Application of Big Data Mining in the Design of Logistics Website

Chunxia Zhai1,*

1Xi'an Traffic Engineering Institute, Xi'an, 710300, Shaanxi, China

*Corresponding author e-mail: zekkie@snnu.edu.cn

Abstract. With the development of the information society, the information storm brought about by big data is changing the way we live, work and think. At present, many existing logistics operation systems have begun to transfer transaction mechanisms to websites due to the influence of networking. Therefore, how to use big data mining technology can complete the analysis of a large amount of data in the logistics website, so as to realize the logistics operation on the website. This article mainly introduces the decision tree method and K-means method. This article uses big data to design the logistics website operation system and establish a potential time-discrete model. The model is solved by the decision tree method, and the design status of the big data mining in the logistics website operation system is evaluated, and historical data is used to revise the model to improve accuracy. The experimental results of this paper show that the application of big data mining in the design of logistics website operation system has increased by 48% and the false report rate has been reduced. However, the large database of my country's big data mining in the design of logistics website operation system has not been established, so that the use of big data cannot be fully implemented.

Keywords: Big Data, Logistics Website, Operation System, Decision Tree Method

1. Introduction

1.1. Background and Significance

With the advent of the era of big data, our lives are full of intelligence and mechanization, which causes everyone to generate a large amount of electronic data every day, and these data must be stored. For the above reasons, it is difficult to imagine how huge the scale of information data stored in the world is [1-2]. Today, the logistics industry is in a period of continuous development[3]. The amount of data stored in the logistics information database has entered the megabyte era [4-5]. At the same time, end users and business managers have higher requirements for data records, which require quick and effective analysis of a large number of data records. However, ordinary database technology
(such as the first communication database system) is far from meeting the needs of users[6].

Data warehouse technology and data mining based on OLAP technology have played an important role in modern logistics information systems. They cannot quickly and effectively analyze and investigate their own knowledge and location knowledge for this purpose, and cannot clarify the needs and needs of the forum population. The dynamics of the target population and company management can be regarded as the right decision, which can provide satisfactory services to classrooms and improve the company's market competitiveness [7-8]. Traditional technology can no longer meet the demand. As a result, data mining is widely used in product optimization, precision marketing, customer retention and return, user behavior analysis, etc. The basis of using the data model has been widely used by various operators [9]. Using data mining technology, telecom operators can analyze and make correct business decisions in a timely manner based on a large amount of complex data, understand industry development trends and analyze key factors affecting product marketing, so as to achieve the purpose of increasing revenue and reducing costs in order to obtain good social and economic benefits, and improve the market competitiveness of operators [10].

1.2. Related Work

The advancement of information technology has led to the rapid development of the logistics industry. At the same time, the emergence of technologies such as the Internet, the Internet of Things, and cloud computing has brought society into the era of big data. Although the logistics of agricultural products in my country has been developed for a long time, the level is low, the price is high, and the loss is huge. By combining current problems and the background of big data, analyze the status quo of agricultural product logistics, and summarize the role of big data in agricultural product logistics and agricultural products in a period of time. Logistics construction procedures [12].

1.3. Main Content

The innovation of this article lies in the decision tree method and K-means algorithm. According to the design scope of big data mining in the logistics website operation system, as a basis for judging the design of big data mining in the logistics website operation system, the analysis and calculation are moved down to the storage layer to reduce computing time while reducing the amount of data transmission, thereby ensuring the data analysis interface Real-time.

2. Big Data Mining in Logistics Website Operation System Design Method

2.1. Decision Tree

As the name implies, the so-called decision tree is a tree structure constructed for strategic decision-making. Compared with other algorithms, some decision trees have obvious advantages: strong interpretability and comprehensibility; it can intuitively reflect the characteristics of the data itself, users do not need to know their knowledge too much, they can understand through explanation The meaning expressed therein. The requirements for data are not high; other technologies often require the processed data to be of a uniform type, but the decision tree can handle data types and common attributes.
Therefore, the data preparation of the decision tree is simpler than other algorithms, and good results can be quickly obtained when dealing with more types of data. The accuracy of the determination model of the decision tree can be measured, and the decision tree is a white box model. As long as any model can generate a decision tree, it is easy to infer the corresponding logical process. The calculation amount of the decision tree is small and can be easily converted into classification rules. As long as the root tree of the decision is followed up to the node of the leaf, the unique classification result can be determined. As long as you follow the root node of the decision tree all the way down to the leaf nodes, you can determine the only classification result. Because of its simplicity, intuitiveness, easy grasp, strong applicability, qualitative and quantitative analysis, and the simple and clear analysis of the business risk and direction of the enterprise, the decision tree method has gradually attracted the attention and application of modern enterprises. The decision-making level of the enterprise.

