# Application and practice of data mining techniques

**Yizhou Liu**

Electronic and Electrical Engineering, University of Birmingham, Birmingham, United Kingdom, B15 2TT


liuyz0410@163.com

**Abstract.** Data mining is a relatively recent phenomenon, but it has emerged as a result of the rapid advancement of information technology and the associated technologies and computer functions. In order to facilitate social production and life, advances in computer data mining, big data, and cloud storage technologies have been made. Data mining, or the process of systematically extracting useful information from large data sets, has many potential applications and can be applied in many different fields. Using a literature review method, this paper defines key terms associated with big data technology and demand, data mining technology, before outlining several widely used techniques for mining this type of information and expanding on the future directions of application in a number of different domains. Values hidden in data are becoming increasingly apparent in the modern world. The growing field of data mining offers significant help in finding a solution to this issue. Exploring the unique content and application practice of data mining technology in the era of big data may provide some reference for social development when taken together with the real situation.

**Keywords:** Data Mining Techniques, Big Data Era, Techniques, Applications.

## 1. Introduction

In the era of Big Data, a massive amount of data floods our computers, networks, and lives. Government organizations, academic institutes, and companies have invested substantial resources in gathering and preserving data. Only a small part of this data will ever be used, as in many cases the amount of information is very large, yet there is both value data and garbage data due to a lack of data integrity. Massive amounts of data have a special significance and crucial knowledge information is concealed behind the data, and are consequently characterized by incompleteness, unpredictability, and ambiguity. Almost all disciplines of business, research and engineering need to understand massive, complex, information-rich data collections. In the business sector, corporate and consumer data is being recognized as a strategic advantage. The ability to extract the usable knowledge contained in this data and act on that knowledge is becoming increasingly vital in today's highly competitive environment. The overall process of utilizing computer-based approaches, including new technologies, to discover knowledge from data is known as data mining.

This paper, through a method of literature evaluation, begins with an account of big data technologies and needs, and concepts. related to data mining techniques, then describes common approaches to computer data mining techniques in terms of decision trees, neural networks, correlation analysis, visualization techniques, cluster analysis, genetic algorithms, and feature analysis, and finally describes

directions for the application of computer data mining techniques in areas such as healthcare, finance, and computer network security.

Data mining is an iterative process in which progress is determined by discovery, either by automatic or manual techniques. Data mining is the search for novel, valuable, and non-trivial information in a big amount of data. It is a collaborative effort between humans and computers. Exploring the specifics and application practices of data mining technology in the era of big data can deepen people's understanding of the technology and enable this technology to be applied more rationally to create greater value for society.

## 2. Overview of data mining techniques

### 2.1. Big data technology and demand

The characteristics of big data are its size, velocity, complexity and diversity, value density and veracity. The ability of big data technology to absorb vast amounts of data and to specialize in processing those data in order to provide value-added, profitable data is a major benefit of this technology. As more and more data is created in this age of information, big data technologies will become more and more crucial. The user's perspective is essential when analyzing the big data analysis requirements.

### 2.2. Concepts related to data mining technology

Data mining technology is the application of various algorithms to massive swaths of data in an effort to unearth previously undiscovered patterns or insights. The real process of data mining is complex and varied, despite its common association with computer science. Statistics, online analytical combining, intelligence retrieval, pattern recognition, machine learning, and other techniques are used to accomplish this. The first step is to define the value of mining by analyzing the shape and characteristics of the data to be processed; the second is to define the relevant criteria and clean up the residual data in accordance with the data's characteristics and mining requirements; and the third is to obtain effective results through deep mining.

Prediction and description are the two most common uses of data mining. Prediction entails making use of particular data fields or variables to speculate on the likely or predicted values of other fields or variables. In contrast, the goal of description is to locate human-understandable patterns within the data [1].

The results of data mining can help professionals in a wide range of disciplines gain a better understanding of the market position because of the technology's ability to transform massive amounts of data into actionable insights and knowledge.

## 3. Common Data Mining Methods

### 3.1. Decision tree data mining methods

According to the decision tree statistical classification method, decision trees are a novel data mining technology that can categorize and predict massive volumes of data. Decision trees can be computed using a variety of approaches, such as SPRINT(Scalable Parallelizable Induction of decision Trees), SLIQ (Supervised Learning In Quest), etc. The method incorporates decision tree benefits such as quick and simple classification methods, the ability to manage massive volumes of data, and the capacity to mine and classify data directly using statistical methods, all while taking a more focused approach to data mining. Because of its reliance on computers, however, the field still has room to grow.

