Research Article

Research on Personalized Book Recommendation Based on Improved Similarity Calculation and Data Filling Collaborative Filtering Algorithm

Yanping Du [iD], Lizhi Peng [iD], Shuihai Dou [iD], Xianyang Su [iD], and Xiaona Ren [iD]

Beijing Institute of Graphic Communication, Beijing Key Laboratory of Digitalized Printing Equipment, Beijing, China

Correspondence should be addressed to Shuihai Dou; doushuihai@126.com

(*Purpose/Significance*). This paper aims at the problems of inaccurate recommendation effect caused by data sparseness and cold start in the traditional collaborative filtering-based book personalized recommendation algorithm. So this paper proposes a collaborative filtering recommendation algorithm which improves the similarity solution method and the filling method of missing data. (*Method/Process*). By considering the influence of the user's common rating book collection on the similarity calculation, the average rating value of all books is used as the threshold, and the user's common rating weight is introduced into the user's similarity calculation. As for data filling, according to the user's average rating, the basic attributes such as the age and gender of users are coded, and then Euclidean distance is initially calculated, making hierarchical clustering on users. What's more, Shope-one algorithm is used to calculate the filling value of the former m similar users , and add the weight value of the degree simultaneously to get the final filling value, so as to improve the data filling method. (*Result/Conclusion*). Experiments were carried out with the data set of Book-Crossing Data set through *Python*. The experimental results show that the improved collaborative filtering algorithm has a significantly improvement in the accuracy and quality of book recommendation.

1. Introduction

With the rapid development of cloud computing, Internet of Things, Internet and other technologies and Web 2.0, in many scenarios such as the network and digital libraries [1–3], a large amount of information data is constantly emerging from book resources, which increases people's access to the required book information difficulty, resulting in 'information overload [4, 5]. How to solve the 'information overload' and help users quickly find the book resources that meet their individual demand is an urgent problem to be solved [6, 7], because book recommendation algorithms can realize 'one-to-one' service according to users' personal preferences, to provide users with personalized services book recommendations.

According to different recommendation strategies, book recommendation algorithms can be divided into: collaborative filtering-based recommendation algorithm [8],

association rule-based recommendation algorithm [9], content-based recommendation algorithm [10]and hybrid recommendation algorithm [11], etc. The advantages and disadvantages of these four types of algorithms in book personalized recommendation are shown in Table 1.

Amazon [12, 13]and Dangdang [14] chose to apply the collaborative filtering-based recommendation algorithm for book recommendation services. The book recommendation based on collaborative filtering recommendation algorithm does not require in-depth analysis of the knowledge and content of book resources. It only needs to analyze the similar characteristics of users or the analysis of book borrowing records to recommend book information that users may be interested in. And the recommendation based on collaborative filtering Algorithms can handle complex structures and is technically easy to implement. Therefore, collaborative filtering recommendation algorithms are widely used. However, with the rapid increase of the number

TABLE 1: Comparison of advantages and disadvantages of book recommendation algorithms.

Book recommendation algorithm	Advantage	Disadvantages
Content-based recommendation algorithms	Easy to implement, simple method, high recommendation quality, and real-time	The algorithm for extracting content is complex and difficult to process
Recommendation algorithm based on association rules	High recommendation quality	Difficulty in data mining, low computational efficiency, poor personalized recommendation
Collaborative filtering recommendation algorithms	Simple method and can handle complex structures, and better personalized recommendation	Cold start problems, data sparse problems
Hybrid recommendation algorithm	Comprehensive use, high recommended quality	The algorithm is complex and the calculation is difficult

of books, the recommendation algorithm collects limited information from users in the initial stage of the system, and the traditional collaborative filtering recommendation algorithm is prone to problems such as cold start and data sparseness.

In response to the problems of data sparseness and cold start in the traditional collaborative filtering algorithm in book recommendation, domestic and foreign researchers have proposed different improvement methods. The traditional solution mainly uses the average value of the existing data to fill in the missing data, which will bring errors to the prediction results and affect the accuracy of the recommendation. In this regard, Wang [15] proposed a book recommendation algorithm based on collaborative filtering and interest, which uses the user's interest as an important measure to improve the accuracy of book recommendation. In order to effectively improve book recommendation, Zhang [16]. proposed a collaborative filtering algorithm based on time sequence, considered the student's book borrowing time and book circulation time. The experimental results showed that the accuracy of book recommendation is effectively improved. Guo [17]et al. designed a collaborative filtering recommendation algorithm based on R-RBM and Top N, which solved the problems of data sparseness, cold start and inability to mine the deep-level features of readers' personalized information in the recommendation system. Gao [18]. used the decision tree algorithm to predict the recommended books, and proposed a user-based collaborative filtering recommendation algorithm. It was confirmed by experiments that the recommendation accuracy of the algorithm is much better than the traditional collaborative filtering algorithm, which has made a big improvement. Liu [19]. used collaborative filtering algorithm for data mining of efficient book management data, and generated a book recommendation model. Noor et al. [20]. proposed a collaborative filtering method of probabilistic keywords in book recommendation system to solve the problem of data sparseness in collaborative filtering recommendation algorithm, which effectively improved the recommendation performance. Wang and Liu [21]. used association rules and collaborative filtering algorithm to mine the information of book borrowing records, and obtained the correlation of borrowing hobbies among users and the correlation between books. In summary, researcher's improved methods for the book recommendation algorithms have alleviated the problems of data sparseness and cold start in collaborative

filtering recommendation algorithms for book recommendation to certain, but the amount of calculation has been increased significantly.

In view of the shortage of similarity calculation and data filling in the traditional book recommendation based on collaborative filtering algorithm, this paper improves the original data filling method by using hierarchical clustering and Slope-one algorithm to make full use of rating data, and avoids single or several user's inaccurate data filling , thus improving the filling accuracy. Then, the similarity calculation between books is added to the similarity calculation to improve the accuracy of the similarity calculation. And through the empirical analysis of the algorithm, it is verified that the improved book recommendation method based on collaborative filtering recommendation algorithm has significantly improved recommendation accuracy compared with the traditional book recommendation based on collaborative filtering recommendation algorithm.

2. Book Recommendation Based on Collaborative Filtering Algorithm

The basic idea of the user-based collaborative filtering algorithm is to calculate the similarity between users according to the user's preference information and historical activity records in the system, obtain the similarity matrix between users, and find the similar neighbor set of the target user. According to the rating data of some books, it predicts the rating of target users on unrated items and selects items with high ratings for recommendation. The principle of book recommendation based on user-based collaborative filtering algorithm is shown in Figure 1.

