Research on the Application Method of Mathematics in Big Data Processing

Linxi Jian[1], Yunxiang Cai[2], Yuxuan Zhang[3]

[1]College of Computer Science and Technology, Jilin University, Changchun, Jilin, 130012, China

[2]College of Science, Xi'an University of Engineering, Xi'an, Shanxi, 710048, China

[3]College of International Business, Dalian University for Nationalities, Dalian, Liaoning, 116600, China

Keywords: Big Data, Data Mining, Applied Mathematics, Application

Abstract: In recent years, with the rapid development of networks and computers, humans have entered the era of big data. In the current era of information explosion, how to use massive amounts of data is a major problem facing humanity. The use of applied mathematics knowledge in big data is of great significance for the management of big data. Therefore, it is necessary to explore the application of mathematics related knowledge in big data. Big data is an inevitable outcome of the era and the development of science and technology. Applied mathematics is mainly the ability to apply mathematical knowledge. Applying the relevant knowledge of applied mathematics to the processing of big data can promote the overall efficiency of knowledge application and data processing. Based on the overview of big data, this paper analyzes the core technologies and implementation methods of big data, and then explores the specific application direction and prospects of applied mathematics in big data.

1. Introduction

As the world enters the information age, data has penetrated into every corner of people's lives and work, and has gradually become one of the important factors in social production. People have also begun to use big data to change traditional production patterns and consumption patterns, thereby promoting humanity. Society has entered the era of big data. In the era of big data, data mining has become the core of the times. How to dig deep into useful information in big data has become the focus of attention in various industries. Mathematics is an important means of implementing data analysis. When mining big data, it is necessary to apply various mathematical theories and knowledge.

In the age of information, big data has become a familiar term. Big data refers to information assets that cannot be captured, managed, or processed by conventional software tools for a certain period of time, and that require a new model to process. The data has 5V characteristics, namely Volume (large amount), large data capacity; Velocity (high speed), fast data acquisition and processing speed; Variety (diverse), diverse data types; Value (low value density), low data value; Veracity (Authenticity), data quality is high and true. Relevant statistics show that by 2018, the global data volume will increase by 1.8ZB, and the global per capita data resources will exceed 200GB. In recent years, with the continuous improvement of information technology, the widespread application of information technology and information resources, global annual data growth. The rate is as high as 50%, and the world has entered the era of big data.

Data mining is the key content of big data, and it is an emerging discipline that is gradually formed along with the emergence of big data. For data mining, the definition of the method differs depending on the focus. The earliest definition of data mining refers to the comprehensive extraction of potentially useful information in the data. At present, data mining has become a professional activity, not only a statistical model technology, but also a deeper knowledge discovery. Data mining has been widely used in various fields and industries, including education, scientific research, marketing, manufacturing, telecommunications, Internet industry, etc., especially in the field of commercial artificial intelligence research, by extracting potentially useful information in

big data. Or knowledge to provide an objective data basis for business decision-making and construction.

2. Big Data Overview

2.1 Specific Concepts

With the rise of the network and the popularization of the use of computer technology, it has greatly promoted various industries in the society, and also produced a lot of data in various industries. For example, customer experience data, sensory data, healthcare, surveillance technology, optical observations, and financial transactions. Due to the continuous generation of data in many fields, data is flooding. In this environment, people gradually began to pay attention to the concept of big data. With the big data entering the people's field of vision, its characteristics are gradually being understood. In the overall cognition of big data, it has no fixed structure, and needs to analyze the data in real time. These features are gradually accepted by people. Therefore, in the overall development of big data, it is imperative to establish a new system framework, so that big data can be the acquisition, storage, and transmission are satisfied.

2.2 Challenges Brought by Big Data

At this stage, the arrival of the era of big data has brought enormous development space and development opportunities to various fields. At the same time, it has also brought enormous challenges to the work of traditional type data management. First, in the era of big data, data generation and consumption in each line has become easier and faster, posing great challenges for data collection and integration management. Second, in the era of big data, the most obvious feature of data is "big", which brings great challenges to data management and storage, and software and hardware development work becomes more urgent. Third, due to the characteristics of big data privacy, complexity, real-time, scale and heterogeneity, it poses great challenges for data analysis, query and prediction.

