Computer Science and Information Systems 18(4):1359–1378 https://doi.org/10.2298/CSIS200917026Z

The Application of E-commerce Recommendation System in Smart Cities based on Big Data and Cloud Computing

Yiman Zhang

College of Oujiang, Wenzhou University, Wenzhou, 325035, China
00194012@wzu.edu.cn

Abstract. In the era of big data, the amount of Internet data is growing explosively. How to quickly obtain valuable information from massive data has become a challenging task. To effectively solve the problems faced by recommendation technology, such as data sparsity, scalability, and real-time recommendation, a personalized recommendation algorithm for e-commerce based on Hadoop is designed aiming at the problems in collaborative filtering recommendation algorithm. Hadoop cloud computing platform has powerful computing and storage capabilities, which are used to improve the collaborative filtering recommendation algorithm based on project, and establish a comprehensive evaluation system. The effectiveness of the proposed personalized recommendation algorithm is further verified through the analysis and comparison with some traditional collaborative filtering algorithms. The experimental results show that the e-commerce system based on cloud computing technology effectively improves the support of various recommendation algorithms in the system environment; the algorithm has good scalability and recommendation efficiency in the distributed cluster, and the recommendation accuracy is also improved, which can improve the sparsity, scalability and real-time problems in e-commerce personalized recommendation. This study greatly improves the recommendation performance of e-commerce, effectively solves the shortcomings of the current recommendation algorithm, and further promotes the personalized development of e-commerce.

Keywords: e-commerce, personalized recommendation, cloud computing, big data

1. Introduction

Under the background of big data era, the development of e-commerce is relatively rapid, and the trading volume in this field shows geometric growth [1]. Online shopping has become an indispensable part of people's life. Due to the increasing variety and quantity of goods on e-commerce websites, when a wide range of goods are provided, it provides users with more choices, and causes the problem of information overload [2]. In the face of massive information, personalized recommendation has become one of the most effective means to solve the problem of information overload, and it is a hot spot in the academic and e-commerce circles [3]. In this context, e-commerce recommendation system comes into being. It can capture key data from rich data information, mine potential customers for businesses, expand sales scope, and provide commodity recommendation for old customers to expand user groups [4]. As the continuous improvement of user demand, the recommendation quality of e-commerce recommendation cannot meet the requirements of users and businesses. The continuous development and expansion of e-commerce lead to the diversified e-commerce mode. On the one hand, the increasing amount of product data leads to the untimely data processing of recommendation system, and users cannot quickly and accurately search for the products they want [5]; on the other hand, the user's demand becomes more and more diversified, so that the recommendation system cannot recommend the products that users are potentially interested in, and the recommendation content is not diversified enough [6]. With the wide use of e-commerce recommendation system in various websites, a large number of user' browsing records and purchase records have been accumulated in the database. The huge amount of data and the complexity of data structure are beyond the load of ordinary single-machine programs. However, high-performance computers are expensive, which forces the original calculation and storage model to be upgraded and improved [7]. How to make the recommended content generated by the system closer to the needs of users has been the core issue in this field.

The emergence of cloud computing is just a good solution to this problem, and the cloud computing framework based on ordinary computers is more suitable for the needs of data processing in the era of big data. Cloud computing is a kind of distributed computing. In the case of increasing data volume and unstructured data and semi-structured data, cloud computing only needs to dynamically expand data storage resources and data computing resources to maintain the timely response of the recommendation system [8]. Wang et al. (2018) used cloud computing and high performance computing (HPC) technology to implement large-scale RS data management and data for dynamic environmental monitoring, effectively solving the problem of data processing in remote sensing [9]; in order to solve the problems of data sparsity and cold start in the recommendation system, Mezni and Abdeljaoued (2018) proposed a cloud service recommendation system based on collaborative filtering. The experimental results confirmed the effectiveness of the method [10]; Mahmood et al. (2018) developed a service selector system based on the advantages of cloud providers' computer trust, and implemented a multi-agent system approach. This method can fully use the data processing advantages of cloud computing to provide better agent-based intelligent cloud solutions for end users [11]; Jiang et al. (2019) proposed a cloud computing slope algorithm based on the fusion of trusted data and user similarity. The experimental results on Amazon dataset suggest that the recommended algorithm is more accurate than the traditional algorithm [12]. The above studies show that cloud computing has strong performance in various fields, especially in data processing. However, there are few studies on the application of cloud computing in e-commerce recommendation system.

Therefore, based on the analysis and research of collaborative filtering, content filtering recommendation and association rule algorithm and other key technologies, the specific implementation method of recommendation algorithm improvement and optimization based on cloud computing technology is proposed. Finally, the improved algorithm is empirically analyzed through the establishment of experimental environment. This study improves the processing capacity of personalized recommendation algorithm for big data, provides users with real-time, intelligent and accurate recommendation information of goods or logistics services, enhances the shopping experience of users, improves the marketing, sales and customer relationship management capabilities of e-commerce websites, promotes the development of regional small and medium-sized logistics enterprises, relieves the pressure of inventory and traffic, and reduces the cost of logistics services.

This exploration is divided into five parts. The first part is the introduction, which mainly puts forward the scientific problem that users are difficult to obtain key knowledge due to massive Internet data, makes a comparative analysis of previous research algorithms, and further puts forward the research content; the second part is the methods, which mainly introduces the problems faced by the current personalized recommendation algorithm, proposes data mining technology, designs e-commerce recommendation system based on cloud computing, and proposes data and computer configuration to verify the model; the third part is the results and discussion, which mainly introduces the comparison of recommendation efficiency of e-commerce recommendation algorithm based on cloud computing, the comparison of recommendation performance and scalability of improved e-commerce recommendation system; the fourth part is the discussion, which compares the personalized recommendation algorithm in previous studies with the method proposed in this exploration, and further puts forward the possibility of future application; the fifth part is the conclusion, which gives a detailed description of the main contributions and limitations of this study.