In Figure 1, user A's preference is similar to user B's, user A prefers books A, B, and C, and user B prefers books A and C, so the algorithm can recommend book B to user B.

The traditional book recommendation based on collaborative filtering algorithm mainly includes the following three stages:

(1) Information collection: It mainly collects the user's basic information, book information, borrowing information of books and user evaluation data about books.

(2) Construction of the user rating matrix: The purpose is to reduce the time cost by converting the relevant data into dictionary format, and directly construct
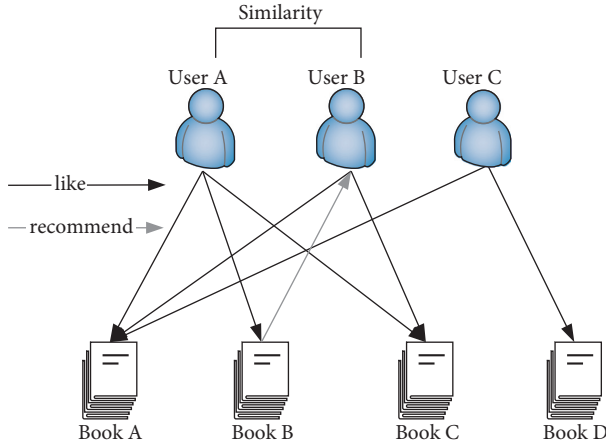
FIGURE 1: Schematic diagram of book recommendation based on user-based collaborative filtering algorithm.

the query conditions according to the recommendation information when recommending, and read the relevant data. The most important part of this process is the calculation of similarity between users, that is, the similarity of users is calculated according to the books that are jointly evaluated by users, sorted by similarity, and the nearest neighbor is taken as the target and recommended to the user. The most common method for calculating the similarity of traditional collaborative filtering-based book recommendation is cosine similarity [22], Pearson similarity method [23]. In the case of high sparsity of the rating matrix, the calculation result of the Pearson similarity method is poor, and the cosine similarity is simple and opposite. Therefore, this paper uses the improved cosine similarity to calculate the similarity between users.

The cosine similarity is mainly obtained by the cosine angle between the two users, and there is an inverse relationship between the two, that is, the smaller the similarity, the larger the angle; the greater the similarity, the smaller the angle. The calculation formula of cosine similarity 1 is as follows:

$$sim(a, b) = \cos(a, b) = \frac{\sum_{n \in N_{ab}} R_{a,n} R_{b,n}}{\sqrt{\sum_{n \in N_{ab}} (R_{a,n})^2} \sqrt{\sum_{n \in N_{ab}} (R_{b,n})^2}}. \tag{1}$$

In formula (1), a, b represent respectively user a and user b, $R_{a,n}$ is user b's rated for book n, N_{ab} is the set of common rating for books by users a and b, N_a is the collection of book rating for user a , N_b is the collection of book rating for user b.

(3) Book recommendation list generation: After determining the target neighbor users according to the similarity, it is necessary to predict the score of the target book by the neighbor users to generate the book recommendation result of the target user, and realize the personalized service of book

recommendation for the user. Commonly used predictive scoring methods include average scoring method [24] and offset weighted scoring method [25].

The traditional book recommendation process based on collaborative filtering algorithm is shown in Figure 2.

3. Improvement of Book Recommendation Based on Collaborative Filtering Algorithm

The traditional book recommendation based on collaborative filtering algorithm only calculates the similarity between users according to the user's basic information and the user's rating of the book, and then recommends the books with high ratings to the user. With the exponential growth of users and the number of books in the book recommendation system, the user's rating data will be very sparse, which will have a great impact on the accuracy of similarity calculation, resulting in a larger error in the accuracy of book recommendation results, which affects the recommendation quality. According to the problems of data sparsity and cold start in book recommendation based on traditional collaborative filtering algorithm, this paper proposes improvements in similarity calculation and data filling.

3.1. Improvement of Similarity Calculation Method. Similarity calculation is the key to finding the nearest neighbor set. The traditional user similarity calculation is to directly calculate the similarity between different users based on whether the target user has the same interest and hobbies for a certain book, different historical scoring data and so on. In this paper, aiming at the problem that there are very few common scores of books between two neighboring users in the traditional similarity calculation method, this paper introduces the user's common scoring weight factor to improve the accuracy of the cosine similarity result. User common rating weight is defined as:

$$\overline{N}_{ab} = \sum_{a,b=0}^{n} \frac{N_a + N_b}{2}. \tag{2}$$

The improved user similarity calculation method is:

$$COSU_{a,b} = \text{norm}(N_a) * \text{norm}(N_b) * \text{norm}(N_{ab}), \tag{3}$$

$$SimU_{a,b} = 0.5 + 0.5 * COSU_{a,b}. \tag{4}$$

In formulas 2, 3, 4, $SimU_{a,b}$ represents the similarity between users a and b , $\overline{N}_{ab}$ is the average score of users a and b for the jointly rated books.

After the improvement and fully consideration, the influence of the common scores between users on the similarity calculation can effectively improve the similarity calculation between users.

3.2. Improvements of Data Filling. The historical users in book recommendation can not score all books, and there will be a large amount of null data between the constructed user-
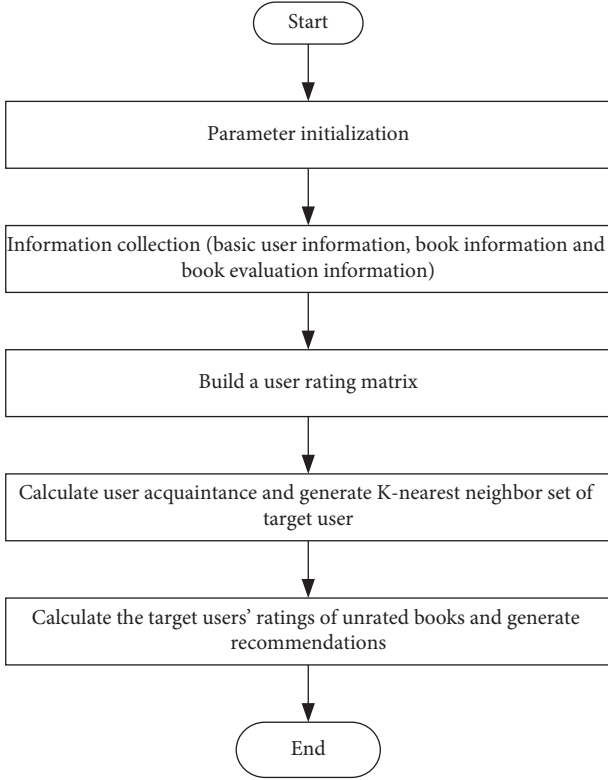
Figure 2: Traditional book recommendation process based on collaborative filtering algorithm.

