

Research Article

Hybrid Algorithm Based on Content and Collaborative Filtering in Recommendation System Optimization and Simulation

Lianhuan Li ¹, Zheng Zhang ², and Shaoda Zhang ³

¹School of International Education, Nanyang Medical College, Nanyang, Henan 473000, China

²School of Computer and Software, Nanyang Institute of Technology, Nanyang, Henan 473000, China

³Cofee Medical Technology Company Limited, Shenzhen, Guangdong 518101, China

Correspondence should be addressed to Zheng Zhang; zhangzheng@nyist.edu.cn and Shaoda Zhang; zhangsd@stu.xmu.edu.cn

Received 12 April 2021; Revised 25 April 2021; Accepted 10 May 2021; Published 19 May 2021

Academic Editor: Yi-Zhang Jiang

Copyright © 2021 Lianhuan Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper explores and studies recommendation technologies based on content filtering and user collaborative filtering and proposes a hybrid recommendation algorithm based on content and user collaborative filtering. This method not only makes use of the advantages of content filtering but also can carry out similarity matching filtering for all items, especially when the items are not evaluated by any user, which can be filtered out and recommended to users, thus avoiding the problem of early level. At the same time, this method also takes advantage of the advantages of collaborative filtering. When the number of users and evaluation levels are large, the user rating data matrix of collaborative filtering prediction will become relatively dense, which can reduce the sparsity of the matrix and make collaborative filtering more accurate. In this way, the system performance will be greatly improved through the integration of the two. On the basis of the improved collaborative filtering algorithm, a hybrid algorithm based on content and improved collaborative filtering was proposed. By combining user rating with item features, a user feature rating matrix was established to replace the traditional user-item rating matrix. *K*-means clustering was performed on the user set and recommendations were made. The improved algorithm can solve the problem of data sparsity of traditional collaborative filtering algorithm. At the same time, for new projects, it can also predict users who may be interested in new projects according to the matching of project characteristics and user characteristics scoring matrix and generate push list, which effectively solve the problem of new projects in “cold start.” The experimental results show that the improved algorithm in this paper plays a significant role in solving the speed bottleneck problems of data sparsity, cold start, and online recommendation and can ensure a better recommendation quality.

1. Introduction

The recommendation system excavates the resources that users may be interested in or need according to the interest characteristics of different users from the mass information and makes recommendations [1]. As a BI [2] platform based on massive data mining, it is considered to be one of the most effective tools to solve the information explosion. In essence, the recommendation system evaluates the likes of some products that the user has never touched by analyzing the resources selected by the user and feedbacks the products with the highest likes among the predicted results to the user.

Nowadays, most recommendation systems tend to sacrifice recommendation quality in order to meet users'

requirements for real-time recommendation. However, the long-term unsatisfactory recommendation quality will reduce users' satisfaction with the website. In addition, some core algorithms of the recommendation system also have problems such as sparse data, narrow application scope [3], and cold start [4]. For a website with a large number of items, users will only have comments on a small part of them, and different users will choose different categories of items for evaluation, which leads to the sparsity of the user evaluation item matrix and seriously affects the quality of recommendation. In addition, the recommendation algorithm is mostly based on the user's evaluation items to make predictions, so for newly registered users, because there is no historical information, the system could not know their

interests and hobbies to make recommendations. For newly launched projects, because no users have made comments on them, they will not be pushed by the system. This is what we call the cold start problem.

In order to solve the above problems and better serve users, this paper studies the recommendation system based on the hybrid algorithm of content and collaborative filtering and its core algorithm, which solves the speed bottleneck and data sparsity problems to a certain extent, ensures the recommendation quality, and promotes the development and application of the recommendation system.

2. Related Work

At present, there are quite a few e-commerce systems using recommendation technology to improve the revenue of enterprises. According to the recommendation technology adopted, the e-commerce recommendation system can be divided into two types: rule-based recommendation system and information filtering recommendation system, and the information filtering recommendation system can be divided into content-based filtering recommendation system and collaborative filtering recommendation system.

The rule-based recommendation systems, such as IBM Web Sphere [5], allow system administrators to formulate rules according to users' static characteristics and dynamic attributes. In essence, a rule is an if-then statement, which determines how to provide different recommendation services under different circumstances. The advantage of a rule-based system is that it is simple and direct, but the disadvantage is that the quality of rules is difficult to guarantee and cannot be updated dynamically. In addition, as the number of rules increases, the system will become more and more difficult to manage. Content-based filtering systems such as Personal WebWatcher [6], Syskill & Webert [7], Letizia [8], CiteSe [9], IFWEB, WebACE [10], Elfish, and WebPersonalizer use similarity between resources and user interests to filter information. The advantages of the system based on content filtering are simple and effective, but the disadvantages are that it is difficult to distinguish the quality and style of the content of resources, and it could not find new resources that users are interested in, but can only find resources similar to users' own interests. Collaborative filtering systems, such as WebWatcher [11] and Let us Browse [12], use the similarity between users to filter information. The advantage of the collaborative filtering system is that it can find new information of interest for users. Failures are two very difficult problems to solve. One is sparse, that is, when the system starts to use, due to system evaluation, resources have not yet received enough system, and it is difficult to use evaluation to find similar users. The other is scalability. As the number of users and resources of the system increases, the performance of the system will become lower and lower.

The specification of the corpus provided by TREC can be used [13]. They extract feature vectors from the topic description and cases in the document as the initial user demand model and interact with the training set for each topic

to set the initial threshold. Then, you use the test set to determine whether the documents in the test set are greater than a certain topic of the Fujian value condition (if the condition is met). Then, feature vectors are extracted from the positive and negative example document sets to update the initial model, and a filtering system based on the vector model is established [14]. The research on Chinese text filtering technology uses vector space model, demand, and text matching technology based on user demand example keywords in the Chinese text filtering model they proposed and uses vector angle cosine as the similarity coefficient [15]. Based on the text representation of the text, starting from the information of the text logic level, the text structure analysis technology is introduced to improve the matching efficiency of documents and queries in the retrieval of text fragments [16]. In addition, they combined the content-based filtering method with the collaborative filtering method to build a hybrid model of text filtering, so as to make better use of user evaluation information. It proposes an information filtering model based on Bayesian network BMIF, describes the basic structure of information filtering, and provides 6 kinds of nodes to describe the relationship between information filtering events [17]. On this basis, it provides a variety of BMIF usages, uses the vocabulary knowledge represented by BMIF, combines automatic learning with manual interaction, and combines collaborative filtering with content filtering.

