. An example of pattern discovery is analyzing retail data to identify seemingly unrelated products that are often purchased together. Other pattern detection issues include detecting fraudulent credit card transactions and identifying anomalous data that may represent data entry key errors.

The most common techniques used in data mining are:

- **Artificial neural networks:** nonlinear predictive models that are learned through training and are structurally similar to biological neural networks.
- **Decision tree:** tree structure representing a set of decisions. These decisions generate rules for classifying datasets. Specific decision tree methods include Classification and Regression Trees (CART) and Automated Chi-Square Interaction Detection (CHAID).
- **Genetic algorithms:** Optimization techniques that use processes such as genetic combining, mutation, and natural selection in design based on evolutionary concepts.
- **Nearest Neighbour:** A technique for classifying each record in a dataset based on the class combination of the k closest matching records in the historical dataset (where k ^{3} 1). Sometimes called the k-nearest neighbour technique.
- **Rule Induction:** Extract useful if-then rules from data based on statistical significance.

III. ROOTS OF DATA MINING

A. Statistics

The most important row is Statistics. There can be no data mining without statistics, since statistics are the basis of most data mining techniques. Statistics encompasses concepts such as regression analysis, standard distribution, standard deviation, standard deviation, discriminant analysis, cluster analysis, and confidence intervals, all of which are used to study the data and relationships between data. These are the building blocks that support more advanced statistical analysis. Of course, at the heart of today's data mining tools and techniques, classical statistical analysis plays an important role.

B. Artificial Intelligence and Machine Learning

The second longest family of data mining is artificial intelligence and machine learning. AI relies on heuristics rather than statistics and attempts to apply human thinking (e.g. processing) to statistical problems. Because this method requires massive computer processing power, it was impractical until the early 1980s when computers began to provide useful functionality at reasonable prices.

AI found some applications in the high-end science/government dark market, but the supercomputers needed in this era put AI out of reach for almost everyone. Machine learning can be considered an evolution of AI because it combines AI heuristics with advanced statistical methods. It allows computer programs to learn more about the data they are studying and then apply what they learn to the data.

C. Databases

Third category is that of databases A large amount of data needs to be stored in the repository and needs to be managed. Learn about databases. The first data was managed in records and fields, then in various models such as hierarchies, networks, etc.

The relational model has long served data storage needs. Other advanced systems that emerged were object relational databases. But in data mining, the amount of data is too large, so we need a dedicated server. We call this term a data warehouse. Data warehouses also support OLAP operations applied to them to support decision making.

IV. CURRENT TRENDS AND APPLICATIONS

Data mining is formally defined as the important process of identifying valid, new, potentially useful, and ultimately understandable patterns in data [2]. The field of data mining is rapidly evolving due to its wide applicability, scientific achievements, advancements and understanding. Many data mining applications have been successfully implemented in various fields such as fraud detection, retail, healthcare, finance, telecommunications, and risk analysis ETC. Too few to mention. Increasing complexity and technological advancements in various fields present new challenges for data mining; various challenges include different data formats, data from different locations, advancements in computing and network resources, research and scientific fields, increasing business challenges, etc. Advancements in data mining and various integrations and influences of methods and techniques have shaped current data mining applications to meet various challenges, current trends in data mining applications are:

A. Fight against terrorism

After the attacks of September 11, many countries adopted new laws to fight against terrorism. These laws allow intelligence agencies to effectively fight terrorist groups. The United States has launched the Total Information Awareness Program, which aims to create a large database that integrates all information on the population. Similar projects have also been launched in European countries and other parts of the world. There are several problems with this program,

a. Due to database heterogeneity, the target database must handle text, audio, image, and multimedia data.

b. The second problem is the scalability of the algorithm. Execution time increases with data size (which is large). For example, 230 cameras were placed in London to read vehicle license plate numbers.

With approximately 40,000 cars passing by the camera every hour, the camera must recognize 10 cars per second, placing a heavy load on both hardware and software.

B. Bioinformatics and Disease Therapy

The second most important application trend involves the development and interpretation of biological sequences and structures. Data mining tools are rapidly being used to find genes associated with the treatment of diseases such as cancer and AIDS.

C. Web and the Semantic Web

The Web is the hottest trend right now, but it's not structured. Data mining helps organize the web, known as the Semantic Web. The underlying technology is the Resource Description Framework (RDF) for describing resources. FOAF is also a tagging support technology widely used by Facebook and Orkut. But there are still problems like merging all RDF statements and handling wrong RDF statements. Data mining techniques have played an important role in the transformation of the Web into the Semantic Web.

D. Business Trends

Today's business environment is more dynamic, so businesses need to be able to respond faster, be more profitable, and deliver high-quality services than ever before. Here, data mining is used as the underlying technology to make customer transactions more accurate, fast and meaningful. Classification, regression and cluster analysis data mining techniques are used in current business trends. Almost all enterprise data mining applications today rely on classification and prediction techniques to support business decisions, creating powerful business intelligence (BI) systems.

