Design and Application of Cross-border E-commerce Selection Platform for Data Visualization under Big Data Technology

Na Wang
Xi'an University of Science and Technology, Xi'an, Shanxi Province, 710054

sunmayicheng@163.com

Abstract. This paper completes the design and implementation of a cross-border e-commerce selection platform for data visualization under big data technology. Hadoop ecosystem and Spark computing framework are used as ETL data processing engines of big data platform, and the development language of application system is Java, and SSM is selected as the development framework. This paper focuses on the collaborative filtering recommendation algorithm based on items to help users determine their choices. This system combines big data technology with echarts visual analysis set to summarize data information and generate data reports, which is convenient for users to make product selection decisions quickly and accelerate the development of e-commerce platform.

Keywords: big data technology; hadoop; Data visualization; E-commerce selection

1 Introduction

There are many cross-border e-commerce platforms in China, but the degree of development is different, so there is a big gap in the research on the application effect of big data technology. Due to the serious data asymmetry among e-commerce companies, it is difficult to deeply explore the market environment, which affects the accuracy of product selection and is not conducive to product sales. The selection of cross-border e-commerce operation is critical, and the combination of big data mining technology and visualization technology can realize the rapid selection of massive commodities. Therefore, the author of this paper believes that a cross-border e-commerce selection platform for data visualization under big data technology should be developed. It is hoped to effectively reduce the operating cost of enterprise big data technology system and improve the overall competitiveness of enterprises in the market. [1]

2 Key technologies

2.1 Hadoop

Hadoop is a distributed computing development software with high reliability and scalability based on big data technology. Hadoop has many advantages, such as high computing speed, low cost, high expansibility and enhanced reliability, through its simplified programming model and the way of cross-computer clusters. Hadoop technology consists of two core technologies, namely HDFS big data file storage and mapreduce parallel computing. Hadoop eco-

MSIEID 2022, December 09-11, Chongqing, People's Republic of China

DOI 10.4108/eai.9-12-2022.2327617

system is composed of core technology components and other key components. Other service subprojects include Hive data warehouse, mahout machine learning library, database HBase, coordination service zookeeper and resource manager Yarn. [2]

2.2 Spark

The operation mechanism of Spark calculation engine is shown in Figure 1. After the task is submitted, it is necessary to initialize the application environment, apply for the computing resources used by the Executor, and run the SparkContext program at the same time. Among them, the Executor is deployed in each node, and the return status of the Executor is performed by heartbeat. The mechanism of SparkContext needs to construct a DAG acyclic graph by combining RDD objects. The ring graph of DAG is split into several separate stage by using the scheduler of DAG and using two types of boundaries of wide dependence and narrow dependence. At the same time, all tasks of stage will be repackaged into a new Tastset, and the task scheduler will be used to execute the tastset, tastset. When all task and stage are executed, the processing of spark will be completed, and the system will release computing resources to end the process. [3]

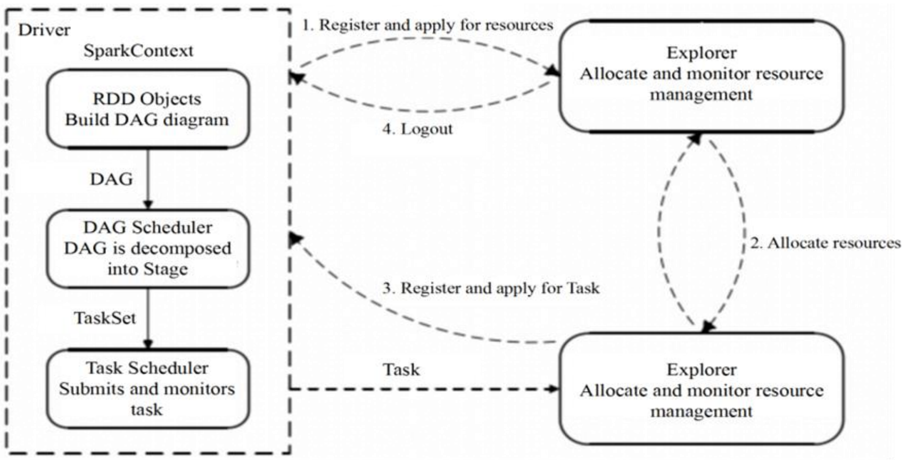

Fig. 1. Spark operation mechanism

2.3 Development environment

The basic experimental environment of the visual cross-border e-commerce selection platform is divided into two parts: one is the cluster construction of big data ETL processing platform, and the other is the construction of web application system. Hadoop version used in this paper is 2.7, and spark cluster uses 2.4.4. On this basis, the shufle module code is reconstructed and packaged as jar by maven tool to replace and deploy hadoop and spark cluster code. According to the data volume of cross-border e-commerce selection data, this experiment chooses to build a cluster composed of six nodes, one of which is the master node, and the master of spark and namenode of hadoop are deployed, and the other five are the slave nodes of the cluster, which are both datanode of hadoop and worker of spark cluster. The cluster construction scheme is heterogeneous, and Scala is 2.11.12. Operating system is linux cent OS7. [4]

3 Requirements analysis

The overall architecture of the system is based on B/S structure, which is divided into front-end interface layer, business logic layer and data layer. The front-end is responsible for the display of information database browsing, data visualization report, data search engine and other functions. In addition, there is an administrator-side admin, which is responsible for the screening and adjustment of merchants, stores and category information. The business layer uses SSM framework to connect the front-end page, realizes the business logic of various functions in the system, and completes the data reading and storage of the database in the process of system access. The data layer uses mysql to store the data of the business logic layer and uses Elastic-Search to complete the construction of the search engine. The offline information is stored in HDFS in hadoop, and the algorithm training selects spark-Mllib in Spark. [5]

4 Functional implementation

After logging in, users can see the main functional modules including sales database, product information, data analysis and visualization. The important indicators of e-commerce product selection analysis report are the selection of consumer groups with high consumption frequency, products with high psychological urgency of consumers, product attributes, price comparison, etc. Products with high psychological urgency of consumers can be judged according to the sales volume of classified users in different seasons, regions, and consumer age groups. This system takes Company B as an example, and its visual report of temperature and sales volume analysis is shown in Figure 2. According to the chart, the system can help users analyze the distribution of selected commodities in different regions and the selection of commodities in different sales periods.

