

Design of Hybrid Recommendation Algorithm in Online Shopping System

Yingchao Wang¹, Yuanhao Zhu¹, Zongtian Zhang¹, Huihuang Liu^{1,*} and Peng Guo²

¹Hunan University of Finance and Economics, Changsha, China

²University Malaysia Sabah, Kota Kinabalu, Malaysia

*Corresponding Author: Huihuang Liu. Email: wangyingchao425@foxmail.com

Received: 15 April 2021; Accepted: 11 September 2021

Abstract: In order to improve user satisfaction and loyalty on e-commerce websites, recommendation algorithms are used to recommend products that may be of interest to users. Therefore, the accuracy of the recommendation algorithm is a primary issue. So far, there are three mainstream recommendation algorithms, content-based recommendation algorithms, collaborative filtering algorithms and hybrid recommendation algorithms. Content-based recommendation algorithms and collaborative filtering algorithms have their own shortcomings. The content-based recommendation algorithm has the problem of the diversity of recommended items, while the collaborative filtering algorithm has the problem of data sparsity and scalability. On the basis of these two algorithms, the hybrid recommendation algorithm learns from each other's strengths and combines the advantages of the two algorithms to provide people with better services. This article will focus on the use of a content-based recommendation algorithm to mine the user's existing interests, and then combine the collaborative filtering algorithm to establish a potential interest model, mix the existing and potential interests, and calculate with the candidate search content set. The similarity gets the recommendation list.

Keywords: Recommendation algorithm; hybrid recommendation algorithm; content-based recommendation algorithm; collaborative filtering algorithm

1 Introduction

With the continuous development of Internet technology and the continuous improvement of social material life, in order to save unnecessary time and improve the quality of life in this fast-paced society, personalized recommendation has become one of the core functions of the Internet and is used in various industries. E-commerce(e-commerce) is a commercial activity that uses the combination of microcomputer technology and network communication technology. E-commerce sites use recommendation algorithms to provide users with the most likely products to purchase [1]; friend recommendations on social networking sites provide users with potential friend attention; video sites recommend users most likely to click on the video [2]; Content recommendation on news websites provides users with the most interesting news information [3]. Personalized recommendation technology is the solution, which is also one of the manifestations of Internet intelligence [4].

2 Related Works

There are three main recommendation algorithms at this stage, content-based recommendation algorithms, collaborative filtering algorithms, and hybrid recommendation algorithms. The content-based recommendation algorithm [5] is based on specific text information to make recommendations [6]. Collaborative filtering algorithms are divided into memory-based methods and model-based methods. At the earliest, it was mainly based on user evaluations, and found similar items and users by traversing the



information of the whole or part of the evaluation matrix [7]. However, with the development and application of data mining technology in machine learning, new ideas have been provided for collaborative filtering algorithms. By establishing a learning model, using the model to use the evaluation matrix as experimental data to make intelligent recommendation predictions. However, the above two recommendation algorithms have fatal problems that affect the final recommendation effect. Therefore, in order to avoid the shortcomings of the two algorithms and take advantage of their advantages, the researchers proposed a third algorithm, the hybrid recommendation algorithm, which is based on Content recommendation algorithm and collaborative filtering algorithm are combined together [8]. The mainstream of hybrid recommendation algorithm is based on the combination of content and collaborative filtering, and has been widely used in actual recommendation systems. At present, most of the actual recommendation systems are hybrid recommendation systems that use the fusion of multiple recommendation algorithms [9].

In recent years, researchers have also made new progress in the research of hybrid recommendation algorithms. Jeffery D. Weir proposed a recommendation system based on the tag model in 2016 [10]. In view of the sparseness and cooling of the data of the social network system, researchers have proposed a hybrid recommendation algorithm based on probability matrix, which combines the user trust network with the method of probability matrix factorization [11]. Association and user rating matrix to evaluate and calculate the relationship between potential users and current users. Aiming at the problem that the collaborative filtering algorithm only uses user rating information as the recommendation basis, and ignores user comments and recommended item tags, thereby reducing the accuracy of recommendation, researchers have proposed an improved hybrid recommendation algorithm based on stack noise reduction autoencoder [12]. Based on the behavior and attributes of the purchased goods, researchers have proposed a hybrid recommendation algorithm based on the Gaussian model probability matrix decomposition [13]. In the recommendation system of online courses, Liang Ying uses migration learning algorithms to initially solve the problem of data sparsity [14].

