Research on Data Mining Security under the Background of Big Data Era

Kaiping Zhang

(College of Information Engineering, Zhengzhou University of Industrial Technology, Xinzheng 451150, China)

595163495@qq.com

Abstract. Since people entered the information age, the volume of data has been increasing exponentially. Especially after 2012, big data has become a hot word. With the rapid development of high technology such as Internet, cloud computing and artificial intelligence, the amount of data people need to process has also become an astronomical figure. How to effectively deal with these data and obtain valuable information has become the key point in the fields of scientific research, enterprise, economic market and other fields, and the key is data mining. This paper discusses the definition of data mining, the related technology application of data mining and the subsequent privacy and confidentiality.

The development of the Internet, the popularization of computer communication technology and the popularization of Internet shopping have made the traditional real economy more and more "moving" to the network, which makes the mass data such as personal information, communication information and consumption information become the data resources that can be traced, These data resources have evolved into a very important research topic in information science. However, while we enjoy the convenience brought by the information technology, there are also some problems, such as the speed of information processing and the speed of data growth; the extraction technology of information needs to be improved. The authenticity of information is difficult to confirm; The confidentiality system and technology of personal privacy need to be improved urgently. Therefore, how to obtain valuable information quickly and efficiently in big data. Further using the laws reflected by these information to guide the market economy, the social trend, and the discovery of a series of hidden knowledge from the mass data, it has promoted the rapid development of data mining technology.

What is Data Mining

The essence of data mining is to discover hidden knowledge contents in databases, which is a major research topic in the field of artificial intelligence and data processing in today's society. It is called "mining" because the knowledge sought by this technology is previously uncertain, potentially valuable and hidden. Data mining is a decision support process. It is based on large data, statistics, recognition technology, automatic learning, AI intelligence and so on. It automatically analyzes data from all walks of life, and summarizes a certain relationship network to find hidden patterns or potential to help the decision-makers to guide the development of the market, reduce the risk and make the correct prediction. From a technical point of view, data mining needs to extract knowledge and information hidden in a huge, incomplete, defective, completely recognized, and random application data that is hidden, but not known before, but contains potential value. Among them, the source of the data must be true, the amount of data is large, the information extracted must be the interest of the target user group, and the actual operation should be [1].

For raw data, it can be from all walks of life databases, such as structured data in relational databases. Some semi-structured data such as text and video data can also be scattered and unrelated data scattered over the network. The ways to extract knowledge include deduction, induction, mathematical models, and non math ematical models. The extracted knowledge can be used in information retrieval, quick query, process control, and so on, and can also be used to maintain data itself. Therefore, data mining is a hybrid technology, which can make full use of data from simple query to complex deduction, from simple surface knowledge to deep hidden knowledge, which are

the key technical knowledge of decision-making and guidance in today's society.

Common Techniques for Data Mining

(1)Classification. Classification is to find a group of data objects with common characteristics in the data source, and divide them into different classes according to the predetermined pattern, so the purpose of extracting data is to connect the data objects in the data source to a particular category by classification. For example, it can be applied to customer classification, analysis of customer attributes and characteristics, satisfaction survey, purchasing power and purchase intention prediction. If a car manufacturer divides the customers in the data source according to the hobby of the car, the frontline salesman can take the initiative to sell the corresponding products directly to such customers and improve the efficiency greatly.

(2) Clustering. The analysis method is to divide a group of data sources into several levels according to whether they are similar and similar, and to make a wide and detailed similarity judgment system, which makes the data sources of the same category as similar as possible, and the data sources of different categories are as small as possible. Cluster analysis is suitable for many aspects of market sales, such as finding customers, exploring potential customers, maintaining contact with customers, analysis of commodity cycle, selling pre judgment and so on.

(3) Association. Association is a rule that describes the relationship between data in a data source. In a simple way, it is the occurrence of some items in a class of data sources, which may lead to a corresponding reflection in another type of data source, resulting in a result, and the knowledge that is hidden in the data source is also found to be [2]. For example, when maintaining a customer relationship, it can be excavated according to the customer data source of the enterprise, find valuable information from a large number of relations exchange records, and find the potential key nodes that affect the market. This has important reference to the price prejudgement, the trend of sales, the resolution of the customer group, the planning of the overall direction of sales, and so on. Test value.

(4)Regression Analysis. This method uses the characteristics of data sources in time, and maps a data category to a predictive variable function, and explores the relationship between variables from mathematical theory. The main points of the study are the trend of data variables, the trend of pre judgment, the relationship between data categories and so on.

(5) Characteristics. Data sources can also classify data according to their characteristics, and extract features from these features. Such formulas can present the overall characteristics of the data source. For example, the first line salesman can quickly analyze a series of factors, main features and secondary features of customer loss by extracting the data characteristic of customer loss, which can prevent the occurrence of customer loss in advance.

(6) Analysis of Variation and Deviation. The deviation in the data source contains a lot of hidden knowledge, such as abnormal phenomena in the data classification, the abnormal phenomenon in the pattern, the actual and the expected deviations, and the analysis of the deviation is intended to explore the value difference between the actual results and the reference. In the crisis early-warning system of large enterprises, the managers should pay more attention to abnormal phenomena and rules. The analysis and deep mining of these deviations can be used to sniff the crisis in advance and make the enterprises respond in advance to avoid the bad consequences.

(7) Web Page Mining. With the popularity of the Internet, more and more people begin to interact through network communication, which leads to extremely rich data on the Internet. Through data mining on the Internet, we can collect various information related to economy, finance, rival enterprises, supply and demand of human resources, potential customer groups and so on. It concentrates on the internal and external factors which have great influence or potential impact on the enterprise. According to the results of data mining, it can prejudge and warn the crisis and make rational use of it, so as to achieve the purpose of overall operation and appropriate decision.