Applications

As data mining matures, new and increasingly innovative applications continue to emerge. Although a wide variety of data mining scenarios can be described. For the purposes of this article, data mining applications fall into the following categories:

- Healthcare
- Finance
- Retail
- Telecommunications
- Text and Web Mining
- Higher Education

Healthcare

Genetic function to identify and study the human genome. Recent research on DNA analysis has uncovered the genetic causes of many illnesses and disabilities as well as methods for diagnosing, preventing and treating illnesses.

Finance

Most banks and financial institutions offer a wide range of banking (such as checking, savings, and business and personal transactions), credit (such as business, mortgage, and auto loans) and investment services (such as mutual funds). Some also offer insurance services and inventory services. Financial data collected by the banking and finance industry is often relatively comprehensive, reliable, and of high quality, which is convenient for systematic data analysis and data mining. It could also help detect fraud by tracking groups of people who cause accidents to collect insurance money.

The Retail Industry

The retail industry collects a large amount of data on sales, customer purchase history, merchandise shipped and consumed, and service records.

The amount of data collected continues to grow rapidly, particularly due to the increasing ease, availability and popularity of web-based business or e-commerce. The retail sector provides rich resources for data mining. Retail data mining can help identify customer behaviour, uncover customer buying habits and trends, improve the quality of customer service, improve customer retention and satisfaction, increase commodity consumption rates, design more efficient commodity transportation and distribution policies, and reduce abatement costs.

Telecommunications

Telecommunications has evolved rapidly from the provision of local and long distance telephone services to the provision of many other integrated communication services including voice, fax, pager, cellular telephone, imaging, email , the transmission of computer and network data and other types of data traffic. . The integration of telecommunications, computer networks, the Internet and many other methods of communication and computing is underway. Moreover, with the deregulation of the telecommunications industry in many countries and the development of new computer and communication technologies, the telecommunications market is booming and highly competitive. This has created a huge demand for data mining to help understand related businesses, identify telecommunication patterns, detect fraudulent activity, better utilize resources, and improve service quality.

Text Mining and Web Mining

Text mining is the process of finding certain keywords or key phrases in a large number of documents.

By performing a text search among thousands of documents, various relationships between documents can be established. However, using text mining, we can easily infer certain patterns in reviews that can help identify commonalities in customer perception not captured by other survey questions. An extension of text mining is web mining. Web data mining is an exciting new field that integrates data and text mining into websites.

Improve the website through intelligent operations, such as recommending relevant links to consumers or recommending new products.

Web mining is particularly exciting because it makes possible tasks that were previously difficult to perform. They can be configured to monitor and collect data from multiple locations and analyze data at one or multiple locations. For example, search engines work on the principle of datamining. Higher Education One of the challenges facing higher education today is predicting the development trajectories of students and graduates.

Which student will study a particular course program?

Who needs extra help to graduate? Meanwhile, other issues, such as enrolment management and graduation timing, are forcing colleges to find new, faster solutions. Institutions can better address these student and alumni concerns through data analysis and presentation. Data mining has quickly become a highly desirable tool for uncovering and understanding hidden patterns in large databases using today's reporting capabilities.

V. CONCLUSION

In conclusion, the future of data mining is bright, and the next few years will bring many new developments, methods and techniques, which is not optimistic.

Moreover, better technological integration and the application of data mining techniques can lead to new types of data processing and new applications. As the types of data and information we have access to have increased, so has the amount and type of data mining that can be performed. While some analysts and domain experts warn that data mining could follow the path of artificial intelligence (AI) and ultimately fail to achieve the commercial success previously predicted, the field of data mining is still young and its possibilities are still limitless. . Data mining has the potential to become one of the key technologies by expanding the applications it can use, integrating techniques and methods, broadening its applicability to general business applications, and making programs and interfaces more accessible to end users. Become the domain of the new millennium.

VI. REFERENCES

[1] Bedard, T. Merrett et J. Han, "Geospatial Data Warehouse Fundamentals for Geographic Knowledge Discovery," par H. Miller et J. Han (eds.), in Geographic Data Mining and Knowledge Discovery, Taylor and Francis, 2001.

[2] Chakrabarti, S., “Data Mining for Hypertext”, SIGKDD Explorations, 1 (2), January 2000. this Dom. "Discovering Distributed Full-Text Resources by Example", Proceedings of the 25th VLDB (International Conference on Very Large Databases), Edinburgh, Scotland, 1999.

[3] Cheung, C. Hwang, A. Fu et J. Han, "Efficient Rule-Based Induction for Feature-Oriented Data Mining", Journal of Intelligent Information Systems, 15(2): 175-200, 2000. Delmater, R.