The recommendation algorithm, as the key of the recommendation system, is the target of researchers' in-depth study. The excellent algorithm not only requires stable and accurate operation but also must contact the application environment with a certain universality, so it can meet the needs of the user key to see the effective use of the algorithm, because the method recommended by the algorithm will be different, which requires researchers to study rational selection through the experiment and targeted trade-off for the field. After 20 years of development of the recommendation system, scholars have made use of knowledge in different fields to improve the recommendation algorithm from multiple perspectives and put forward different recommendation algorithms [18]. At present, the recognized recommendation algorithms include collaborative filtering recommendation, content-based recommendation, knowledge-based recommendation, and hybrid recommendation algorithm [19]. The collaborative filtering recommendation algorithm is the earliest recommendation algorithm [20]. The characteristic of this algorithm is to find similar users of target users and then recommend products or information for target users by using similar users. The recommendation advantage of the collaborative filtering algorithm is that, under the condition of relatively dense rating data, collaborative filtering can bring better recommendation effect for users than other algorithms. However, the resource required by this algorithm is the user-project rating data, so it does not need to know the user information or make clear the project information. Because the information of the user and the project is not easy to obtain, the algorithm has a low complexity table and a wide range of application fields [21]. However, the reliance on scoring also becomes a defect of the

collaborative filtering algorithm. When the scoring matrix is sparse, the recommendation effect is greatly reduced. In view of the imperfection of the collaborative filtering recommendation algorithm, scholars began to improve collaborative filtering from multiple perspectives. The LDA topic model was introduced into the collaborative filtering algorithm [22], and the potential factor vectors of users and projects were obtained by using the LDA topic model. The sparse matrix was reduced by dimensionality reduction of the scoring matrix, and then, the similarity between users and projects was calculated in the low-dimensional space. Collaborative filtering is improved by defining the similarity of user attributes in social networks [23] and giving the calculation method of attribute similarity. Considering the length and dimension of the space vector comprehensively, the improved cosine similarity is used to find similar users [24]. Based on practical experience, the expert trust factor of users is introduced into the algorithm [25]. Considering the structural relationship between users, a modeling method for time-series behavior among users is proposed [26]. Some scholars seek similar users by effectively analyzing the trust relationship between users [27]. Some other scholars use relevant theories and methods to cluster items and predict unrated items in the same category to densify the scoring matrix in order to mitigate the adverse impact of scoring sparsity on the collaborative filtering algorithm [28].

Aiming at the problems of the current recommendation system and combining the advantages and disadvantages of content filtering and collaborative filtering, a content-based filtering is proposed. In the hybrid mode based on the combination of collaborative filtering recommendation technology, the recommendation filtering technology is fully utilized. The advantages of the two recommendation technologies make up for each other's shortcomings, improve the performance of the recommendation system, help companies improve the quality of customer service, enhance the cross-selling capabilities of the e-commerce system, improve corporate competitiveness, and provide customers with more accurate and personalized services. Using the collaborative filtering recommendation method recommended by the user data set and then based on the content reuse recommendation method, the recommended data set is filtered by establishing a model interested in the recommended data set. In order to realize that the content-based recommendation method is used to optimize the results of collaborative filtering recommendation, the purpose of this hybrid is recommended, but the purpose is very strong.

3. Hybrid Recommendation Algorithm and Recommendation Strategy Based on Content and Collaborative Filtering

3.1. Research on the Hybrid Recommendation Model Based on Content and Collaborative Filtering. The schema framework of the hybrid mode recommendation algorithm based on content and collaborative filtering is shown in Figure 1.

The whole recommendation is divided into two modules, namely, content filtering recommendation

module and collaborative filtering recommendation module, both of which are invisible to users. The preparation process of the dataset of the recommendation algorithm is as follows: first, the user's interest is extracted from the shopping history data of the user and the topic vector and feature vector preprocessed by the network log, and the data processing is established based on the recommendation module of content filtering. Then, based on user interest characteristics, user rating data, and current access sequence data, a recommendation module based on collaborative filtering is constructed to extract the nearest neighbors of the user and the nearest neighbors of the current access sequence (item). Then, it integrates two recommendation weighted sum calculation modules for the similarity calculation model of mixed recommendation (i.e., recommendation processing) to generate the recommended top visit sequence. The web server recommends the sequence to the user and accesses the recommendation sequence on the user, the adaptive adjustment of the recommended model, and the idle speed value of the feedback information to obtain the best recommendation data.

In order to realize personalized recommendation service, we must first collect the user's personal information and establish the user's interest characteristic description model. The ratio of the time spent browsing a web page to the number of characters on the page can effectively reveal the user's interest, which is related to the categories of information, and these categories are determinable and relatively stable. Users browse the Web information including each page clicks, stay time, access, and order, and each page URL can be found in the Log of the proxy server, and the user visited Web pages can be found in the Cache of the proxy server, so you can go through the Web mining way to get the user's interest.

