Retraction

Retraction: Research on E-commerce Recommendation System Based on Big Data Technology (J. Phys.: Conf. Ser. 1883 012159)

Published 9 September 2022

This article has been retracted by IOP Publishing following an allegation that raises concerns this article may have been created, manipulated, and/or sold by a commercial entity. In addition, IOP Publishing has seen no evidence that reliable peer review was conducted on this article, despite the clear standards expected of and communicated to conference organisers.

The authors of the article have been given opportunity to present evidence that they were the original and genuine creators of the work, however at the time of publication of this notice, IOP Publishing has not received any response. IOP Publishing has analysed the article and agrees there are enough indicators to cause serious doubts over the legitimacy of the work and agree this article should be retracted. The authors are encouraged to contact IOP Publishing Limited if they have any comments on this retraction.

Retraction published: 9 September 2022
Research on E-commerce Recommendation System Based on
Big Data Technology

Zhenjie Wan*
Huizhou Economics and Polytechnic College, Huizhou, 516057, Guangzhou, China
*Corresponding author’s e-mail: wanzhenjie@hzcollege.com

Abstract. With the continuous development of e-commerce and the increasing number of users, more and more scholars have devoted themselves to the research of recommendation algorithms for e-commerce platforms. How to quickly dig out information or products that users are interested in from massive data is a research category in the field of recommendation systems. The emergence of the Spark memory computing platform can provide technical support for improving the efficiency and real-time performance of the recommendation algorithm. This article first analyzes the requirements of the e-commerce recommendation system, and designs the overall architecture of the system. Through the analysis of the system, a recommendation system that conforms to the e-commerce system is designed, and a stream computing framework is used to implement a recommendation system that can meet offline and online recommendations combined recommendation system.

1. Introduction
The number of online shopping users is increasing with the improvement of the online shopping environment, which makes more and more scholars invest in the research of recommendation systems for e-commerce platforms. E-commerce recommendation systems can provide users of e-commerce websites with an intelligent and personalized shopping experience, making users tend to buy more goods, thereby improving user satisfaction [1]. Major e-commerce companies have developed recommendation systems based on users' historical behavior records to recommend similar products that they have bought before or products of potential interest to meet users' needs and increase their desire to buy. Although certain results have been achieved in the field of recommendation systems for e-commerce platforms, there is still room for improvement [2]. There are still some remaining issues that need to be resolved urgently, such as data sparsity issues, real-time issues of recommendation algorithms, and user interest transfer issues. In view of some of the problems existing in the above e-commerce platform user recommendation methods, this article conducts an in-depth study. The article studies the direction that can be optimized and proposes an improved recommendation algorithm. From the results, the research method in this paper improves the data sparsity and cold start problem to a certain extent, and improves the accuracy of recommendation [3].

2. Spark architecture and recommendation system analysis

2.1. Spark architecture
The Spark architecture is mainly analyzed from its operating architecture. After the system application is deployed, a series of services will be started for program operation scheduling [4]. The specific processes and service roles in the operation process are as follows:
• Driver: The client driver, which can be understood as a client APP, is used to convert submitted tasks into RDD and DAG, and communicate and schedule with the Cluster Manager;

• Cluster Manager: The Spark cluster manager is responsible for system resource allocation and management. It allocates resources such as memory and CPU on each Worker. It is not responsible for the resource allocation of each Executor in the Worker. Currently, the cluster managers supported by Spark include: Standalone, YARN, Mesos, EC2, etc.;

• Worker: The task execution node is responsible for creating an Executor, further assigning resources to the Executor, and synchronizing resource information to the Cluster Manager;

• Executor: The process of executing tasks, responsible for task execution and keeping information synchronized with Workers and Drivers.

The relationship between each part is shown in Figure 1.