2. Methods

2.1. Personalized Recommendation System

As the most important part of e-commerce personalized recommendation system, personalized recommendation technology largely determines the type and performance of e-commerce personalized recommendation system. At present, according to the different recommendation methods, personalized recommendation can be divided into collaborative filtering recommendation, recommendation based on association rules, recommendation based on user statistics, and combined recommendation technology. The traditional personalized recommendation system consists of three parts: behavior record, analysis module and recommendation algorithm [13]. Figure 3 shows the specific workflow. Among them, the behavior record module is used to collect the user's behavior information (browsing and rating); the model analysis module mainly uses the information data collected by the behavior record module to analyze, obtain the potential user preferences and the degree of liking, and establish the corresponding user preference information model; according to the recommendation model, the recommendation algorithm module finds the products that the user may like from the product set based on the user's preference information, and then recommends it to the user.

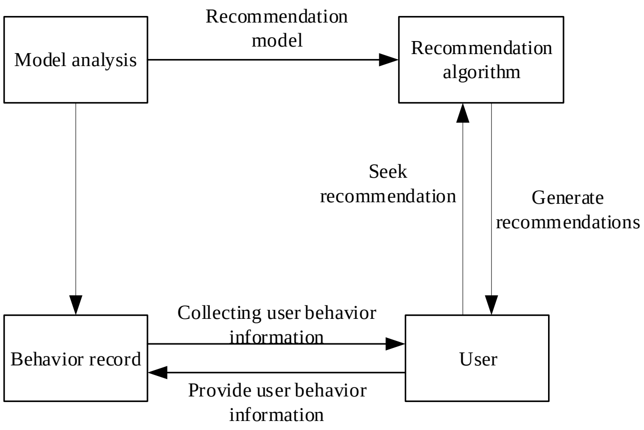

Fig. 1. Internet of things architecture

From the perspective of e-commerce, the recommendation system can be regarded as application software that can help e-commerce websites recommend commodities. User's behavior information data are collected, and they will be analyzed through statistical analysis, machine learning and other analysis methods. The commodities that users may be interested in are found from the product database and recommended to users, which can improve the sales level of e-commerce, increase user stickiness, promote consumption, and promote the development of e-commerce at last.

2.2. Big Data

Big data can be simply regarded as data with an extremely large scale. Since big data itself has an abstract concept, its definition has not been completely unified. First, big data is considered to be a dataset that takes more time to acquire, manage and process data with software tools than can be tolerated, but this definition is too unilateral and does not reflect the characteristics of big data. Then, the famous big data 3V model is proposed, which believes that big data has three characteristics: massive, diverse, and high-speed. Later, with the continuous development of the eras, it is proposed that big data also has the characteristics of value and authenticity [14].

2.3. Design of Recommendation System for E-commerce Based on Cloud Computing

(1) The framework of the recommendation system for e-commerce based on cloud computing: the e-commerce recommendation system based on cloud computing is constructed in a hierarchical structure, which can be divided into four layers from top to bottom: application layer, recommendation engine layer, cloud computing platform layer, and data source layer. The four layers are independent and mutual-restricted with

each other. They use the interface to interact information with each other, and design the level interior through the modular idea, so as to ensure the performance advantages of high cohesion, low coupling and easy to expand in the system architecture [15]. The data source layer mainly stores the data information on the e-commerce website, including the original data from different machines in various forms. After the integration of these data, the recommendation system obtains the information characteristics of users and commodities. The data information is preprocessed by the cloud computing platform layer. The information stored in the data source layer has the characteristics of multi-source, heterogeneity, and multi-type, and these characteristics lead to the increase of data noise; therefore, it is necessary to preprocess and filter the data to remove the noise before they are used. Data preprocessing includes five processes: data extraction, data cleaning, data conversion, data mapping and data integration. Different extraction methods are used to extract the corresponding characteristic data. Finally, the unified structure is used to store the data. The cloud computing platform layer mainly uses distributed computing and distributed storage systems to process and calculate data, which is mainly completed by the Hadoop platform. Figure 2 shows the overall architecture of the distributed recommendation system.

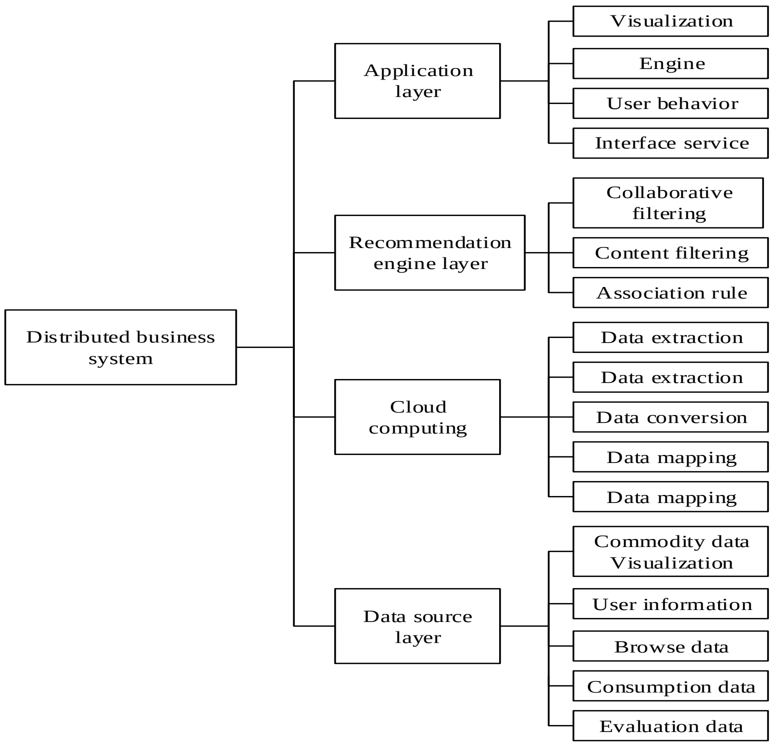

Fig. 2. Overall architecture of distributed recommendation system

The recommendation engine layer is the core layer of the recommendation system. It uses some general algorithm interfaces in the cloud computing platform to construct a recommendation algorithm and recommendation strategy into a recommendation engine that can operate independently. According to the diversified recommendation requirements, a corresponding recommendation engine is designed for each different recommendation requirement, which solves the scalability of the recommendation system. The recommendation engine based on collaborative filtering technology, content filtering calculation and association rule is designed.

