An Analysis of Data Mining and Application in the Era of Big Data

Yue Zhang, Li Du

Chengdu Neusoft University, Dujiangyan Sichuan, China, 611844

Keywords: era of big data; data mining; application

Abstract: With the development of computer technology and Internet technology, human society has stepped into the era of big data. The purpose of using data mining technology in the era of big data is to mine valuable information in a large amount of information, integrate information together, and then deal with this information, so as to effectively enhance the value of the data. This paper describes the meaning and characteristics of big data and data mining technology, analyzes the data mining technology methods, and studies the application of data mining in the era of big data.

1. Introduction

With the development of computer technology and information technology, there will be a large amount of data information every day, and such effective collection, storage, analysis and application of these data are issues that we should study in depth. The characteristics of big data include large number, diversity, high speed, and value. In 2011, many countries around the world believed that the world has entered the era of big data. In order to make effective use of data information, it is necessary to make full use of data mining technology, use this technology to find valuable data information, and organize and use this information so as to promote the development of various industries. Therefore, the use of data mining technology in the era of big data has important implications for economic development.

2. Overview of Big Data and Data Mining Technology

2.1 The meaning of big data

The meaning of big data. Unlike other definitions, big data is not a specific concept, different fields have different understandings of big data. Wikipedia considers that big data refers to data integration that cannot be handled using database management tools [1], this data information are characterized by massiveness and complexity. Usually, big data refers to a collection of data that cannot be collected, managed, and processed in a short time using traditional IT technologies and software and hardware tools.

2.2 The characteristics of big data

In general, the characteristics of big data can be summarized as follows:

First, a large number: The volume of data is relatively large, and the data information that needs to be processed is increased from TB to PB-level data. Second, diversity: There are many kinds of data, mainly including semi-structured and unstructured data, such as video, web pages, texts, and traditional structured data. Again, high speed: The processing speed is faster, and big data can be analyzed in real time, this is also the difference between big data and traditional data mining technologies. Finally, the value: the low value of the density, the scientific use of low-density value data information, and accurate analysis of these data information [2], will provide users with greater commercial and social value.

3. The Meaning and Characteristics of Data Mining

3.1 The meaning of data mining

Data mining is the method and technology of excavating rules and extracting knowledge from a large amount of data information. Data mining has an important connection with the database, so it can be called database knowledge discovery [3]. In simple terms, data mining is the process of finding information and knowledge that already exists and people do not understand but have some value in a large amount of data information.

3.2 The characteristics of data mining

First, when dealing with information in various industry databases, data mining techniques can be used, the information in the database is preprocessed and structured data information.

Second, use knowledge in mathematical statistics to build mining models, develop mining algorithms, find valuable data information, and tap the relationships between them.

Third, the scientific construction mining model has important implications for displaying user behavior.

4. Research on Data Mining Technology Methods

At this stage, many enterprises in our country face the problem of "the massive amount of data and the lack of information" after they have centralized data information. Therefore, data mining technology has emerged at the historic moment. Using data mining technology can effectively solve this problem. Data mining is not a traditional technology, but a new discipline that has been developed for decades. The use of data mining technology is to use artificial intelligence technology. In simple terms, data mining technology needs to be based on the development of technology to achieve data innovation. Therefore, data mining technology has application value, using this technology can enable managers to accurately and quickly find effective information and make correct judgments, it effectively improves work efficiency and can effectively enhance the competitiveness of enterprises. The focus of data mining technology is to analyze data information, solve various problems through different analysis methods, and achieve the purpose of mining data values.

4.1 Cluster analysis

This analysis method is used to classify these data information into several categories. After that, each plate is formed so that the data information not related to each other becomes related data information, and then valuable data information is obtained from these data information and used [4]. Using this method can not be classified according to the type of data, attributes, so this analysis method is more suitable for psychology, statistics and other aspects.

4.2 Artificial neural networks

This method uses a large amount of data information for analysis, using an artificial neural network data analysis method requires building a data model first. Therefore, using this analysis method can be classified according to the requirements of the data, so the application of this analysis method in life and work is relatively extensive.

4.3 Correlation analysis

Some data information has the hidden characteristics. Therefore, using ordinary data analysis methods can not fully tap and use these data information. At this time, it is necessary to use correlation analysis methods to study whether there is correlation between data and information, thereby reducing manual work pressure. Usually, correlation analysis methods are used to analyze data information when there is a certain purpose. Therefore, this method is suitable for work that requires high accuracy of data information.

4.4 Characteristic data analysis

With the human society stepping into the information age, the number of network data has continuously increased, and the source of data information is relatively wide. How to classify data based on the characteristics of data information is an important part of data management. The artificial neural network data analysis technology in the above is also one of them. In addition, some data analysis methods use the computer to classify the data[5], seek out the regularity between data and information to analyze the characteristics of the data, then can classify these data information.

5. Data Mining Process

5.1 Preparatory works

Before data mining technology is used to mine data information, data information needs to be prepared. Only when data information is prepared, can the data mining work be performed smoothly and improve the quality and speed of data mining. How to prepare data information? This requires that the original data be processed in advance. The main reason is that the original data information is not suitable for using data mining technology, and it needs to go through the process of data selection, data cleaning, etc. And converts the original data into data information that can use data mining technology.

