

Application of Collaborative Filtering Algorithm in Mathematical Expressions of User Personalized Information Recommendation

Yufeng Qian*

School of Science, Hubei University of Technology, Wuhan, 430068, China

ARTICLE INFO

Article History

Received 07 Sep 2019
Accepted 24 Nov 2019

Keywords

Collaborative filtering algorithm
User personalized information
Mathematical expression
User similarity measure

ABSTRACT

In order to solve the shortcomings of the existing mathematical expressions of the user's personalized recommendation search and to increase the accuracy of the user's personalized information recommendation, in this study, the collaborative filtering algorithm, the related theory of fuzzy sets, and the evaluation methods were introduced at first. Then, the establishment of collaborative filtering model and working principle, user similarity measure, generation of mathematical expression recommendation list, design of mathematical expression recommendation model and implementation of function were introduced. Finally, the application of the collaborative filtering algorithm in the user's personalized information recommendation mathematical expression was verified by collecting data. The results showed that the recommendation mathematical expression model based on collaborative filtering algorithm had higher accuracy in collecting user personalized information, and the accuracy of the system was higher than other recommendation algorithms, which can better meet the individual needs of users. This research can provide a fast and convenient way for users to search for personalized information, and has a good guiding significance for the retrieval and recommendation of informational scientific and technical literature.

© 2019 The Authors. Published by Atlantis Press SARL.

This is an open access article distributed under the CC BY-NC 4.0 license (<http://creativecommons.org/licenses/by-nc/4.0/>).

1. INTRODUCTION

With the rapid development of the Internet era, the generation and interaction of data has become increasingly diverse. Under this new trend, various data search engines such as Google, Baidu, and 360 searching have emerged. Their appearance has brought great convenience to people's lives, enabling people to quickly and accurately search for target information, effectively promoting the development of all walks of life. However, with the increasing depth and breadth of data volume and the increasing complexity of people's data search, traditional target-based retrieval and presentation methods can no longer meet the needs of the current society. Sometimes, because people are not fully aware of their information needs, or the complexity of information needs, they can't directly abstract the target words for search. At this time, how to balance the relationship between the candidate information set and the target user, and how to provide personalized customized information content for the user with demand have become a new problem to be solved in the field of search engines [1].

Under the new demand, recommendation technology came into being. As a technical means to effectively solve this problem, it has been widely applied in various Internet platforms, especially in the e-commerce industry and social networking industry [2]. Users' demand search is no longer limited to the traditional target word

search feedback. More importantly, with the rapid development of personalized recommendation mode, people's search habits have developed into a more advanced personalized way. Personalized recommendation system is a bridge between users and information, which can help users find the information they need in the huge network system. It can actively guide customers to explore resource information that they are interested in but do not understand, and intelligently recommend items that users are really interested in according to their historical behavior information, which is a supplement to uniform resource locator (URL) classification navigation and search engine tools [3]. This technological transformation saves the time spent searching information, improves the Internet experience of users, and increases their dependence on websites. Meanwhile, for the e-commerce platform, it can also promote the consumer behavior of users and promote the sales of e-commerce [4]. Recommendation algorithm is the key of personalized recommendation system. Its performance determines the recommendation efficiency and quality of the recommendation system. Collaborative filtering algorithm is the most mature algorithm at present. It does not need to consider the content of recommended items, and can provide new things for users to recommend, which has less interference for users to visit the website and is easy to implement [5,6]. With the popularization of the Internet, scientific, and technological papers have been transformed from paper storage to network storage, with a large number, and it will take a lot of time to retrieve literatures [7]. As an important part of

*Email: yfqian@whu.edu.cn

scientific and technological papers, mathematical expressions can be used as one of the key words in the retrieval of scientific and technological papers, but they contain images, Word, PDF, Tex LaTeX, and other formats, with complex structure, variable form, and difficult retrieval [8]. Although the current research on mathematical expression retrieval has made some progress, these retrieval systems do not contain personalized recommendation models [9].

Therefore, in this research, it studies the application of collaborative filtering algorithm in mathematical expression retrieval model, aiming at improving the current situation of user personalized information in mathematical expression retrieval and providing effective retrieval approaches. It is expected to provide a fast and convenient way for users to search personalized information.