(2) The cloud computing platform – Hadoop: it is a kind of distributed system infrastructure, which is supported by cheap computer cluster hardware to deal with massive data. Hadoop platform makes it easier for users to develop distributed programs; it can use cluster hardware to achieve massive data storage and high-speed computing, which has excellent scalability and high reliability [16]. Hadoop distributed file system (HDFS) and basic execution unit (MapReduce) of distributed computing tasks are the core components of the Hadoop platform. The HDFS is at the bottom of the Hadoop platform, which is mainly used to store files in all data nodes in the cluster. MapReduce is used to process massive data [17]. Figure 3 presents the architecture of the Hadoop platform.

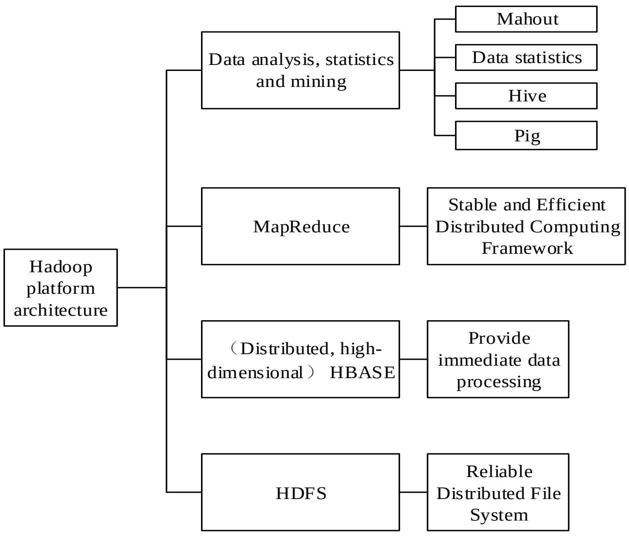

Fig. 3. The architecture of the Hadoop platform

HDFS is mainly used for the storage of data files and adopts the master/slave architecture, providing storage services for high-throughput, reliable and scalable large data files of upper layer distributed computing tasks. MapReduce is a programming model for parallel computing of large datasets, and it is easy to use and understand; the use of this programming model does not require users to understand its distributed and parallel programming, and the development of the program can be realized by using map function and reduce function [18]. MapReduce is used to process big data, which is realized mainly through the idea of dividing and ruling. When cloud computing technology is used to design a recommendation system of e-commerce, the concurrency elements involved in the traditional algorithm need to be found out. These parallel tasks are opposite to each other, so that the distributed computing method can be used directly; however, for the serial tasks, they should be decomposed as much as possible; then, the parallel tasks are found to calculate them. Figure 4 is the processing flow of tasks with MapReduce.

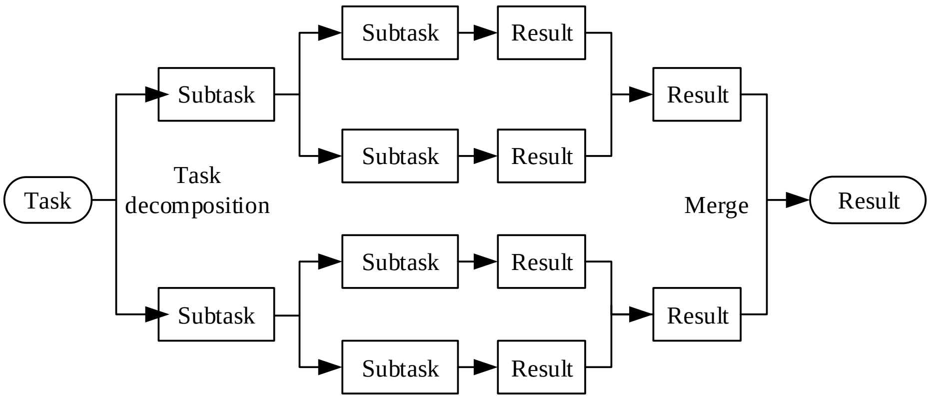

Fig. 4. Flow chart of task processing based on MapReduce

2.4. Personalized Recommendation Algorithm Based on Cloud Computing

(1) Collaborative filtering recommendation algorithm based on cloud computing: the collaborative filtering recommendation algorithm can be divided into user-based collaborative filtering (UserCF) and an item-based collaborative filtering algorithm (ItemCF) [19]. The widely used recommendation algorithm in e-commerce system is collaborative filtering algorithm, which simulates the scene of mutual recommendation between people in real life, uses the user's behavior characteristics in their historical information data to calculate the user's similarity, and uses the similarity data to recommend the product information to the user. $C=\{c_1,c_2,\ldots c_n\}$ is regarded as the set of all users in the system, and $S=\{s_1,s_2,\ldots s_n\}$ is a set of all products. The score of user c for the unevaluated product s is rc,s,. Then, the score is calculated as follows.

$$r_{c,s} = \frac{1}{N}\sum_{\hat{c}\in\hat{C}} r_{\hat{c},s} \tag{1}$$

$$r_{c,s} = k\sum_{\hat{c}\in\hat{C}} sim(c,\hat{c}) \bullet r_{\hat{c},s} \tag{2}$$

$$r_{c,s} = \bar{r}_c + k \sum_{\hat{c} \in \hat{C}} sim(c, \hat{c}) \bullet (r_{\hat{c},s} - \bar{r}_{\hat{c}}) \tag{3}$$

$\hat{C}$ represents the similarity set of c, k is a standardization factor, and $sim(c,\hat{c})$ represents the similarity between targeting user c and similar user $\hat{c}$. $\bar{r}_c$ represents the average score of user c, and $\bar{r}_{\hat{c}}$ represents the average score of the user $\hat{c}$.