3 Technical Foundation

3.1 Collaborative Filtering Algorithm

3.1.1 Fundamental

The basic idea of the collaborative filtering algorithm is to make recommendations for current users based on the historical purchase behavior of users with the same interest. In short, from the massive data, mine users with the same interests as you, make them your neighbors, and then generate a recommendation list based on what they like and push it to you. From this, it can be seen that the collaborative filtering algorithm has two core problems:

- a) Identify users with the same interests as current users;
- b) Generate a recommendation list and push it to the current user.

There are three major steps to implement collaborative filtering algorithms:

- a) Collect user's likes;
- b) Find similar users and items;
- c) Calculate data and generate recommendation list.

3.1.2 Problems with Collaborative Filtering Algorithms

Data sparsity is the biggest problem that collaborative filtering algorithms will inevitably face. In an actual commercial recommendation system, the number of users and their corresponding purchase items are very large, and users tend to base their evaluation on only a few purchase items. With the increase of users and corresponding purchased items, due to the limitation of computing resources and computing speed, the efficiency of the collaborative filtering algorithm will be greatly reduced after users and corresponding purchased items increase to a certain number, so that the actual demand cannot be met.

3.2 Content-Based Recommendation Algorithm

3.2.1 Fundamental

The content-based recommendation algorithm calculates the similarity between items based on the items that users liked in the past to generate a recommendation list. The core of the algorithm is the accuracy of the pairwise similarity of items. There are three steps to implement the algorithm:

- a) Extract feature values of past items;
- b) Use feature data to build user preference models;
- c) Generate a recommendation list by comparing the candidate items with the preference model.

3.2.2 The Problem of Content-Based Recommendation Algorithm

The problem of the diversity of recommended objects is the primary problem of content-based recommendation algorithms. The preference model is derived from the learning of items that users like in the past, so there are hidden dangers in the original data. At the same time, establishing an association model between users and items and extracting main features is a process that consumes a lot of time and manpower. In addition, if the user rarely purchases and purchases, the details of the data obtained will have a great impact on the browsing history of the recommendation result. The comparison of content-based recommendation and algorithmic collaborative filtering algorithm is shown in Table 1.

Table 1: Comparison of content-based recommendation algorithm and collaborative filtering algorithm

Recommendation algorithm	Advantage	Disadvantage
Disadvantage	The recommendation results are intuitive and there is no cold start problem	The establishment of an association model between users and items requires a lot of time and manpower; new user problem; sparsity problem
Collaborative filtering algorithm	Realize cross-interest recommendation, with a high degree of automation	Cold start; data sparsity problem; scalability problem

3.3 Hybrid Recommendation Algorithm

3.3.1 Fundamental

Collaborative filtering algorithms and content-based recommendation algorithms can be complementary. The collaborative filtering algorithm does not have the problem of recommending the diversity of items, and the content-based recommendation algorithm does not have the problem of cold start. Naturally, these two recommendation algorithms can be merged into a new hybrid recommendation algorithm for better recommendation results.

3.3.2 Algorithm Types and Advantages

This method improves the traditional content-based method to obtain the user's existing interest, obtains the user's potential interest through the collaborative filtering of feature words, and mixes the user's existing interest with the potential interest to obtain a mixed user interest model. Similarity calculation, recommend products that may be of interest to different users. Compared with the previous method, this article takes into account the needs of user preferences for diversity and personalization, more fully explores the potential interests of users, and improves user's click-through rate on recommended products.

4 Hybrid Recommendation Algorithm in Online Shopping System

The hybrid recommendation system model mainly consists of three parts: the user's existing interest model, the potential user's interest model, and the hybrid recommendation algorithm model. First of all, here are some definitions that need to be explained.