2. LITERATION REVIEW

At present, the research on the direction of user recommendation algorithm is mainly focused on recommendation model. Bobadilla *et al.* [10] summarized the recommendation system and collaborative filtering methods and algorithms, provided the original classification for these systems, and determined the implementation direction of the social domain. Toledo *et al.* [11] summarized the use of fuzzy tools in recommendation systems to detect more common research topics and research gaps, so as to propose future research directions for promoting the current development of fuzzy-based recommendation systems. Liu *et al.* [12] set out to reduce the user's search cost in the overloaded network environment, and used the Pearson correlation coefficient to evaluate the interest of users for commodity without record to form a sparse matrix for interest degree evaluation, and a neighbor set with high similarity to the target user was analyzed. The research focused on user interest similarity as the eigenvalue research, which greatly improved the accuracy value of the previous recommendation algorithm and improves the recommendation quality. Xiao [13] proposed an improved collaborative filtering algorithm, the goal was to optimize the similarity calculation method, using K-means-based clustering algorithm to select Top-N by user score as weight, and then the research idea of clusters was used to aggregate the similarity of users in the cluster, so as to realize the personalized recommendation of users. In order to improve the defect that the traditional collaborative filtering algorithm only relies on a single factor for calculation, Li *et al.* [14] proposed a URA algorithm for comprehensive scoring prediction and sorting prediction

based on the traditional PMF algorithm and x-CLiMF algorithm. The improvement point was mainly to share the characterization factors of the traditional algorithm, and use multiple factors to optimize, so as to improve the accuracy of the algorithm. Liao *et al.* [15] proposed a personalized recommendation cold boot algorithm based on the research status of collaborative filtering recommendation algorithm. The main idea was to make an investigation of the new users to obtain their interest preference, and the near-neighbor set was obtained by using the cosine similarity calculation method to improve the accuracy of the recommendation algorithm. Zhang *et al.* [16] proposed a method based on decision tree to improve the calculation of similarity, and obtain the optimal scoring model, and they added time weights in the calculation to avoid the problem of user interest transfer. The algorithm also improved the accuracy value of the recommendation technology, introduced a time specific decision factor, and obtained a better recommendation effect.

3. DESCRIPTION OF RELATED CONCEPTS

3.1. Coordinated Filtering Algorithm

The basic principle of the collaborative filtering algorithm is to refer to the user's preference for personalized selection of information, select neighbor users similar to the target user's personalized hobbies, and recommend similar content to the neighboring users that are similar to the personalized hobbies but have not been contacted [17]. This kind of recommendation method is not only simple in principle, but also can be matched to a high-accuracy personalized hobby with high precision. Based on the above thought principle, the collaborative filtering recommendation algorithm is widely used.

According to the recommendation content of the collaborative filtering recommendation algorithm, it can be divided into two categories: recommendation based on user and recommendation based on project. The user-based recommendation is based on the user's historical personalization, and it is suitable for the situation where the difference of the applicable data volume and the number of users is relatively large. The project-based recommendation is based on the attribute characteristics of the content itself, which is more suitable for the case where the recommendation user data is small and the data, is relatively stable [18,19]. Figure 1 shows the recommendation systems and recommendation methods for each system.

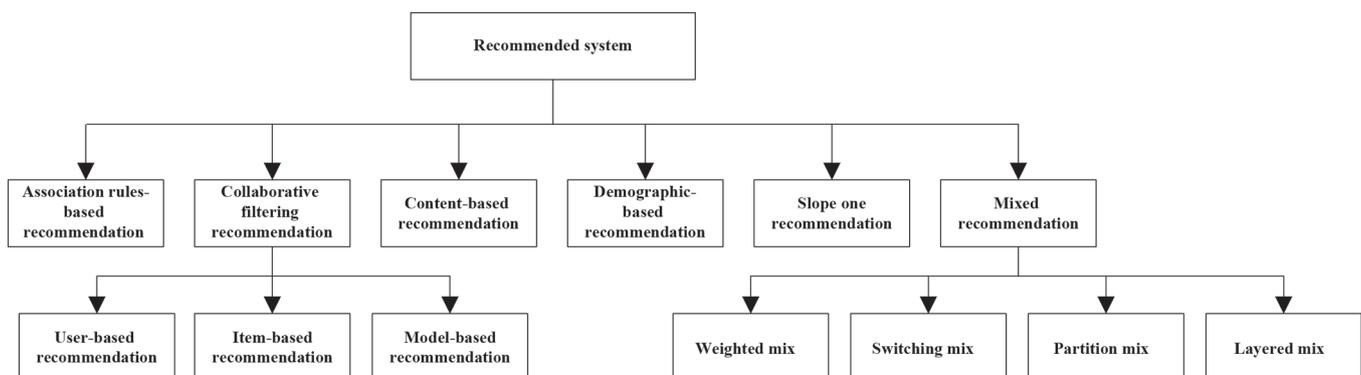


Figure 1 | Common recommendation systems and recommended methods for each system.

3.2. Fuzzy Set

3.2.1. The concept of fuzzy sets

Fuzzy sets are created by Zade’s introduction of traditional fuzzy set theory. Fuzzy sets are usually used to express the situation where the position is not clear. The logical truth value of fuzzy logic is $[0, 1]$ [20].