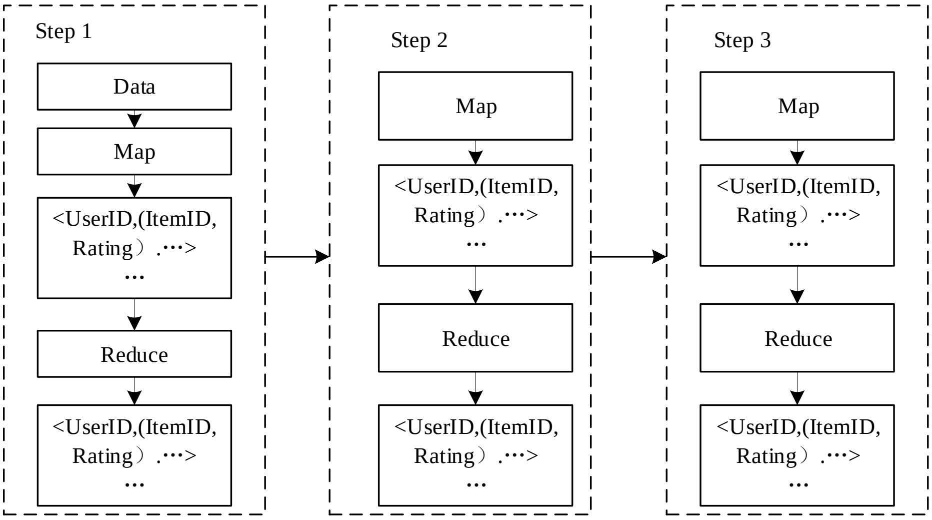

Fig. 5. Flow chart of system filtering recommendation algorithm improved by cloud computing technology

The principle of the collaborative filtering recommendation algorithm improved by cloud computing is the same as that of the traditional collaborative filtering recommendation algorithm. However, compared with the traditional collaborative filtering recommendation algorithm, the performance efficiency of the algorithm is improved because of the enhanced ability of distributed parallel computing.

(2) Content filtering recommendation algorithm based on cloud computing: the content-based recommendation (CBR) method recommends objects with similar attributes to users according to their selection objects, and Figure 6 shows the recommendation process based on content filtering.

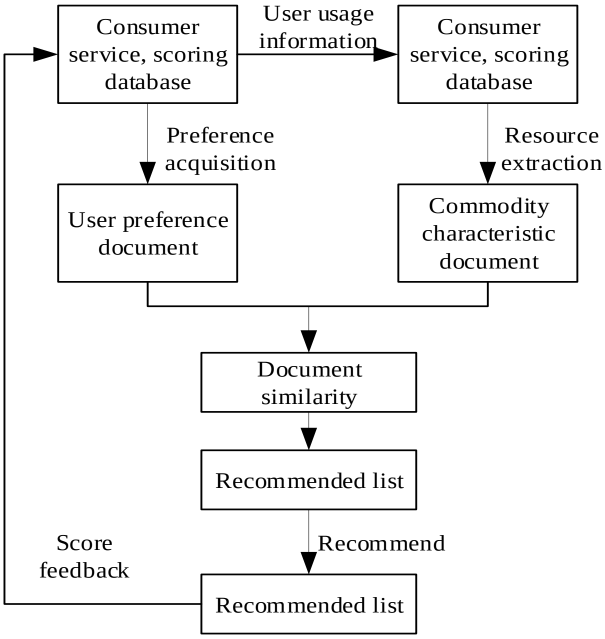

Fig. 6. Recommendation flow chart based on content filtering

Algorithm recommendation flow based on content filtering is as follows. First, according to the user's scoring database, the attribute features of the item are extracted, and the feature documents of the item are constructed. Combined with the user's historical information, the user's preference document is constructed, and the similarity between the two documents is calculated to find items similar to the user's preference and recommend them to the user. Both feature documents and user preference documents are built based on the content recommendation algorithm, and the content information of the item is represented by the vector space model [21]. If item i has k attributes, wij is used to represent the weight of the j-th attribute of item i; therefore, the contentProfile (i) of item i can be expressed in the following ways.

$$contentProfile(i) = \{w_{i1}, w_{i2}, \cdots, w_{ik}\} \tag{4}$$

The user's preference information can be obtained by decision-tree, Bayesian classification algorithm, neural networks, and other machine learning algorithms. The importance of the j-th attribute to user u is represented by wuj, and the userProfile (u) can be calculated by the following equation.

$$userProfile(i) = \{w_{u1}, w_{u2}, \cdots, w_{uk}\} \tag{5}$$

Finally, cosine similarity is used to calculate the similarity between item i and user u in the item document and user preference document. The calculation method is as follows.

$$sim(u,i) = \cos(u,i) = \frac{\sum_{j=1}^{k} w_{u,j} w_{ij}}{\sqrt{\sum_{j=1}^{k} w_{u,j}^2}\sqrt{\sum_{j=1}^{k} w_{i,j}^2}} \tag{6}$$

MR-CBR, an improved content filtering recommendation algorithm based on cloud computing technology, can be regarded as a mutually independent and parallelizable process when user's preference documents are calculated, and can be regarded as a MapReduce. When the similarity between two document information is calculated, it is also an independent and feasible calculation process and regarded as another MapReduce, and there is a serial relationship between the two MapReduces [22]. Figure 7 presents the flow of the improved content filtering recommendation algorithm based on cloud computing.

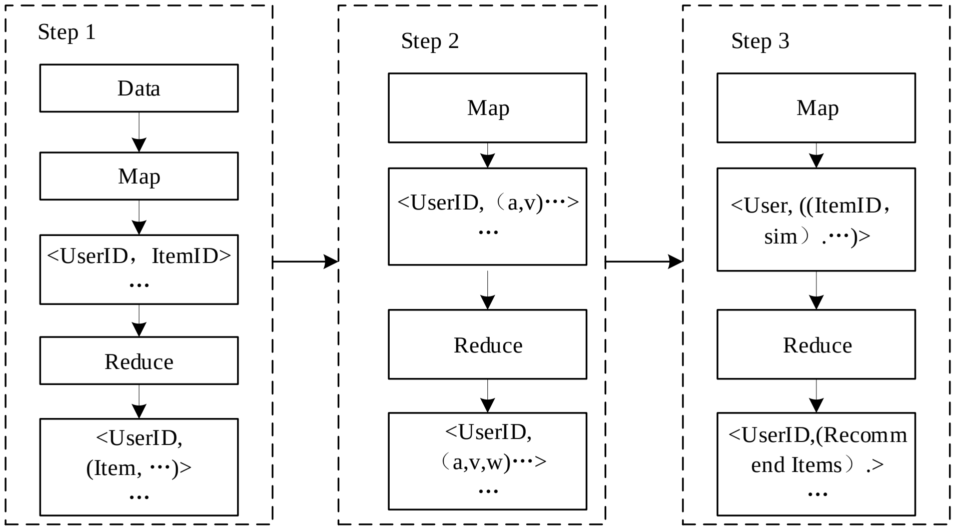

Figure. 7. The flow chart of the improved content filtering recommendation algorithm based on cloud computing

