Application of Data Mining Technology in Economic Statistics

Yao Chen(✉)

Yunnan Technology and Business University, Kunming 651701, Yunnan, China

Abstract. At present, the domestic social and economic development is constantly improving, and during this period, there are many economic data reflect the uncertainty. Relevant departments make a classification of these data by means of statistical calculation, and then make a summary of the economic development in a period of time, so as to plan the next economic activities. However, in the unified calculation of these data, we must use a reasonable trial method, and the current unified calculation method has certain uncertainty, unable to get accurate results. If the data mining technology is used reasonably, the above problems will be solved. This paper discusses the application of data mining technology in economic statistics.

Keywords: Equipment support · Audit information · Data mining · Quick recommendation

1 Introduction

In the context of the rapid development of domestic economy, the statistical work of economic data of personnel in relevant units is becoming more and more complex, and there are more and more data categories. However, there is obvious uncertainty when using the traditional statistical method to sort out and calculate these data. This brings great difficulty for the following economic activities such as data analysis. With the help of data mining technology, we can improve the traditional way, not only to calculate the accurate data, but also to expand the depth and breadth of data statistics. In this way, we can get the practical data, and provide the guarantee for the orderly development of economic activities [1].

The so-called data mining technology is formed through the effective integration of multiple related professional knowledge. We can sort out and analyze the complex data, and through a series of sorting and analysis process, we can calculate the data with practical effect. It can be said that data mining technology belongs to the unified computing technology that can accurately process complex data.

M. A. Jan and F. Khan (Eds.): BigIoT-EDU 2021, LNICST 392, pp. 447–453, 2021.
https://doi.org/10.1007/978-3-030-87903-7_54

2 The Advantages of Data Mining Technology in Economic Statistics

2.1 With Strong Comprehensive Ability

In today's era when the domestic economy is greatly improved, it is necessary to accurately analyze and calculate the economic data. However, the standards for data projects are different in different working groups of relevant departments. In the real environment of various economic activities, usually involves a lot of management departments. During the period of making some choices and decisions, different management departments must take accurate data as the basis. However, different management departments will reflect different management methods, so the perspective of data requirements will be different. If the data mining technology is used reasonably, the data can be processed from different angles, so as to meet the different needs of different management departments for data, and can account for different departments. Summary work to provide the basis. For example, in the financial department of an enterprise, when accounting, the accounting personnel need to summarize the data. In this process, we must use data mining technology to format the data stored in the computer. This is the basis of accounting and summary. And this advantage is reflected by data mining technology. It can be said that data mining technology can play a comprehensive role in economic activities [2].

2.2 The Effect of Actual Statistics is Strong

Reasonable use of data mining technology can make a deep analysis and collation of complex data, in order to calculate the data with practical effect. For example, the application of data mining technology to a large number of complex data collation can make a reasonable and accurate calculation and collation of a large number of data. In order to make the complicated data more clear. Improve the efficiency of management. And this technology can also make a deeper and wider collation of the current data, and fully guarantee the reliability of the statistical data [3].

2.3 The Applicability of Data Mining Technology is Strong

To make accurate accounting and summary of economic data, the main purpose is to provide basis and guidance for economic activities in various real environments. However, different economic activities have different statistical angles. If the data mining technology can be used effectively, it can further improve the traditional calculation method. And can meet the statistical work of many units. It can be said that data mining technology belongs to the unified calculation method with the function of sorting out. The reasonable application of this technology to the economic data calculation can guarantee the integration of high-value information, so as to meet the needs of different angles of different economic activities. It can be said that this technology has good applicability.

3 Data Mining Technology

3.1 Basic Concepts of Data Mining

With the change of communication mode and the development of storage technology in recent years, a large number of data are generated and recorded in every bad section and every moment of life and production. From the state to enterprises, they gradually realize the value of data, which will become a symbol of national strength and wealth in the future. Any of our activities will produce data, and these intangible data, like historical relics, can discover the value of cultural relics in specific fields by using effective archaeological discoveries. The process of value exploration and discovery of data as "historical relics" is data mining. 2. Just as people can predict some familiar affairs according to their long-term work experience, In the information age, data mining is an important link to find, extract or capture knowledge from the daily accumulated massive data [4].

According to the results of the requirement analysis of the equipment support information system, this chapter will carry out the specific design and implementation of the system. This project uses BS architecture, Java background language, bootstrap foreground language and the corresponding data mining technology to design the system, so as to meet the functional requirements of remote access, friendly interaction and quick recommendation. The main contents of equipment support information system design include: system overall framework design, database design, core function module design and user support information recommendation design. In the process of core function module design and user support information recommendation design, the business process is expanded and described, and the function interface is implemented.

