Design and Implementation of Precision Marketing System Based on Data Mining Technology

Na Yang(✉)

Chongqing College of Architecture and Technology, Chongqing, China
34233143@qq.com

Abstract. In view of the low-efficiency marketing problem caused by customer orientation ambiguity in many enterprises' marketing strategies, this paper studies the calculation of customer value based on the existing customer behavior characteristics data, comprehensive utilization of hadoop data processing software, accurate dynamic portrait of consumers by using K-means clustering algorithm, and construction of user labeling model. At the same time, python web crawler technology is used to obtain the characteristics of potential users. The software is developed with java language and SSM framework. From the perspective of systematic precision marketing, this paper puts forward a rational analysis on the marketing scheme design of enterprises, and helps the marketing decision-makers of enterprises to improve their marketing benefits.

Keywords: data mining · K-means algorithm · Big data technology · precision marketing

1 Introduction

In the current era, due to the rapid development of network information technology, society is fully integrating with Internet technology to achieve digital development. Under the influence of technology, the market competition of enterprises is becoming more and more fierce. More and more consumers will choose consumer products according to the marketing content of the network, but the customer orientation ambiguity in marketing strategies of many enterprises in the current market leads to inefficient marketing problems. Therefore, implementing accurate marketing is an effective way to improve the operating conditions of many small and medium-sized enterprises.

Precision marketing refers to the establishment of customer database through the Internet and scientific analysis, so as to provide enterprises with a more accurate marketing communication model. Precision marketing requires enterprises to make full use of data processing and analysis technology provided by big data technology, combined with data mining algorithm technology to achieve it.

Big data technology can process and integrate a large amount of data in a short time. Big data technology produces the data value chain of data generation, collection, storage

J. Yen et al. (Eds.): ICBIS 2023, AHCS 14, pp. 686–693, 2023.
https://doi.org/10.2991/978-94-6463-198-2_71

and analysis. Therefore, big data provides the foundation and technical platform for the implementation of precision marketing. At the same time, based on precision marketing positioning, the enterprise has established a personalized service model, achieving the goal of high efficiency and low cost. Data mining technology can help small and medium-sized enterprises predict the potential needs of customers according to the historical business data within the enterprise, so as to achieve the purpose of improving the marketing decision-making level of enterprises. Making full use of data mining technology can help enterprises to mine the most suitable and valuable information and knowledge from the massive customer and market information, and promote the promotion of the core competitiveness of enterprises. According to the above analysis, the author of this paper thinks that the precision marketing system should be designed and implemented by combining data mining technology. The system can not only screen, delete and query the customer data, but also expand the function of extracting and analyzing valuable information and knowledge from the data, which can greatly improve the precision marketing ability of enterprises [1].

2 Technical Summary

2.1 Data Mining

Data mining technology refers to the process of extracting and optimizing valuable information and knowledge from huge data. The main components of data mining technology platform are modeling, data and algorithm. The most closely related technologies of data mining include machine learning, decision support, statistics and data warehouse. The key content of the data mining platform studied in this paper is the core algorithm used in data mining. The commonly used analysis methods of data mining include grouping method, clustering method, complex data type mining method, prediction method, classification method and estimation method. Classical algorithms of data mining include C4.5, K-means, SVM, Apriori, EM, PageRank, Adaboost, KNN, Nave Bayes, Cart [2].

2.2 Python Crawler Technique

Before the advent of web crawler technology, it was undoubtedly a huge workload to basically use manpower to collect information on the Internet, and countless developers were troubled by this work. Therefore, web crawler technology came into being. Web crawler technology refers to the technology of automatically extracting web information by a class of programs according to Internet rules. Python is the most commonly used development language for this technology. The principle of web crawler technology is realized by setting up new crawling rules and setting the URL of the portal. The basic workflow of this technology is shown in Fig. 1 [3].