(3) An improved association rule recommendation algorithm based on cloud computing: association rules are based on mining the correlation between items from many data. It can analyze the transaction data of commodities, find out the commodities that are purchased frequently at the same time from the data, generate corresponding rules, and recommend for users based on the current behavior data of users [23]. The item set is set to I={i1,i2,…in}, the commodity transaction database is represented by D, and each transaction is represented by T. T is a subset of commodity items, and A is regarded as an item set when A is less than or equal to T. It can be considered that trade T contains A. The calculating method of ($\sup port(A \Rightarrow B)$) is as follows.

$$\sup port(A \Rightarrow B) = P(A \cup B) = \frac{\left|\{T : A \cup B \subseteq T, T \in D\}\right|}{|D|} \quad (7)$$

The $confidence(A \Rightarrow B)$ is calculated as follows.

$$confidence(A \Rightarrow B) = P(B \mid A) = \frac{\left|\{T : A \cup B \subseteq T, T \in D\}\right|}{\left|\{T : A \subseteq T, T \in D\}\right|} \quad (8)$$

It is difficult to extract rules due to the sparsity and high dimension of data in the association rule recommendation algorithm, resulting in quality instability. In addition, the offline building time of rule in this algorithm is relatively long. Meanwhile, the algorithm will increase the management difficulty with the increase in the number of rules.

For the improved association rule recommendation algorithm (MR-FP) based on cloud computing, it is not necessary to build frequent tree for the whole transaction set in the construction of a frequent tree. The frequent tree of the frequent term conditions is constructed with each calculated node, and the corresponding conditional frequent tree is established. Then, the final solution of the algorithm is obtained by combining conditional frequent trees, and the association rules of the frequent term set are obtained. Finally, according to the association rules, products are recommended to users [24].

2.5. Construction of a Ccomprehensive Evaluation System for Distributed Recommendation System

To verify the advantages and disadvantages of the distributed recommendation systems in three kinds of e-commerces, a comprehensive evaluation system of the corresponding distributed recommendation system will be established [25]. The following principles should be followed in indicator selection, as shown in Table 1.

Table 1. Selection principle of comprehensive evaluation system index

number	principle	selection method
one	objectiveness	comprehensive and easy to quantify
two	systematicness	can reflect essential features
three	independence	an indicator reflects a single element
four	science	scientific and reasonable to meet the theory of statistics, economics, e-commerce and other related disciplines

Precision, efficiency, coverage, diversity, and novelty are regarded as evaluation indicators of the e-commerce recommendation system. When the precision of the recommended system is calculated, the mean absolute error (MAE) can be used as the judgment method for the precision of the system score prediction. The calculation method is as follows.

$$MAE = \frac{\sum_{u,i \in T} \left| r_{ui} - \hat{r}_{ui} \right|}{\left| T \right|} \tag{9}$$

In the above equation, r_{ui} refers to the actual score of user u for product *i*, and $\hat{r}_{ui}$ refers to the predicting score of user u for the product i.

Users' preferences for recommended items can be calculated by the prediction precision recommended by TopN, expressed by two indicators of precision and recall, and fitted by F1-Score. The calculation method is as follows.

$$\Pr ecision = \frac{\sum_{u \in U} \left| R(u) \cap T(u) \right|}{\sum_{u \in U} \left| R(u) \right|} \tag{10}$$

$$\operatorname{Re} call = \frac{\sum_{u \in U} \left| R(u) \cap T(u) \right|}{\sum_{u \in U} \left| T(u) \right|} \tag{11}$$

$$F1 - Score = \frac{2 * \Pr ecision * \operatorname{Re} call}{\Pr ecision + \operatorname{Re} call} \tag{12}$$

In the above equations, *R(u)* refers to the user's behavior and the number of items in their recommendation list, and *T(u)* refers to the number of items the user likes.

Efficiency refers to the time consumed by the algorithm in calculating and processing, and the calculation of this indicator can be measured by the time consumed by the algorithm operating on the computer.

Coverage refers to the widespread degree of items recommended by e-commerce recommendation system to users. The calculation method is as follows.

$$Coverage = \frac{\left| U_{u \in U} R(u) \right|}{\left| I \right|} \tag{13}$$

Where *U* refers to the user set of the system, and *R(u)* refers to the list of products recommended to users.

Diversity is represented by the dissimilarity of products in the recommendation list and the dissimilarity of the list recommended to different users. Hamming distance (HM) is used to express the dissimilarity of recommendation lists of different users *u* and *v* [26], and the calculation method is as follows.

$$HM_{u,v} = 1 - \frac{R(u) \cap R(v)}{R(u) \cup R(v)} \tag{14}$$

In the above equations, *R(u)* and *R(v)* refer to the list of products recommended to users u and v. When *R(u)* and *R(v)* are identical, the value of HM is 0, and when there is no overlap, it is 1. When the value of HM is larger, the system diversity is higher.

Novelty means that the recommendation system recommends some non-popular new products to users, and the novelty of the recommended items can be evaluated according to the average popularity of the recommendation list. When the average popularity of products in the recommendation list is smaller, the novelty of the recommendation system is stronger.

The e-commerce recommendation system is built through the Hadoop platform. Because the number of nodes of the platform can be increased and decreased flexibly, the recommendation effect of the distributed recommendation system and in the single machine environment is compared by the acceleration ratio (R). The calculation method of *R* is as follows.

$$R = \frac{Ts}{Tc} \tag{15}$$

Ts in the above equation represents the operating time of the recommendation system in a single machine environment, and *Tc* represents the operating time of the distributed recommendation system.

