Construction and Application of Precision Marketing System of E-commerce Platform under the Background of Big Data

Yuan Wang[1]

258812604@qq.com

[1]Jiangxi Vocational College of mechanical & Electrical Technology, Nanchang, Jiangxi, China

Abstract: In order to promote the vigorous development of China's Internet industry e-commerce market, help e-commerce enterprises to use big data technology to operate scientifically, so that e-commerce enterprises can achieve sustainable development, gain more profits and more room for growth, and build an accurate marketing system of e-commerce platform. This system uses Python web crawler technology to obtain information data, and carries out data analysis on Hadoop platform. Combined with tableau data visualization technology, it intuitively presents the data required by e-commerce enterprises. Big data precision marketing system can help e-commerce enterprise managers to judge customers' psychology and behavior, understand customers' needs and consumption preferences, and accurately analyze the traffic volume and transaction volume data generated by new and old consumers in different preferential means and different periods to improve user loyalty. So as to increase the user return rate, stabilize the existing user base, improve the marketing plan of enterprises, improve the sales performance of enterprises, and help enterprises obtain more ideal economic benefits.

Keywords: big data; e-commerce; precision marketing; Hadoop; network platform construction

1   Introduction

With the rapid development of Internet and Web technology, the e-commerce market has become increasingly prosperous, and the traditional production and life style of human beings has undergone earth-shaking changes. The convenient and fast shopping experience has made many consumers increasingly favor "online shopping", and online shopping has become an indispensable and important part of people's lives. However, with the increasing variety and quantity of goods on e-commerce websites, the information overload caused by online shopping provides people with rich and colorful product information, which leads users to get lost in a large number of commodity information spaces when choosing goods, which seriously restricts the further development of e-commerce, and users have higher and higher requirements for personalized information services [10]. At the same time, modern enterprises are used to adopting the traditional marketing mode, which can't meet the changing needs of the market

environment, so they can't get the corresponding information. Enterprises don't realize the importance of consumer data information, so they don't collect complete data information. As a result, their marketing strategies and market decisions are unreasonable, and they can't provide personalized marketing services, resulting in low consumer demand and consumer conversion rate. When the information technology level of human society and the research of computer science and technology reach a certain level, the continuous evolution of traditional Internet, mobile Internet and social media makes the data scale show a geometric surge. According to IDC Digital Universe Report, the amount of data generated by mobile devices, such as smart phones and tablets, has increased dramatically, which has doubled the scale of the digital universe in recent years. In 2020, it has reached 40ZB, with a per capita rate of over 5247GB, but only 0.4% of the global data has been analyzed. The increasing scale of data promotes the development of big data technology, and we have to admit that the "big data era" has arrived [3]. The huge data contains a lot of valuable information. If all walks of life do not standardize and effectively sort out these data as soon as possible, it will inevitably affect the formulation of future development decisions. In view of this situation, we should actively use big data technology to extract, sort out and mine data, and give full play to the role of data in the development of modern society. If an enterprise doesn't have enough data support in its development process, and its ability in data analysis is insufficient, it will inevitably lead to a lack of persuasiveness in its development decisions and a deviation in its development direction. Similarly, the enterprise managers of e-commerce platform should integrate and analyze a large amount of data for the precise marketing plan, judge the psychology and behavior of customers, understand their needs and preferences, improve the current enterprise marketing strategy, and promote the economic benefits of enterprises. In marketing management activities, if an enterprise lacks data and information support, it can't accurately understand and predict the market development, and it can't make a scientific and reasonable marketing plan. Although many enterprises have mastered the use of big data technology to achieve their marketing goals, they only stand in their own marketing perspective, and do not have an in-depth understanding of customers' needs and consumption intentions for accurate analysis. Blindly placing large-scale advertisements will make users feel resentful because of these invalid information, which will not only cause users' troubles, but also bring adverse effects on the brand image of the enterprise, bring adverse effects on the marketing effectiveness of the enterprise, and make a large number of potential consumers lose. Therefore, how to lock the target consumers of e-commerce websites and provide differentiated services to improve user satisfaction has become a top priority for e-commerce website merchants [5].