The optimal feature items are those words with the largest mutual information amount with the related text set $Rel(Q)$. The logarithmic mutual information amount between the words and the related text set is calculated by the following formula:

$$\log NI(\alpha, rel(Q)) = \log \left(Q \left(\alpha | \alpha \in \frac{rel(Q)}{Q(\alpha)} \right) \right). \quad (1)$$

The cosine similarity between user preference document and project document is

$$\text{sim}(\beta, e_j) = \cos(e_j, e_c) = \frac{\sum_i \alpha_i \varepsilon_{ij}}{\sqrt{\alpha_i^2} \sqrt{\varepsilon_{ij}^2}} \quad (2)$$

The higher the calculated similarity, the more preference the users have for this feature. The biggest problem facing TF-IDF is the choice of features. The content category of users is based on the similarity between user interests, that is, the similarity between user feature vectors. Here, the commonly used method of cosine of included angle is selected. The similarity of user interests is

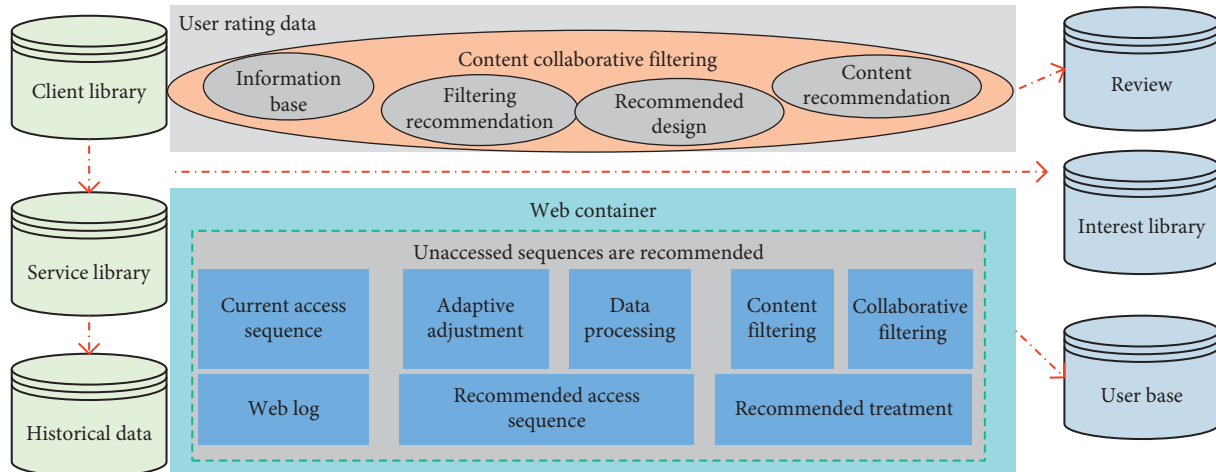


FIGURE 1: Schematic diagram of a hybrid recommendation model based on content and collaborative filtering.

$$\text{sim}(U, V) = \sum_i (u_i, v_i) * \frac{1}{\sqrt{u_j^2} \sqrt{v_i^2}} \quad (3)$$

Clustering is carried out according to the similarity between user feature vectors so that users with similar interests can be classified into one group for easy processing. Meanwhile, for new product information documents, a list of recommended users can be obtained by judging their categories. It is assumed that the classification of user sets is controlled manually, so the recommender system clustering method can be adopted.

After the initial recommendation model has been established and the threshold has been set, the similarity between the text about the product introduction in the product information database and the model vector of a user's interest topic can be calculated. If the similarity value is greater than or equal to fujian, it is considered to be associated with the user's interests. You can recommend to

the user through the Web server and then determine whether the recommendation is valid. The user can adaptively modify the model vector or adjust the proximity value according to the judgment result so that the recommended performance can be realized. The system is continuously improved to better serve the user.

If the recommended access sequence is judged to be related to the user's interest, it will browse its relevant information, and then, the recommended access sequence will become the current access sequence. When adjusting the model vector, the interest topic vector can be extracted from the current access sequence, and the feature vector can be extracted from the user's shopping history data and the Web log (the Web log has changed accordingly). The new model vector is obtained by the weighted sum operation of the topic vector and feature vector. Let the weights be A' , B' , C' , and D' , respectively, then

$$pf(Q) = A' * pf_0(Q) + B' * pf_1(Q) + C' * pf_2(Q) + D' * pf_3(Q). \quad (4)$$

Further, the more the items are rated by two users, the more similar their preferences would be if both users rated them higher or lower than the average rating of the two users. To put it simply, the more items both users like and dislike, the more similar the two users' interests and preferences are to some extent. For example, Table 1 shows the rating of some users on items. The scoring rule is 5-point scale, and the higher the rating, the greater their preference for the item. Now, let us calculate the rating prediction of User3 on Item3.

We consider a user's rating of an item greater than or equal to its average rating as a favorite and a user's rating less than its average rating as a dislike. The reason why the average score is chosen is that some users' score above 3 points means they like it, while some users' score above 4 points means they like it. Therefore, compared with the average score, it can better reflect the user's liking degree. Therefore, user interest similarity can be integrated into the

traditional similarity calculation, and the improved similarity formula of user u and user v is defined as follows:

$$\text{sim}(u, v) = \alpha * \text{PCsim}(u, v) + (1 - \alpha) * \text{Interest}(u, v). \quad (5)$$

3.2. Improved Content and Collaborative Filtering Algorithm Recommendation System Based on K-Means Clustering. A collaborative filtering recommendation algorithm based on K-means clustering is proposed. The new algorithm has two components: offline and online. When offline, first of all, users are clustered according to their characteristic data to form several clustering clusters. When online, the clustering cluster to which the target user belongs to is determined according to the similarity between the target user and each clustering center, so as to find the nearest neighbor in the cluster. Then, based on the preference of the nearest

TABLE 1: User rating data.

	Item 1	Item2	Item3	Item4	Item5
User1	3	4	4	6	
User2	2	5	2	4	2
User3	1	3	1	3	2

neighbor group to the project, we can predict the interest preference of the target users and finally get the recommendation.

The specific idea is to apply K -means clustering to collaborative filtering. For the whole user space, the similarity between users and the clustering center is calculated according to users' purchasing habits and scoring characteristics (that is, the user-item scoring matrix), and the clustering cluster is assigned to users according to the principle of nearest distance, thus the whole user space can be divided into several small groups. Based on the scoring characteristics of all users in each cluster, a virtual user is generated for each cluster. As the representative of all users in the cluster, the rating of the virtual user to the project can be the average of all users in the cluster to the project. At this point, the project ratings of all virtual users form a new search space (virtual user-project rating matrix), which replaces the original user-project rating matrix. When online recommendation is made, it only needs to calculate the similarity between target users and all virtual users, determine the cluster to which the target users belong according to the similarity level, search for neighbors in the cluster, and generate recommendation. The algorithm flow is shown in Figure 2.