Figure 1. Spark running architecture diagram

2.2. Recommendation system analysis
After more than 20 years of development, the recommendation system has achieved certain results. During these 20 years, many researchers have adapted to the needs of different fields. The recommendation system has been widely used in many fields such as e-commerce, news, geographical location and so on. For example, in the field of e-commerce, the function of the recommendation system is to recommend corresponding items to users based on existing information, such as item information, user information, and user behavior information [5]. The purpose of personalized recommendation service is to present valuable product information to consumers, thereby increasing the turnover of the platform. The main content of personalized recommendation service is to actively send information about these products or services that have not been discovered by consumers to consumers. Stimulate consumers' desire to buy, and finally realize the purpose of transaction. Customer groups have different interests. Strengthen the product coverage of the personalized recommendation system of e-commerce companies to ensure that the recommended products are recognized by the recommended customers and bring higher-quality customer group experience to the recommended customers [6]. The organization chart of the e-commerce personalized recommendation system is shown in Figure 2.
3. E-commerce recommendation system requirements analysis and overall architecture

3.1. E-commerce recommendation system demand analysis

3.1.1. Functional Requirements. According to the division of users and system functions, the main functions of users of this system include: users can log in to the system by entering a user name and password; provide registration function for new users, the system provides unique identification for each user, some basic information is stored in it; after the user enters the homepage, he can see the products recommended by the system for the user, presented in the form of a list, and the user can click on any product to browse the details according to his preference; in the product detail page, users can score or tag the product [7].

The following functions need to be realized inside the system: data storage function module. The system processes massive scoring and tag data through the big data platform, and completes offline recommendation display. The algorithm module starts after the system receives the data, and updates the recommendation results in real time. When a new user logs in for the first time, the system can only provide non-personalized product recommendations for the user.

3.1.2. Performance requirements. The most important thing in a quotient recommendation system is the accuracy and timeliness of the data. Since the system needs to process massive amounts of data and is accompanied by a mixture of multiple recommendation algorithms, the recommendation system has extremely high requirements for data accuracy and fast response time. It is embodied in the following aspects: First, the system needs to have good stability. While it can ensure the long-term uninterrupted operation of the system, it can quickly recover from server disorder caused by certain emergencies. This is also a distributed architecture [8]. Secondly, the system runs faster than under normal circumstances, which is the basic requirement for a big data computing platform. Finally, the system should have high fault tolerance, use redundant data storage methods, automatically save multiple copies of data, and be able to automatically redistribute failed tasks.

3.1.3. WEB interface requirements. As an intermediate medium between the user and the system, the interface provides convenience for user operations. The user only needs to get the recommended information through the familiar interface, and does not need to understand the complex operating logic of the system. This article has the following requirements for the design of the system's operation interface:

- The operation page is clear, the functions are easy to find, and the recommended products should not be too many.
- The system page mainly includes the user login registration page, recommendation page and product detail page. The functional modules displayed in it should be as clear and layered as possible.
• The content of the page should encourage users to do more operations such as scoring and tagging to facilitate the collection of information for later recommendations.

The design flowchart of user registration and login module is shown in Figure 3.

![User registration and login flowchart](image)

Figure 3. User registration and login flowchart

3.2. Overall architecture design of e-commerce recommendation system

The real-time recommendation system based on Spark in this paper is mainly divided into two parts. The architecture diagram is shown in Figure 4. The offline computing part and the real-time online computing part. The system is an architecture that can combine offline and online computing. It can solve some of the real-time requirements of the recommendation system, and it can also make recommendations using batches of offline data [9]. Since the computing platform used is the Spark memory computing framework, it is faster and more versatile than Hadoop, and has a variety of RDD programming models to facilitate development. One advantage compared to Storm is that it can perform offline query of batch data and compare with Combine other components.

![Architecture diagram of e-commerce recommendation system](image)

Figure 4. Architecture diagram of e-commerce recommendation system
The whole system consists of two subsystems: Offline recommendation and online recommendation. It can be seen from the data flow that both subsystems are performed in accordance with the steps of data input, intermediate processing, and result data output. In order to increase the real-time performance of the system, high-performance designs are used from these three levels. In addition, online and offline are performed under the Spark platform, so the interaction between online and offline modules can also be completed through Spark.