book model, which will directly affect the accuracy of user similarity calculation, resulting in a plummeting recommendation quality. To face this problem, it is necessary to pre-process the data, filter and delete some user data without a large number of scores, and fill in the missing part of the data with specific data to ensure data integrity. The general data filling method does not significantly improve the final result. In order to effectively improve the quality of user recommendation, this paper uses the hierarchical clustering algorithm and the improved Slope-one algorithm to improve the data filling method.

In order to improve the efficiency of users' data analysis of book-related information and reduce the sparsity of the score matrix, a hierarchical clustering of users is established according to the user's historical record information on books [26, 27], and users are divided into high-rating (U_o) and low-rating (U_p) and medium rating (U_n) three user groups to achieve dimensionality reduction processing of data and reduce the amount of calculation. That is, if the average user's rating for books is above 4, it is U_o, then if the average user's rating for books is less than 2, it is U_p, and finally, if the average user's rating for books is between 2 and 4, it is U_n.

The clustering process is as follows:

$$\begin{aligned} &\text{If } \overline{r_{n,i}} > = 4r_n \in U_o, \\ &\text{Else If} \overline{r_{n,i}} < = 2r_n \in U_p, \\ &\text{Else } 2 < = \overline{r_{n,i}} < = 4r_n \in U_n. \end{aligned} \tag{5}$$

Where $\overline{r_{n,i}}$ is the average rated of book n by user i.

Through the process of hierarchical clustering, each group of data after clustering is filled with missing data to reduce the dimension of the data and reduce the amount of calculation. The similarity between users in the same group needs to be considered before filling. The traditional data filling method only uses the historical scoring data to solve the Euclidean distance as the similarity value, but the newly added users have no historical scoring information and cannot calculate the Euclidean distance. Therefore, this paper introduces user attributes , which is added to the similarity calculation.

The user information is encoded using One-hot coding, that is, the user's score for each book, with 0 means no score or a rating of 0. The encoded user information is obtained, and the similarity value is obtained. Using the Slope-one algorithm to calculate the filling data for the missing values of the first m users, and adding the weight of the similarity, the filling value is finally obtained. The selection of the padding value needs to be within a suitable range. If the selected value is too large or too small, the filling accuracy will be affected.

The basic idea of the Slope-one algorithm is a linear algorithm that calculates the difference between the user's scores on different books, and predicts the score of another book according to the user's score on a certain book. The specific calculation formula is as follows:

$$\overline{R(a,b)} = \frac{\sum_{i \in N_{ab}} (r_{ai} - r_{bi})}{\text{card}(N_{ab})}, \tag{6}$$

$$T_{bj} = r_{aj} - \overline{R(a,b)}., \tag{7}$$

In formulas (5), (6), $\overline{R(a,b)}$ represents the average score difference between users a and b for each jointly rated book; r_{ai} is the rating value of user a for book i, r_{bj} is the rating value of user b for book i;card(N_{ab}) is the total number of books rated both a and b,P_{bj} is the filling data information of user b's rating of book j,r_{aj} is the filling data information of user a's rating of book j.

The formula for filling the value of each missing value:

$$D(u,v) = \sqrt{\sum_{i \in I_{u,v}} (r_{u,i} - r_{v,i})^2}, \tag{8}$$

$$a_{u,v} = \frac{U_u \cap U_p}{U_u \cup U_p},$$

$$\text{Sim}_{a,b} = D(u,v) \times a_{u,v}, \tag{9}$$

$$T_{bj} = r_{aj} - \frac{\sum\limits_{i \in N_{ab}} (r_{ai} - r_{bi})}{\text{card}(N_{ab})}, \tag{10}$$

$$T_{wj} = \frac{\sum_{u=1}^{n} (\text{Sim}_{u,w} \times T_{u,j})}{\sum_{u=1}^{n} \text{Sim}_{u,w}}. \tag{11}$$

In formulas (5), (6),$D(u, v)$ is the encoded Euclidean distance,$r_{u,i}$is the vector composed of the user's information code and the score for book i,$a_{u,v}$ is the proportion of items scored together,$Sim_{a,b}$ is the similarity between the user and the added user's specific information and scoring together,$T_{u,j}$is the specific padding value for user u and book j in terms of missing data, and T_{wj} is the final padding value.

Data filling process:

Step 1. According to the basic attributes of users in the data set, perform hierarchical clustering on users, and divide all users into three different user categories.

Step 2. According to different user categories, use formula (9) for the user groups to calculate the similarity in the same group, and find the neighbor user set;

Step 3. Use formula (10) to fill in the missing values in the system;

Step 4. Introduce the similarity calculation weight to optimize the filling data;

Step 5. Repeat the above steps until the entire user data set is populated.

In summary, hierarchical clustering is performed on users according to user ratings, which achieves the effect of initial dimensionality reduction, and at the same time, the accuracy of similarity calculation after clustering is improved to a certain extent. The basic information of the user is added when calculating the similarity, which effectively alleviates the cold start problem. The accuracy of the similarity is further improved by adding the weight of the common rating.

3.3. Prediction Score. The set of neighbor users of the target user is obtained by similarity calculation as $U = \{u_1, u_2, u_3, \cdots, u_n\}$, according to the score of each neighbor user on the target book, the filling formulas (10) and (11) are used to predict the score of the target book [28, 29]. The prediction result of the i-th target user is, and then use the prediction formula (12) to predict and score the final target book, complete the prediction of all target book scores, and form the final book recommendation result.

According to the calculation of similarity, the sore prediction formula is as follows:

$$r(u, g) = \overline{r_u} + \frac{\sum_{k \in \overline{U}} sim(u, k)\left(r_{k,i} - \overline{r_k}\right)}{\sum_{k \in \overline{U}} |sim(u, k)|}. \tag{12}$$

In formula (12),$\overline{r_u}$ is the average rating of all books by user u,$\overline{r_k}$is the average rating of all books by neighbor user k.