### 3.2. Neural network

Data mining issues are ideally suited to the simulation of human brain function that neural network analysis provides. Based on the research done on neural networks in computers, this method simulates human neural networks to model how the brain stores and processes data [2]. This allows for the analysis of complex data and the generation of new input, processing, and output modules. Utilizing computer

networks in data processing allows for the standardization of computational data information and adjustments to computational outputs [3]. Adaptive, error-tolerant, and supporting parallel processing processes, the approach is a major focus of data mining studies.

By examining the mapping relationship between data attributes and data values, computer data mining techniques operating under decision trees and neural networks can produce a unified output of data, a key strength of decision trees. Researchers may now employ decision trees and neural networks to do predictive analysis of data, making this method the de facto standard [4].

### 3.3. Association analysis method

Everything has been working together, and certain data sets have commonalities despite having various classification standards, etc. Using this property as a springboard, data mining can begin. Association analysis is a technique for data mining that exploits the connections between datasets in order to draw forth relevant insights. After data has been collected, this technique can be used to verify its authenticity and enhance its quality. The concept has practical applications in business, where it may be used to optimize products by learning more about how those items are used and by whom, thereby inspiring the company to work to enhance product features and efficiency.

### 3.4. Visualization techniques

Statistics may be used more effectively with data mining techniques, and it is possible to guarantee that the outcomes will be up to par. However, data statistics frequently include obtuse elements; these features can be shown more intuitively through the use of visualization techniques like graphs and scatter plots. By combining visualization and computer data mining techniques, we can better demonstrate the benefits of both methods and so increase their efficiency. The practical analysis shows that high-dimensional data is difficult to visualize and present in a chart, yet these charts can lead the development of visualization techniques and are technically creative.

### 3.5. Cluster analysis

Cluster analysis is a sort of data mining that uses a variety of criteria to break data into a number of related groups for more in-depth examination. An increased degree of standardization and hierarchy can be achieved in the data because to this method's ability to effectively organize disparate data. Useful knowledge can be unearthed in the midst of seemingly unrelated data kinds in this way. One clear drawback of this approach is that it cannot ensure the precision of the data classification [5]. Customer group mining is only one example of where this technology is finding useful, but it also has applications in areas like psychology, health, and data identification.

### 3.6. Genetic algorithms

Genetic algorithms are a highly integrated randomized search approach inspired by biological natural selection and genetic principles. Integrating data mutation processes and combining data are key to making computers capable of intelligent learning. As a part of the data processing procedure, irrelevant data is discarded while high-quality data is kept for use in generating novel data permutations. The optimization procedure for data can improve the information's flexibility. Genetic algorithm theory is a standardized processing method that can aid in the advancement of data mining technology. Properly positioned data mining technology, with decision data as the core technology, broadens the scope of technology development applications, necessitates a low knowledge base of decision-makers, and demonstrates the benefits of data mining and the creation of entirely new technological measures.

### 3.7. Feature analysis

Incorporating the data's specific content to filter out keywords, the feature analysis method enables data to be classified according to its characteristics; the keywords can then be used to refer to a set of data, and the keyword analysis results can be used to refer to the characteristics of a set of data. Effectively cutting down on data processing time and alleviating stress for workers is the goal of this strategy.

However, there are a number of drawbacks when using this approach, such as mistakes in the selection of keywords that do not represent the total data, which can reduce the accuracy of data mining [6]. Its primary application at the moment is in the building industry [6].

## 4. Specific applications of data mining

### 4.1. Applications in the medical field

Data mining has had a profound effect on the progression of scientific inquiry, clinical care, and the ability to detect and treat disease. For patients, the deep mining and analysis of medical big data can provide a new way for drug development and disease diagnosis, helping to solve medical problems and controlling medical costs; for medical researchers, the analysis of medical big data can provide a new way for drug development and disease diagnosis, benefiting the field as a whole [7].

There is a great deal of data generated in the medical field, and it is all interconnected. The application of association rule mining algorithms allows for the extraction of the related knowledge in the data, which in turn helps with the diagnosis and treatment of diseases, the summarization of factors that lead to diseases, public health, and health monitoring. Obesity, for instance, is found to have a strong correlation with CVD mortality when massive amounts of electronic medical records were mined for connection rules between established risk variables and death.

Data mining applied to healthcare's large data can be used to track outbreaks and prepare for them. With the use of integrated disease surveillance and response programs, reliable forecasting of transmission routes and times of transmission is possible thanks to data mining techniques applied to enormous volumes of data, thereby effectively reducing the probability of epidemic infection.