3. Results and Analysis

3.1. Comparison of Recommendation Efficiency of E-commerce Recommendation Algorithm Based on Cloud Computing

The recommendation efficiency of the algorithm is compared mainly through the comparison of the operating speed of the algorithm. The operating time of the algorithm is inversely proportional to the computing ability of the algorithm. The shorter the operating time of the algorithm is, the stronger the computing ability of the algorithm is, and the higher the recommendation efficiency of the algorithm is. The recommended efficiency of several algorithms is compared, and the comparison results are shown in Figure 8. It suggests that when the amount of input data information increases, the running time of the algorithm under different nodes increases slowly, which shows that the greater the amount of information to be calculated is, the slower the calculation and processing speed of the algorithm are. However, the running time of the algorithm under 7 nodes is better than that of 5 nodes, the running time of algorithm with 5 nodes is better than that of 3 nodes, and the running time of algorithm with 3 nodes is better than that of 2 nodes. It shows that the in the distributed platform, the more the nodes are, the stronger the computing power of the algorithm is, and the faster the execution speed of the algorithm is. Therefore, the improved algorithm runs fast. The recommendation system has high recommendation efficiency.

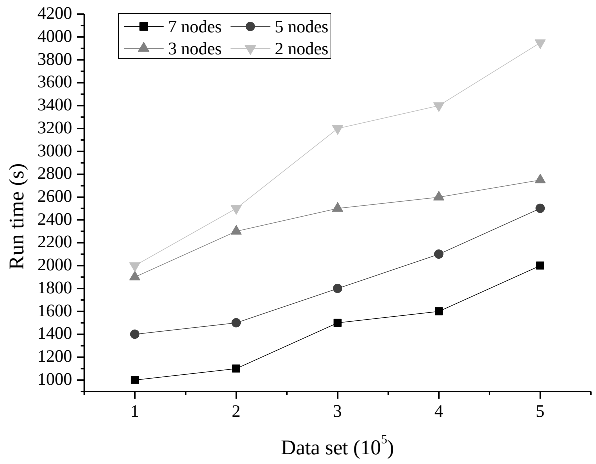

Fig. 8. Analysis chart of algorithm recommendation efficiency based on cloud computing improvement

3.2. Recommendation Performance Comparison of E-commerce Recommendation System Based on Cloud Computing iImprovement

The comprehensive evaluation system is used to analyze the performance of the e-commerce recommendation system based on cloud computing, and the analysis results are shown in Figure 9. It reveals that as far as the accuracy indicator is concerned, the final evaluation results are accuracy rate, recall rate and the average value of F1 in five experiments, because the three recommendation engines of collaborative filtering, content filtering and association rules are all based on TopN recommendation. The F1 values are 7.2%, 8.4% and 5.6%, respectively. It suggests that the content filtering recommendation engine is superior in recommendation accuracy; in terms of efficiency indicator, 5000 users are randomly selected for offline calculation, and the running time of the three is 12s, 23s and 42s, respectively, which shows that the collaborative filtering recommendation engine is the best in this respect; in terms of coverage indicator, the proportion of items recommended for all users by the three recommendation engines is counted. Coverage rate is 45.6%, 68.4% and 72.1%, respectively; in terms of diversity indicator, the recommended list of the three engines for the users at a certain time is selected, and the Hamming distance of the user pairs used. The diversity degree of the three recommendation engines is 74.6% 86.2%, and 64.5%, respectively. In terms of novelty indicator, the average popularity of items in the recommended list of the three recommendation engines is 15.2%, 28.6% and 15.6%, respectively. The above research results prove that the performance of content filtering recommendation system based on cloud computing is the best, and the system filtering recommendation system based on cloud computing has the least running time and the highest efficiency.

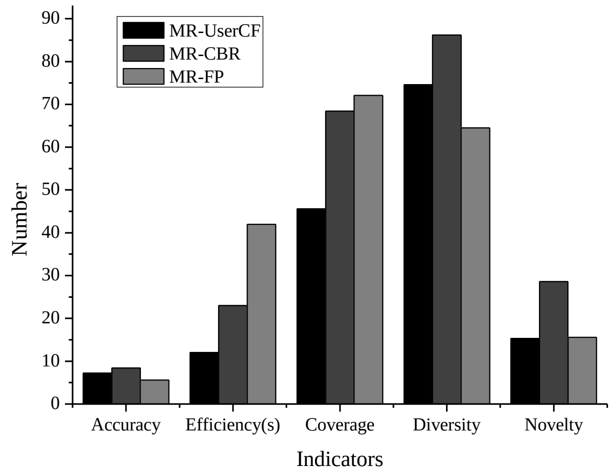

Fig. 9. Comprehensive evaluation indicator results chart of different recommendation algorithms

3.3. Scalable Performance Comparison of E-commerce Recommendation System Based on Cloud Computing

The R of several recommendation systems in a single machine environment (node number is 1) and distributed recommendation system (node number is at least 2) is compared, and the comparison results are shown in Figure 10 and Figure 11. It shows that the acceleration ratio of the algorithm increases with the increasing number of cluster nodes. The change is very fast in the early stage and slow in the later stage. The acceleration ratio of the algorithm with 7 nodes is larger than that of the algorithm with 5 nodes. The acceleration ratio of the algorithm with 5 nodes is larger than that of the algorithm with 3 nodes. The acceleration ratio of the algorithm under 3 nodes is larger than that of the algorithm under 1 node. It shows that in the distributed platform, the more the nodes are, the stronger the computing power of the algorithm is, the better the recommendation effect of the recommendation system is. However, the experiment suggests that when the number of nodes is less than 5, the acceleration ratio changes linearly. When the number of nodes is greater than 5, the acceleration ratio changes slowly with the increasing number of nodes. It reveals that simply increasing the number of cluster nodes cannot infinitely improve the performance of the algorithm.