3.4. Improved Book Recommendation Process. The improvement of the collaborative filtering algorithm is mainly divided into three steps: Firstly, for the improvement of data filling method, build a user-book model, perform

FIGURE 3: Workflow of the improved algorithm in the book personalized book recommendation system.

hierarchical clustering on users, encode the basic information of users and initially calculate the Euclidean distance, and the introduce of Shope-one algorithm to fill in the missing data; secondly, the improvement of similarity calculation method, considering the influence of the common score of neighboring users on the similarity calculation, an improved similarity calculation method is introduced to calculate the similarity of neighboring users; finally, the improved collaborative filtering algorithm is used to predict users' ratings of unselected books for recommend books to users.

The book recommendation process based on the improved collaborative filtering recommendation is shown in Figure 3.

4. Analysis of Experimental Results

Firstly, the experimental environment, the selected data set and the experimental evaluation indicators involved in this paper are briefly introduced, and then the method proposed

TABLE 2: RMSE values of neighbor users with different numbers of clusters before and after algorithm improvement.

Number of clustered users	RMSE value of traditional collaborative filtering recommendation algorithm	RMSE value of improved collaborative filtering recommendation algorithm
5	5.903	3.469
10	5.878	4.184
15	6207	4.123
20	6.355	3.86
25	6.47	4.243
30	6.339	4.252
35	6.333	4.263
40	6.408	4.444
45	6.34	4.637
50	6.234	5

in this paper and the method before improvement are compared and analyzed according to the data set.

4.1. Introduction to Experimental Environment and Data Set.
The computer configuration used in this study is Intel Core i5-6200 CPU with 4 GB of running memory. The operating system is Windows 7 64 bit, the programming language is *Python* language, the version is *Python* 3.8, and the editor is Anaconda Jupyter Notebook.

The data used in this article comes from the Book-Crossing Data set released by the Free University of Germany. The Book-Crossing Data set is composed of the scores of 278,858 users in the Book-Crossing community, including 1,149,780 book scoring data of about 271,379 books, with a score of 1∼ 5 points, the Data set contains 3 categories: BX-users, BX-books, and BX-book ratings. In this experiment, on the one hand, the experimental results are obtained from the MASE value of the improved similarity calculation and data filling; on the other hand, the final experimental results are obtained by calculating the precision, recall and F-measure of the book recommendation results.

4.2. Algorithm Evaluation Index and Experimental Results

4.2.1. Predictive Scoring Accuracy of the Algorithm

(1) Experimental Evaluation Index. The performance evaluation indexes of the recommended algorithms mainly include mean square error (MSE), root mean square error (RMSE), square absolute error (MAE), precision, recall and so on. The accuracy of the predicted score refers to the difference between the predicted score of the recommendation algorithm and the actual score of the user. This paper uses RMSE as the evaluation index. The smaller the RMSE value, the better the accuracy of the prediction results.

Assuming that the set of predicted scores for the book by the recommendation system is $\{p_1, p_2, p_3, \cdots, p_n\}$, and the actual score of the book is $\{q_1, q_2, q_3, \cdots, q_n\}$, then the RMSE can be expressed as formula (12).

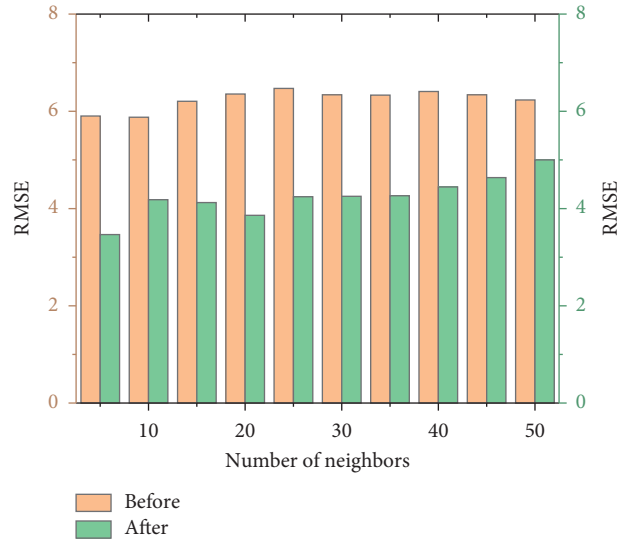$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(p_i - q_i)^2}{N}}. \tag{13}$$

FIGURE 4: RMSE values of nearest neighbor users with different number of clusters.

In formula (12), p_i is the predicted rating, q_i is the actual rating, and N is the total number of predicted books.

(2) Experimental Results and Analysis. Method 1: Selected different number of neighbor user, which is 5, 10, 15, 20, 25, 30, 35, 40, 45, and 50 to analyze the experimental results. In the experiment, users in the data set were randomly selected for prediction and scoring experiments, and the method of similarity calculation and data filling before and after the improvement of personalized book recommendation based on collaborative filtering algorithm was adopted. The results are shown in Table 2.

Figure 4 shows the change of RMSE value. The abscissa represents the number of neighboring users with different numbers of clusters, and the ordinate represents the RMSE value of the predicted scoring result. It can be seen from the figure that the RMSE value of the improved algorithm is always lower than that of the traditional algorithm, and the RMSE value of the improved algorithm is reduced by an average of 32% compared with the traditional algorithm. Because the smaller the RMSE value, the better the recommendation effect of the algorithm is. The proposed

TABLE 3: RMSE values of users with the same number of clusters and different neighbors before and after the improvement of the algorithm.

Number of clustered users	RMSE value of traditional collaborative filtering recommendation algorithm	RMSE value of improved collaborative filtering recommendation algorithm
100	5.801	5.082
100	5.878	5
100	6.02	5.378
100	6.282	5.846
100	6.483	5.841
100	6.426	5.568
100	6.474	6.414
100	6.466	5.898
100	6.252	5
100	6.119	4.966
Average	6.2201	5.4993

improvements in similarity calculation and data filling are better than the unimproved methods, and the improved recommendation algorithm is better than the traditional recommendation algorithm in the cold-start environment.

Method 2: Method 1 mainly conducts experiments from neighboring users with different cluster numbers. In order to further verify that the similarity calculation and data filling method proposed in this paper is better than the unimproved method, we select 100 neighbors with the same number of clusters and different neighbors. The user conducts multiple experimental results analysis, and the experimental results are shown in Table 3.