Mixed interest model (HM): the weight vector obtained by combining the user's existing interest model and potential interest model according to certain rules, denoted as $W_3 = \{w_{31}, w_{32}, w_{33}, \dots, w_{3n}\}$, where W_{3i} is the feature The corresponding weight in the word sequence is called the mixed interest model(The model construction process is shown in Fig.1).

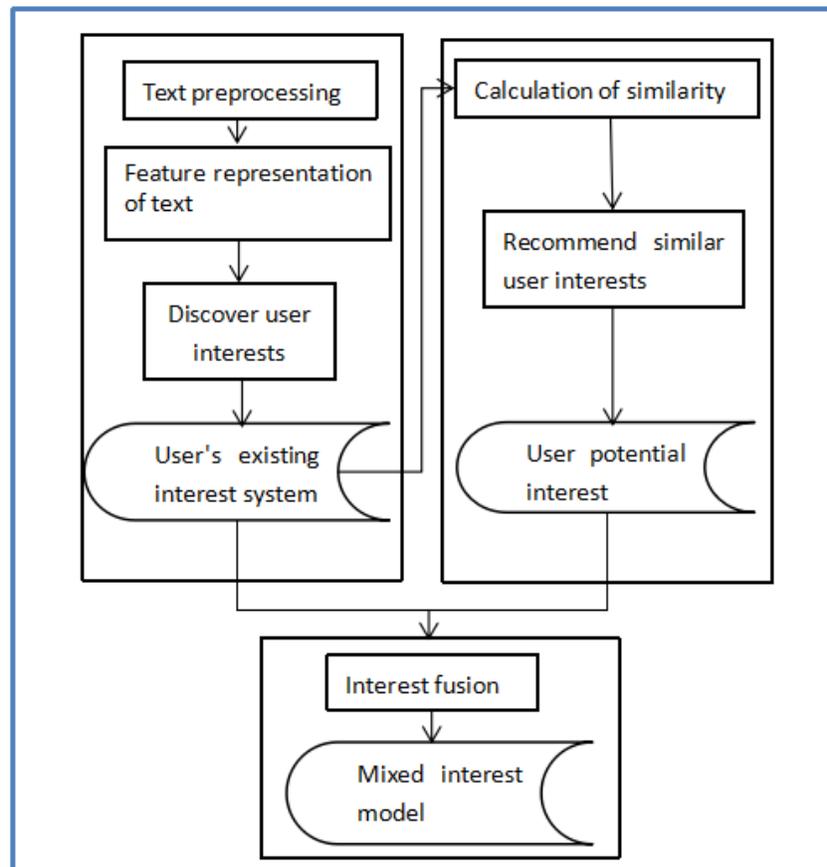


Figure 1: Construction process of mixed interest model

In the mixed interest model, the user's existing interest model can be divided into the content-based recommendation algorithm based on the principle. The user potential interest model belongs to the content of the collaborative filtering algorithm, so this is the basic framework of the mixed interest recommendation model.

4.1 User's Existing Interest Model Design

The searched content should be structured before establishing the user's interest model. The typical solution is TF_IDF (term frequency_inverse document frequency) notation. This notation uses weights to measure the importance of words. TF (term frequency) word frequency, that is, the proportion of a specific word in all words in the text information. In all the text content, "the", "these", "those", and similar common words with no real meaning usually have a high frequency. Therefore, in order to improve the accuracy of the extracted feature words, IDF (inverse document frequency) inverse document

frequency concept. Assuming that a certain word is relatively rare, but appears more frequently in a certain document, these words are likely to reflect the characteristics of the article, and most likely are the characteristic words of the article content. A new weight parameter is set for this situation, IDF, which is inversely proportional to the commonness of the vocabulary. The weight calculation method of TF_IDF is $TF * IDF$,

$$[\text{freq}(i,j)/\text{sum}(k,j)] * \log[N/n(j)], \quad (1)$$

$TF = \text{freq}(i,j)/\text{sum}(k,j)$, where $\text{freq}(i,j)$ is the number of occurrences of word i in the search content set d_j ; $\text{sum}(k,j)$ is all in the search content set d_j the number of vocabulary. $IDF = \log[N/n(j)]$, where N is the total number of searched content sets, and $n(j)$ is the number of searched content that has appeared in vocabulary i .

