Design of Rural E-commerce Customer Data Mining System

Xiaowei Huang1,*

1Xiaowen Huang, College of Economics and Trade, Shandong Management University, 250357

*Corresponding author e-mail: huangxiaowen@sdmu.edu.cn

Abstract. Customer relationship management based on data mining helps to strengthen the cultivation of high-quality customers and improve customer loyalty and loyalty. China's e-commerce industry market is becoming more and more prosperous, with the data volume increasing exponentially. Mining the purchasing behavior data of e-commerce industry customers has become a major way for major e-commerce platforms to increase sales volume. Traditional mining algorithms cannot meet the needs of big data analysis. According to different purposes, we will combine mining algorithm and cloud computing platform to mine the potential information of these data and provide beneficial guidance for enterprise operation. This paper starts with the impact of big data on rural e-commerce customers. At the same time, this paper analyzes the current situation of rural e-commerce and proposes the design of rural e-commerce system based on data mining [1].

Keywords: Data Mining, Customer Relationship Management

1. Introduction to Data Mining

Data mining is a hot topic in the field of artificial intelligence and database. From the perspective of technology, data mining is the process of extracting the knowledge needed by the industry or company from big data by using a series of relevant algorithms and technologies. This valuable underlying knowledge and information is hidden in big data. Previously unknown, extracted knowledge can be represented in a variety of ways, such as concepts, rules and forms. Compared with traditional statistical analysis, data mining has the following characteristics:

(1) Have stronger ability to process big data, and can use data mining tools without too professional statistical background [2];

(2) From the perspective of use and demand, data mining tools are more in line with the needs of the business community;

(3) Analyze from different angles of data mining and statistical analysis. The ultimate purpose of data mining is to facilitate the use of enterprise end users rather than statisticians for detection.

In data mining, the data source should be obtained according to the actual requirements, which may be of different structure types, such as semi-structured. When mining data, there are some duplicates and errors in the original data set to be cleaned and deleted. To convert to a data format that matches a computer processing format. Data set algorithm processing involves data visualization, artificial
intelligence and other technical fields. The extracted potential information is helpful for decision makers to manage and make decisions.

2. Data mining process
Data mining is the iterative process of exploring the special relationship between data in a large number of data sets and discovering patterns, rules and output values. The general execution process suitable for data mining has the following steps:

**Table 1. Data mining process.**

| Step                      | Description                                                                 |
|---------------------------|-----------------------------------------------------------------------------|
| State the problem         | According to the actual demand and mining purpose, some initial assumptions of unknown correlation are proposed. Select the appropriate algorithm to develop a detailed data mining scheme. |
| Data collection           | There are usually two different methods: One is the planned generation of some data under the control of the requirement personnel; The other is that the requirements person cannot influence the data generation process. |
| Data preprocessing        | Preprocessing can improve the standard and consistency of data and the efficiency of data mining algorithm [3]. |
| Data mining               | The appropriate data mining method is selected according to the mining target. Select the appropriate mining algorithm according to the data source and user demand. |
| Analytical model          | According to the initial mining target, select the relevant Angle to verify the accuracy of mining results. Visually present valuable information to the user. |
| Application stage         | Apply the mined information to the real world.                               |

3. Methods of data mining
In the process of data mining, the knowledge and technology of many subjects are constantly integrated into data mining. At present, data mining methods and algorithms have been presented in various forms. With different mining purposes as the classification criteria, data mining can be divided into:

3.1. Cluster analysis
Cluster analysis is the process of classifying abstract objects into similar objects. Clustering results in great similarity among the same cluster and great difference among different clusters. Clustering can be used to classify customers into the same level, and customers with similar consumption records can be classified into the same level, so as to regularly recommend products and improve the sales volume of the company.

3.2. Classification
Categorical analysis is usually a search for common ground among items in a data set. Build data items into different categories at different points. The purpose of classification is to map each data item to a given category. Unlike clustering, classification divides objects according to a given category.
3.3. Deviation analysis
Bias analysis, also known as comparative analysis, is designed to find valuable differences between detected and referenced data and to filter out a large number of uninteresting patterns. The data is usually analyzed based on some statistical methods, such as density, distance, etc. Different data information is extracted from the database, including transactions different from the standard class and abnormal fluctuation between two adjacent data items, so as to find the cause of the abnormal data.