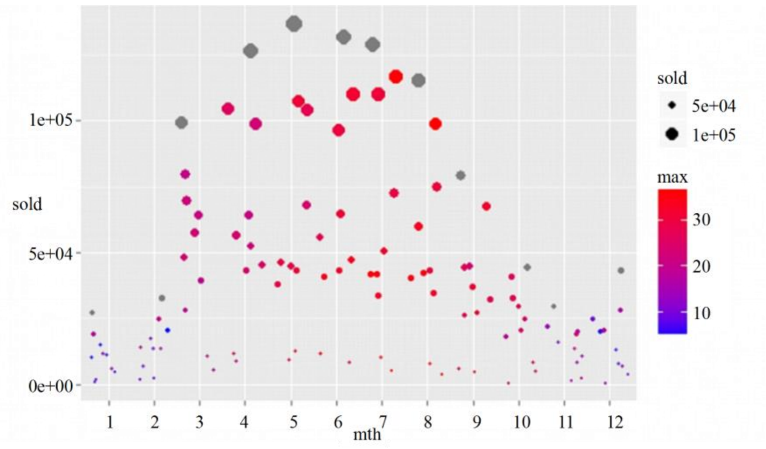

Fig. 2. Temperature and sales volume analysis chart

According to the existing data, the internal relationship between commodities and consumer groups is found to determine the appropriate products. This part of the function adopts collaborative filtering recommendation algorithm. Starting with the historical purchasing behavior data of existing consumer groups, we can find similar users and commodity sets according to the information, and calculate more suitable products. The core function of the data cross-border e-commerce selection platform is to help select products. Therefore, the main algorithm is the project-based collaborative filtering recommendation algorithm. The specific formula is shown in Formula 1. Among them, $\underset{i}{\rightarrow}$, $\underset{j}{\rightarrow}$ are product scoring vectors, U(i,j) is a collection of users who have browsed, purchased and scored i, j in historical behavior, and r_{ui}, r_{uj} represent the evaluation scores made by user u on product i, j.

$$\cos(i,j)=\frac{\vec{i}\,\vec{j}}{\left|\vec{i}\right|\left|\vec{j}\right|}=\frac{\sum_{u\in U(i,j)} r_{ui}^2 \cdot r_{uj}^2}{\sqrt{\sum_{u\in U(i,j)} r_{ui}^2}\sqrt{\sum_{u\in U(i,j)} r_{uj}^2}} \tag{1}$$

The main method is to use cosine similarity formula to calculate the similarity of two commodity data sets. Then, according to the consumer's historical consumption behavior, the top N recommendations are made to determine the most appropriate related products. This algorithm is suitable for the case where the attributes and characteristic data of the existing products of e-commerce enterprises are clear.

In the process of system operation, new product information will be put into storage, and timely data update should be carried out in commodity search engines to avoid the consequences of decision-making mistakes caused by data missing. Therefore, this paper uses canel tool to synchronize the commodity information entered in relational database mysql and hadoop cluster. Canel regards itself as one of the mysql nodes by simulating the mysql interactive protocol, so that it can send a request to the master node, so that Canel can get the Bin_ log object according to a fixed time, and finally persist it in the index list of the commodity library. [6]

After testing, the average click-through rate of products on the shop pages of Company B is 3%, while the click-through rate of selected products selected by using the data visualization cross-border e-commerce selection platform studied in this paper is 8%, which increases by 5% under the same business process mode. The overall sales conversion rate of cross-border B2C e-commerce company B is 10% in 2021, while the average conversion rate of products under the operation mode guided by this system is 26%, a substantial increase of 16 numerical percentage points. Experiments show that the cross-border e-commerce product selection platform studied in this paper can effectively help cross-border e-commerce companies greatly improve their marketing efficiency.

5 Conclusion

The application of big data in cross-border e-commerce selection is at an initial stage in this paper, and there is still a long way to go for in-depth research on big data. The work of other parts in the knowledge supply chain of commodity selection in the process of e-commerce sales also plays an important role in the sales performance of e-commerce. However, due to the limited space of this paper and the limited ability and energy of the author, only the detailed design and implementation of the product selection link can be made.

References

[1] Diao Shengfu, Feng Liru. Application of Big Data in Digital Transformation of Cross-border E-commerce. China Statistics.2020.05.

[2] Liu Yuanyuan. Discussion on the Selection Strategy of Cross-border E-commerce Platform under Big Data. Business Information.2020.07.

[3] Wu Yuhao. Big Data Helps Precision Marketing: A Study on the Selection Logic of E-commerce Live Broadcasting. News Culture Construction 2021.2.

[4] Yi Meilian. A Summary of Cross-border Electronic Commerce's Product Selection Research. Market Modernization.2021.06.

[5] Deng Zhichao. Analysis Strategy of Cross-border E-commerce Platform Selection Based on Big Data. Special Zone Economy.2019.06.

[6] Cheng Jie. Research on Commodity Recommendation Algorithm Based on Spark Big Data Platform. Heilongjiang University.2020.05.