The author believes that according to the above analysis, in order to meet the needs of today's e-commerce platform enterprise managers, an e-commerce precision marketing system based on Hadoop platform web technology and web crawler technology came into being under the background of big data. This e-commerce precision marketing system obtains consumers' interest preferences by analyzing users' behaviors, which is a powerful way to solve the problem of information overload, effectively improving the marketing accuracy of merchants and reducing marketing costs.

2   Key technology

2.1 Python crawler technology

Web crawler is a program that automatically crawls the content information of web pages on the Internet according to certain crawling rules. Python is the most commonly used language. For the data on the Internet, it will be a huge workload if we collect this massive amount of information manually. In order to solve this problem, the web crawler came into being. When the crawler users set the crawling rules and the URL of the crawling portal, the web crawler will automatically start crawling and save the captured information. The basic workflow of web crawler is as follows: First, select a part of the seed URL. Put these URLs into the URL queue to be crawled. Take out the URL from the URL queue to be grabbed, download the webpage content corresponding to the URL, and then store the content data. In addition, these URLs are put into the crawled URL queue. Analyze the URL in the crawled URL queue. Extract a new URL and put it in the URL queue to be grabbed. At present, there are many stable crawler frameworks, so we don't manually implement the crawler's bottom framework. The main work focuses on the realization of the crawler's upper crawling logic and the construction of distributed crawler. In this paper, Scrapy, an open source crawler framework, is chosen, because Scrapy crawler framework developed earlier, is relatively stable and has perfect community support. As shown in Figure 1, it is a schematic diagram of the network crawler [6].
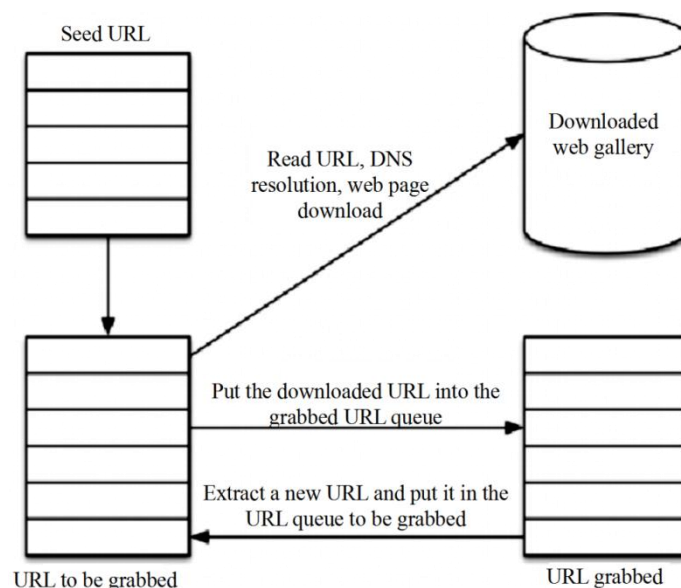

Figure 1: Schematic diagram of the network crawler

2.2 Hadoop processing platform

As an open source project of Apache Software Foundation, Hadoop was founded by Doug Cutting in 2004 based on the research of three academic papers published by Google. It is written in Java language with the characteristics of distributed, platform independence and