Since hospital data sources are not complex, the benefits of heterogeneous data processing in data mining technology have not been realized; and since the major hospital systems have not yet formed an interconnection, the amount of data in a single hospital has not yet reached the entry level of big data, there are still fundamental problems with data mining technology. But as our culture and technology advance, the volume of medical data will increase exponentially, and the benefits of data mining technology applied to medical big data will become more apparent, leading to larger economic and social benefits [8].

### 4.2. Applications in the financial sector

Risk management, business correlation analysis, customer relationship management, fighting and avoiding economic crime, boosting core competitiveness of firms, and building data warehouses are some of the most prominent uses of data mining in the financial industry. Data mining techniques, such as "feature selection" and "attribute correlation computation," can be used to sort through potential variables and zero in on the ones that matter. Companies can learn about market trends, shifts in customer demand, and other patterns lurking in their transaction data by using data mining techniques to examine massive amounts of this data and build applicable mathematical models. Data mining technology allows businesses to collect and analyze the data created routinely by their customers' interactions with financial institutions in order to better understand and predict their customers' actions and behaviors. Companies can utilize this data to better understand their consumers' needs and preferences, increase their product's market share, and boost their business's competitiveness. Market conditions, potential customers, supplier details, and competitors are just some of the external elements that data mining may help businesses understand and improve [9].

Since the financial sector relies on big sums of money, there are a plethora of financial crimes such as fraud, money laundering, and currency counterfeiting. Data mining techniques that seek for anomalies, including big and frequent changes in cash flow between a small number of persons in a short period of time, and anticipate their risk level can offer decision-makers with useful information for avoiding and preventing dangers.

*4.3. Application in the field of computer network security*

Dealing with computer network viruses is the primary responsibility of those responsible for maintaining network security. Data mining technology can quickly identify computer virus data information, crawl data information, examine data information, and analyze abnormal problems in a computer network system to determine if a virus is present, and then implement the necessary defensive and preventative measures against the virus to ensure the data information is secure. Adopting such encryption technology enables effective enforcement of defenses to be carried out, and the security settings collecting technology that is a part of data mining allows for the development of unique rules and frequent content columns in each operation.

Anderson proposed a data mining–based security maintenance solution for networks in the 1980s. Real-time monitoring and control of computer virus infection, warning of security-related concerns, and subsequent risk mitigation are the primary goals of such a system [10].

## 5. Conclusion

The advancement of data mining technology is crucial because it helps people in many ways, including by providing for their material and spiritual needs and by fostering social progress and development. The effective application of data mining technology can have a significant impact on the growth of a number of different industries; as such, those fields should devote a great deal of attention to it, integrating it with their own realities in order to rationally apply a variety of data mining technologies in order to efficiently acquire the information they require in order for their development to obtain the requisite data support and ultimately accomplish the goal of advancing scientific knowledge. The target is a rise in rational and efficient decision-making.

People's privacy and security cannot be guaranteed under computer data mining technology, so we need to actively study the connotation and application methods of data mining techniques so that this double-edged sword can be used fairly to provide a strong boost to social progress.

## References

[1]    Mehmed Kantardzic. Data Mining: Concepts, Models, Methods, and Algorithms. Hoboken, N.J.: John Wiley, 2011:2.
[2]    Pingrong Li. Data mining techniques and applications in the era of big data[J]. Journal of Chongqing Three-Gorges College, 2018, 25(3):45-47.
[3]    Abdu Aini Abdukhuvsuli. Research on the development of computer data mining technology and its application. Computer Knowledge and Technology. 2022,18(03):28.
[4]    Weicheng Meng. Analysis of the development of computer data mining technology and its application. Electronic World. 2022, (01):77.
[5]    Ronglong Gu, Wenjie Zhao, Lei Wang. Application of data mining techniques in the era of big data. Technology Innovation and Application. 2022,12(05):177
[6]    Peng Zhang. Data mining techniques and practices in the era of big data. China New Communication. 2021,23 (22):69
[7]    Mintao Chen, Fangjun Kuang. Research on the application of data mining technology in medical big data [J]. Computers and Telecommunications, 2017(11):34-36.
[8]    Xiujuan Zheng. Research on the application of data mining techniques in medical big data. Computer Knowledge and Technology, 2020:35.
[9]    Tingting Yang. Data mining in the financial industry. Journal of Economic Research. 2021, (15):58-59.
[10]   Yong Lin. Application of Data Mining Technology in Computer Network Security Maintenance. Changjiang Infocomm. 2021,34(10):144