Assuming that M is a non-empty set and the membership function of F is μ_F , then the mathematical expression of the definition of fuzzy set F is Equation (1):

$$F = \{ \langle m, \mu_F(m) \rangle \mid m \in M \} \tag{1}$$

$$\mu_F : M \rightarrow [0, 1] \tag{2}$$

where, $\mu_F(m)$ takes a single value in $[0, 1]$, $\mu_F(m)$ is the degree to which the element m in M belongs to F . Any mapping of W in the interval of $[0, 1]$ is μ_b , which is the membership function of fuzzy subset, and its expression is as follows:

$$\mu_b : X \rightarrow [0, 1] \tag{3}$$

$$w \rightarrow \mu_b(w) \tag{4}$$

where, μ_b can be regarded as the membership degree of w relative to $A(w)$.

In the traditional recommendation system, such as the coordinated filtering recommendation system, the user’s evaluation of a certain content is limited to the determined numerical score, that is, “0” means dislike; “1” means like. Such evaluations often fail to accurately reflect the user’s preference for a certain content. Based on this situation, the concept of fuzzy sets was introduced into the recommendation system in this study, and the membership relationship in the traditional classical mathematics set was expanded. The membership degree of the elements in the set that can only take certain values (such as “0,” “1,” “-1,” etc.) was expanded to take any value in the interval $[0, 1]$. Using such an approach can reflect the user’s preference for a certain content more clearly and effectively to a certain extent by evaluating the uncertainty in the decision making of the indicator.

3.2.2. Similarity measure of fuzzy sets

In this study, the fuzzy set recording method was adopted in the user’s personalized search, because the similarity measure of the fuzzy set can be used as an important index to describe the similarity or closeness of the two fuzzy sets.

Suppose there are two fuzzy subsets A and B on the domain X , and the closeness between A and B was defined as

$$D(A, B) = \frac{1}{2} [A \cdot B + (1 - A \otimes B)] \tag{5}$$

where, $A \cdot B$ represents the inner product of A and B . The larger the D value is, the higher the similarity between the two fuzzy sets is, and the more similar the personalized preferences of the corresponding two users are.

Different preference behaviors of users are collected, the fuzzy score matrix of “user-equation” is formed by combining membership function, and the similarity among users is calculated.

3.3. Evaluation Method

There are many evaluation criteria in the recommendation system, and there are performance differences among them. In this research, the measurement methods and evaluation methods corresponding to commonly used evaluation and measurement indicators are statistically analyzed, as shown in Table 1. In addition to the evaluation indicators in Table 1, there are also coverage, fineness, robustness, and trust methods.

According to the nature of the experimental data in this study, it is necessary to select a method that is sensitive to error performance. Therefore, the root mean square error (RMSE) method will be used in statistical accuracy measurement method, RMSE can better reflect the precision of experimental data measurement.

In order to facilitate the calculation of the gap between the predicted score and the multi-feature comprehensive score, the prediction scores of mathematical expressions were expressed using set, as the Equation (6):

$$PU_i \in \{PU_1, PU_2, \dots, PU_n\} \tag{6}$$

The set of multi-features comprehensive scores was expressed as Equation (7):

$$P_i \in \{P_1, P_2, \dots, P_i\} \tag{7}$$

The difference between the predicted score and the multi-feature comprehensive score was calculated as Equation (8):

$$\begin{aligned} RMSE &= \sqrt{\frac{(PU_1 - P_1)^2 + (PU_2 - P_2)^2 + \dots + (PU - P_n)^2}{n}} \tag{8} \\ &= \sqrt{\frac{\sum_{i=1}^n (PU_i - U_i)^2}{n}} \end{aligned}$$

RSME indicates the degree of deviation between the predicted data and the actual data. Therefore, the larger the RSME value, the more the prediction data deviates from the actual data, and the lower the recommendation accuracy; the smaller the RSME value, the smaller the deviation between prediction data and the actual data, and the higher the recommendation accuracy.

Table 1 | Measure and evaluation methods for measurable indicator of recommendation.

Measurable Indicator Type	Classification of Measure Methods Type	Measure Method
Top-n recommendation	Statistical type	Evaluation of absolute error, root mean square error, absolute error of normalized evaluation
Score prediction	Decision support type	Recall rate, accuracy rate, roc (rate of change)

4. MATHEMATICAL EXPRESSION RECOMMENDATION MODEL

4.1. Operating Principle of Collaborative Filtering Recommendation Model

In the recommendation mathematical expression model of the collaborative filtering algorithm, the scoring matrix of the “user - mathematical equation” can be generated according to different personalized requirements of user. The scoring matrix consists of three parts, namely user set U , item parameter (i.e., mathematical expression), and $m \times n$ scoring matrix Q_{ixj} . The illustration of user set, item parameter, and scoring matrix Q_{ixj} is shown in Equations (9-11):

$$U = \{1, 2, \dots, n\} \tag{9}$$

$$\text{Item parameter} = \{s_1, s_2, \dots, s_m\} \tag{10}$$

$$Q_{ij} = \begin{cases} 1, \text{ Preference record} \\ \text{NULL, No record} \end{cases} \tag{11}$$

Table 2 is the score table of “user item parameters.” The search at the marked position “Point_{ij} = 1” indicates that the user has searched for the item parameter, which is denoted as the preference record. If there is no record, it is denoted as “NULL.” As can be observed from Table 2, the item parameters searched by the user 1 are $S_1, S_3,$ and S_4 , and the item parameters searched by the user 2 are S_2 and S_4 , and the item parameters searched by the user 3 are $S_1, S_3, S_4,$ and S_5 .