Figure 5 shows the change of RMSE value of 100 neighbors with the same number of clusters selected each time. The abscissa represents the 100 neighbors in different intervals in the data set, and the ordinate represents the RMSE value of the predicted scoring result. It can be seen from the figure that the RMSE value of the improved algorithm is always lower than that of the traditional algorithm, and the RMSE value of the improved algorithm is reduced by an average of 12% compared with the traditional algorithm. Because of the smaller the RMSE value, the better the recommendation effect of the algorithm is. The proposed similarity computation improvements and data padding improvements are compared to the unimproved methods, and the results are consistently better than the unimproved traditional methods.

4.2.2. The Accuracy of the Recommended Results

(1) Experimental Evaluation Index. Based on the results of using different numbers of neighbor users from 5 to 50, and the results of selecting 100 different neighbor users in different intervals each time, the RSME prediction error of the improved algorithm is smaller than that of the traditional algorithm. In order to evaluate the correlation prediction accuracy of the recommendation results in this algorithm, this paper selects three evaluation indicators related to the effect of book recommendation, namely accuracy, recall, and F-measure. These three evaluation indicators are used to compare book recommendation based on traditional collaborative filtering algorithm with the improved
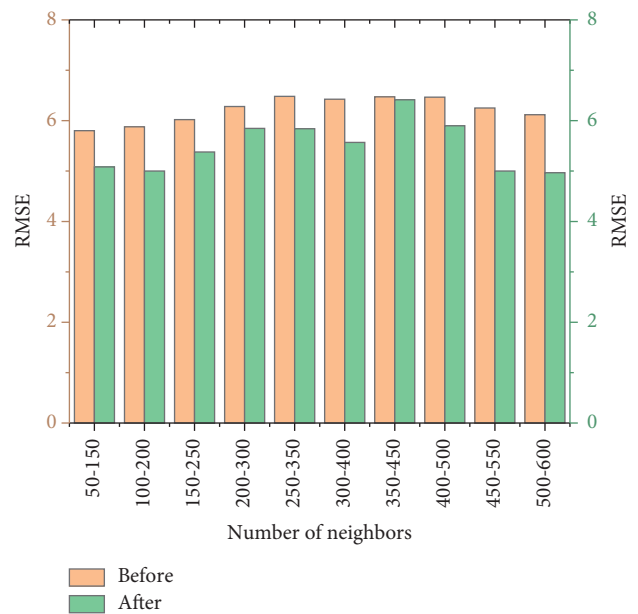

FIGURE 5: RMSE values of users with the same number of clusters but different neighbors.

collaborative filtering algorithm-based book recommendation is compared.

The accuracy rate indicates the accuracy of the recommendation in the recommendation list, that is, the proportion of the total number of books that are successfully recommended. The recall rate indicates the proportion of the books that users recommend accurately in the selected experimental data set, that is, the proportion of successfully recommended books to the user's interest. Among them, the calculation formula of precision rate (14) is shown, and the formula of recall rate (15).

$$\text{Precision} = \frac{\sum_{u \in U} |R_u \cap I_u|}{\sum_{u \in U} |R_u|}, \tag{14}$$

$$\text{Recall} = \frac{\sum_{u \in U} |R_u \cap I_u|}{\sum_{u \in U} |I_u|}. \tag{15}$$

In formulas (13) and (14), R_u is the target user U_i recommends R books, and I_u is the collection of books that he likes during the experiment.

F-measure is calculated by the combination of precision and recall. When there is a contradiction between precision and recall, F-measure is usually used to evaluate the efficiency of the recommendation algorithm. The higher the F-measure value, the higher the effectiveness of the algorithm. The F-measure calculation formula is shown in (3)-(4).

$$F = \frac{2 \times recall \times precision}{recall + precision}. \tag{16}$$

(2) Experimental Results and Analysis. Method 1: For the book personalized recommendation results based on collaborative filtering algorithm, the main influence is the number of clustered nearest neighbor users. In the experiment, the values of nearest neighbor users are selected as 5, 10, 15, 20, 25, 30, 35, 40, 45 and 50 respectively to analyze the experimental results.

Figure 6, Figure 7, and Figure 8 show the experiments of the precision, recall, and F-measure of the improved book recommendation based on collaborative filtering algorithm and the traditional book recommendation based on collaborative filtering algorithm. As a result, the abscissas all represent the number of neighboring users with different numbers of clusters, the ordinates in Figure 6 represent the precision of the predicted scoring results, the ordinates in Figure 7 represent the recall rates of the predicted scoring results, and the ordinates in Figure 8 The ordinate represents the F-measure value of the predicted scoring result. As can be seen in Figure 6, Figure 7, and Figure 8, the precision, recall, and F-measure of the traditional book recommendation results are always lower than those of the improved book recommendation results. rate, F-measure value. When the number of clusters reaches 15, the precision, recall, and F-measure value all reach an optimal value, indicating that the recommendation effect is the best at this time. It can be seen that the precision, recall, and F-measure have experienced a process of first increase and then decrease with the increase of the number of clusters of neighboring users, which conforms to the general law of precision, recall, and F-measure value, indicating that this paper The proposed improvements to the collaborative filtering-based recommendation algorithm can effectively improve the quality of book recommendation.

Figure 9 is a comparison chart of the precision, recall, and average value of F-measure of each index of neighboring users with different numbers of clusters. The average precision rate is about 52% higher than that of traditional book recommendation, and the average recall rate is higher than the traditional book recommendation , which is improved by about 59%, and the average F-measure value is about 72% higher than that of the traditional book recommendation. It can be seen that the improved algorithm in this paper has achieved a very significant improvement in the recommendation effect. The accuracy and intelligence of the
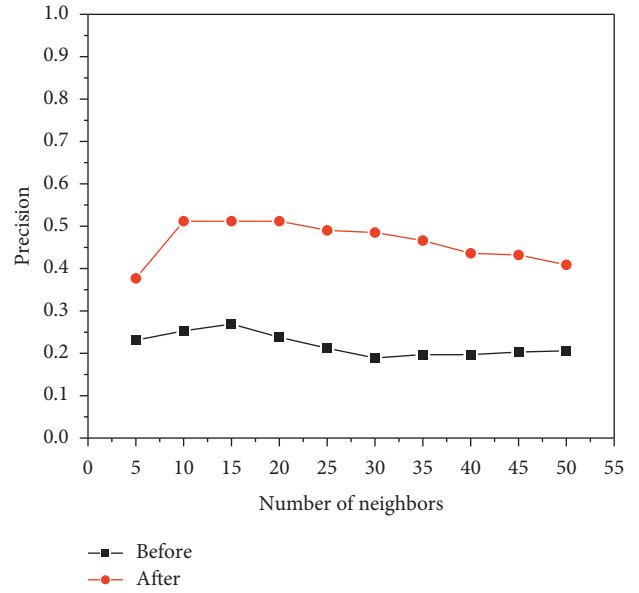