portability, and can well support data-intensive applications. Hadoop belongs to the form of Platform as a Service. So far, Hadoop has developed many subprojects including Ambari, HDFS, MapReduce, Hive, HBase, Pig, Zookeeper, Flume, Mahout, etc. The two most important core technologies are distributed file system HDFS and parallel programming model MapReduce. HDFS, as a file system management storage, is used to store large-scale data sets, and MapReduce, as a computing engine, is used to compute large-scale data sets in parallel on distributed systems. On the premise of lacking the basic knowledge of distributed processing software framework, users can conveniently and quickly make full use of the advantages of clusters on Hadoop distributed computing platform to develop and run distributed applications that can process large-scale data sets, and perform high-speed operation and storage. Therefore, Hadoop has good scalability, high efficiency, high reliability and high fault tolerance. Hadoop platform structure diagram is shown in Figure 2 [4].
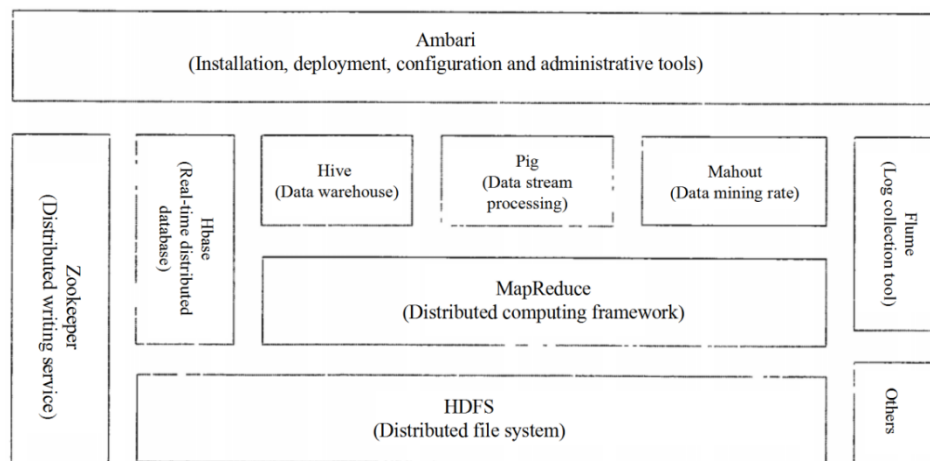

Figure 2: Structure diagram of the Hadoop platform

2.3 Tableau data visualization

Tableau is a commonly used data visualization software in the field of big data, which can flexibly and quickly generate a variety of graphics based on massive data, intelligently transform a large amount of data into interactive charts, and visualize conveniently and quickly. Tableau can quickly extract data by using convenient data interface, which is compatible with many mainstream databases (sql, aws, hadoop, SAP HANA), and even can directly connect sales reports and Google analytics traffic analysis. Starting from version 10, the function of connecting json files has been added. Tableau realizes the function of updating all reports from metadata regularly, and can realize cross-platform visualization among web pages, mobile phones and tablet. In addition, Tableau can easily add users and set user groups in user management, and can automatically complete user management in termainal through tabcmd [7].

2.4 Development environment

The implementation of the system is based on Linux system, and the operating system used is CentOS 8.2. Two programming languages are used, one is Java1.8, which is mainly used for the

development of log collection module and distributed computing module, and the other is Python 3.8, which is mainly used for the development of distributed crawler module. Hadoop version 3.3.1, Hive version 3.1.2. The version of Scrapy is 2.5, and the version of Redis is 3.2.9. Tableau Dasktop 2021.4 is used to visualize the data. Use the remote connection server function to connect to Cloudera Hadoop on Tableau. Finally, the visualization results are displayed in JAVA language by using the Web technology on the Eclips framework. Figure 3 shows the execution command code for creating a new Hive directory on Hadoop cluster platform.

```
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 -ls /tmp/hive
```

Figure 3: Execution command code of new Hive on Hadoop

3   Requirement analysis

3.1 Functional requirement

The precision marketing system of e-commerce platform based on the background of big data can help e-commerce companies to establish their own data centers and collect omni-channel customer data. Through intermediate technology, information from different sources and different communication protocols are completely integrated, so that different types and formats of digital resources can be seamlessly connected. And using data management to review data integrity, and using tableau to design interactive charts, data visualization can be carried out conveniently and quickly, which is convenient for e-commerce enterprise marketing managers to analyze more intuitively. By planning marketing activities, data can be gradually improved and supplemented. In order to make the e-commerce enterprise managers intuitively understand the consumer data that is beneficial to marketing, this system classifies the consumer data carefully, and designs four modules: consumer information data, consumption habit data, consumer consumption frequency and amount data, and store access data. The basic static information and dynamic behavior data of consumers determine the key consumer groups and the range of potential customers. According to the length of store page stay, the number of visits to e-commerce stores and the record of the length of time from visit to transaction, observe the initial results brought by different application marketing schemes and make timely adjustments. According to the recorded data of consumer buy-back rate, buy-back period and accumulated consumption amount, a membership system is set up to give preferential treatment, thus increasing the brand stickiness and loyalty of consumers. According to the distribution of consumption period and preference of consumption discount, the priority of promotion stimulus and the sales price of commodities are adjusted in stages by choosing the most suitable preferential way for store marketing. Business managers can effectively help e-commerce

enterprises to improve their user loyalty, increase the return rate of users, improve their sales performance, stabilize the existing user base, and greatly reduce their operating costs according to the relevant types of data of consumers provided by e-commerce precision marketing system, thus promoting the healthy and rapid development of e-commerce enterprises in the fierce market competition with increasingly serious commodity homogenization and increasingly complex information [2].