By analyzing the search behaviors in Table 2, it is found that the search behaviors of user 1 and user 3 are very similar, and they can be regarded as nearby users. According to the search behavior of item parameter S_5 of user 3 in the figure, the personalized interest preference of user 1 can be predicted through collaborative filtering algorithm, and it can be calculated that user 1 may also be interested in item parameter S_5 of user 3, so that item parameter S_5 can be recommended to user 1. The algorithm principle of its search behavior is shown in Figure 2.

Table 2 | “User-mathematical expression” score sheet.

U	S ₁	S ₂	S ₃	S ₄	S ₅
1	1	NULL	1	1	(Recommend)
2	NULL	1	NULL	1	NULL
3	1	NULL	1	1	1

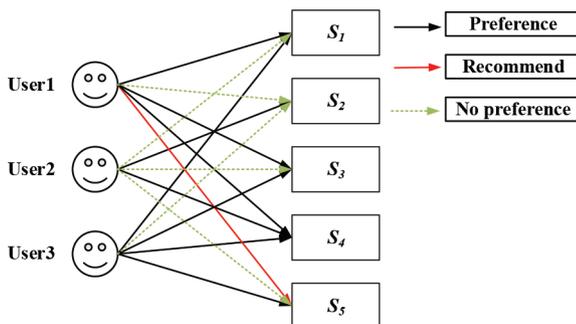


Figure 2 | Basic principles of user-based collaborative recommendation.

The recommendation process of mathematical expressions consists of three parts: collecting personalized preferences of user, similarity measures between users and recommendation prediction scores. Its working principle is shown in Figure 3. As can be observed from Figure 3, the user’s personalized preferences start from the pre-processing of the formula, and variable substitution was used to obtain a sample of the formula, and the user’s personalized preferences were collected by browsing time, click volume, collection, and download. Then, the similarity measure was performed, the scoring matrix was obtained according to the collected user’s personalized preference, the neighboring user set was obtained from the closeness, and the predicted score was finally carried out to generate the recommendation list.

4.2. User Similarity Measure

4.2.1. User similarity measurement method

In the study of user similarity measurement methods, the main methods are Euclidean distance, Manhattan distance, cosine similarity, Pearson correlation coefficient, and so on. The main idea is to measure the similarity by the distance factor by calculating the distance between the features of the thing. If the distance is smaller, the similarity is greater, and vice versa. Collaborative filtering algorithm research is generally based on the “user - formula” scoring matrix to calculate the similarity between user [21].

The closeness measure metric fuzzy set method used in this study was optimized for the traditional “user - formula” scoring matrix. The multi-feature factor was used to calculate the user closeness according to the user’s preference. The higher the calculated value, the higher the user similarity, and the closer the user’s preference. Equation (12) shows the fuzzy score calculation method for multi-feature factors in closeness calculation:

$$D(U_a, U_b) = \frac{1}{2} [U_a \cdot U_b + (1 - U_a \otimes U_b)] \tag{12}$$

$$= \frac{1}{2} [(p_{a1} \wedge p_{b1}) \vee (p_{a2} \wedge p_{b2}) \vee \dots \vee (p_{am} \wedge p_{bm}) + 1 - (p_{a1} \vee p_{b1}) \wedge (p_{a2} \vee p_{b2}) \wedge \dots \wedge (p_{am} \vee p_{bm})]$$

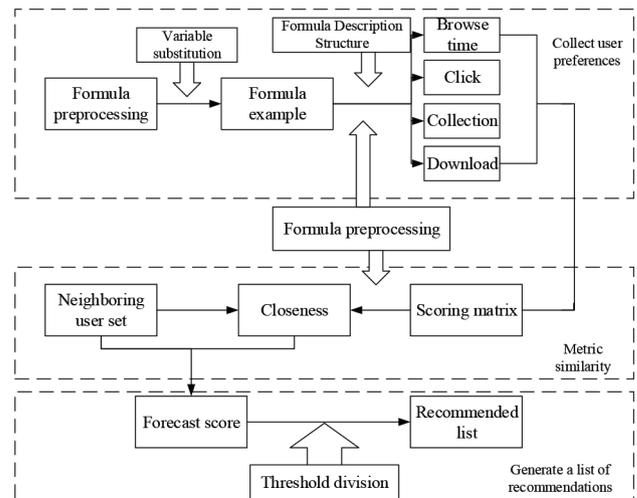


Figure 3 | Basic structure diagram of mathematical expressions.