For a large recommendation system, there will be a lot of user and project data. However, only a small fraction of the total project space has been evaluated by users, which is known as the data sparsity problem. By clustering, the data dimension can be greatly reduced. After neighborhood users are identified, the degree of preference of target users for unrated items can be predicted based on neighborhood users' preferences. The prediction scoring formula for the project is as follows:

$$\text{pre} = \gamma_{\alpha} + \frac{\sum (\gamma_{ij} - \gamma_i) * \text{sim}(\alpha, i)}{\sum \text{sim}(\alpha, i)}. \quad (6)$$

When the hybrid recommendation algorithm starts to run, this article first uses the judgment conditions to process the user score data on the item. The total score of the project is less than 20 users. It is recommended to use them directly based on the content filtering method, which is also considered to be collaborative. The filtering recommendation algorithm for low-scoring data is effective in the fact that there are few user recommendations. The content-based recommendation algorithm is directly used to recommend similar items for users through item feature matching. This will result in mediocre recommendations, but it will also avoid the risk of invalid recommendations due to collaborative filtering of similar users not being able to find them. Of course, the value of 20 here is not fixed. In practical application, it can be adjusted according to the situation.

In addition, in the process of establishing the hybrid algorithm model proposed in this paper, the calculation

method is not rigid with the traditional algorithm, but improved or innovated on the basis of the traditional calculation method, which is mainly reflected in the following points:

- (1) A method to optimize the user similarity calculation formula by using project heat was proposed
- (2) In order to present users' preferences more stereoscopic, the table-oriented feature extraction is carried out in the content-based recommendation algorithm, and the square-one method for calculating users' similarity using the interest model is presented
- (3) According to the characteristics of the algorithm in this paper, a method to derive the weight coefficients of different features by using variance is proposed

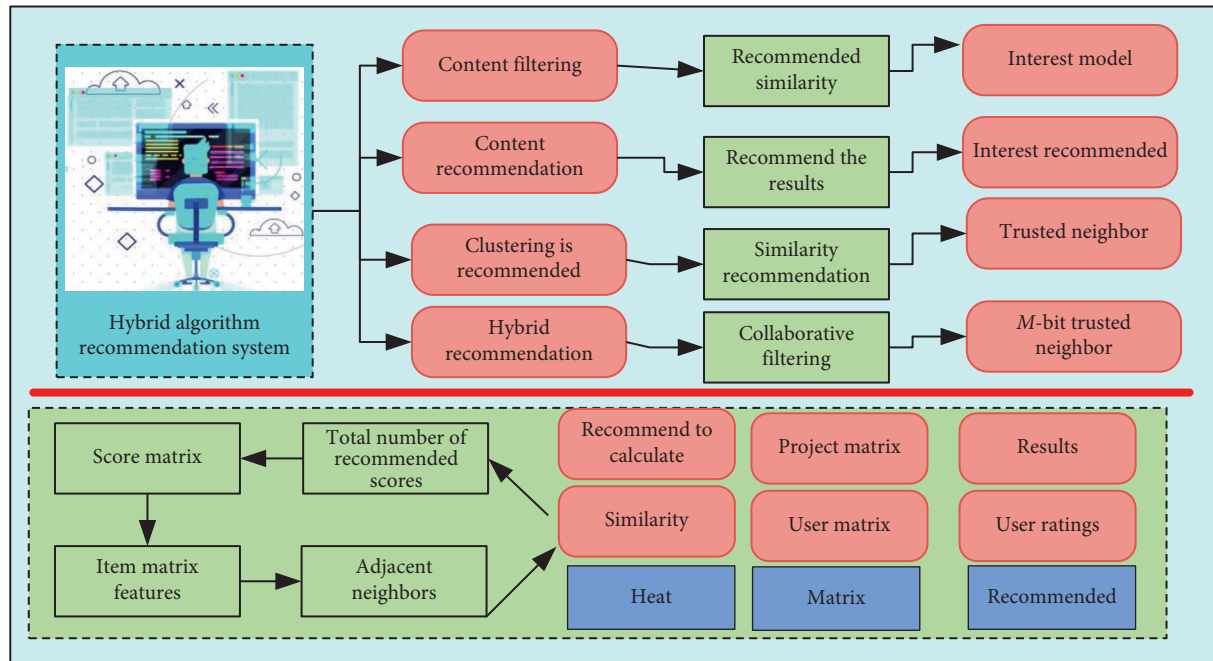
The purpose of the content-based recommendation algorithm is to effectively filter out the third category of users whose interests are different from those of the target users, and the work required in this process generally includes three steps. The first step is to extract project features. The second step is to establish the user interest model.

3.2.1. Extract Project Features. For nonliterary items, it is a difficulty to propose item features, but it is also the key to the effectiveness of the algorithm. Tags have the characteristics of complete classification and concise description. Therefore, the choice of labels as the feature description of the project can mark the characteristics of the project well from the point of view of user preferences and can also properly reflect the user's interest demands. In order to better describe the project, this paper first extracts a number of general attributes of the project, then, on this basis, refines the different attributes, and gets the different project characteristics of each attribute; the same kind of project features are included in the same project attribute.

After extracting the features of the project, the next step is to construct the user's interest model, which can also be said to be the interest vector that represents the user's interest and is composed of the project attributes and the project features contained in the attributes.

3.2.2. Establish User Interest Model. The user's interest model is composed of the project attributes and the features contained in the attributes, so for the user, what kind of project can be used to establish their interest model is a question. At this point, the user-project scoring matrix comes in handy. To build a user interest model, it is necessary to count the items that users like and extract their signature features.

3.2.3. Calculate the Similarity between Users. After extracting item features and establishing the interest model for users by using item features and scoring matrix, the next step is to consider how to use the user interest model to calculate the similarity between users. The calculation of similarity can often be transformed into distance visually.



Euclidean distance can be used to calculate the distance between points in the multidimensional space.