Equipment support information table is used to store the questions and answers submitted by users, which is the most important data in the system. In order to ensure the security of the system and introduce the security mechanism in the future, the picture content is stored in the table in Base64 format instead of separately. Equipment support information status includes saved, submitted, unanswered, answered, failed and passed. At the same time, with activiti5 calling process progress ID, the equipment support information flow is realized. User support data is the core data of data mining in the system, which should have the required user information, equipment information and other data content, such as the number of user organization, equipment number, support object number and equipment parameters. Equipment parameter is the type of equipment parameter corresponding to special character segmentation. Local audit table is used to store local audit information, including user number, user name, user organization number, log type and operation time. When more data is accumulated, it is helpful to achieve faster query recommendation through system upgrade and time series data mining.

3.2 Common Algorithms of Data Mining

According to the data mining tasks in different application scenarios, the data mining algorithms are generally different, and sometimes the mining algorithms need to be modified according to the research data objects.

Logistic regression analysis can explain the relationship between the discrete dependent variable and the independent variable. Generally, the binary variable, i.e. the dependent variable, has two values. For example, Y(1) represents the positive result and Y(0) represents the reverse result. At this time, the independent variable x is a group of explanatory variables of the dependent variable, and the logarithm of the approximate response ratio of X to the dependent variable Y = 1:

$$\text{Logit}(P) = \ln\left(\frac{P}{1-P}\right) = \alpha + \beta \tag{1}$$

In the formula, a is the intercept and β is the slope. In the multi classification relationship, the value of dependent variable is not limited to two kinds, that is, there are multiple values of Y.

Typical discriminant function is based on Bayes theory. It analyzes and calculates the correlation between the observed values of each category, and then establishes the corresponding classification rules, so as to realize the re classification of the original sample. By comparing the original class with the predicted class, the classification and discriminant accuracy of the initial sample are finally determined. It can be expressed as:

$$\mathrm{X}^i = \left({x_1}^i, {x_2}^i, \ldots {x_m}^i\right), i = 1, 2, \ldots k \tag{2}$$

Based on the deficiency of multiple logistic regression prediction, the classification method based on discriminant function is used for further classification prediction.

4 Application of Data Mining Technology in Economic Statistics

4.1 The Method of Data Integration Lays a Foundation for the Accurate Statistics of Economic Data

With the continuous development of social economy, the relevant data is bound to increase correspondingly, and the categories are also more abundant, which makes the data statistics work more difficult. If the data mining technology is reasonably used, the data in the computer database can be checked and sorted, which ensures the efficiency of data identification and provides guarantee for the orderly development of data statistics work.

The equipment support information system is not only the data provider, but also the data demander. The acquisition, storage and utilization of all kinds of data in the support process is very important to the system. Combined with the characteristics of all kinds of important data, this paper focuses on the following factors to design the database of the system: 1) considering the scalability and adaptability of the system, it is necessary to fully reserve the field and table relationship. 2) Because the data in the system needs to connect with other systems or equipment, it is necessary to ensure that the user support data meets the standard audit data format proposed by this topic, which is the data requirement for data mining of user support data, and also the basis for establishing the standard equipment support information database. 3) Equipment

support information is sensitive. In order to facilitate the subsequent adoption of data encryption and other security measures, increase the redundancy appropriately, and store the text and pictures in the support information in the form of text in the database, so as to ensure the consistency of data and avoid the mismatch of graphics and text in equipment information call.

4.2 The Pretreatment Method Should Be Applied in It

The reason why data mining technology can do efficient unified calculation for complex data is mainly due to its intelligent function, so when there is deviation in the initial unified data, it will inevitably lead to the uncertainty of the unified data. But in the specific work, the initial statistics data are often uncertain and imperfect, so with the help of preprocessing, these uncertain and imperfect data can be completely deleted to ensure the accuracy and integrity of the data by means of data mining technology. The pretreatment method should be used, as shown in Fig. 1.