2.3 Limitations of the Actual Environment

In the past, data management and analysis used a relational database aspect management system as the basis, and this management system has outstanding advantages in processing structured data. However, it is difficult to function during the processing of semi-structured and unstructured data. Not only is it lacking in dealing with new data structures, but it is also lacking in hardware storage. Although the actual demand can be satisfied by purchasing and increasing the storage capacity, the cost will also increase, and the overall production cost of the hardware will be increased. Big data processing and storage expand the hardware to improve storage capacity and process efficiency. These are things that rdbms can't achieve in the past, which limits the overall development of big data.

3. Big Data Core Technology and Method

At present, under the specific bearing of computer-related technologies, the processing basis of big data is gradually improved and updated based on the traditional computer technology, which is changing the world little by little. Moreover, the processing technology of big data can process the actual data requirements of users through visual, efficient and real-time processing of massive data, and distribute the data on demand. With cloud technology, big data collection, processing, transmission, and applications can be realized. In production and life, you can handle large amounts of unstructured, semi-structured, and structured data with non-traditional tools.

First, the work of big data collection and preprocessing. An important feature of big data is that data sources are diverse, and the diversity of data has important performance in addition to the formation of data diversification, and also has important manifestations in the content and structure

of data diversification. Big Data The first step in this technology is to collect and pre-process data to provide high-quality data for the smooth implementation of subsequent processes.

Second, the processing of big data is a difficult task, in which data accumulation is an important aspect, and the effective accumulation of data requires the storage and management of data. Therefore, during the application of big data technology, it is necessary to overcome the difficulties of large storage scale, complicated management and high service level requirements.

Third, big data security and hidden dangers. Today, in the era of data explosion, this poses a serious threat to people's privacy. During the management of big data, it mainly faces the following problems: First, under the era of big data, information security is more complicated. Second, the security issues during the use of big data. Third, the analysis of higher companies and groups with big data presents significant security challenges. Fourth, the sharing of big data. Fifth, the data has dynamics. Sixth, there is no guarantee of relevant laws and regulations. Seventh, the convergence challenge brought by multiple data. Eighth, privacy data based on location is extremely easy to expose.

4. The Use of Mathematics in Data Mining

Applied mathematics is based on the theory of disciplines and basic methods, and gradually develops the ability to solve practical problems with the knowledge acquired and computer-related technologies, thus providing teaching, research and research for economic, educational and social science and technology. Production and management of these high-end talents for practical application, management and development research. Big data refers to a collection of data that is difficult to capture, process, and manage through conventional tools within a tolerable range and time, and is specialized in computer-related disciplines. During the processing of big data, more knowledge of applied mathematics can be incorporated, and the utility data can be utilized in the actual processing of big data.

4.1 Data Processing

To implement big data mining, the data must first be collected and initially processed. Big data is time-sensitive. When processing data, the time requirements are very strict. Therefore, in an effective time, the data must maintain high efficiency while ensuring good processing results. If the original data is noisy or incomplete, not uniform, etc., the data should be pre-processed to improve the accuracy of the data; if the data volume or indicator set is too large, you can select some important and research results. Relevant data, or some key indicators that fully reflect the research results.

In the process of data processing, many commonly used or classical analysis methods are usually used in mathematics. The most commonly used methods are statistical methods such as descriptive analysis, regression analysis, and correlation analysis. The regression analysis method is usually combined with the correlation analysis method. Based on the correlation analysis, by observing the quantitative relationship between two or more related variables, the corresponding mathematical model is established to utilize the known amount. Infer the unknown. The main purpose of regression analysis is to use the sample data to estimate the parameters, and then to build a mathematical model to test, judge or predict parameters or models. In addition, many mathematical theories are used in data processing. For example, the measure theory combines two or more monotonic measures by operation to construct a new monotonic measure. When implementing data dimensionality reduction processing, the measurement theory can be applied. Compared with the traditional principal factor analysis method, this analysis method can effectively ensure the integrity of the data, thus retaining more relevant information and improving the scientific results of the analysis results.