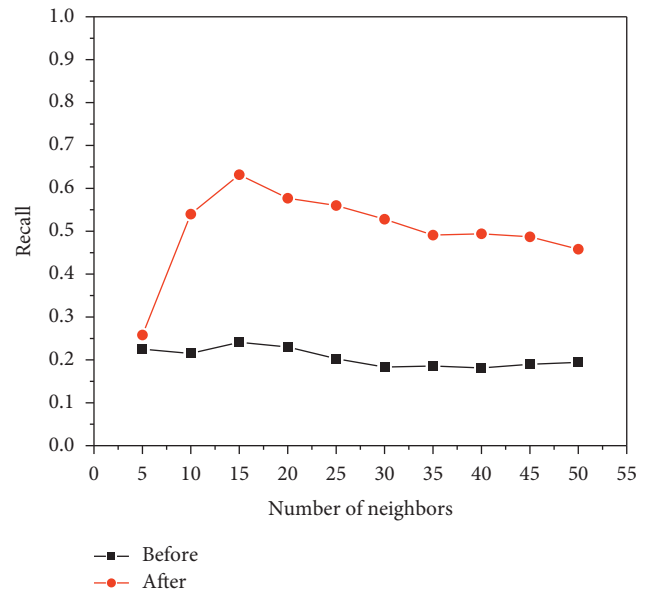FIGURE 6: The precision of neighbor users with different numbers of clusters.

FIGURE 7: The recall rate of neighbor users with different numbers of clusters.

recommendation provide support for users to provide more personalized information services.

Method 2: Method 1 is mainly to verify the neighboring users with different numbers of clusters to conduct experiments. In order to further verify that the improved personalized book recommendation based on the collaborative filtering recommendation algorithm is better than the unimproved algorithm, in the experiment, 100 users are selected each time. The experimental results are analyzed for users with the same number of clusters and different neighbors.

Figure 10, Figure 11, Figure 12 show the experiments of the precision, recall, and F-measure of the improved book
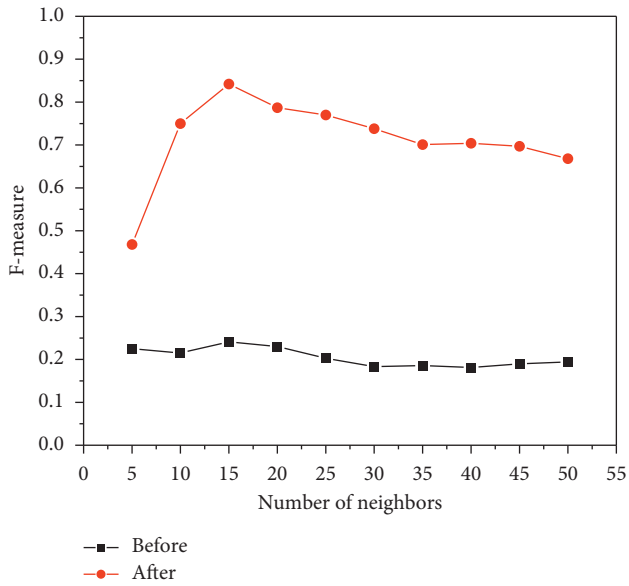
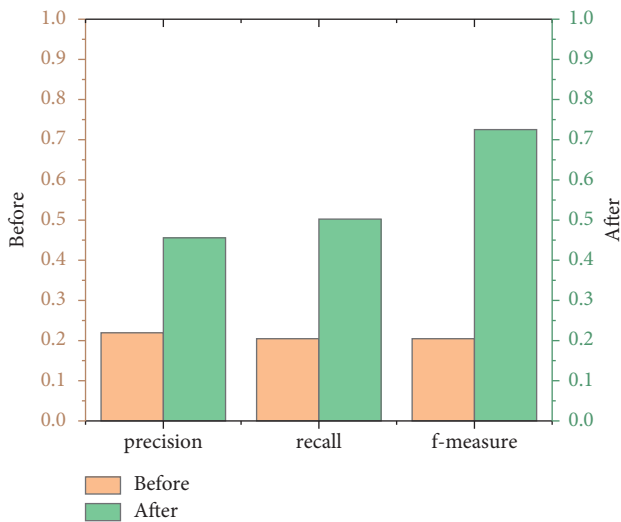FIGURE 8: F-measure values of neighbor users with different numbers of clusters.

FIGURE 9: Average values of various index of neighboring users with different numbers of clusters.
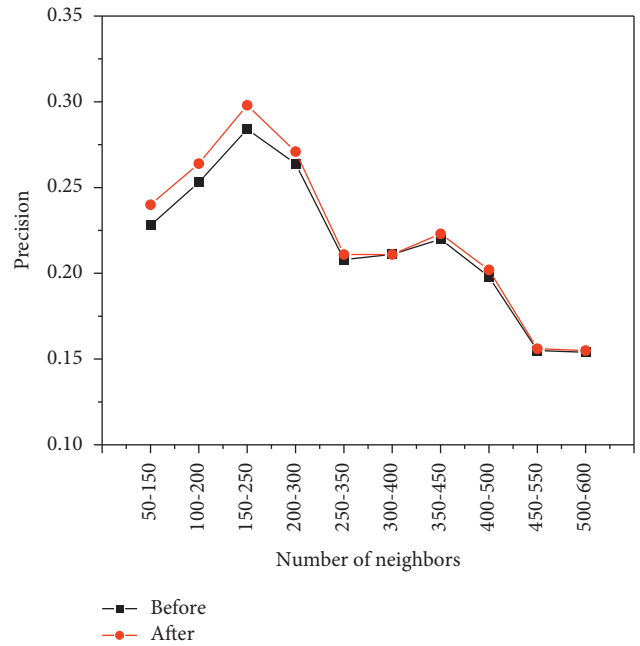