Replace all the features of the items that users like in the user's interest model with the corresponding number of movies containing this feature. That is, a new representation of the user interest model is obtained by summarizing the user-item attribute preference vector directly. The reason for this is that you do not need to consider what features the user prefers in the following steps, just what the user likes and dislikes about the project. The new form is shown in Table 2.

4. Test Experiment

In this paper, we use the MovieLens dataset and the scientific literature experimental dataset to test the performance of the mixed mode recommendation algorithm, the user-based collaborative filtering recommendation algorithm, and the content filtering recommendation algorithm, respectively.

We selected 12,500 scoring data from the user rating database as the experimental dataset, which contained 248 users and 1120 movies in total, among which each user rated at least 20 movies, with the score value being an integer from 1 to 5. The higher the value, the higher the user’s preference for the movie. Each record in the dataset describes information such as user ID, item ID, user rating value of the item, and timestamp. Different user ID numbers represent different users, and different item ID numbers represent different movies.

In order to measure the sparsity of the entire dataset, we introduce the concept of the sparsity level, which is defined as the percentage of unrated items in the user rating data matrix.

Experiment 1. Using the MovieLens dataset, check the performance comparison of the two algorithms under

different training set test set proportions, that is, considering different data sparsity degrees. The number of nearest neighbor users is set to 30. The experimental results are shown in Table 3 and Figure 3.

From the experimental results, the average absolute deviation of the MAE value of the mixed mode recommendation algorithm is smaller than that of the collaborative filtering recommendation algorithm based on user values. In the line chart of the MAE line of the mixed mode recommendation algorithm based on the user collaborative filtering algorithm, it is below the MAE line. As the training set data increases, the gap between the two continues to shrink. This means that the recommendation effect of the mixed mode recommendation algorithm is better than that of the recommendation algorithm based on user collaborative filtering, but this advantage decreases with the increase of the training set.

Experiment 2. Using the MovieLens dataset, check the performance comparison of the two algorithms under the condition that the proportion of the training set test set is constant, i.e., the sparsity degree, and the size of the nearest neighbor set is different, and consider the influence of the size of the nearest neighbor set on the algorithm performance.

When the partition coefficient $x = 0.8$ was selected (it can be seen from Experiment 1 that the prediction effect was better when $x = 0.8$), the experimental results are shown in Table 4 and Figure 4.

It can be found from the experimental results that the average absolute deviation MAE value of the mixed mode recommendation algorithm is smaller than the MAE value of the recommendation algorithm based on user collaborative filtering, and the MAE broken line of the mixed mode

TABLE 2: Numerical representation of the user interest model.

Number of attributes/features/users	Attribute 1	Attribute 2	...	Attribute m
User 1	23 13 9 6	21 13 14 8	...	6 2 3 4
User 2	17 11 12 8	12 13 9 8	...	6 4 2 4
...
User n	14 9 7 2	8 5 5 3	...	6 7 5 4

TABLE 3: Experimental results of different data sparsity degrees.

Proportion of training set (x)	Mean absolute deviation MAE	
	Recommendation algorithm based on user collaborative filtering	Mixed mode recommendation algorithm
0.2	0.8345	0.7868
0.3	0.7981	0.7653
0.4	0.7753	0.7575
0.5	0.7591	0.7382
0.6	0.7493	0.7297
0.7	0.7383	0.7242
0.8	0.7304	0.7204
0.9	0.7252	0.7144

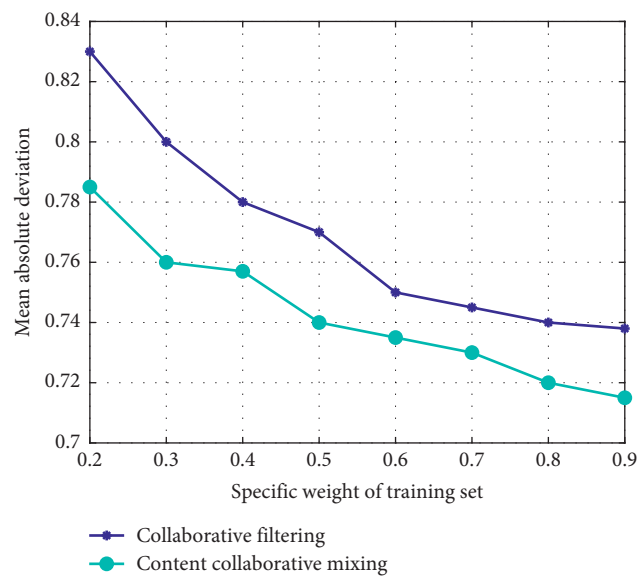


FIGURE 3: Experimental results of different data sparsity degrees.

TABLE 4: Experimental results of different sizes of nearest neighbor sets.

Proportion of training set (x)	Mean absolute deviation MAE	
	Recommendation algorithm based on user collaborative filtering	Mixed mode recommendation algorithm
10	0.7265	0.7163
20	0.7233	0.7117
30	0.7212	0.7103
40	0.7235	0.7115
50	0.7274	0.7152
60	0.7322	0.7227
90	0.7451	0.7284
160	0.7263	0.7164

recommendation algorithm is also below the MAE broken line of the recommendation algorithm based on user collaborative filtering, which also indicates that, under the

condition of the same data sparse degree, the recommendation effect of the hybrid mode recommendation algorithm is better than that of the user collaborative filtering

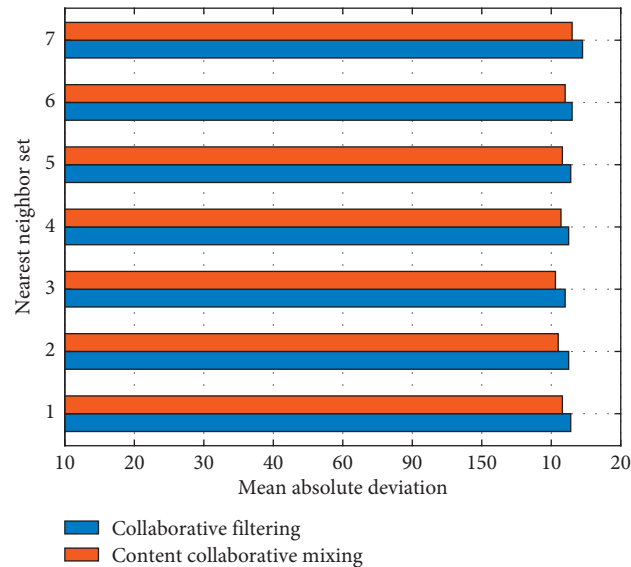


FIGURE 4: Experimental results of different sizes of nearest neighbor sets.

recommendation algorithm. With the increase of the number of the nearest neighbor sets, the recommendation effect will decrease, indicating that the more accurate prediction value can be obtained within a small range of the nearest neighbor values.