5.2 Data mining

Data mining can extract some valuable data information from data information. Therefore, when mining data, it is necessary to first clarify the goals of data mining [6], and select appropriate analysis methods based on data information, use scientific analysis methods to analyze data information, and find the connection between data & information to obtain valuable data information.

5.3 Model evaluation and analysis

In simple terms, this section needs to explain and analyze the structure of mining data, select the law of significance or value, and explain the analysis results in a way that people can understand.

6. The Application of Data Mining in the Era of Big Data

6.1 The functional scope of data information resources after data mining

At this stage, the data mining technology with informatization characteristics is widely used in practical work. Through researching the aspects of social economy, culture, etc., scientifically arranging production activities of enterprises, explaining the changes in the nature and value of substances in the production activities of enterprises, and searching for the characteristics of changes in data related to substances, so that enterprises can rely on these data information[7], scientific adjustment of the enterprises' production and operation goals, direction, and enhance economic benefits. The use of data mining technology to analyze data information can expand the scope of functions of data information, provide a basis for the development of enterprises, and have a certain role in improving the competitiveness of enterprises.

6.2 The use of data mining technology

6.2.1 Data mining technology in the Internet

The Internet contains a lot of text, picture information, voice information and other information, but also contains a lot of link structure information and usage information. By mining these information, structures, records, etc., valuable data information can be obtained, so that users can query these data information, explore authoritative pages and optimize the site, thereby effectively improving the user's access speed and aggregating multiple similar users. The use of data mining technology in the Internet is mainly demonstrated in the following points:

First, web search: Using data mining technology to process the data information in the Internet can extract the data information needed by the user in a large amount of network data information by refining the characteristics of the target data, word segmentation processing, clustering, etc.

Second, network intrusion detection system: When checking whether an unauthorized user uses a computer or an authorized user of a computer system for illegal access, this function can be implemented by analyzing the time series mode mining method to analyze network data packets and system logs.

6.2.2 The application of data mining technology in marketing

Compared with other industries, the application of data mining technology in marketing is relatively early. The main reason is that the use of data mining technology in this industry can fully tap the user's data information, and then obtain some valuable data information, for example: using data mining technology to analyze the consumer's consumer behavior, which has an important role in the development of new products, provide users with effective services, improve the competitiveness of enterprises.

6.2.3 The application of data mining technology in scientific research

With the development of technology, data mining technology has been widely applied in all walks of life, and has also been applied in scientific research. The main role of data mining technology in scientific research is to assist research, use data research techniques to explore the laws and knowledge between data and information, and lay the foundation for subsequent research work, such as: using data mining techniques to analyze DNA data. Therefore, data mining technology has great application value in scientific research.

6.2.4 The application of data mining technology in manufacturing industry

With the development of economy, people's living standards have been continuously improved, and the requirements for product quality are also increasing. Therefore, the manufacturing industry needs to improve its own production technology and improve product quality. The manufacturing industry can use data mining technology to research the relevant data information of the product, extract related data with insufficient products, and relevant data information of production efficiency, and provide data support for scientific research on the problems of products and the reasons for low production efficiency. Manufacturing companies can propose effective measures to solve these problems, improve product quality while improving the company's production efficiency, and then increase the economic efficiency of enterprises, and promote the stable and healthy development of the manufacturing industry. This shows that the use of data mining technology in manufacturing has an important role.

6.2.5 Data mining technology in the telecommunications industry

With the development of the telecommunications industry, the number of users is constantly increasing, and telecommunications companies need to provide high-quality technical services for these customers. However, as telecommunication technologies and services are a more complex mixed carrier, it has seriously affected the service quality of telecommunication enterprises' technical services. The use of data mining technology in the telecommunications industry can effectively solve this problem. The main reason is that this technology can scientifically analyze telecommunication data information, find the rules in a large amount of data information, providing data support for technical services provided by telecommunication companies, through the analysis of these data, to continuously improve telecommunications technology services, provide users with excellent services, thereby promoting the stable development of telecommunications companies.

7. Conclusion

In a word, with the development of science and technology, the daily generated data information is also increasing. Using data mining technology can quickly and accurately find valuable data.

Therefore, using data mining technology in the era of big data has important significance for social development. How to improve the quality of data mining technology is an important direction to promote the development of data mining technology. I believe that in the coming period, data mining technology will be more complete, and processing speed will also be improved.

References

[1] Xunwei Xia. The application of data mining technology in power in the era of big data[J]. Communication World, 2016(7):115-115

[2] Xufeng Hu. On the Effect of Data Mining on Managerial Accounting Functions in the Era of Big Data[J]. Financial Industry (Academic Edition), 2016(4):188-188.

[3] Ping Duan. The Application of Data Mining in Banks in the Era of Big Data[J]. Science and Technology Economic Guide, 2017(6):31.

[4] Ang Li. The Application of Data Mining in Banking CRM System in Big Data Era[D]. Zhengzhou University,2016.

[5] Ruijuan He. The Application of Data Mining Technology in Enterprises in the Era of Big Data[J]. Network Security Technology and Application, 2016(12):90-91.

[6] Jiayu Yang. Research on Application of Data Mining Technology in Big Data Era[J]. Engineering Technology: Full-text version, 2016(8):00242-00242.

[7] Wen Yu. Data Mining Technology and Enterprise Informationization in the Background of Big Data Era[J]. Communication World, 2016(2):62-62.