First, the developer selects a certain amount of seeds according to the requirements and saves the corresponding URLs. Then, a URL queue to be grabbed is set by the algorithm to save the selected URLs. After that, the program starts to download the contents corresponding to these URLs and grab the key information, and the processed URLs will be saved in the new grabbed queue. At the same time, DNS resolution data

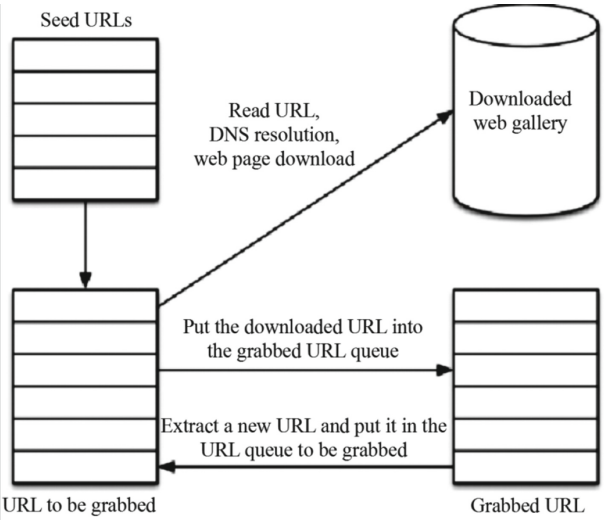

Fig. 1. Schematic diagram of web crawler

and webpage download data generated by URL resolution will also be saved in the downloaded webpage database. Part of the URLs saved in the captured URL queue will be returned to the URL queue library to be captured for reprocessing if the processing is incomplete. Scrapy is chosen as the framework software of web crawler technology in this paper, because it is an open source authoritative and widely used stable framework.

2.3 K-means Algorithm

K-means algorithm is a data mining technology algorithm proposed by Macqueen, which is based on data division and clustering. The basic principle is to select K cluster centers at first, and divide all data points into each cluster range according to the nearest principle. Then, the center of the cluster is divided again according to the mean value. The function measures the clustering quality by Sum of the Squared Error (SSE) and error square. Determine a sample data set $S = \{x_1, x_2, ..., x_n\}$ containing n multidimensional data objects. First, determine K cluster centers $z_1, z_2, \ldots, z_k$, and then calculate the Euclidean distance d between each cluster center and the sample data. The calculation formula of d is shown in Formula (1) [4].

$$d(x_i,x_j) = \left(\left|x_{i1} - z_{j1}\right|^2 + \left|x_{i2} - z_{j2}\right|^2 + ... + \left|x_{ip} - z_{jp}\right|^2\right)^{\frac{1}{2}} \quad (1)$$

Then, the sample data is distributed to the cluster center with the minimum Euclidean distance. Then, the center points of each cluster are recalculated according to the mean value, $z_j = (z_{jm1}, z_{jm2}, ..., z_{jmp})$, and finally, the steps are repeated continuously. The above process is repeated until the cluster center is no longer changed.

```
cd$HADOOP_ HOME #Enter Hadoop home directory
bin/hadoop fs -mkdir -p /user/hive/warehouse # Create directory
bin/hadoop fs -chmod -R 777 /user/hive/warehouse #The newly created directory is
given read and write permission
bin/hadoop fs -mkdir -p /tmp/hive/#Create a new /tmp/hive/ directory
bin/hadoop fs -chmod -R 777 /tmp/hive #Directory grant read-write permission
Check whether the directory was created successfully with the following command
bin/hadoop fs -ls/user/hive
bin/hadoop fs -1s/tmp/hive
```

Fig. 2. Execution command code of creating a new Hive on Hadoop

2.4 Development Environment

In this paper, the author briefly introduces the related technologies of platform development and use. The big data precision marketing system uses Hadoop as a big data server cluster to process data and store it in MySQL database, and uses JavaWeb technology to develop the corresponding application platform.

According to the data volume and overall operation requirements of the system, this paper chooses to build a Hadoop3.3.1 cluster with three nodes, including a master node and two slave nodes. The main node is named namenode, and the secondary node is named datanode. Then, the distributed collaboration system zookeeper-3.4.1, distributed file system HDFS 2.6.5, flume1.9.0, Hive 0.13.1 and Hbase2.6.5 are installed and deployed in these three nodes synchronously, and the initial construction of hadoop cluster is completed. The cluster will be developed under Linux system. This paper selects Centos6.5 Server release version of Linux operating system. The version of the web crawler framework Scrapy is 2.5, and Python3.8 is chosen as the development language [5].

The front-end development tool used in the JavaWeb application of this system is boomstrap + jquery, and the development language is JavaScript + HTML + CSS. The back-end Java development tool is IDEA 2021.1.3 (Ultimate Edition), the development environment is JDK 1.8, and the J2EE framework of Tomcat + Spring MVC + Spring + MyBatis is is used in the implementation of this system. The development language is Java, and MySQL 8.0.28 is selected to help manage data. The execution command code for creating a new Hive on Hadoop is shown in Fig. 2. Through the introduction of the above key technical theories, the overall environment for the development of big data precision marketing system, the configuration of related software and tools are determined, and the technical feasibility of the overall project is also clarified.