2.2. K-Means Method

Cluster analysis can be used as an independent tool for data extraction, and this article summarizes the characteristics of each category, such as the customer segmentation of Apple mobile phone users. Similarly, aggregation analysis can also be used as a preprocessing step for other data extraction algorithms. The process of K-means algorithm is as follows:

The number of k clusters can be predetermined according to needs, and then k objects are randomly selected. Each object at the beginning represents average or middle-class value. For the remaining objects, calculate the distance to each cluster center and divide it into "provide it to the nearest cluster". Then recalculate the average of each cluster. Repeat this process until the criterion function converges.

The average calculation method is generally:

$$M_j = \frac{\sum_{j=1}^{n} H_{j}E_j}{\sum_{j=1}^{n} H_{j}}, j = 1,2,\ldots,k$$

(1)

Where $M_j = 1,2,\ldots,k$ is the cluster center; $k$ is the given number of clusters; $K=$ is the weight. When the value is 1 or 0, that is, when $=1$, it means $H_{j}$ belonging $M_j$, otherwise it does not belong. The cluster at this time is called hard clustering.

When the value $H_{j}$ is between the continuous interval (0,1), it is called fuzzy clustering. $M_j$ is used to describe the degree of continuous membership of the data object to the cluster center. clusters

The general convergence criterion adopts the definition of the square error criterion as follows:
\[ F = \sum_{j=1}^{s} \sum_{q=B_{j}} \left| q - n_{j} \right|^{2} \tag{2} \]

The sum of the error values of all data objects in the database is F, q is a point in the space, represents a given object, and is the average of the cluster (q and are more balanced). The function of this structure is to make emerging groups as independent as possible. According to a given S value, grouping can give two completely different results: \( s = 1 \), that is, the result obtained is meaningless, that is, all information is divided into a button; on the contrary, \( s = n \), that is, record The number of objects in the actual data set is also meaningless. The number of other groups depends on the value of \( s \). There is no clear and fast rule to follow for the price of S. Usually, the best way is to check different values.

3. Big Data Mining in Logistics Website Operation System Design Experiment

3.1. Big Data Mining in Logistics Website Operation System Design Data Collection

The data used in this paper is the frequent item set mining data repository in the network public data test set, and the following two sets of data are obtained as the original data set. As shown in Table 1:

| Database   | Trans | Items | Size     | int |
|------------|-------|-------|----------|-----|
| Accidents  | 340183| 656   | 46.8 MB  | 114 |
| webdocs    | 5267656| 785   | 52.6 MB  | 881 |
| varchar    | 1792082| 486   | 22.4 MB  | 365 |

3.2. Big Data Mining in the Functional Design of Logistics Website Operation System

Commercial websites can also get customer relationship management capabilities. The so-called customer relationship management is actually the use of data mining technology for enterprises and customers. Commercial websites can also implement customer relationship management functions. The so-called customer relationship management is actually the management of the interaction between the enterprise and the customer. From a design perspective, webmasters need to be customer-oriented and try their best to meet the individual and diverse needs of users. Applying data mining technology to mining customer information will help website administrators understand the historical transaction information of customers and products, thereby understanding customer usage trends and product popularity. Therefore, the application of data mining technology can provide companies with more opportunities to sell products to existing customers, and provide a database for companies to make product analysis decisions. In addition, customer relationship management that uses data mining technology to share and automatically manage business data can also help companies complete business analysis, which will bring difficulties to data analysis and analysis. For this reason, it is also
necessary to use data analysis technology to analyze data about the behavior of website visitors and draw conclusions about customer background information from these data.

When a commercial website is designed, only a commercial software platform with a business logic foundation can be used for data mining to achieve the purpose of extracting user data. Therefore, in order to develop the structure of commercial websites, it is necessary to create a data mining system combined with commercial publications to use data mining technology. During the design process, the database system can be configured in three areas: data storage, data processing and data display. At the same time, the user access part of the site should be regarded as customers, and the data mining system should provide support for user access and interaction. After the customer submits the request, the system needs to decompress the requested business logic, obtain the data from the data store, and then return the processed result to the customer.

4. Big Data Mining in Logistics Website Operation System Design Analysis

4.1. Big Data Mining in Logistics Website Operation System Design Data Analysis

The system was tested in a 10 Gigabit network environment of a backbone node in Chengdu. In the network deployment plan of this system, it was used as a 10 Gigabit network collector, the system was used as a database server, and the disk array system was used as data storage and backup, and used as a data analysis server. To ensure processing speed and data security between the web server, database, 10G acquisition and storage system, optical storage networks are needed, and optical switches are used to provide high-speed interconnection for them. This paper implements the telecom operator's network operation support system in the 10G network environment, and proposes a fast matching algorithm to realize real-time server collection at the 10G rate. The result analysis of the collected data is shown in Figure 1:
4.3. Big Data Logistics Website Operation Analysis

The overall architecture of the big data logistics website operation analysis module will focus on "customer insight-resource status analysis-resource optimization strategy-marketing. The platform will divide the big data logistics website operation investment into one-time resource investment and monthly allocation investment, and then pay for marketing resources Sort out and construct an evaluation model. One-time resource input includes remuneration cost, terminal cost, and physical cost; monthly amortized cost includes communication cost, settlement cost, fee increase cost, integral cost, etc. The system formulates corresponding service marketing projects based on production indicators and tasks as well as resource usage plans, and rolling adjustment of resource usage plans according to resource implementation, rolling adjustments need to be reviewed by the office/department/company budget administrator.