FIGURE 10: The precision of users with the same number of clusters but different neighbors.
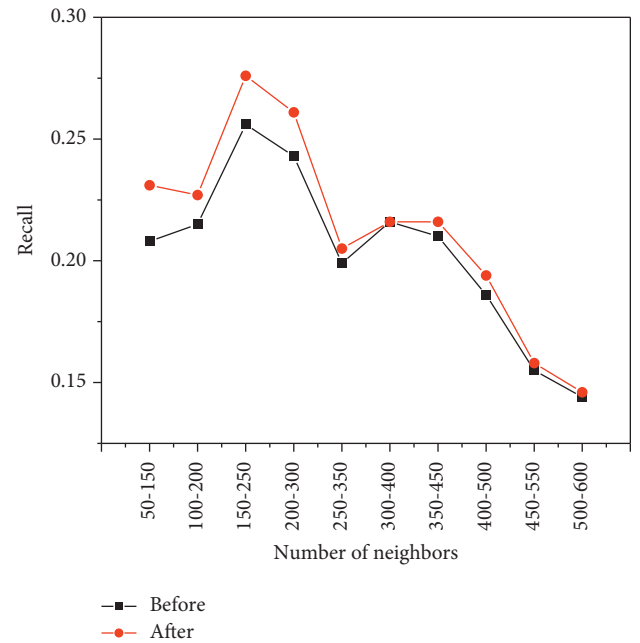
FIGURE 11: The recall of users with the same number of clusters and different neighbors.

recommendation based on collaborative filtering algorithm and the traditional book recommendation based on collaborative filtering algorithm. As a result, the abscissas all represent 100 neighbors with the same number of clusters in different intervals in the data set, the ordinates in Figure 10 represent the precision of the predicted scoring results, and the ordinates in Figure 11 represent the recall of the predicted scoring results. rate, and the ordinate in Figure 12 represents the F-measure value of the predicted scoring result. As can be seen in Figure 10, Figure 11, and Figure 12, the accuracy, recall, and F-measure of traditional book recommendation results for 100 neighboring users in different intervals are always lower than those of the improved books' precision, recall, and F-measure value of the

recommended results are in line with the general rules of precision, recall, and F-measure value, which further shows that the improvement of the recommendation algorithm based on collaborative filtering proposed in this paper can effectively improve the quality of book recommendation.

Figure 13 is a comparison chart of the precision, recall, and average value of F-measure for the same number of clusters and different neighbors. It can be seen that the
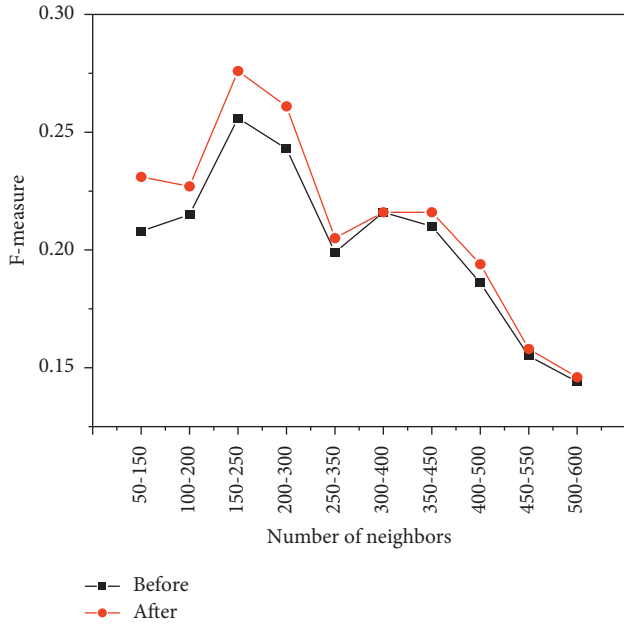
FIGURE 12: The F-measure values of users with the same number of clusters and different neighbors.
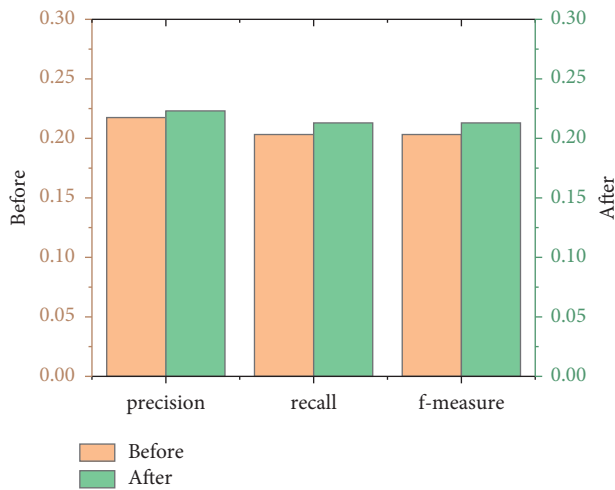

FIGURE 13: Average values of each index for users with the same number of clusters and different neighbors.

improved algorithm in this paper has 100 users with the same number of clusters in different intervals. The improvement of the recommendation results above further shows that the improved results of the book recommendation based on the collaborative filtering algorithm proposed in this paper are better than the traditional book recommendation based on the collaborative filtering algorithm.

5. Conclusion

In order to improve the accuracy of book recommendation and predict the needs of users' books, the traditional book recommendation based on collaborative filtering algorithm

has data sparse and cold-start problems in similarity calculation and data filling. In this paper, the slop-one algorithm is used to improve the way of data filling, the common score of neighboring users is introduced to improve the similarity calculation. And a book recommendation based on the improved collaborative filtering algorithm is proposed. The basic information, book information and evaluation information are hierarchically clustered, and the missing data values are filled, and then the similarity calculation is performed, which can easily solve the problems of data sparseness and cold start encountered by traditional algorithms. Through the verification of the algorithm improvement based on *Python*, the experimental results show that under the sparse data, compared with the traditional book recommendation based on collaborative filtering algorithm, the improved book recommendation based on collaborative filtering algorithm has smaller prediction error, precision, recall, F-measure value is higher. Therefore, the improved book recommendation based on collaborative filtering algorithm solves the problems of data sparseness and cold start, and helps to provide users with high-quality book recommendations and realizes personalized services. However, the improved algorithm in this paper also has shortcomings. On the one hand, with the passage of time, users' interest in books may change, and historical data is time-sensitive, which may result in poor book recommendation results. Compared with the previous improvement, book recommendation consumes a longer time, and which will be important research directions for the next step.