In the optimization process of the nearest neighbor selection, the filter parameters of similarity are set to eliminate the users or items that are not very similar, and then, the best value of the filter parameters is determined by combining the improvement of the score prediction part. Figure 5 shows the effect of similarity filter parameters on recommendation quality.

As can be seen from Figure 5, if the filter value of similarity is too small, that is, the filter parameter of similarity is too small and does not filter the nearest neighbor, it does not significantly improve the recommendation performance. With the increase of similarity filtering, the MAE value gradually increases, indicating that the similarity filtering parameter has gradually started to play a role in filtering out users who are not very similar. When the parameter is set to 5, the MAE value reaches the maximum and then begins to decline, that is, the recommendation performance is getting higher and higher. The collaborative filtering recommendation algorithm based on content recommendation quality is slightly higher than that of the collaborative filtering recommendation method. If the content filter-based recommendation method is not used in the case of “false neighbors,” it is also effective to find similar users by optimizing the Pearson correlation coefficient; therefore, introducing heat project to optimize the Pearson correlation coefficient method is effective.

The following four experiments were used to compare the performance of the three algorithms under different sparsity degrees, and the training set and test set were randomly selected from the dataset according to a certain proportion. Here, the number of nearest neighbors is chosen as 4 to test the performance of MAE values of the three algorithms under different sparsity.

Figure 6 shows the comparison of the recommendation quality of the three algorithms under four experiments. Green represents user-based CF, blue represents item-based CF, and red represents the hybrid algorithm proposed in this paper. As can be seen from the figure, under the same other conditions, MAE values of the three algorithms are continuously decreasing with the decrease of dataset sparsity in the four experiments, that is, recommendation quality is continuously increasing. If we carefully look at the figure above, we can find that, in Experiments 1 and 2, the optimization degree of the hybrid algorithm proposed in this paper is higher than that in Experiments 3 and 4. That is to say, the more sparse the dataset is, the more obvious the advantage of using the hybrid algorithm is.

It serves as a reference in the collaborative filtering algorithm and fixes the relationship between M and N . Assuming that compared with the number of neighbors $[FI\ N\ \text{and the change of parameter } M]$, for $M = 2N/3$, the collaborative filtering recommendation algorithm and the proposed hybrid recommendation method have a change in recommendation effect. In addition to comparing the effects of hybrid algorithms and collaborative filtering algorithms, the purpose of this step is also recommended. It also includes finding the optimal value of the number of users N as a fixed value, so as to obtain the relationship between M and N through further experiments. The experimental results are shown in Figure 7.

As can be seen from Figure 7, the recommendation accuracy of the hybrid algorithm proposed in this paper is better than that of the collaborative filtering algorithm. Moreover, with the increasing number of neighbors, the recommendation quality of the two recommendation algorithms both shows a trend of first increase and then decrease. When $N=4$, MAE of the collaborative filtering recommendation algorithm is the minimum, and the quality of the recommendation algorithm is the best. When the relationship between M and N is fixed and $M = 2N/3$, the recommended quality is the best. The

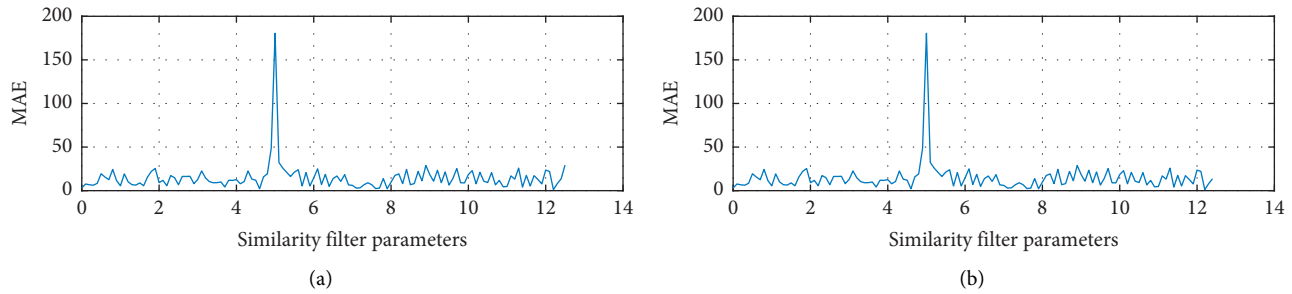


FIGURE 5: Influence of similarity filtering on user recommendation quality.

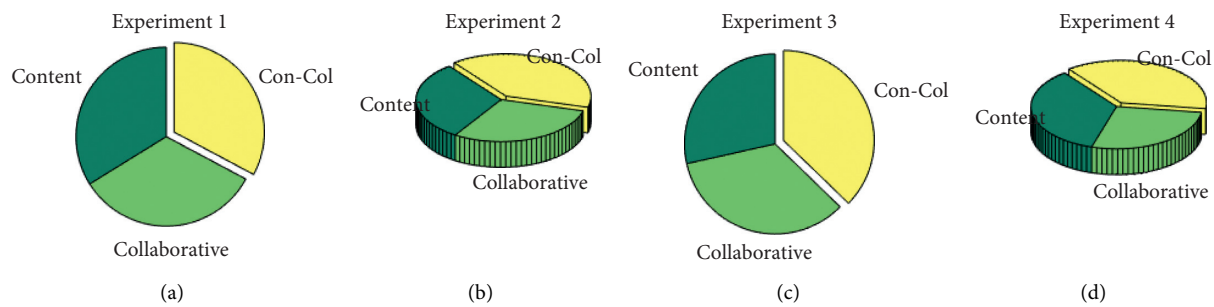


FIGURE 6: Performance of algorithms with different sparsity degrees. (a) Experiment 1. (b) Experiment 2. (c) Experiment 3. (d) Experiment 4.