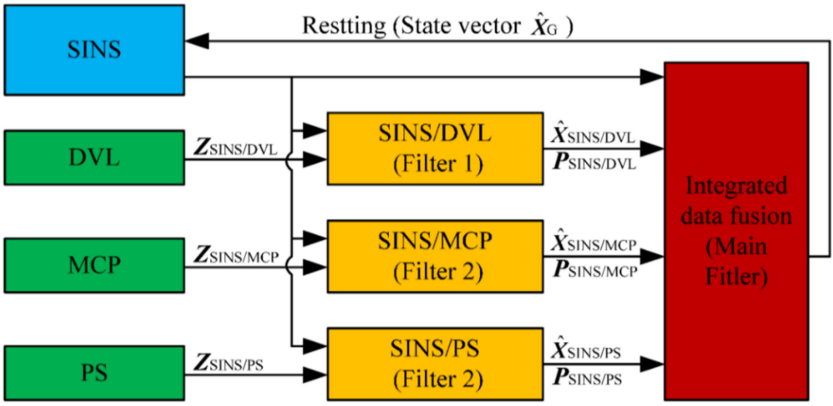

Fig. 1. Pretreatment should be used

4.3 Regression Analysis Method Can More Clearly Reflect the Data Statistics

Regression analysis method can show the relationship between variables more clearly. This method can more directly analyze the current commodity market share, sales and other data statistics. Taking the simple linear regression analysis method as an example, the qualitative variable is set as X, the dependent variable is set as y, and a is the intercept and B is the correlation coefficient. At this point, we can get the linear equation as $y = a + bX$. At this time, the results of different dependent variables are substituted into the formula, which can intuitively analyze the variables.

The system uses Java language to complete the main development of the application. User and role management, equipment information management, equipment support information management, business process management, statistical analysis and data

management constitute the basic functions. Role management provides the functions of creating, editing, deleting and permission control of roles in the system. User management provides users with various roles to add, edit and delete. Equipment information management provides the addition and editing of equipment attribute parameter types and corresponding items. Equipment support information management provides support information creation, answer, approval and recommendation.

Timeliness of data. The interface for obtaining and importing the user support information of different equipment systems is reserved in the system, which is convenient to obtain the audit log in time and ensure that the data in the system is kept up-to-date. The choice of data mining algorithm. The core function of the system is to find valuable information in the user support data, then use the system filtering algorithm to analyze the user similarity, and use the association rules to analyze the support area between equipment. According to the characteristics of user information, equipment information, audit information and other support data, this paper improves the traditional Apriori algorithm to meet the requirements of application scenarios.

Regression analysis method can more clearly reflect the statistical nature of the data, and the specific analysis is shown in Fig. 2.

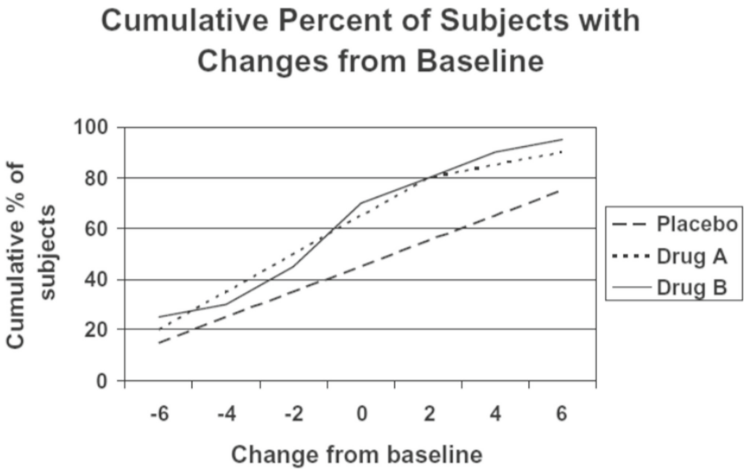

Fig. 2. Regression analysis method reflects the statistical nature of data

5 Conclusion

Data statistics can play an irreplaceable role in the social and economic development. The traditional statistics can not meet the needs of social and economic development. With the help of data mining technology, the traditional statistical method can be improved, which provides a guarantee for the realization of accurate data. It can be said that the use of data mining technology can make a clear arrangement and analysis of the current economic data, so as to obtain accurate and reliable data, and provide the basis for future economic activities.

References

1. Wang, S.: Wang Shuan's Theory and Practice of Foreign Military Information Construction. PLA Press, Beijing (2008)
2. Jiawu, H., Qiuxing, G.: Support measures and characteristics of US military equipment package in Iraq war. Foreign Mil. Sci. **8**, 40–42 (2003)
3. Yao, H.: Overview of the development of US military logistics information system. Mod. Mil. (05), 87–91 (2016)
4. Zhao, J., Deng, J., Liu, Y., Jungle, T.: Digital construction of information support for ordnance equipment. Logist. Technol. **36**(10), 142–145 (2017)