Data Availability

The data for all Figures used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] Y. H. Chun and M. K. Cho, "An empirical study of intelligent security analysis methods utilizing big data," *Journal of Logistics, Informatics and Service Science*, vol. 9, no. 1, pp. 26–35, 2022.

[2] J. Xu and H. K. Kim, "A service system study on the effects of the influencers' characteristics on purchase intention," *Journal of Logistics, Informatics and Service Science*, vol. 9, no. 1, pp. 136–155, 2022.

[3] J. Jang and S. Kyun, "An innovative career management platform empowered by ai, big data, and blockchain technologies: focusing on female engineers," *Journal of Logistics, Informatics and Service Science*, vol. 9, no. 1, pp. 274–290, 2022.

[4] M. Zhang, T. Liu, X. Wang, and L. Sun, "Searching similar books based on student's preference for personalized education," *Science Journal of Education*, vol. 5, no. 2, pp. 60–65, 2017.

[5] B. A. Alyouzbaky, R. D. Hanna, and S. H. Najeeb, "The effect of information overload, and social media fatigue on online consumers purchasing decisions: the mediating role of technostress and information anxiety," *Journal of Logistics, Informatics and Service Science*, vol. 8, no. 2, pp. 19–46, 2021.

[6] S. Y. Lee, "Blockchain technology-based medical information system for personalized medical services," *Journal of System and Management Sciences*, vol. 11, no. 2, pp. 122–133, 2021.

[7] C. Ciurea, L. Pocatilu, and F. G. Filip, "Using modern information and communication technologies to support the access to cultural values," *Journal of System and Management Sciences*, vol. 10, no. 2, pp. 1–20, 2020.

[8] B. L. Velammal, "Typicality-based collaborative filtering for book recommendation," *Expert Systems*, vol. 36, no. 3, Article ID e12382, 2019.

[9] F. Gao, D. Chen, and T. Yan, "Personalized evaluation based on content recommendation and collaborative filtering algorithm," *Journal of Anhui University (Natural Science Edition)*, vol. 46, no. 2, pp. 22–29, 2022.

[10] Q. Wang, Z. Mao, B. Wang, and L. Guo, "Knowledge graph embedding: a survey of approaches and applications," *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, no. 12, pp. 2724–2743, 2017.

[11] C. Duan, J. Zhang, B. He, and Z. Chen, "Deep hybrid recommendation algorithm integrating attention mechanism [J]," *Computer Application Research*, vol. 38, no. 9, pp. 2624–2627+2634, 2021.

[12] Y. Ding, X. Zhang, Q. Sang, M. Jin, and H. Zhang, "Recommendation of collaborative filtering learning resources integrating learners' social network," *Modern Educational Technology*, vol. 26, no. 2, pp. 108–114, 2016.

[13] Z. P. Li, J. L. Zhang, H. J. Zhang, and G. W. Hua, "Optimal selection of movable shelves under cargo-to-person picking mode," *International Journal of Simulation Modelling*, vol. 16, no. 1, pp. 145–156, 2017.

[14] H. Zhang, C. Zhang, and Q. Zhou, "A study on book recommendation considering similarity of book recommendation lists," *Books and Information*, no. 3, pp. 128–134, 2018.

[15] Z. H. Wang and D. Z. Hou, "Research on book recommendation algorithm based on collaborative filtering and interest degree," *Wireless Communications and Mobile Computing*, vol. 2021, Article ID 7036357, 7 pages, 2021.

[16] F. Zhang, "A personalized time-sequence-based book recommendation algorithm for digital libraries," *IEEE Access*, vol. 4, pp. 2714–2720, 2016.

[17] X. Guo, Y. Gao, and Y. Lin, "Research on personalized recommendation system for libraries based on RT-RBM collaborative filtering," *Journal of Taiyuan Normal University (Natural Science Edition)*, vol. 9, no. 04, pp. 59–64, 2020.

[18] 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, Article ID 052099, 2019.

[19] Y. Liu, "Data mining of university library management based on improved collaborative filtering association rules algorithm," *Wireless Personal Communications*, vol. 102, no. 4, pp. 3781–3790, 2018.

[20] I. Noor, I. Syachrudin, M. K. Sophan, and S. Wahyuni, "Enhancing the performance of library book recommendation system by employing the probabilistic-keyword model on a collaborative filtering approach," *Procedia Computer Science*, vol. 157, pp. 345–352, 2019.

[21] Z. Wang and K. Liu, "Book recommendation system based on collaborative filtering and association analysis," *Modern Commerce and Industry*, vol. 40, no. 35, pp. 209–211, 2019.

[22] L. Wang, F. Pan, F. Qiu, and H. Chen, "Collaborative filtering recommendation algorithm integrating relative differences in scores," *Small Microcomputer System*, vol. 43, no. 7, pp. 1388–1393, 2022.

[23] N. Guo, B. Wang, Y. Hou, and P. Chang, "Collaborative filtering recommendation algorithm integrating social network features," *Computer Science and Exploration*, vol. 12, no. 2, pp. 208–217, 2018.

[24] D. Lin and B. Yang, "A recommendation score prediction method based on user interest concept lattice," *Small Microcomputer System*, vol. 41, no. 10, pp. 2104–2108, 2020.

[25] C. Sun, Y. Zhou, and T. Yan, "Research on personalized recommendation method based on hybrid collaborative filtering," *Journal of Central China Normal University (Natural Science Edition)*, vol. 54, no. 06, pp. 956–962, 2020.

[26] H. K. Chaitra and K. R. Suneetha, "Applying spectral clustering algorithm to group users by interest," *Journal of System and Management Sciences*, vol. 12, no. 1, pp. 363–382, 2022.

[27] A. Hussein, F. K. Ahmad, and S. S. Kamaruddin, "Cluster Analysis on covid-19 outbreak sentiments from twitter data using K-means algorithm," *Journal of System and Management Sciences*, vol. 11, no. 4, pp. 167–189, 2021.

[28] K. Li, M. Wang, Z. Yu, and L. Zhai, "An improved collaborative filtering recommendation algorithm integrating trust value and semi-supervised density peak clustering," *Small and Microcomputer Systems*, vol. 41, no. 08, pp. 1613–1619, 2020.

[29] J. Janekova, J. Fabianova, and M. Fabian, "Assessment of economic efficiency and risk of the project using simulation," *International Journal of Simulation Modelling*, vol. 18, no. 2, pp. 242–253, 2019.