Given the search content set $D = \{d_1, d_2, \dots, d_n\}$, and the feature word set $S = \{s_1, s_2, \dots, s_n\}$, the search content set can be expressed as corresponding to the feature word set S A vector space model, $d_i = \{w_{i1}, w_{i2}, \dots, w_{ij}, \dots, w_{ik}\}$, where w_{ij} represents the weight of the feature word s_j in the search content set d_i , if w_{ij} is 0, it means the search content There is no feature word s_j in the set d_i . Therefore, the search content set can be equivalent to a weight matrix:

$$DM = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1k} \\ w_{21} & w_{22} & \cdots & w_{2k} \\ \vdots & \vdots & & \vdots \\ w_{n1} & w_{n2} & \cdots & w_{nk} \end{bmatrix}$$

(2)

4.2 User Potential Interest Model Design

The difference between the user's potential interest model and the user's existing interest model is that the user's potential interest cannot be found directly through previous searches. This paper proposes to use collaborative filtering algorithm to solve the problem. The traditional collaborative filtering algorithm finds similar users through a scoring matrix. Usually, different users will buy or browse the same items but purchase different items. It is difficult for these users who browse the same items to be classified as similar users. To solve the above problems, we only need to calculate the similarity $\text{sim}(u,v)$ of the search content of different users.

4.2.1 Similarity Calculation

The core part of the collaborative filtering algorithm is to find users with the same interest, and its efficiency and results largely determine the efficiency and results of the collaborative filtering algorithm. The method of measuring the similarity of the search content of user i and user j is as follows: first obtain all items of the weights of the feature words in the search content set of users i and j , and then calculate the similarity of users i and j through the similarity measurement method, denoted as $\text{sim}(i, j)$.

Usually there are three calculation methods for $\text{sim}(i,j)$: cosine similarity calculation method, Pearson correlation coefficient, Euclidean metric. This article uses cosine similarity calculation method.

Suppose user u searches for content set $D_u = \{du_1, du_2, \dots, du_i, \dots, du_m\}$, $EM_u = (w_{1u1}, w_{1u2}, \dots, w_{1uj}, \dots, w_{1uk})$, user v searches for content set $D_v = \{dv_1, dv_2, dv_j, \dots, dv_s\}$, D_u, D_v are both subsets of D , $EM_v = (w_{1v1}, w_{1v2}, \dots, w_{1vj}, \dots, w_{1vk})$, the content similarity of users u, v is as follows:

$$\text{sim}(u, v) = \cos(u, v) = \frac{EM_u \bullet EM_v}{|EM_u| \times |EM_v|} \quad (3)$$

4.2.2 Recommend Interest Words of Similar User Groups and Build Models

Through the above algorithm, the similarity between the search content of the current user and all other users can be calculated, and the n users with the highest similarity to the current user can be arranged as neighbor groups. The collaborative filtering algorithm recommends the existing interest model of users in the neighbor group to the current user, which is the potential interest model of the user.

4.3 Design of Hybrid Recommendation Algorithm Model

After obtaining the user's existing and potential interest models, the two interest models are combined according to the rules, and the similarity is calculated with the candidate recommended item set, given the similarity threshold a , and the recommendation result is checked.

Suppose user u 's $EMu=(w1u1, w1u2, \dots, w1uj, \dots, w1uk)$, $PMu=(w2u1, w2u2, \dots, w2uj, \dots, w2uk)$, $HMu=(w3u1, w3u2, \dots, w3uj, \dots, w3uk)$, candidate recommended content set $d=(wd1, wd2, \dots, wdj, \dots, wdk)$. The calculation method of $w3$ is:

$$w3 = \max \{w1, w2\}$$

where $\max\{\}$ represents the larger value of $w1$ and $w2$. Finally, use the cosine method to calculate the similarity between d and HM , and check the recommended results.

Algorithm 4: HM construction algorithm and recommendation result generation are as follows:

Input: user u 's EMu , PMu , candidate search content set d , threshold a ;

Output: the recommendation result of candidate search content set d for user u .