3 Requirement Analysis

3.1 Functional Requirements

The users of precision marketing system based on data mining technology are managers of small and medium-sized enterprises. The adapted enterprise should have a local information management system to facilitate data import and assist managers in marketing

analysis. The main functional modules of the system are user login, data import, data management, data query, data analysis and visual report.

3.2 Overall Design

The precision marketing system based on data mining technology establishes a one-stop data collection, analysis, processing and visualization system from top to bottom according to the needs of enterprises. The main functions of data collection, data storage, data cleaning, data query and data analysis are supported by hadoop ecological cluster, and visualization is realized by javaweb technology [6]. The overall design of the data processing module function of the system is shown in Fig. 3. Firstly, the data is collected from three sources. One is the collection of local enterprise server data by flume, the second is the URL data collected from the product details page by python web crawler technology, and the last is the access to Taobao, Weibo and other shared data through external JDBC interface. These data will be preliminarily cached in HDFS distributed storage. The data of the crawler set is stored by redis. The data calculation module is implemented by mapreduce. Mapreduce analyzes the preliminary data and manages the crawler results, and uses K-means clustering algorithm to classify customers, and then combines the data mining technology of association rule algorithm to achieve the portrait of customers. The processed data will be saved in HDFS and hive.

The overall design of javaweb of this system chooses B/S mode and adopts MVC for development. The architecture is designed and developed by the traditional three-tier architecture of J2EE, which is the control layer, the business layer and the persistence layer. The business logic design of the core function of the whole system is developed by spring, the control layer is used to design the interactive function of client display, which is designed by springmvc, and the data persistence layer uses mybatis.

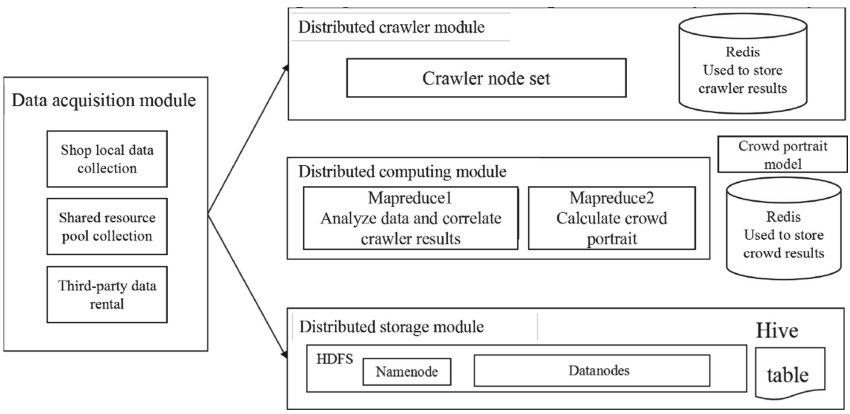

Fig. 3. System background function architecture diagram

4 Functional Implementation

In order to effectively distinguish high-quality customers from general customers and serve them with differentiated marketing strategies, this paper takes the customer data of an Internet e-commerce company before 2021 as the research object. Customer value needs to be calculated before clustering analysis of customers. The formula is $T = R_1 \times q + R_2 \times c + R_3 \times d + R_4 \times m$. Where T is the customer's value matrix, and (R_1, R_2, R_3, R_4) refers to the eigenvalue obtained by using the component analysis method. is the difference between the last shopping and the average number of days between shopping, q is the purchase quantity of customers in the latest period, m is the average consumption amount of customers in the latest period, and is the purchase conversion rate of customers in the latest period [7].

In this paper, gender, age, academic degree and marriage are selected as the customer segmentation criteria. Firstly, the modeling data is preprocessed, and the customer data of closed accounts, abnormal customer data and so on need to be deleted. After the above indicators are quantified, the bigger the indicator value, the more favorable it is for e-commerce enterprises. For example, in terms of age, customers under 25 years old are quantified as 4, 3 for 25–35, 2 for 36–50, and 1 for those over 50 years old.

The processed and cleaned quantized data are imported into the system, and the K-means clustering algorithm calculation process is realized by using script code. After the clustering process is completed, it is verified according to the customer value matrix of each category. Take the customer nicknamed blue_ram as an example to calculate the value of this customer. $V = 0.875 \times D + 1.101 \times Q + 1.024 \times M + 0.999 \times C = 0.875 \times 34 + 1.101 \times 1 + 1.024 \times 49 + 0.999 \times 0.0986 = 81.126$. The final clustering results are shown in Table 1 [8].