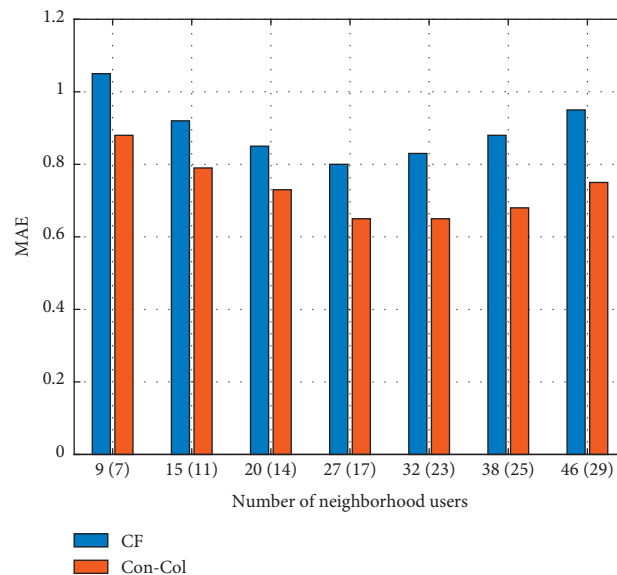


FIGURE 7: The trend of recommended quality when the number of neighbors changes.

reasonable selection of the number of similar users has great influence on the recommendation quality of the recommendation algorithm.

As can be seen from Figure 8, among the three mixed recommendation algorithms, the recommendation algorithm proposed in this paper has the lowest MAE value, that is, the validity of the recommendation algorithm is the best.

Moreover, with the increasing number of neighbors N , the advantage becomes more and more obvious. In addition, after the MAE of the three algorithms reaches the lowest, the IVIAE bar graph of the hybrid algorithm proposed in this paper rises more gently, indicating that compared with the other two recommendation methods, the hybrid recommendation algorithm proposed in this paper is insensitive to

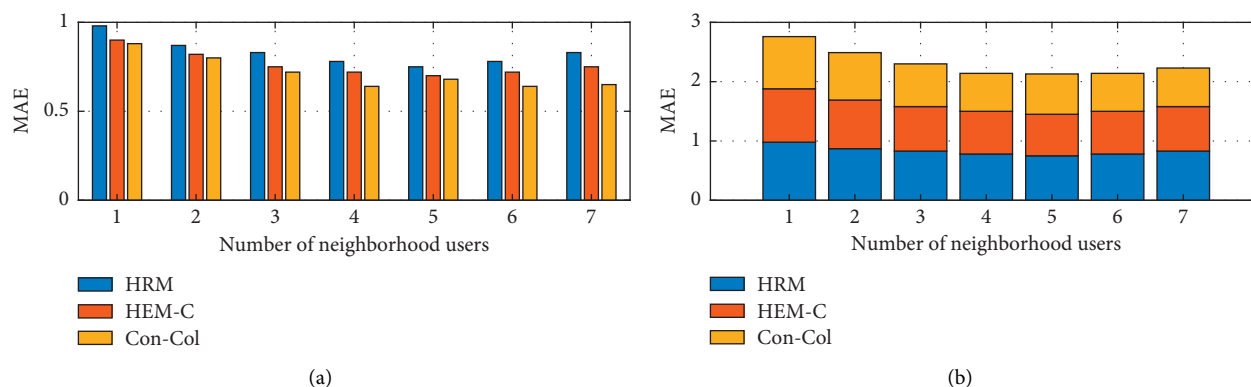


FIGURE 8: Comparison of recommendation validity of mixed recommendation methods.

the change of the number of neighbors, and the recommendation is more stable.

5. Conclusion

Combining the advantages and disadvantages of content filtering and collaborative filtering, we proposed a hybrid recommendation technology based on content filtering and collaborative filtering and studied the workflow, user characteristic description, data processing algorithm, and recommendation strategy of the hybrid recommendation technology in detail. This method not only makes use of the advantages of content filtering but also can carry out similarity matching filtering for all items, especially when the items are not evaluated by any user and can be filtered out and recommended to users, avoiding the problem of early level. At the same time, this method also takes advantage of the advantages of collaborative filtering. When the number of users and evaluation levels are large, the user rating data matrix of collaborative filtering prediction will become more dense, which can reduce the sparsity of the matrix and make collaborative filtering more accurate. The performance of the mixed mode recommendation technology based on content and collaborative filtering is verified. We designed experiments using the MovieLens dataset and the scientific literature experiment dataset to test the system performance of the mixed mode recommendation technology, user-based collaborative filtering recommendation technology, and content-based filtering recommendation technology, respectively. The experimental data show that the performance of the hybrid mode recommendation technology based on content filtering and collaborative filtering is better than that of the two technologies.

Data Availability

No data were used to support this study.

Consent

Informed consent was obtained from all individual participants included in the study references.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Acknowledgments

This work was supported by Henan Soft Science Research Program Project (212400410192): Research on the Application of Recommendation Algorithm Based on Multivariate Collaborative Filtering in Medical Practice Qualification Examination.