In Equation (12), the degree of closeness between two users is denoted by $D(\underline{U}_a, \underline{U}_b)$. The inner product and outer integral of the fuzzy set determined by the multi-characteristic factor of user a and user b are respectively $\underline{U}_a \cdot \underline{U}_b$ and $\underline{U}_a \otimes \underline{U}_b$. Operator \vee means take large and operator \wedge means take small.

4.2.2. Selection of neighbor user sets

According to the derived calculation equation, the top M neighbor users with the highest closeness to the target user U_q can be calculated from the plurality of neighbor users. After ranked by the closeness, the higher ranking of user indicates that the closer the target user is, that is the closer the hobby is.

Based on the scoring fuzzy set method based on multi-feature factor, the top M users with the highest closeness to the target user U_q can be calculated as neighboring user sets, and it was marked as H_{U_q} . As expressed in Equation (13), U_{i_g} represents the user with the top g highest degree of closeness to the target user, $g \in \{1, 2, \dots, M\}$:

$$H_{U_q} = \{U_{i_1}, U_{i_2}, \dots, U_{i_g}\} \tag{13}$$

4.3. Generation of Mathematical Expressions Recommendation List

4.3.1. Weighted prediction scores for unrecorded mathematical expressions

The scoring matrix based on multi-feature factor was used to calculate the similarity between users, and then the score of the unrecorded formula was calculated according to the weighted prediction scoring mechanism, and the recommendation result list was obtained after the threshold screening.

Based on the “user - equation” scoring matrix and the degree of interest of the user on the equation, it can be divided into two types: scored and no score records, wherein no score record indicates that the user has no search record or is not interested in the equation. Equation of no score record is expressed as S_n . Figure 4 shows the predictive scoring of the unrecorded equation by the neighbor user of the target user U_q .

The scoring matrix shown in Figure 4 is a multi-feature factor fuzzy set method adopted in this study, and the target user U_q and its neighbor set users were scored without a record expression, wherein p_{i_g} represents the unrecorded score of the neighbor centralized user

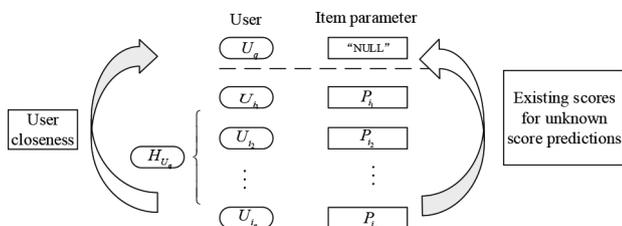


Figure 4 | Score prediction chart for unrecorded mathematical expressions.

U_{i_g} and the target user U_q . The closeness of the neighbor set user and the target user is used as the weight of the multi-feature factor comprehensive score. According to the closeness of each neighbor set user and the target user as the weight of the multi-feature factor comprehensive score, the weighted prediction score $PV(U_q, F_n)$ of the target user and its neighbor set users for the unrecorded formula F_n was obtained by the Equation (14). In the equation, $N(U_q, U_{i_a})$ is the close degree of the first i_a neighbor user to the target user, $P(U_{i_a}, F_n)$ indicates the score in the scoring matrix of the neighbor user for the formula without a record of the target user.

$$PV(U_q, F_n) = [D(U_q, U_{i_1}) \cdot P(U_{i_1}, F_n) + D(U_q, U_{i_2}) \cdot P(U_{i_2}, F_n) + \dots + D(U_q, U_{i_g}) \cdot P(U_{i_g}, F_n)] = \sum_{a=1}^g [D(U_q, U_{i_a}) \cdot P(U_{i_a}, F_n)] \tag{14}$$

4.3.2. The generation method of interest lists

The source of the traditional interest list is the formula of the target user’s neighbor user set [22]. In this study, two kinds of interest list generation methods were used for comparative analysis, as shown in Figures 5 and 6, which are respectively an interest list of determined number of and a list of interest recommendations generated by the threshold filtering method.

For the interest list of determined number, the basic principle is to arrange the priority according to the score, and taking the first k formulas as a list of interest recommendations. As shown in Figure 5, for the selected item center target ①, the k value was selected as 4. Therefore, the four equations ②, ③, ⑤, and ⑥ in the Figure 5, as the sequence with the highest score, are the recommended results of the interest list of the target item. According to the distribution of points in the figure, it can be observed that points with discrete distribution will cause relatively large errors. Therefore, it is necessary to extract a certain number of recommended mathematical expressions. When there are not enough highly rated mathematical expressions near the item ①, the less highly rated mathematical expressions are forced to be recommended. However, users are not

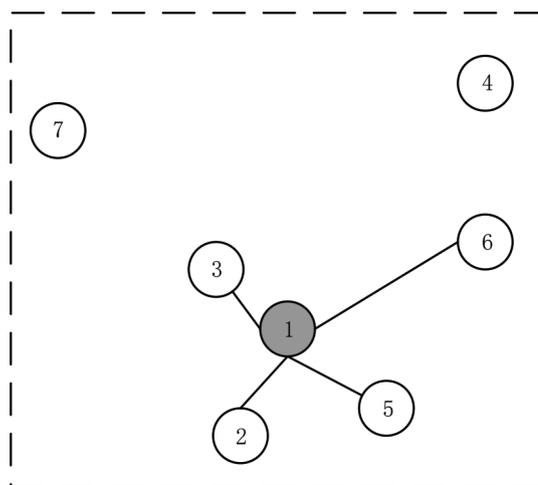


Figure 5 | Interest list of fixed number.

necessarily interested in mathematical equations with low scores, which will affect the recommendation effect.