Begin

$multi = 0$, $moduleu = 0$, $moduled = 0$, $isRecommend = 0$;

$HMu = (w3u1, w3u2, \dots, w3uj, \dots, w3uk)$;

 for each $s_j \in S$

$w3uj = \max(w1uj, w2uj)$;

 end for;

 for each $S_j \in S$

$multi = W1uj * W1vj$;

$moduleu = W1uj * W1uj$;

$modulev = W1vj * W1vj$;

 end for;

$sim(u, v) = multi / (\text{sqr}(moduleu) * \text{sqr}(moduled))$;

 if $sim(u, d) > a$

$isRecommend = 1$;

 end if;

End

5 Experimental Results and Analysis

5.1 Data Representation

In the hybrid recommendation system introduced in this article, recommendation results are generated according to the different weights of each feature word in different search content sets. The weight matrix can be represented by an $m \times n$ matrix. The m row represents the search content set of m users, the n column represents n feature words, and the element W_{ij} in the i -th row and the j -th column represents the weight of the j -th feature word in the search content set of the i -th user. The weight matrix is shown in Table 2.

Table 2: Data Representation

	Term ₁	Term _k	Term _n
Content ₁	W ₁₁	W _{1k}	W _{1n}
.....
Content _j	W _{j1}	W _{jk}	W _{jn}
.....
Content _m	W _{m1}	W _{mk}	W _{mn}

5.2 Experimental Evaluation Index

After recommending the recommendation results to users, the following results are obtained (the recommendation results are shown in Table 3):

Table 3: Recommended results

	Recommend	Not recommended
Visited	Recommend this item and was visited (rv)	The item was not recommended but was visited (nv)
Not visited	Recommended not visited (rn)	The item is neither recommended nor visited (nn)

According to the several possibilities of the result, the accuracy rate (or precision rate, Precision) and coverage rate (or recall rate, Recall) are usually used as the index of algorithm evaluation. The calculation formula for accuracy is:

$$precision = \frac{rv}{rv + rn} \tag{4}$$

Indicates the ratio of the number of recommended and visited items to the total number of recommended items, and the coverage rate calculation formula is:

$$Recall = \frac{rv}{rv + nv} \tag{5}$$

Represents the ratio of the number of recommended hits to the total number of items visited by users in the test set.

In fact, the precision and recall are in conflict with each other. If you increase the number of recommended items, the coverage rate increases, but at the same time the accuracy rate decreases. Therefore, the two are usually given a comparable weight and combined into a comprehensive measure F to evaluate the quality of recommendation. The larger the F value, the higher the recommended quality. Calculated as follows:

$$F = \frac{2 \times Recall \times precision}{Recall + precision} \tag{6}$$

5.3 Experimental Program

The data in this article is taken from Datacastle's user browsing data set, randomly selected 1,000 users, and the number of recommended items from 10 to 60 for experiments. In the process of discovering the potential interests of users, it is necessary to determine the size of the neighbor group, that is, the number of users with similar interests. In order to facilitate the evaluation of performance, we fixed the size of the neighbor group to 35 people, and the similarity algorithm uses the cosine calculation method. The final result of the recommendation algorithm is to generate N items that users may be most interested in, so that users can choose, and examine different algorithms under different N values (based on content recommendation algorithm, collaborative filtering algorithm, this article introduces the algorithm) The accuracy, coverage, and F value.

Therefore, in the case of different recommended items (N),

The accuracy, Coverage rate and F value of each algorithm is as follows:

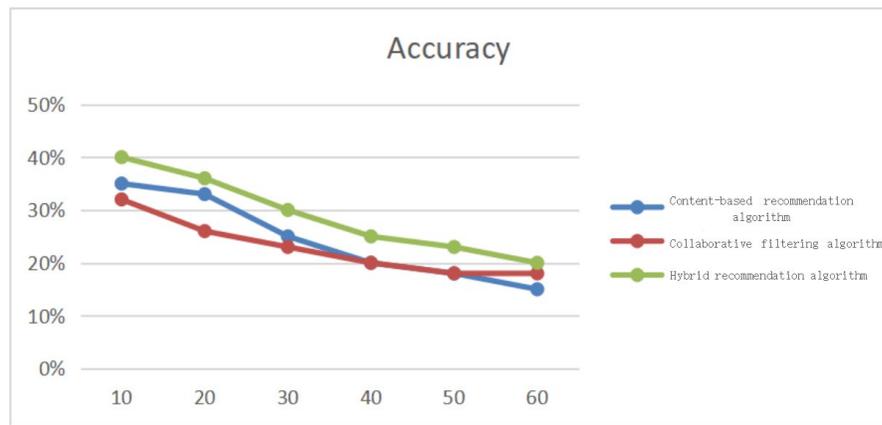


Figure 2: Accuracy

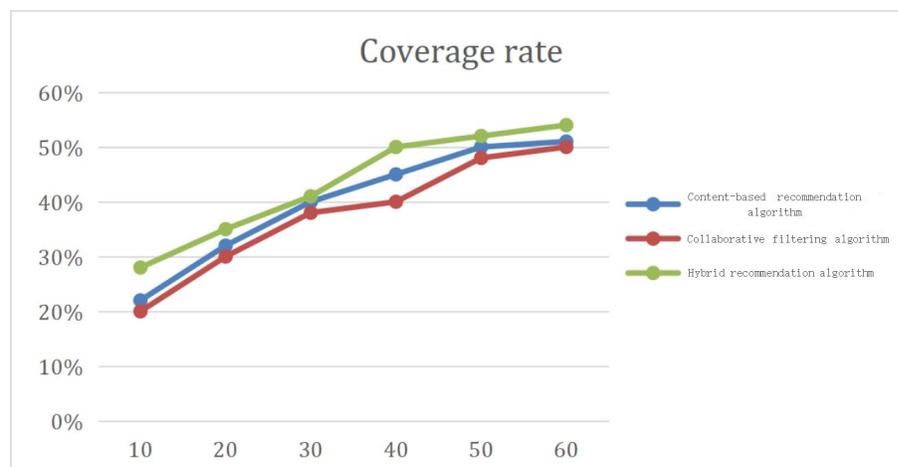


Figure 3: Coverage rate

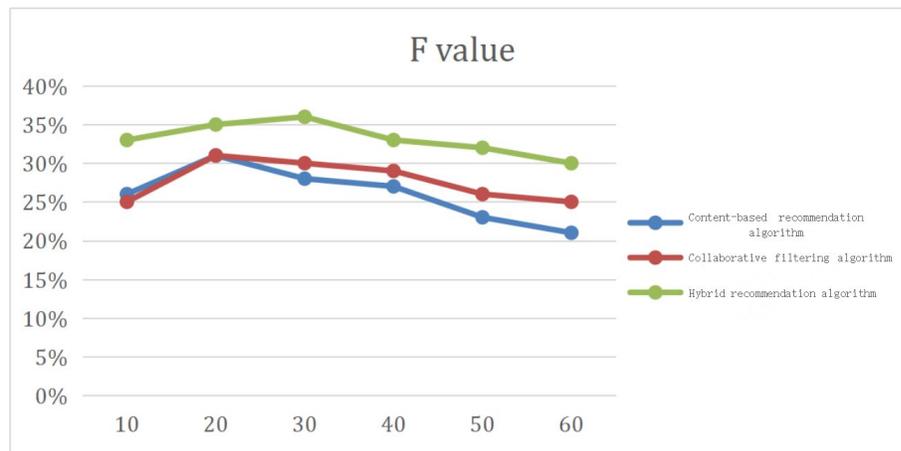


Figure 4: F value

From the above experimental results (Figs. 2–4), we can draw the following conclusions:

1. Accuracy and coverage are two reciprocal parameters. As the number of recommended items increases, accuracy decreases and coverage increases.
2. As a comprehensive parameter weight, F has a peak value as the number of recommended items increases, and then slowly decreases.
3. According to Fig. 4, the algorithm introduced in this paper is better than collaborative filtering algorithm and content-based recommendation algorithm, and there is no cold start problem.

6 Summary and Outlook

The recommendation system has gone through a long period of research and development and has achieved remarkable results. The function of personalized recommendation system is mainly manifested in three aspects: 1) Turn e-commerce viewers into buyers; 2) Improve the ability of e-commerce websites to cross-sell; 3) Improve user experience and increase user loyalty.