According to the above table, the final results are divided into four categories, and the number of customers in each category presents an inverted pyramid structure pattern, which is in line with the group characteristics of e-commerce customers. Customers can be divided into general customers, potential customers, key recovery customers and high-value customers as a whole. The classification of various customers is shown in Fig. 4. The first type of customers are between 23 and 28 years old, mainly with bachelor degree, mostly unmarried and mostly male. Their annual income is between 30,000 and 80,000, the consumption amount is mostly less than 1,000, and the number of transactions is about 2–3 times. Obviously, the first type of customers are mainly those college graduates who have just entered social work. Their income is not high and

Table 1. Cluster center results

	1	2	3	4
Gender	1.50	1.56	1.43	1.57
Age	3.64	3.04	3.65	2.63
Academic degree	3.36	3.68	3.53	2.48
Marriage	1.23	1.90	1.43	1.18

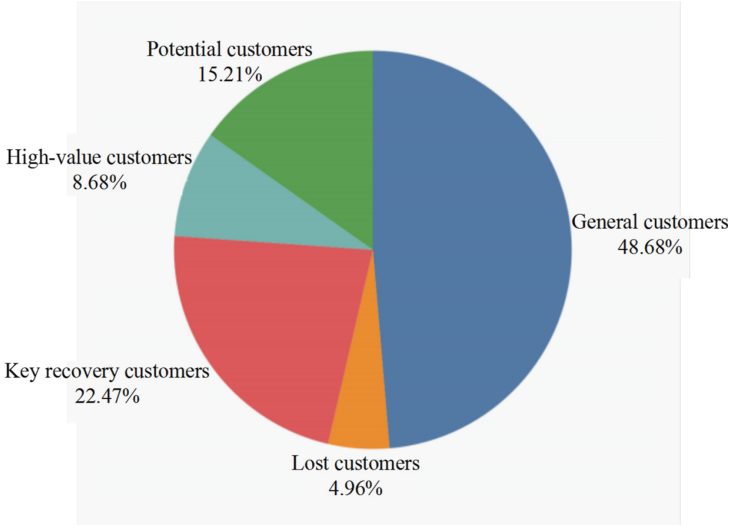

Fig. 4. Proportion distribution of various types of customers

their deposits are low, but they are keen on consumption. The initial growth customers who belong to e-commerce are also the general customers of this enterprise [9].

5 Conclusions

Faced with the challenges of technological development and market competition, enterprises in the Internet era should realize that data mining technology has a significant impact on the level of enterprise precision marketing. Paying attention to data mining of enterprise customers can effectively enhance the accuracy of enterprise marketing, and make it conform to the opportunities and challenges brought by market competition. Because the author's ability and energy are limited, there are still many shortcomings in the research of precision marketing system based on data mining technology, which need to be improved later. Enterprise marketing system needs to strengthen data security management, and the security of this system needs to be strengthened, especially to prevent internal leakage of customer information. For the external data obtained, there is a lack of audit mechanism for the legality of the data source.

References

1. Wu Liwen, Cai Shaolin. (2021) Research on Precision Marketing Path of Agricultural Products Based on Data Mining. Agricultural Economy.08.
2. Tang Huixiang, Chang Xiao, Song Laimin. (2020) Research on Taobao Precision Marketing Strategy Based on Data Mining. Natural Science Journal of Harbin Normal University.03.
3. Zheng Shufang. (2019) Research on the Application of K-MEANS Clustering Algorithm in the Precision Marketing of Communication Operators. Jilin University.11.

4. Liu Dinglong. (2017) Precision Marketing Strategy Based on Big Data Mining. Low Carbon World.05.
5. Yang Bo. (2018) Research on Precision Marketing Based on Big Data Mining. Xiamen University.09.
6. Yang Donghong, Shi Yingjian, Lei Ming, He Congxi. (2018) Correlation Analysis between Big Data and Enterprise Precision Marketing. Journal of Shenyang University of Technology(Social Science Edition).04.
7. Li Yang. (2022) Analysis and Research on Precision Marketing of Automobile Market Based on Association Rule Algorithm. Journal of Guiyang University.03.
8. Chen Yiyun. (2017) Application of Data Mining Technology in Precision Marketing. China Science and Technology Information.07.
9. Lin Qingpeng. (2016) Research on Precision Marketing Strategy Based on Big Data Mining. Lanzhou University of Technology.04.