For threshold-based recommendation list of interest, the interest list generated by the threshold filtering method is different from the previous interest recommendation list methods of fixed number, and its principle needs to ensure that the score is within a certain range, thereby improving the recommendation effect. As shown in Figure 6, for the selected item center target ①, the selected recommendation result falls within the range in which the center object R is in the radius. Therefore ②, ③, and ⑤ will be the recommendation result of ①. This method is an optimization of the former method and it improves the quality of the interest list recommendation.

Figure 7 is a flow diagram of the generation of an interest list of determined number and a list of interest recommendation lists generated by the threshold filtering method.

Through the above analysis, the threshold filtering method was adopted in this study to generate the recommendation results. First, the formulas with no record of the recommendation users were ranked by scores; then, filtering and screening were performed according to the set thresholds; finally, several formulas whose

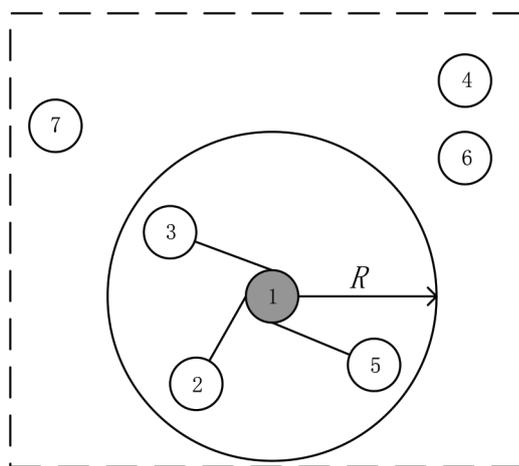


Figure 6 | Threshold-based recommendation list of interest.

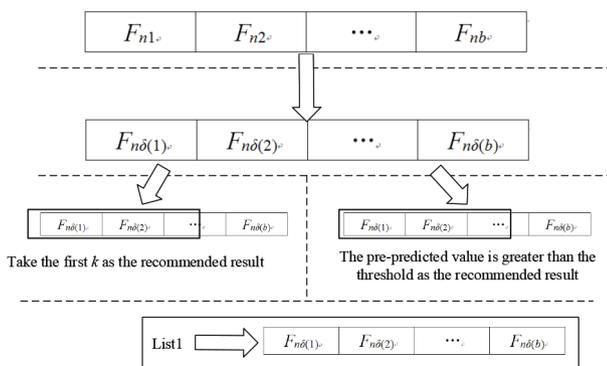


Figure 7 | Mathematical expression recommendation list flow chart generated by two methods.

scores were greater than the threshold were obtained as the interest recommendation list of the users to be recommended.

4.4. Mathematical Expression Recommendation Model

The model design mainly consists of three parts: user search module, behavior information collection module, and expression recommendation module. The user search module was mainly responsible for the background module operation after the user logs in to the system, match of the information index, and the return of the index result to the user interface for user interface search; the behavior information collection module mainly included that after the user logs in, the user identity was identified, the historical behavior information of the user was indexed, the information was classified, and the corresponding behavior membership function was used to calculate the information, and the multi-feature fuzzy score set was obtained to form a “user - formula” scoring matrix. The recommendation module was mainly to select the things with high interest according to the behaviors of the user after logging in to the system, then perform the similarity calculation to find the set of neighboring users, and obtain the list of interest recommendation through the operations of predictive scoring, sorting, threshold screening, etc., of the formula with no record, and finally realize personalized recommendations when users log in again. The design model diagram is shown in Figure 8.

After the user logged in to the system model, the user identity was first identified, the user’s historical behavior information was retrieved according to the user’s identity, and the mathematical expression hierarchy was used to classify the presented formulas with user preference, and the search result that satisfies the condition was returned to the user viewing interface. In this interface, the user saved operations such as browsing, selecting, and clicking on the returned results. The information was saved as a basis for the next similarity calculation, score prediction, and interest list recommendation. In this study, the multi-feature factor fuzzy set method was adopted to collect the multi-class information to form a “user - formula” scoring matrix, and the selected formula with no record was predicted and scored, and then the threshold screening was used to obtain the recommendation list of mathematical expressions.