3.2 Global design

The precision marketing system of e-commerce platform has established a one-stop data analysis system driven by data from top to bottom. The background of the system needs to include data collection, data storage, data cleaning, data query, data analysis and data modeling functions. Among them, using Hadoop for a series of operations is the core technology of this system, so the author of this paper describes the overall architecture of Hadoop's data processing function in detail. The system architecture is divided into four modules according to functions, namely, data acquisition module, distributed storage module, distributed crawler module and distributed computing module. Data acquisition module, which deploys a data acquisition client on the enterprise store server. The data acquisition client automatically collects all kinds of internal business product development data, product sales data, channel data, customer basic information data, and all-round and multi-dimensional user behavior data and logs acquired by Taobao, Weibo, WeChat and other mobile Internet shared databases and data rented by third-party data centers. The client needs to realize the function of sending data and original logs to HDFS, and then to distributed crawler module. The distributed crawler module captures the received data. In the distributed storage module, the original files will be written into the specified directory of HDFS and stored as files. Hive payload stores parsed logs in the form of tables. Distributed computing mainly realizes two functions: First, it analyzes the original log file, correlates the crawler results, and saves the analysis results into Hive. The second is to correlate the analyzed results with the crowd portrait model, analyze the crowd portrait, and save the results into Redis. The system architecture diagram is shown in the following Figure 4 [8].
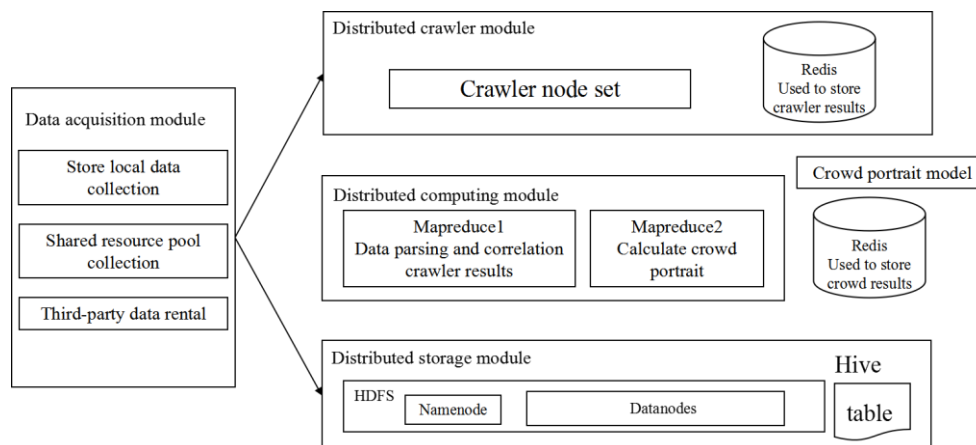

Figure 4: System background function architecture diagram

4 Function realization

In order to make the managers of e-commerce enterprises intuitively understand the consumer data beneficial to marketing, this system classifies the consumer data carefully, and divides it into four modules: consumer information data, consumption habit data, consumer consumption frequency and accumulated consumption amount data, and store visit records.

4.1 Consumer information data

This part of data includes basic static information of consumers such as age, gender, hobbies, educational background, marital status, and location, as well as dynamic behavior data generated by consumers every day, such as which videos they watch, which websites they visit, which things they buy, which places they have been to, and so on. Based on these data, the data system will build a data model through Hadoop to build portraits of consumer groups. The crowd portrait model based on consumer information data can help enterprise marketing managers accurately determine the key consumer groups and the range of potential customers of products according to the characteristics of consumer groups, so that e-commerce enterprise managers can better understand consumers and deeply consider consumers' needs from the perspective of consumers. E-commerce managers can therefore strengthen the propaganda investment of key consumer groups in promotion and decide the style direction of marketing activities [1].