3.2.1. Offline system architecture design. The offline recommendation part mainly completes the training process of the offline model, and the training data in the offline mode is massive. Therefore, its model training takes a long time, but these offline data often have a good effect on analyzing user preferences and giving recommendations. According to the architecture diagram, we can know that the modules in the offline architecture are mainly divided into the following three parts:

- Data warehouse part. In a big data environment, the establishment of a high-performance data warehouse can meet the data query and loading requirements under the recommendation system. This part mainly uses HDFS, Parquet and other technologies to package and store data.
- Recommendation engine part. The recommendation engine part mainly uses various recommendation mechanisms to train the recommendation model. When the recommendation engine works, it accepts the input source at the input, and then trains the recommendation model based on the input data to predict the user’s preferences, so the recommendation engine can recommend its possibilities to the user interested items and content.
- The result is output. This part mainly outputs and stores the recommendation results obtained by the recommendation engine, and stores them in the HBase high-performance database. It is recommended to query related recommendation lists directly from HBase.

3.2.2. Online system architecture design. The online recommendation part is mainly for recommendation in a real-time dynamic data environment, so it is necessary to receive and send the data stream in real time, and be able to process the calculation of the data stream. Faced with the dynamic update of data, the recommendation model can be dynamically updated in time to accurately generate real-time recommendations to users. According to the architecture diagram, online recommendations can be divided into the following design parts:

- Kafka real-time data stream part. The Kafka system mainly implements real-time data publish-subscribe, which is divided into three roles: producer, broker, and consumer. The producer implements the simulation of real-time scoring data, the broker stores the messages from the producer, and finally uses the consumer to obtain real-time streaming of the messages.
- Spark Streaming stream computing part. Spark Streaming receives data from Kafka data sources and performs stream calculations, and uses dstream to complete real-time analysis of real-time data and give real-time recommendations. When performing stream calculations, the model can be updated according to the update cycle. At the same time, operations such as storage and filtering of stream data can also be implemented.
- Stream data output part. This part mainly stores real-time data, uses the existing data interface to read and write the stream data to the database. At the same time, it also needs to transmit the generated stream data to the data warehouse for the collection of historical data and the next use.

4. Conclusion
The accuracy of the personalized recommendation service will have a direct effect on the sustainability of the service. E-commerce companies should send useful data and information around the needs of users. Further mining of the potential needs of consumers can ensure the efficiency of content recommendation services and help e-commerce companies to tap more potential needs of consumers. For e-commerce platforms, it can ensure the accuracy of product keywords, improve the ordering rules of product content, and can deeply discover the personalized connection between consumers and products. This System uses content association rules to filter valuable content, and
intelligent analysis algorithms such as consumer sentiment analysis can introduce the data into the personalized content recommendation function after computing the data. This operation is mainly through data mining technology, in order to realize the accurate analysis of user needs of the e-commerce platform and provide personalized recommendation service functions for them.

References
[1] Hu X S 2021 Impact of pushing commodity information online on consumers’ purchase intention under the background of big data. Business Economics Research, (02):64-66.
[2] Kuang W B 2021 Logical principles, ethical issues and regulatory strategies of intelligent algorithm recommendation technology. Journal of Shenzhen University (Humanities and Social Sciences Edition), 38(01):144-151.
[3] Shen Y M 2021 Innovation of e-commerce service model in the era of big data[J]. Trade Fair Economy, (01):23-25.
[4] Xie B, Peng X and Yin G, et al. 2021 The method and environment of intelligent software development based on big data. Big Data, 7(01):3-21.
[5] Liang J F 2020 Research on application of product personalized recommendation algorithm based on collaborative filtering. Journal of Hebei Software Vocational and Technical College, 22(04):22-25+41.
[6] Tao Q F 2020 Research on e-commerce personalized recommendation system based on Apriori algorithm. Journal of Chongqing University of Science and Technology (Natural Science Edition), 22(06):62-64.
[7] Guo J 2020 Analysis of e-commerce data analysis based on machine learning. Communication World, 27(05):192+211.
[8] Chen X, Wang B and Zeng F Q. 2020 Review of research on commodity matching big data recommendation methods. Computer Engineering and Science, 42(01):36-45.
[9] Chen J 2019 Research on marketing strategy of e-commerce platform based on big data technology. Fujian Tea, 41(11):18.