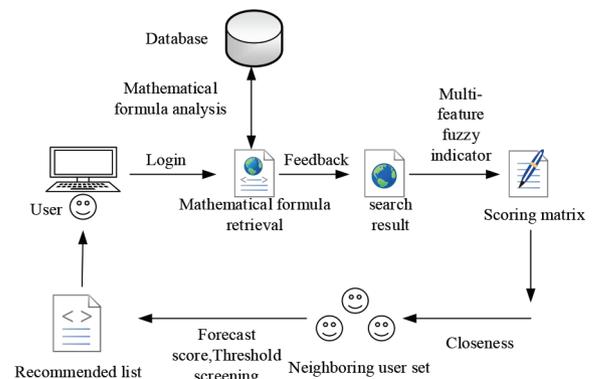


Figure 8 | Structural diagram of the mathematical expression recommendation model.

5. DISCUSSION

5.1. Analysis of Results

In this experiment, the user can input the expression in the browser side recommendation by the mathematical expression according to his or her own personal preference. After the system analyzes his or her personalized behavior according to the user’s historical login, it will display the mathematical expressions of interest in the interface of the recommendation module. The effect model of the mathematical model is shown in Figure 9.

The recommendation results and user input are placed in the same interface, so that the historical user can get personalized recommendation service before entering the mathematical formula, which is in line with the user’s habits, while the corresponding recommendation results will not appear for new users.

The experimental data in this study was calculated by counting the 200,000 scores of 3012 mathematical expressions by 130 users. In the experiment, the number of neighboring users n and the threshold β were set to calculate the RMSE, and the result was analyzed to get the optimal solution. Table 3 shows the RMSE of the user’s personalization in the mathematical expression under different parameters.

In Table 3, it can be observed that the computer has been encountering errors when $\beta = 0.8$, indicating that the threshold is too high, and there are few conditions that meet the recommendation target after the score is close to the full score. As it is a fuzzy score, the comprehensive score range is $[0, 1]$. When $\beta = 0$, it means that there is no rating limit, and all expressions are recommended. The closer β is to 1, the higher the recommendation threshold. And when the threshold is $\beta = 0.6$, RSME has been in a stable state. In the range of values of the threshold (0.6, 0.8), it can be observed that the value of RSME was the smallest. When the threshold was

0.7, the threshold channel minimum value and the number of users is 12, indicating that the recommendation system can approximate the interest recommendation effect of the mathematical expression.

5.2. Results of Discussion

In this study, by collecting the experimental data and evaluating it to adjust different parameters, it was concluded that the recommendation precision was different under different thresholds, indicating that if the threshold is too high, there are few conditions that meet the recommendation goal after the score is close to the full score. And when the threshold was 0.6 RSME has been in a stable state. In the range of values of the threshold (0.6, 0.8), it can be observed that the value of RSME was the smallest. When the threshold was 0.7, the threshold channel minimum value and the number of users are 12, indicating that the recommendation system can approximate the interest recommendation effect of the mathematical expression. Finally, it was concluded that the recommendation system has the best recommendation effect on the mathematical expression when the threshold is 0.7, and the accuracy of the system was higher than other recommendation algorithms, which can better meet the user’s individual needs.

6. DISCUSSION

In order to solve the shortcomings of the existing mathematical expressions in the user’s personalized recommendation search, in this study, the collaborative filtering algorithm, the related theory of fuzzy sets, and the evaluation methods were introduced at first. Then, the establishment of the collaborative filtering model and the design of the mathematical expression recommendation model and the realization of the function were introduced. Finally, the data experiment showed that when the threshold is 0.7, the recommendation system has the best recommendation for the interest of mathematical expressions. And the collaborative filtering algorithm was applied to the user’s personalized information to promote the recommendation of mathematical expression, it was found that the mathematical expression recommendation model based on the collaborative filtering algorithm had higher accuracy in collecting user personalized information, and the accuracy of the system was higher than other recommendation algorithms, which can better meet the individual needs of users.

This research can provide a fast and convenient way for users to search for personalized information, and has a good guiding significance for the retrieval and recommendation of informational scientific and technical literature.

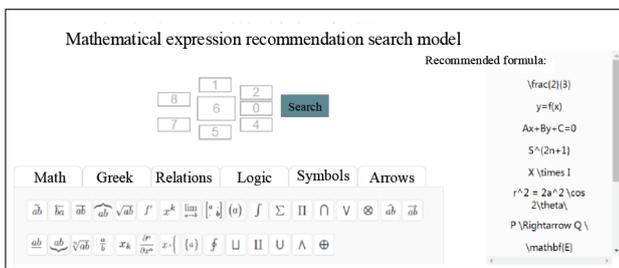


Figure 9 | Mathematical expression recommendation model interface renderings.

Table 3 | RMSE of user personalization in mathematical expressions under different parameters.

β	$n = 4$	$n = 8$	$n = 12$	$n = 16$
0.1	1.9041	2.1439	2.5113	2.7539
0.2	1.9041	2.1439	2.5113	2.7539
0.3	1.9041	2.1439	2.5113	2.7539
0.4	1.9041	2.1439	2.5113	2.7539
0.5	1.9041	2.1439	2.5113	2.7539
0.6	1.9041	2.1439	2.5113	2.7539
0.7	1.9041	1.5406	1.0612	Error
0.8	Error	Error	Error	Error

Note: Error indicates that the computer has an error in the calculation process.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

We would like to thank you for following the instructions above very closely in advance.

