Computer Fraud and Security
ISSN (online): 1873-7056

Research on Collaborative Filtering Recommendation Method Based on User Behavior Analysis

Xuewei Li*, Wei Hong
School of Intelligent Sports Engineering, Wuhan Sport University, Wuhan 430079, Hubei
**Corresponding Author.*

Abstract:
In online resource search systems, users who lack clear query or viewing objectives often have a certain degree of blindness. Users hope to obtain recommended items from other users with similar behaviors, because the filtered items are more likely to be accepted by new users. Based on this, the paper constructs a knowledge graph using the ontology concept architecture of the project as the knowledge representation foundation. It conducts in-depth research on user behavior around user behavior ontology, project resource ontology, user relationship graph instances, user similarity, and rating matrix completion, in order to find the target user's similar neighbor set. Based on the user's collaborative filtering recommendation algorithm, the predicted rating of the recommended project is calculated and Top-n projects are recommended to the target user. The experimental evaluation results indicate that the collaborative filtering recommendation method based on user behavior analysis proposed in this study has good recommendation quality in online resource recommendation systems.

Keywords: user behavior, collaborative filtering recommendation, knowledge graph, user similarity

INTRODUCTION

Online resource search and recommendation system is a kind of information service system that makes it easier for users to find their most needed user profiles among the huge amount of user resources. Through the analysis of the user's previous user behaviour, it recommends the project resources that similar users like for the target users by means of collaborative user filtering, so that the target users perceive more user common sense and sense of integration, thus enhancing their user satisfaction.

This study takes the conceptual architecture of project resource ontology as the knowledge representation basis to construct a knowledge graph, conducts in-depth research around user behaviour ontology, project resource ontology, user-project resource relationship graph instance, user similarity, and scoring matrix complementation, etc., and calculates the similarity between the target user and similar neighbours so as to obtain the project resources that meet the needs of the users according to user-based collaborative filtering method, and proposes the User Collaborative Filtering Recommendation based on Learning Behavior Analysis (LBA-UCF) method. project resources that meet the needs of users, and proposed User Collaborative Filtering Recommendation based on Learning Behavior Analysis (notated as LBA-UCF).

RESEARCH STATUS

Currently, collaborative filtering recommendation is the most successful paradigm applied in the commercial world, and it is also the typical of the most research results in the academic world. Research on recommendation methods based on user behaviour analysis mainly calculates the similarity between users, so user-based collaborative filtering recommendation is used. Recommending items, commodities or resources based on the general preference of similar users has a certain social nature, such as 'users who have watched video A have also watched video B, C, D...' or 'users similar to you have also watched video E, F, H...' in video websites. ...', all belong to this type of recommendation. The advantage of this method is that it only requires the user to rate the item, does not require user characteristics and item characteristics, easy to understand, and the accuracy of the recommendation results is also higher. However, the disadvantages are obvious, such as the cold-start problem, the data sparsity problem, and the over-reliance on historical data, as well as the long-tailed and cold items that seldom get attention. Among them, as the number of online user system resources continues to increase, the sparseness problem becomes more significant, and if the ratings are too sparse, the collaborative filtering recommendation accuracy will be seriously affected, even lower than the non-personalized recommendation system. In addition, a small number of malicious ratings can also lead to data distortion, which affects the recommendation effect. For this reason, many researchers have proposed various methods to improve data sparsity.

Gong[1]improved the collaborative filtering recommendation algorithm by combining the double K-means of user clustering and item clustering, the main point is to use Smooth scores to fill the scoring matrix in user clustering, and experiments proved that this algorithm effectively reduces data sparsity and improves the accuracy of recommendation.Salehi[2] proposed a new method of calculating the similarity of the recommendation system, which uses the tree state structure analysis method to build a preference model for the users, which has provided some reference value for many subsequent studies.Quan[3] based on the study of users' personalized characteristics on the collaborative filtering recommendation system. system, the method is to use the tree state structure analysis method to build a preference model for the user, which provides a certain reference value for a lot of research after that.Quan[3] based on the study of the user's personalized characteristics of the impact of collaborative filtering recommendation algorithm, respectively, from the personality Features and personality Rating two different dimensions to compute the user Similarity, to generate similar user recommendation items for the target users to generate recommendation items for similar users. Yao Jinbo[4] based on the analysis of machine user correlation algorithms, the PCA dimensionality reduction method is used to downscale the highly sparse rating matrix to solve the problem from the perspective of the opposite of data filling. Zhao Wei[5] first classified users based on K-means clustering algorithm, and then calculated to generate different Top-N recommendation lists based on each category of user interest features respectively. Wang Xuelong[6], on the other hand, proposed a collaborative filtering recommendation algorithm for cloud mode, which is optimized by fusing the relevant algorithms of behaviour, similarity, and relevance on the basis of the association clustering calculation of user behaviour.