4.2 Data Mining

Data mining is the core of big data. Big data seems to be complex and irregular. To obtain effective or valuable information from huge amounts of data, data mining must be implemented. Data mining is applied, engineered, aggregated, and cross-cutting. In the process of implementing data mining, mathematics plays an irreplaceable role. In data mining, commonly used mathematical methods include neural network, association analysis, cluster analysis and decision tree. The most important one is cluster analysis, which is also widely used in other industries and fields, including Psychology, medicine, statistics, marketing industry, data identification, etc.

The cluster analysis method is based on the "physical class", according to certain criteria, the objects with greater correlation are classified into one class, and the differences between different types of objects are maximized, so that the data set or key indicators are carried out. Scientific grouping. In the process of implementing data mining by cluster analysis, the mathematical methods applied mainly include grey relational analysis, objective function ambiguity and interval value algorithm. Firstly, using gray correlation analysis, the correlation between data is judged by comparing the geometric shapes between geometric curves. The closer the geometric shapes are, the larger the correlation is, and the smaller is vice versa. Gray correlation analysis implementation data mining is mostly used in data analysis where the sample data volume is small, or the sample is incomplete, such as data analysis that is missing due to historical reasons, or the sample data is not uniform due to the rapid update rate of the sample. Data analysis of the situation.

Secondly, using the objective function ambiguity, standardizing the data and then calibrating, and establishing a fuzzy matrix, then implementing direct clustering or fuzzy equivalence matrix to implement clustering of data sets or key indicators, and also using the maximum tree method Or network method for clustering. The objective function fuzzy has the characteristics of high efficiency, large scalability and high processing dimension, which is the key method used in the data mining process. In fact, the cluster analysis method with fuzzy objective function is also very common in people's daily life. Various data analysis, data mining, and even image processing are applied to this method to cluster data sets or key indicators. The objective function fuzzy method is relatively scientific, and its application is relatively mature, which is an effective method to solve the clustering problem.

Finally, using the interval value algorithm, some "comparative" data that can be transformed in the data mining process, or data with a fixed range of values can be analyzed. Interval value algorithm is a commonly used mathematical method. Its application in data mining is mostly reflected in the mining and analysis of incomplete system information. In the implementation of data mining, there are three types of interval value algorithms: number and interval clustering, interval and interval clustering, and matrix and interval clustering. The number and interval clustering methods are the most commonly used. Efficient, accurate, and true statistical analysis of incomplete system information. When determining the interval value, it can be determined by experienced experts or statistical methods.

5. Conclusion

Big data is an inevitable result of historical development, and it is also an inevitable trend of people's overall demand for data information in the new era. While big data meets the needs of people's digital information, it also proposes various technologies. The corresponding challenge. Up to now, big data has not only referred to pure data, it is a technology, but also an application. In order to truly do the application of big data, we should first firmly grasp the theoretical basis of its application, start from the early collection of data, conduct in-depth analysis, and form a conclusion. In the whole process, big data, whether it is processing or mining, needs the support of mathematical theory, including its later application, and also needs to be applied to mathematics knowledge. Therefore, in the process of implementing big data mining, we should pay full attention to the role of mathematics and apply mathematical knowledge reasonably. Applied mathematics

plays an important role in the management of big data, and its application of a broad application space should be valued by people.

References

[1] Qiao Xiaofei, Ma Kaili. Analysis of the relationship between mathematical theory and empirical phenomena from the perspective of big data [J]. Journal of Systems Science, 2019 (01): 61 - 64+70

[2] Lv Xiaoguang, Jiang Le, Cheng Qingsong. Analysis of the training mode of data science and big data technology professionals [J]. Journal of Huaihai Institute of Technology (Humanities and Social Sciences Edition), 2018, 16 (09): 132 – 135.

[3] Wu Chusheng. On the use of mathematics in big data mining [J]. Enterprise Technology and Development, 2018 (08): 140 – 141.

[4] Kong Linghai. Research on Talent Classification Training Strategy Based on Employment Data Analysis——Taking Mathematics and Applied Mathematics of Jilin Engineering and Technology Teachers College as an Example [J]. Journal of Jilin Engineering and Technology Teachers College, 2018, 34 (02): 63 - 65.

[5] Li Jinxi. Research on the development prospect of mathematics + big data [J]. China's strategic emerging industry, 2018 (20): 7 - 8.

[6] Huai Jinpeng. Data that can create value is “big data” [J]. Software and Integrated Circuits, 2016 (08): 12.