REFERENCES

- [1] E. Guruceaga, A. Garin-Muga, G. Prieto, Enhanced missing proteins detection in NCI60 cell lines using an integrative search engine approach, *J. Proteome Res.* 16 (2017), 4373–4390.
- [2] A.Y. Maa, B. Wojciechowski, K.J. Hunt, Early Experience with Technology-Based Eye Care Services (TECS), *Ophthalmology.* 124 (2017), 539–546.
- [3] A. Klačnja-Milićević, M. Ivanović, B. Vesin, Enhancing elearning systems with personalized recommendation based on collaborative tagging techniques, *Appl. Intell.* 48 (2018), 1519–1535.
- [4] V. Subramaniaswamy, G. Manogaran, R. Logesh, An ontology-driven personalized food recommendation in IoT-based healthcare system, *J. Supercomput.* 75 (2019), 3184–3216.
- [5] N. Friedli, Z. Stanga, A. Culkin, Management and prevention of refeeding syndrome in medical inpatients: an evidence-based and consensus-supported algorithm, *Nutrition.* 47 (2017), 13–20.
- [6] J. Zhang, J. Tang, C. Ma, Fast and flexible top- k similarity search on large networks, *ACM Trans. Inf. Syst.* 36 (2017), 1–30.
- [7] D. Ozsoyeller, Multi-robot symmetric rendezvous search on the line with an unknown initial distance, *IEEE Trans. Robot.* 29 (2018), 1366–1379.
- [8] R. Kumar, S.J. Darak, A. Yadav, Channel selection for secondary users in decentralized network of unknown size, *IEEE Commun. Lett.* 21 (2017), 2186–2189.
- [9] P. Marciulionis, Analysis of space charge distribution in dc corona discharge field computed with finite-element method, *IEEE Trans. Plasma Sci.* 45 (2017), 1698–1703.
- [10] J. Bobadilla, F. Ortega, A. Hernando, *et al.*, Recommender systems survey, *Knowl. Based Syst.* 46 (2013), 109–132.
- [11] R.Y. Toledo, L. Martinez, Fuzzy tools in recommender systems: a survey, *Int. J. Comput. Intell. Syst.* 10 (2017), 776–803.
- [12] C. Liu, J. Shen, L. Nie, *et al.*, Augmented collaborative filtering for sparseness reduction in personalized POI recommendation, *ACM Trans. Intell. Syst. Technol.* 8 (2017), 71.
- [13] C. Yin, L. Shi, R. Sun, *et al.*, Improved collaborative filtering recommendation algorithm based on differential privacy protection, *J. Supercomput.* (2019), 1–14.
- [14] C. Li, J. Luo, W. Xin, Personalized travel route recommendation using collaborative filtering based on GPS trajectories, *Int. J. Digit. Earth.* 11 (2018), 284–307.
- [15] W. Liao, S. Zhang, D. Wang, *et al.*, A recommendation algorithm for elective courses in university based on nearest neighbor model and probability matrix factorization, *J. Liaoning Tech. Univ.* 36 (2017), 976–982.
- [16] J. Chen, Uljji, H. Wang, Z. Yan, Evolutionary heterogeneous clustering for rating prediction based on user collaborative filtering, *Swarm Evol. Comput.* 38 (2017), 35–41.
- [17] C.A.N. Santos, D.L.N. Martins, N.D.A. Mascarenhas, Ultrasound image despeckling using stochastic distance-based BM3D, *IEEE Trans. Image Process.* 26 (2017), 2632–2643.
- [18] E.K. Yeoha, M.C.S. Wonga, E.L.Y. Wonga, Benefits and limitations of implementing Chronic Care Model (CCM) in primary care programs: a systematic review, *Int. J. Cardiol.* 258 (2017), 279–288.
- [19] P.R. Mills, A.E. Weidmann, D. Stewart, Hospital staff views of prescribing and discharge communication before and after electronic prescribing system implementation, *Int. J. Clin. Pharm.* 39 (2017), 1320–1330.
- [20] K.H. Lim, J. Chan, C. Leckie, *et al.*, Personalized trip recommendation for tourists based on user interests, points of interest visit durations and visit recency, *Knowl. Inf. Syst.* 54 (2017), 375–406.
- [21] K.G. Saranya, G.S. Sadasivam, Personalized news article recommendation with novelty using collaborative filtering based rough set theory, *Mob. Netw. Appl.* 22 (2017), 719–729.
- [22] A.M. Yagci, T. Aytekin, F.S. Gurgen, Scalable and adaptive collaborative filtering by mining frequent item co-occurrences in a user feedback stream, *Eng. Appl. Artif. Intell.* 58 (2017), 171–184.