The platform will monitor the marketing activities of each branch in real time. It will evaluate the effectiveness of each marketing resource usage in detail and fully understand the implementation of
marketing activities, including daily customer participation trends, regional implementation, resource usage, activity participation rate, and in-use Information such as the number of customers and the evaluation of marketing effects. Perform statistical analysis on various marketing activities and make statistics on the use of resources. Support resource statistics of a single marketing activity, and support resource statistics analysis for a fixed time period. Effectively help business personnel understand the overall use of resources. The operating results of the big data logistics website are shown in Table 2:

Table 2. Big data logistics website operation results table

| Logistics website    | Confidence | project | Support | Support count |
|----------------------|------------|---------|---------|---------------|
| Jingdong             | 99         | 81      | 120     | 6             |
| Suning Online Market | 87         | 79      | 118     | 9             |
| Alibaba              | 92         | 88      | 109     | 12            |

For the purpose of data mining, the data is formed into a new data table, and the data is further collected according to items such as comorbidities, names, regions, etc., so that the data in the original database is reorganized according to the target requirements. It is beneficial to make decision analysis and can see the inner relationship between data more clearly. The analysis result is shown in Figure 2:

Figure 2. Analysis of big data logistics website operation results

5. Conclusion

Although this paper has made certain research results on the application research model of big data
mining in the design of logistics website operation system, there are still many shortcomings. Based on the decision tree method and the K-means method, there are still many in-depth content worthy of study. There are many steps in the decision-making process of the logistics website operation system for big data mining because of space and personal ability, etc., which are not covered. In addition, the actual application effect of the improved algorithm can only be compared with the traditional model from the level of theory and simulation.

References

[1] Brinch M, Stentoft J, Jensen J K, et al. Practitioners understanding of big data and its applications in supply chain management[J]. The International Journal of Logistics Management, 2018, 29(2): 00-00.

[2] Jula P, Leachman R C. A supply-chain optimization model of the allocation of containerized imports from Asia to the United States[J]. Transportation Research Part E Logistics & Transportation Review, 2011, 47(5): 609-622.

[3] Mason J B, Shrimpton R, Saldanha L S, et al. The first 500 days of life: policies to support maternal nutrition[J]. Global Health Action, 2014, 7(1): 23623.

[4] Fs/EPDF. Minutes of the second annual meeting of the Zonal Monitoring Program for the Northwest Atlantic[J]. Laryngoscope, 2002, 122(2): 189-194.

[5] Yap J Y L, Ho C C, Ting C Y. A systematic review of the applications of multi-criteria decision-making methods in site selection problems[J]. Built environment project and asset management, 2019, 9(4): 548-563.

[6] Aubert B, Ballansat J, Colas J, et al. Development and construction of large size signal electrodes for the ATLAS electromagnetic calorimeter[J]. Nuclear Inst & Methods in Physics Research A, 2005, 539(3): 558-594.

[7] Parry G, Bustinza O F, Vendrell-Herrero F. Servitisation and value co-production in the UK music industry: An empirical study of Consumer Attitudes[J]. International Journal of Production Economics, 2012, 135(1): 320-332.

[8] Yang S, Munson C L, Chen B. Using MSRP to enhance the ability of rebates to control distribution channels[J]. European Journal of Operational Research, 2010, 205(1): 127-135.

[9] Ouma S, Boeckler M, Lindner P. Extending the margins of marketization: Frontier regions and the making of agro-export markets in northern Ghana[J]. Geoforum, 2013, 48(aug.): 225-235.

[10] Zhai S, Duan J, Ai X. Research on the third party logistics system and economic performance optimization based on big data analysis[J]. Boletin Tecnico/Technical Bulletin, 2017, 55(11): 301-308.

[11] Yu J, Chen Y, Jia X, et al. Study on the Construction of Big Data Platform for Medical Supply Chain[J]. Revista De La Facultad De Ingenieria, 2017, 32(5): 225-229.
[12] Zhao Junqing. Research on the Application of Big Data in Chinas Agricultural Products Logistics. Research on the Application of Big Data in Chinas Agricultural Products Logistics[J]. Journal of Zhengzhou Railway Vocational and Technical College, 2018, 030(004):8-11.