In order to solve the problems of collaborative filtering algorithms such as data sparsity, there are many other ideas, such as combining the established research with algorithms such as clustering, regression, graphs, etc. [7,8] or techniques such as matrix decomposition, multi-modal data fusion, and matrix padding [9,10,11] for combined recommendation. In terms of clustering algorithms, division clustering algorithms are often improved by scholars for clustering users due to the advantages of high accuracy and operability. Improvement methods mainly include elbow method [12], profile coefficient [13], spectral clustering [14], coarse clustering [15] and so on.

In a comprehensive analysis, most of the above related studies use the methods of downscaling and clustering, which to some extent improves the impact of data sparsity and cold-start problems on the quality of recommendation, but ignores the fact that the user user behaviour logs and the feedback of operation clicking behaviour can also reflect the user's preference information, and at the same time, the previous studies are less involved in the personalized recommendation that combines with the knowledge graph. Comprehensively, the basis of the data sparsity improvement method in collaborative filtering recommendation, one is the scoring matrix clustering filling to improve the matrix density; the second is the use of scoring matrix dimensionality reduction processing to reduce the matrix volume. These two aspects are not mutually exclusive, and the combination of them can also provide good results.

For this reason, this paper takes the course example of the educational knowledge graph and the ontology conceptual architecture of the course as the knowledge representation basis, combines the explicit rating of user resources by the user and the implicit rating of user behaviors by the user as a way to complement the sparse rating matrix, and puts forward a user collaborative filtering recommendation method based on the analysis of user behaviors, in order to improve the credibility of the recommendation and stimulate the user's user activeness by the user.

METHODS

A Recommendation Model Framework Based on User Behavior Analysis

Based on the review of relevant research results at home and abroad, this study analyzes user behavior interests and user resource characteristics, and proposes a user collaborative filtering recommendation method based on the combination of user behavior interest ontology and user resource knowledge ontology. The system model framework of the LBA-UCF method consists of four parts: ontology library, knowledge graph, collaborative filtering recommendation algorithm, and user resource presentation, as shown in Figure 1.

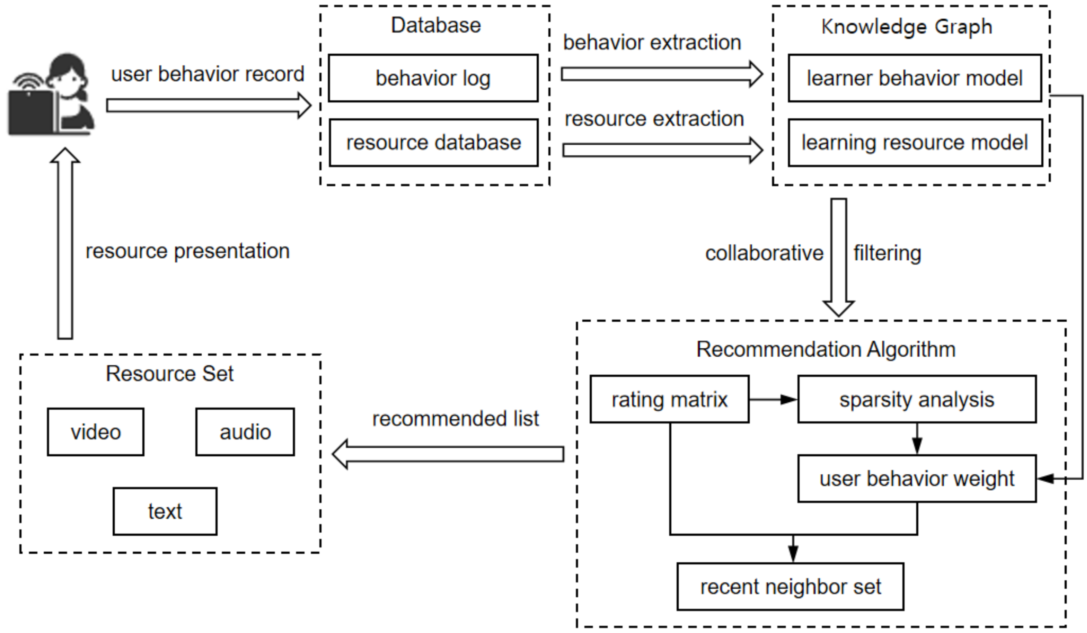

Figure 1. Framework diagram of LBA-UCF model

Semantic Metadata Model

This article adopts the semantic metadata model description method of the semantic web to structurally restructure and describe user behavior data and user resource knowledge data, forming metadata that can describe the data and be understood by machines.