4.2 Store visit record

This part of the data comes from the log records of e-commerce platform operation, including the duration of store page stay, the number of visits to e-commerce stores, and the duration from visit to transaction. Store visits also include new customer visits, old customer visits and their proportion in pie chart. E-commerce managers can deal with the marketing strategy priority of developing new consumers and maintaining old consumers by knowing the data related to store visits and comparing the transaction volume data, observe the initial results brought by different application marketing schemes, make timely adjustments, and improve the profit conversion rate of stores.

4.3 Consumption frequency and accumulated consumption amount

This part of the data includes all the transaction records, repurchase rate, repurchase cycle and accumulated consumption data of the group consumers of the business. Consumers with high repurchase frequency and accumulated consumption amount reflect their high brand dependence on the business. Managers use this part of the data to set up a membership hierarchy, and set discounts according to the level, so as to increase the brand stickiness and loyalty of this part of consumers [9].

4.4 Consumption habit

Click on the consumption habit module to see the consumption time distribution sub-module and the consumption discount preference sub-module. The consumption period module includes data of purchasing frequency of various commodities by consumers in different seasons, months and other stages. Enterprise managers can know the peak period of product sales by looking at this part of data, and adjust the priority of promotion stimulus and the sales price of goods by

stages. The data of consumer discount preference module includes the transaction volume data of various preferential means such as full discount, full reduction and full quantity discount by consumers on major e-commerce platforms. E-commerce enterprise managers can summarize the discount preference of consumers and set the most suitable preferential way for store marketing.

5   Conclusions

This paper takes the precision marketing of e-commerce platform under the background of big data era as the research object, combs the related concepts of big data, data acquisition, data calculation, data modeling, data visualization, web technology, web crawler technology, Hadoop data calculation technology and other technical methods, and summarizes the research status of precision marketing in China. On this basis, it puts forward the research of precision marketing system based on big data. By studying the relationship between big data and precision marketing, this paper tries to find the breakthrough point of the application of big data in precision marketing of e-commerce platform, and then draws the precision marketing strategy based on big data according to the breakthrough point. Big data is the entry point for precision marketing applications, and target customers are selected according to consumer behavior information to help e-commerce platform enterprises push accurate information. And refine e-commerce selection management according to market size and market structure.

Based on the background of big data era, this paper studies precision marketing, which is only qualitative analysis, but lacks quantitative analysis. This is the defect of the paper. At the same time, the paper is not specific to a certain industry, so although it has certain applicability to various industries, the specific situation of different industries is different, and its application strategies are also different. Therefore, I hope that the future research can be based on a certain industry, making the research more in-depth and specific.

References

[1] Cao Gang. (2019) Research on Precision Marketing Strategy Based on Big Data. Beijing University of Posts and Telecommunications.03.

[2] Hu Heping. (2022) Optimization Strategy of E-commerce Marketing Management under the Background of Big Data. Time-honored Brand Marketing.04.

[3] Ji Xiaoyan. (2017) Research and Implementation of E-commerce Personalized Recommendation Based on Hadoop. Lanzhou Jiaotong University.06.

[4] Lin Qingpeng. (2016) Research on Precision Marketing Strategy Based on Big Data Mining.. Lanzhou University of Technology.04.

[5] Li Zhang. (2017) Design and Implementation of Internet Data Marketing System Based on Hadoop. University of Chinese Academy of Sciences.04.

[6] Ma Xiaohong. (2022) Application and Value Overview of Big Data Precision Marketing in E-commerce. E-commerce.01.

[7] Qian Haodong, Su Youhui. (2021) Design and Implementation of Big Data Visualization of COVID-19 Epidemic in Europe. Computer Era.09.

[8]   Wu Wei. (2016) Research on Precision Marketing of Mobile Internet Advertising in Big Data Era. Southwest University.05.

[9]   Yang Xingchen. (2020) Enterprise Customer Relationship Solution Based on Big Data.Science & Technology Information.08.

[10] Zhang Wenhui. (2020) Research on BJ Company's Precision Marketing Strategy under the Background of Big Data. Hebei GEO University.12.