3.4. Data visualization
Data visualization is a scientific and technological research on the visual expression of data. Among them, the visual representation of such data is defined as a kind of information extracted in some summary form, including various attributes and variables of the corresponding information unit. For example as shown in figure 1:

![Data visualization presentation](image)

**Figure 1.** Data visualization presentation

4. Design of data mining system

4.1. The system design
This topic combines the background of big data environment and e-commerce system of beauty industry. It organically combines Hadoop batch processing technology, data mining technology and visualization technology to finally realize a complete set of big data processing system [4]. The overall module block diagram of the subject is shown in the figure 2.
Figure 2. The system block diagram

The overall design of the system is mainly composed of the following stages:

The first stage is the data preparation stage. The data comes from the trading data of the internship company's system platform and WeChat official account, as well as the data information of commodities stored on the platform. These data are stored in the non-relational database MongoDB, and relevant business data are read into THE HDFS by Hadoop according to the business logic requirements. In this process, the data reading and preprocessing are completed [5].

The second stage is the data processing stage, in which data mining and analysis are performed according to different requirements. Distributed computing framework MapReduce combined with mining algorithm to realize parallel processing. Mining and analyzing the business data read by Hadoop and establishing the data analysis model corresponding to the visualization topic.

The third stage is the transformation stage of data processing results. The data is converted into Json format according to the analysis results, so as to prepare for ECharts visual class diagram display.

The fourth stage is visual presentation. SpringMVC integrates Hadoop so that data processed by Hadoop cluster can be directly returned to the front end. Baidu Echarts is applied to display different types of graphics.

4.2. Data preparation

The first prerequisite for studying and analyzing big data is to have big data. The way to have big data is either to collect and aggregate natural data, or to acquire data collected by others. Now, in order to achieve precision marketing, many companies have started to buy customer data from e-commerce and search engine companies in order to accurately discover new customers. Various institutions and departments in the society are constantly generating a large amount of information in real time. This information needs to be processed in a simple way while being accurate and fast enough to satisfy all types of data demanders.

5. Conclusion

This research relies on the rural e-commerce platform, establishes the analysis Angle according to the business logic and designs a set of big data processing system. Through the related theory and technology research, the designed system is realized [6]. However, due to the incomplete data analysis Angle and the diversity of algorithms, there are still many aspects to be improved and perfected in the design process. The following points are the main research directions of this paper:

(1) When the data source is obtained, MongoDB single database is used as the data source. Once the multi-database operation is carried out, the data migration operation is needed, which does not have good scalability. In the case of multiple databases, data migration remains to be studied [7].

(2) Combining Hadoop cloud computing platform with data mining algorithm, the algorithm realized parallel processing and further optimized the algorithm. It can improve the processing efficiency of data and also choose other mining algorithm to compare and analyze the row results.
References

[1] Wu You, Gao Jingyi. Analysis of Unbalanced Classification Problem Set in Data Mining [J]. Electronic Production, 2014(10): 139.

[2] Zhao Junmin, Huang Obituary, Zhang Weiwei. Research and implementation of clustering system based on hadoop and mahout [J]. Fujian computer, 2017, 33(01): 6-7+14.

[3] Chen Zhixiong. Design discussion of Distributed Data Mining System based on Hadoop Platform [J]. Digital Technology and Application, 2017, 01: 179.

[4] Hong Bo, Lv Yanxia, Huang Lei. Research and Implementation of Data Mining Algorithm based on Hadoop Framework in big data Environment [J]. Electronic Design Engineering, 2017, V. 25; No.35707: 41-44.

[5] Sheng-yu, Wang Jing-yu1, Zhang Xiao-lin, Gao Jun-feng. Research on K-Means Clustering Algorithm Optimization based on Hadoop Platform [J]. Journal of Inner Mongolia University of Science and Technology.

[6] Long Hu, Yang Hui. Research on Data Analysis and Visualization in the Context of Big Data [J]. Journal of Kaili University, 2016, 03: 98-102.

[7] Yu Yue. Research on parallel Distributed Association rules Mining Algorithm based on Hadoop Platform [D]. Jilin University, 2017.