Resource metadata model

In the online user domain, user resources have a more complex structure compared to products or projects on the internet, in the form of text, images, audio, and video.The data storage format is mostly unstructured.The process of transforming unstructured user resources into structured associated data requires a semantic information extraction system and a semantic information storage tool to display, which presents the knowledge graph of the domain ontology. Finally, complex reasoning is completed through RDF or OWL description to meet system requirements. Therefore, the user resource metadata model is defined as:

$$\text{IModel} = \{\text{Concept},\ \text{Relation},\ \text{Property},\ \text{Axiom}\}$$

Among them, Concept represents the concept or instance in the resource ontology Onto. This article mainly refers to user resources, which can also be a knowledge point; Relation represents the semantic relationship between concepts, expressed as Relation=(SunclassOf, InstanceOf), where SubclassOf is the hierarchical relationship and InstanceOf is the instance relationship; Property represents the attribute of a resource, denoted as Property(type, mode), type represents the category of resources, such as textbooks, courseware, test questions, etc., and mode represents the modality of resources, such as text, video, audio, etc.

User metadata model

In the field of online users, the ontology of user behavior interests includes basic information obtained in the early stage and behavior information and rating data obtained in the later stage, which together constitute the basic portrait of the user. The behavior information of users needs to be obtained through the analysis of user behavior access log data, forming implicit interest feedback, while rating data is usually considered as explicit interest feedback. In addition, this article adopts user collaborative recommendation, so the user interest ontology metadata model should also include a description of the user's similar neighbor set. Therefore, by using metadata such as basic user information, interests and preferences, and social relationships to abstractly describe users, a more comprehensive user profile can be obtained. Each user can be described using a user ontology model,And it can dynamically update this ontology based on the user's continuous behavior. Define the user metadata model as:

$$UModel = \{Uinfo, \ Interest, \ Neighbour\}$$

Among them,Uinfo=(ID、Name、Sex、Age、Profession…), Fill in by user during registration. Neighbor represents the set of similar neighbors of the target user; Interest represents the collection of users' interest in resources, Interest=$\{<C_n, Degree>|C_n \in Ontology \text{ and } Degree \in [0,5]\}$ include both explicit feedback interest based on ratings and implicit feedback interest derived from a series of user behaviors after entering the system, including user access records, access time, access duration, access frequency, as well as downloads, shares, collections, etc. of resources.

User Behavior Model

It is generally believed that user rating of resources is an explicit feedback of user interest, and this explicit feedback data can be used for direct calculation in collaborative recommendation algorithms. The more user rating data, the higher the accuracy of the recommendations obtained.But the reality is not like that, not every user is enthusiastic about rating or evaluating the resources they visit, it can be said that only a few users have a rating habit. Therefore,the scoring matrices in the system are generally sparse.In the online user environment,the interaction environment of users is relatively positive.Through encouragement and reward mechanisms, most users are willing to rate and evaluate user resources. However, users generally only rate or evaluate related user resources based on their subjects and interests, and cannot evaluate all resources. Therefore, the rating matrix is still sparse.Therefore, online user systems still need to obtain auxiliary information on implicit user ratings to establish user models. This article believes that the time a user stays on a page and the number of clicks can reflect their level of interest in the content of the resource. Therefore, by extracting the user's access log files to the resource, the implicit interest rating of the user can be obtained, and the comprehensive interest level of the user in the resource can be comprehensively processed.

Based on the above analysis, it can be concluded that users' interest in user resources can be measured by explicit and implicit ratings.Due to sparse data,implicit rating,which refers to user behavior information,is needed to assist in judgment when displaying ratings.This study combines browsing behavior and clicking behavior from user network logs to obtain a comprehensive score of user behavior. As shown in Figures 2, 3, and 4, they are the user ontology, user resource ontology, and user user resource knowledge graph fragments, respectively.