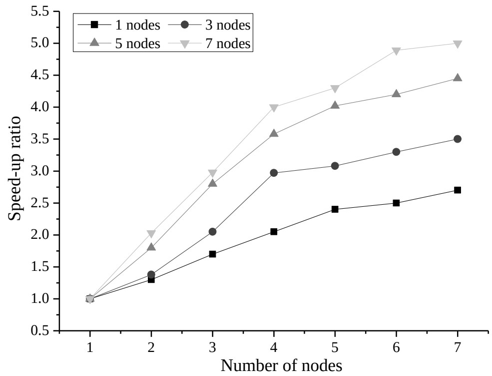

Fig. 10. Acceleration ratio analysis chart of the recommendation algorithm under different cluster nodes

Figure 10 suggests that the R of the distributed recommendation algorithm increases gradually with the increase of the number of nodes. It indicates that the scalability performance of the recommendation algorithm is better than that in the single machine environment. Figure 11 shows that the three distributed algorithms have obvious advantages over traditional algorithms in execution time. In the case of experimental data, with the increase of the number of clusters, the acceleration ratio of the three recommendation algorithms gradually increases, which shows that the distributed algorithm continues to run at this time. When the cluster size reaches 10, the execution efficiency of MR-UserCF algorithm is more than 7 times than that of single machine. In the best case, MR-CBR and MR-FP algorithms can reach 4 times and 3 times. However, due to the scale of experimental data, when the Hadoop cluster nodes are added after reaching the peak of acceleration ratio, the growth rate slows down with the increase of nodes, but the slope decreases, which indicates that the growth rate slows down with the increase of nodes. However, in general, the increase of the number of nodes can effectively guarantee the decrease of system running time.

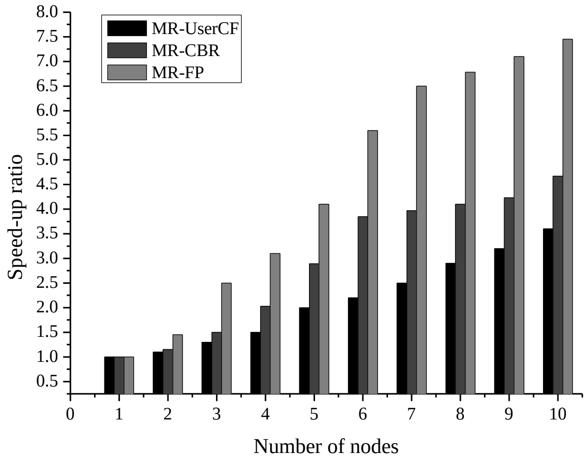

Fig. 11. R analysis chart of different recommendation algorithms

4. Discussion

Cloud computing technology and Hadoop platform are used to improve the traditional e-commerce recommendation algorithm. The recommendation speed of the algorithm significantly increases after the improvement (Figure 8). The main reason is that the more the nodes are in the distributed platform, the stronger the computing power of the algorithm is, the faster the implementation speed of the algorithm is. This is confirmed by the research in Zhu and Bai (2020). They took collaborative filtering recommendation algorithm as an example of data mining platform, and introduced the idea of weighted factor based on project popularity to improve the degree of personalized recommendation system. The improved algorithm improves the performance of personalized recommendation system [27]. Based on the open source cloud computing platform Hadoop, MapReduce parallel framework is used to process massive data sets. Distributed file system HDFS is adopted to store and manage super large files. Compared with the traditional recommendation algorithm, it has obvious improvement in recommendation speed, which can greatly improve the recommendation speed of the recommendation system and improve the user satisfaction (Figure 9). It has also been confirmed in the research of Tang and Cheng (2017). They constructed and replaced the rating matrix based on user preference characteristics. MapReduce parallel computing framework can improve computing efficiency and algorithm scalability [28]. The applicability and scalability of traditional recommendation algorithms for single machine computing in the face of large data sets is systematically analyzed, including the operation mechanism of divide and rule and the method of problem domain partition in MapReduce method of cloud computing. From the perspective of parallel decomposition of tasks and data, the design principles of recommendation algorithm based on cloud computing are proposed. It is found that simply increasing the number of nodes in the cluster cannot improve the performance of the algorithm infinitely, and the increase of the number of nodes can effectively guarantee the decrease of system running time. This is confirmed by Cao et al. (2018). They found that in the process of MapReduce parallelization, the data partition matrix is stored in line segments, the computing load is distributed in each node of the cluster, and the time consumption and partition matrix consumption of moving data matrix is calculated, which can reduce the calculation amount in the execution process, and greatly reduce the consumption of storage space, but the number of nodes cannot be increased too much [29]. Hadoop platform is used to build e-commerce recommendation system, and improve the parallel computing ability of the recommendation system. Therefore, this study is consistent with the existing research results, which shows that the research method is effective and feasible.

5. Conclusion

In the environment of big data, the traditional system filtering recommendation, content filtering, and association rule filtering algorithms are rebuilt on the Hadoop platform through cloud computing technology, and the parallel processing efficiency of several algorithms is improved by MapReduce. A comprehensive evaluation system is established. Its calculation and analysis as well as R prove that several improved distributed recommendation algorithms based on cloud computing are more efficient than those in the traditional single machine environment; moreover, the performance of the distributed recommendation algorithm in R comparison and analysis is excellent. It shows that the improved distributed recommendation system based on cloud computing improves the support of the recommendation algorithm, reduces the operating time of the algorithm, and improves the recommendation efficiency of the algorithm. The content filtering recommendation algorithm based on cloud computing has an excellent performance in precision, coverage, diversity, and novelty. The collaborative filtering recommendation system based on cloud computing has the best operation efficiency and good scalability.

The Hadoop platform is a relatively mature research platform at present. In the future development, the optimization and improvement of this platform is still a research focus, and it is hoped that more traditional algorithms can be improved through this platform to improve the performance of the algorithm; moreover, when cloud computing technology is used to improve the traditional algorithm, MapReduce improves the parallel processing efficiency of the algorithm; however, in coordinating the hybrid systems of different cloud platforms and coordinating the layout of datasets, further research is needed to improve the working performance of the cloud system.