But further efforts are needed. It is undeniable that there are still many difficulties in the recommendation system that have not been broken through, such as extracting accurate user preferences and object characteristics; multi-dimensional research on recommendations; research on the security of recommendation systems. However, we firmly believe that with the development of society, technological advancements With continuous progress, the research on recommender systems will become more and more in-depth, so as to better serve people's material and cultural life.

Acknowledgement: This work was supported in part by the National Natural Science Foundation of China, Grant No. 72073041. Open Foundation for the University Innovation Platform in the Hunan Province, Grant No. 18K103. 2011 Collaborative Innovation Center for Development and Utilization of Finance and Economics Big Data Property. Hunan Provincial Key Laboratory of Finance&Economics Big Data Science and Technology. 2020 Hunan Provincial Higher Education Teaching Reform Research Project under Grant HNJG-2020-1130, HNJG-2020-1124. 2020 General Project of Hunan Social Science Fund under Grant 20B16. Scientific Research Project of Education Department of Hunan Province (Grand No. 20K021), Social Science Foundation of Hunan Province (Grant No. 17YBA049).

Funding Statement: The authors did not receive any specific funding for this study.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- [1] L.Y. Zhang, "Analysis on the development of "Internet +" E-commerce," *Modern Marketing (late issue)*, vol. 14, no. 3, pp. 164, 2018.
- [2] Q. Zhang and Y. Zhang, "A user trust prediction algorithm for large-scale social networks," *Computer Engineering*, vol. 44, no. 8, pp. 1–9, 2018.
- [3] S. T. Shang, M. Y. Shi, W. Q. Shang and Z. G. Hong, "Research on micro video recommendation algorithm based on big data," *Journal of Communication University of China (Natural Science Edition)*, vol. 24, no. 2, pp. 38–45, 2017.
- [4] K. Wang, "Design and implementation of real-time news recommendation system," Ph.D. dissertation, Beijing Jiaotong University, Beijing, 2017.
- [5] G. H. Sun, D. Q. Liu and M. Y. Li, "Overview of personalized recommendation algorithms," *Software*, vol. 38, no. 7, pp. 70–78, 2017.
- [6] Z. F. Jiang, J. Yang, E. Hai-Hong, P. C. Namp and C. Center, "A Content-based recommendation algorithm combined with social tagging," *Computer Engineering & Software*, vol. 36, no. 1, pp. 1–5, 2015.
- [7] X. D. Xiang and Z. X. Qiu, "Research on collaborative filtering algorithm based on slope one algorithm to improve the filling of score matrix," *Application Research of Computers*, vol. 37, no. 5, pp. 1–5, 2019.
- [8] R. Duwairi and H. Ammari. "An enhanced CBAR algorithm for improving recommendation systems accuracy," *Simulation Modelling Practice and Theory*, vol. 60, pp. 54–68, 2016.
- [9] H. L. Yang, S. L. Li and W. J. Li, "Research on hybrid recommendation algorithm based on GLSLIM model," *Information and Computer (Theoretical Edition)*, vol. 24, no. 20, pp. 77–80, 2017.
- [10] C. Cui, M. Q. Hu, J. D. Weir and T. Wu, "A recommendation system for meta-modeling: A meta-learning based approach," *Expert Systems With Applications*, vol. 46, pp. 34–44, 2016.
- [11] F. R. Yang, Y. J. Zheng and C. Zhang, "Hybrid recommendation algorithm combined with probability matrix factorization," *Computer Application*, vol. 38, no. 3, pp. 644–649, 2018.
- [12] S. Yang and J. Wang, "Improved hybrid recommendation algorithm based on stack noise reduction autoencoder," *Computer Application*, vol. 38, no. 7, pp. 1866–1871, 2018.
- [13] H. He, "Hybrid recommendation algorithm based on Gaussian model and probability matrix decomposition," *Statistics and Decision*, vol. 34, no. 3, pp. 84–86, 2018.
- [14] Y. Liang and B. Q. Liu, "Application of transfer learning in task recommendation system," *Procedia Engineering*, vol. 174, pp. 518–523, 2017.