References

- [1] R. N. Behera and S. Dash, "A particle swarm optimization based hybrid recommendation system," *International Journal of Knowledge Discovery in Bioinformatics*, vol. 6, no. 2, pp. 1–10, 2016.
- [2] T. Horváth and A. de Carvalho, "Evolutionary computing in recommender systems: a review of recent research," *Natural Computing*, vol. 16, pp. 441–462, 2017.
- [3] B. Song, Y. Gao, and X.-M. Li, "Research on collaborative filtering recommendation algorithm based on mahout and user model," *Journal of Physics: Conference Series*, vol. 1437, no. 1, pp. 012095–012101, 2020.
- [4] J. Li and Z. Ye, "Course recommendations in online education based on collaborative filtering recommendation algorithm," *Complexity*, vol. 2020, Article ID 6619249, 10 pages, 2020.
- [5] L. Xu and X. Lu, "Collaborative filtering recommendation system based on interest drifting and similar cloud," *Journal of Liaoning Technical University (Natural Science Edition)*, vol. 36, no. 7, pp. 761–766, 2017.
- [6] G. Geetha, M. Safa, C. Fancy, and D. Saranya, "A hybrid approach using collaborative filtering and content based filtering for recommender system," *Journal of Physics: Conference Series*, vol. 1000, pp. 012101–012123, 2018.
- [7] L. Zuping, "Collaborative filtering recommendation algorithm based on user interests," *International Journal of U- & E-Service*, vol. 8, no. 4, pp. 311–319, 2015.
- [8] J. Joaa, S. Bangb, and G. Parka, "Implementation of a recommendation system using association rules and collaborative filtering," *Procedia Computer Science*, vol. 91, pp. 944–952, 2016.
- [9] R. Chen, Q. Hua, Y.-S. Chang, B. Wang, L. Zhang, and X. Kong, "A survey of collaborative filtering-based recommender systems: from traditional methods to hybrid methods

- based on social networks,” *IEEE Access*, vol. 6, pp. 64301–64320, 2018.
- [10] Z. Qianqian and W. Dongliang, “A SVM recommendation IoT model based on similarity evaluation and collaborative filtering of multi-angle knowledge units,” *International Journal of Computers & Applications*, vol. 42, no. 3, pp. 278–281, 2018.
 - [11] Y. Yoo, J. Kim, and B. Sohn, “Evaluation of collaborative filtering methods for developing online music contents recommendation system,” *Transactions of the Korean Institute of Electrical Engineers*, vol. 66, no. 7, pp. 1083–1091, 2017.
 - [12] X. Song, Y. Guo, Y. Chang et al., “A hybrid recommendation system for marine science observation data based on content and literature filtering,” *Sensors*, vol. 20, no. 22, pp. 6414–6428, 2020.
 - [13] Y. Xu and N. Zhu, “Hybrid recommendation algorithm based on long-term and short-term interest and matrix factorization for collaborative filtering,” *Journal of Physics: Conference Series*, vol. 1624, no. 4, pp. 042015–042021, 2020.
 - [14] X. Liu, “A collaborative filtering recommendation algorithm based on the influence sets of e-learning group’s behavior,” *Cluster Computing*, vol. 22, no. 5, pp. 121–131, 2017.
 - [15] J. Shi, Y. Tang, and P. Xu, “A collaborative filtering recommendation method based on clustering and user trust,” *Journal of Computational Information Systems*, vol. 11, no. 18, pp. 6845–6852, 2015.
 - [16] Y. Gao, C. Huang, M. Hu, J. Feng, and X. Yang, “Research on book personalized recommendation method based on collaborative filtering algorithm,” *IOP Conference Series: Earth and Environmental Science*, vol. 252, no. 5, pp. 052099–052105, 2019.
 - [17] H. Xia, J. J. Li, and Y. Liu, “Collaborative filtering recommendation algorithm based on attention GRU and adversarial learning,” *IEEE Access*, vol. 8, pp. 208149–208157, 2020.
 - [18] Z. Romadhon, “Various implementation of collaborative filtering-based approach on recommendation systems using similarity,” *Kinetik Game Technology Information System Computer Network Computing Electronics and Control*, vol. 5, no. 3, pp. 134–146, 2020.
 - [19] M. Hikmatyar, “Book recommendation system development using user-based collaborative filtering,” *Journal of Physics: Conference Series*, vol. 1477, no. 3, pp. 032024–032032, 2020.
 - [20] S. Hattori, R. Misawa, H. Ishikawa, and Y. Takama, “Hybrid recommender system based on collaborative filtering employing personal values-based user model,” *Journal of Japan Society for Fuzzy Theory and Intelligent Informatics*, vol. 29, no. 4, pp. 628–636, 2017.
 - [21] Y. Yan and H. Xie, “Collaborative filtering recommendation algorithm based on user preferences,” *Journal of Physics: Conference Series*, vol. 1549, no. 3, pp. 032147–032155, 2020.
 - [22] N. Ifada, N. F. D. Putri, and M. K. Sophan, “Normalization based multi-criteria collaborative filtering approach for recommendation system,” *Rekayasa*, vol. 13, no. 3, pp. 234–239, 2020.
 - [23] J. Deng, J. Guo, and Y. Wang, “A novel K-medoids clustering recommendation algorithm based on probability distribution for collaborative filtering,” *Knowledge-Based Systems*, vol. 175, no. 7, pp. 96–106, 2019.
 - [24] T. Zhang, “Research on collaborative filtering recommendation algorithm based on social network,” *International Journal of Internet Manufacturing and Services*, vol. 6, no. 4, pp. 343–256, 2019.
 - [25] P. K. Singh, P. K. D. Pramanik, and P. Choudhury, “An improved similarity calculation method for collaborative filtering- based recommendation, considering neighbor’s liking and disliking of categorical attributes of items,” *Journal of Information and Optimization Sciences*, vol. 40, no. 2, pp. 397–412, 2019.
 - [26] J. Wang, A. K. Sangaiah, and W. Liu, “A hybrid collaborative filtering recommendation algorithm: integrating content information and matrix factorisation,” *International Journal of Grid and Utility Computing*, vol. 11, no. 3, pp. 367–378, 2020.
 - [27] B. S. Neysiani, N. Soltani, N. Soltani, R. Mofidi, and M. H. Nadimi-Shahraki, “Improve performance of association rule-based collaborative filtering recommendation systems using genetic algorithm,” *International Journal of Information Technology and Computer Science*, vol. 11, no. 2, pp. 48–55, 2019.
 - [28] J. Yamawaki, Y. Kudo, and T. Murai, “Recommendation method based on interrelationship mining and collaborative filtering,” *Transactions of Japan Society of Kansei Engineering*, vol. 17, no. 4, pp. 481–488, 2018.