6. References

1. Fan, W.; Xu, M.; Dong, X.; Wei, H.: Considerable environmental impact of the rapid development of China's express delivery industry. Resources, Conservation and Recycling, Vol. 126, 174-176. (2017)
2. Swar, B.; Hameed, T.; Reychav, I.: Information overload, psychological ill-being, and behavioral intention to continue online healthcare information search. Computers in Human Behavior, Vol. 70, 416-425. (2017)
3. Xiao, J.; Wang, M.; Jiang, B.; Li, J.: A personalized recommendation system with combinational algorithm for online learning. Journal of Ambient Intelligence and Humanized Computing, Vol. 9, 667-677. (2018)
4. Hwangbo, H.; Kim, Y.S.; Cha, K.J.: Recommendation system development for fashion retail e-commerce. Electronic Commerce Research and Applications, Vol. 28, 94-101. (2018)
5. Subramaniyaswamy, V.; Logesh, R.; Chandrashekhar, M.; Challa, A.; Vijayakumar, V.: A personalised movie recommendation system based on collaborative filtering. International Journal of High Performance Computing and Networking, Vol. 10, 54-63. (2017)
6. Kumar, P.; Thakur, R.S.: Recommendation system techniques and related issues: a survey. International Journal of Information Technology, Vol. 10, 495-501. (2018)
7. Tian, G.; Wang, J.: Recommendation algorithm for mobile E-commerce based on cone depth learning. International Journal of Computers and Applications, 1-6. (2019)
8. Xu, B.; Xu, L.; Cai, H.; Jiang, L.; Luo, Y.; Gu, Y.: The design of an m-Health monitoring system based on a cloud computing platform. Enterprise Information Systems, Vol. 11, 17-36. (2017)
9. Wang, L.; Ma, Y.; Yan, J.; Chang, V.; Zomaya, A.Y.: pipsCloud: High performance cloud computing for remote sensing big data management and processing. Future Generation Computer Systems, Vol. 78, 353-368. (2018)
10. Mezni, H.; Abdeljaoued, T. A cloud services recommendation system based on Fuzzy Formal Concept Analysis. Data & Knowledge Engineering, Vol. 116, 100-123. (2018)
11. Mahmood, A.; Shoaib, U.; Shahzad, S.: A Recommendation System for Cloud Services Selection Based on Intelligent Agents. Indian Journal of Science and Technology, Vol. 11, 1-6. (2018)
12. Jiang, L.; Cheng, Y.; Yang, L.; Li, J.; Yan, H.; Wang, X.: A trust-based collaborative filtering algorithm for E-commerce recommendation system. Journal of Ambient Intelligence and Humanized Computing, Vol. 10, 3023-3034. (2019)
13. Zhuo, R.; Bai, Z. Key technologies of cloud computing-based IoT data mining. International Journal of Computers and Applications, 1-8. (2020)
14. Alsbatin L, Öz G, Ulusoy A H.: Efficient virtual machine placement algorithms for consolidation in cloud data centers. Computer Science and Information Systems, Vol. 1, 36-36. (2019)
15. Tarus, J.K.; Niu, Z.; Mustafa, G.: Knowledge-based recommendation: A review of ontology-based recommender systems for e-learning. Artif. Intell. Rev, Vol. 50, 21-48. (2018)
16. Wu, D.; Xu, Y.; Zhou, Y. Study on the technical architecture of the computer data mining platform based on the hadoop framework. Boletin Tecnico/Technical Bulletin, Vol. 55, 275-281. (2017)
17. Kune, R.; Konugurthi, P.K.; Agarwal, A.; et al.: XHAMI - extended HDFS and MapReduce interface for Big Data image processing applications in cloud computing environments. Software, Vol. 47, 455-572. (2017)
18. Bi Z, Dou S, Liu Z, et al. A recommendations model with multiaspect awareness and hierarchical user-product attention mechanisms. Computer Science and Information Systems, Vol. 0, 24-24. (2020)
19. Yi, L.; Jun, F.; Tong-Tong, W.; et al. An Improved Collaborative Filtering Recommendation Algorithm. computer and modernization, Vol. 32, 204-208. (2017)

20. Li, C.; He, K.: CBMR: An optimized MapReduce for item-based collaborative filtering recommendation algorithm with empirical analysis. Concurrency & Computation Practice & Experience, Vol. 29, 4092-4106. (2017)
21. Gavalas, D.; Konstantopoulos, C.; Mastakas, K.; Pantziou, G.: Mobile recommender systems in tourism. J. Netw. Comput. Appl, Vol. 39, 319-333. (2014)
22. Son, J.; Kim, S.B.: Content-based filtering for recommendation systems using multiattribute networks. Expert Systems with Application, Vol. 89, 404-412. (2017)
23. Hang, L.; Kang, S.-H.; Jin, W.; Kim, D.-H. Design and Implementation of an Optimal Travel Route Recommender System on Big Data for Tourists in Jeju. Processes, Vol. 6, 133-143. (2018)
24. Li, C.; Liang, W.; Wu, Z.; et al.: [IEEE 2018 IEEE International Conference on Web Services (ICWS) - San Francisco, CA, USA (2018.7.2-2018.7.7)] 2018 IEEE International Conference on Web Services (ICWS)An Efficient Distributed-Computing Framework for Association-Rule-Based Recommendation, 339-342. (2018)
25. Bhatta R, Ezeife C I,.: Distributed Data Mining-Based E-Commerce Recommendation System. Computer Systems & Applications, Vol. 18, 33-37. (2009)
26. Kacprzyk, J.; Nurmi, H.; Sławomir, Z.: Towards a Comprehensive Similarity Analysis of Voting Procedures Using Rough Sets and Similarity Measures. Intelligent Systems Reference Library, Vol. 42, 359-380. (2013)
27. Zhuo, R.; Bai, Z.: Key technologies of cloud computing-based IoT data mining. International Journal of Computers and Applications, 1-8. (2020)
28. Tang, H.; Cheng, X. Personalized E-commerce recommendation system based on collaborative filtering under Hadoop. World, Vol. 1, 146-148. (2017)
29. Cao, Y.; Li, P.; Zhang, Y.: Parallel processing algorithm for railway signal fault diagnosis data based on cloud computing. Future Generation Computer Systems, Vol. 88, 179-283. (2018)

Yiman Zhang was born in Wenzhou, Zhejiang, China, in 1983. She received the Master degree from Kyoto College of Graduate Studies for Informatics, Japan. Now, she works at Oujiang College of Wenzhou University. Her research interests include Data mining and E-commerce.

Received: September 17, 2020; Accepted: February 19, 2021.