Application of Data Mining in Marketing.

Data mining technology is widely used in the market sales of large enterprises. Its application is based on the principle of market segmentation. Through the analysis of customer's daily consumption behavior, it can judge its future consumption behavior [3]. Specifically, it is through the extensive collection, processing and processing of all the information about the daily behavior of the consumer, categorized, clear the bias, consumption level and consumption trend of each class of consumer groups, and further determine the future consumption behavior of this type of consumer group, and then take this as the criterion for each class of consumer groups. Directional marketing is similar to "drip irrigation" instead of "drip irrigation", which greatly improves the accuracy of sales, improves the efficiency of sales, and quickly accumulates customers and brings profits for the enterprises.

The consumption data information comes from a variety of channels. Our private information automatically enters the corresponding database when we apply for the debit card, the purchase of goods for registration, and the filling out of the items. Each time we use a credit card, the enterprise can collect our consumption information in the credit card bill calculation stage and record the corresponding time, location, commodity content, and consumption ability. In addition to the enterprise itself collection of relevant information, it can also buy from his business company in the form of purchase[4].

The user information from a variety of channels is reorganized, using a variety of information processing methods such as computer, AI intelligence analysis, model algorithm and other information processing methods to extract the favorable decision and judgment information for the enterprise, which is used for the directional sales of the consumer groups. For example, when a bank system makes data mining for a transaction, a bank user is screened for a double related account, and it is clear that this is the first requirement for the consumer. Then the data mining software will prejudge that the user is going to marry and accordingly can recommend a house mortgage for its orientation. It may even sell this information to the wedding service industry.

In a country with developed market economy, many large enterprises have begun to process deep layer through data mining on the basis of the original information, so as to establish their own advantages, enhance the competitiveness and expand the scope of operation. American Express, for example, has set up a database to record credit card consumption records, and as the volume of business continues to rise, content is constantly updated[5]. The company has made a thorough excavation of these databases and developed a very effective "correlation settlement" strategy. In a word, a user has bought a suit in the use of a credit card. If a pair of leather shoes is bought in the same place, there will be a huge discount. Shops and banks have made more profits since then.

Based on data mining, predictive marketing can recommend goods that are comparable to their previous consumption levels for specific consumers. Kaf food has a database of tens of millions of customer data, which is the feedback from different consumer groups of consumer vouchers and vouchers that have been issued by the company. Kaf food is based on the data mining, sends its preferred food to a specific population, and further recommends the related Kaf diet. Another American reader publishing company has a business database accumulated for decades. Its content is the data of hundreds of millions of customers worldwide. The database runs uninterrupted in 7*24 hours to ensure that the content is updated in real time. It is based on the mining technology of database information. The company's business expansion, from ordinary magazines to professional magazines, from ordinary books to audio-visual products, has greatly improved the company's profits.

The data mining of large foreign companies is also of reference to the domestic market. The staff of our daily contacts often do not distinguish the age, clothes, preferences and transportation grades of the objects, but only one ignorant "identical", which will only lead to inefficiency and can not be reached. Good effect. If the enterprises in the household electrical appliances service industry publicize and sell the pharmaceutical industry in the door of the electrical shop, the efficiency must be improved a lot.

Data Security and Secrecy

While big data bring convenience to people, there are also many security risks, especially personal privacy habits are analyzed by illegal elements.

For example, through the mining of population health data and gene sequencing data, we can predict the health of the national body. Through the data mining of mobile consumption, we can get the financial data of national consumption; through the analysis of social culture, we can draw people's entertainment preferences and ideological and cultural preferences. These results are very important. It may be exploited by lawless elements to further affect the security at the national level.

Because of the interest driven, the sale of personal information has formed a complete industrial chain. According to the current network laws and regulations, these illegal phenomena are still difficult to be renovated. But in a different way, the big data can play the most important role only in the mode of full flow and sharing and exchange. It is necessary to look at the advantages and disadvantages of the big data in today's society with a dialectical view. Only by perfecting the laws and regulations, can the big data play its due role.

Epilogue

With the advent of the era of large data, data mining technology has become an extremely influential tool, subverting the operation patterns of the past various fields, making all walks of life able to grasp the more efficient operation scheme. With the convenience of data mining, the security of personal privacy is also urgent to be solved. Only by establishing perfect laws and regulations, strengthening the protection of data infrastructure, giving full play to the guidance and leading role of standard, can the data mining technology be better for social service.

Reference

[1] Chongchong Cao,Yaohong Kang. Research on Web data mining [J]. Modern Electronics Technique,2007,(4):92-94,97.

[2] Kang Xie. Library bibliographic recommendation based on reader personalized feature data mining[J].ModernElectronicsTechnique,2018,(6):34-36.DOI:10.16652/j.issn.1004-373x.2018.06.008.

[3] Tianying Wang. Data mining analysis of planning supervision surveying and mapping based on BIM concept [J]. Engineering of Surveying and Mapping,2018,(3):58-62.

[4] Shunshun Wang . People's way of thinking in the era of big data [J]. Teaching Reference of Middle School Politics,2017(12):1+4-5.

[5] Quanhong Zhang. Thinking on the related problems of large data technology [J]. Chinese & Foreign Entrepreneurs,2016(05):126

[6] Yu D, Peng L. When does Inferring Reputation Probability Countervail Temptation in Cooperative Behaviors for the Prisoners' Dilemma Game? [J]. Chaos, Solitons & Fractals, 2015, 78: 238-244.