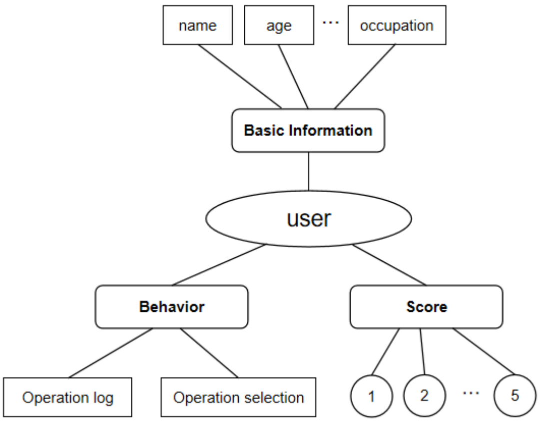

Figure 2 . User Ontology

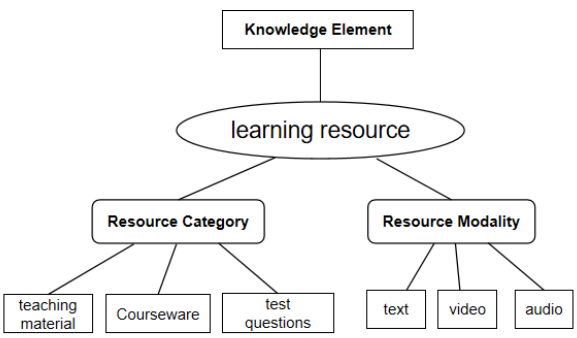

Figure 3. User Resource Ontology

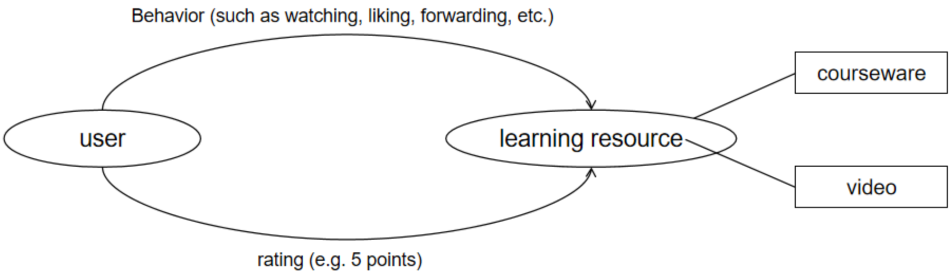

Figure 4. User Resource Knowledge Graph Fragment

Based on the analysis of the user ontology and user user resource knowledge graph examples mentioned above, the user behavior model is divided into browsing behavior interest model and clicking behavior interest model, which are analyzed separately and then comprehensively calculated.

Browsing Behavior Interest Model

Create a sequence set of user resources visited by the user, represented as $\hat{R}(u_i)=\{r_i \mid i=1,2,\cdots,n\}$, where $\hat{R}(u_i)$ represents the resource access series obtained by the i-th user u_i arranged in chronological order of resource access, where r1 is the earliest accessed resource, r_n is the most recently accessed resource. The log file records the user's behavior trajectory in the system, such as the number of times the user uses a certain resource, the time and duration of each use, and other information. This information can effectively reflect the user's preference for resources. The user's browsing behavior interest is measured by the duration and frequency of visits. It is generally believed that the importance of resources to users is directly proportional to the duration of access, with longer access times indicating user interest in the project. Here, we also need to consider the impact of resource information, such as the significant difference in information content between a 300 word text and a 3000 word text, and the different reading times. At the same time, the more times a user accesses a resource, the more likely it is that the user is interested in the resource, so a comprehensive consideration should be made. Therefore, it is considered to obtain the reason why user i stays on user resource j through network log data.The formula for the degree of interest D1 generated by the duration and number of visits is:

$$D_1(U_i, I_j) = \theta \times \frac{\sum_{visit-j} \frac{time(U_i, I_j)}{size(I_j)}}{\max\left(\sum_{visit-q} \frac{time(U_i, I_q)}{size(I_q)}\right)} \qquad \text{Formula(1)}$$

Among them, D_1 (U_i, I_j) represents the interest of user i in resource j due to browsing duration and number of visits, time (U_i, I_j) is the browsing duration, size (I_j) is the information content of resource j, usually represented by the storage capacity of the resource, visit-j is the number of times user i accesses resource j, visit-q is the number of times user i accesses resource q, θ is the modal parameter of the resource, and in this article, θ takes a value of 1.

Click Behavior Interest Model

During the process of using resources, users may also generate some click selection operations, such as bookmarking, sharing, downloading, etc. These behaviors can also reflect the user's preference for resources. In a 5-point scale, it is generally believed that a score of 3 or above indicates interest. Therefore, a scoring value is set for click behavior and converted to interest level I_2, as shown in Table 1.

Table 1. Scoring for Click Selection Behavior

Behavior	Scoring value	Interest level D_2
Collection	3.0	0.6

share	3.5	0.7
download	4.0	0.8
Collect+Download/Collect+Share/Share+Download	4.5	0.9
Collect+Share+Download	5.0	1.0

User Behavior Comprehensive Calculation Model

The weighted sum of interest obtained from operation logs and click selection is calculated, with the influence weight of log operation behavior being α, $0 \le \alpha \le 1$, and the influence weight of click selection behavior being β, $0 \le \beta \le 1$. The calculation formula for user i's predicted rating based on user behavior analysis of resource j is as follows:

$$D = 5 \times (\alpha \times D_1 + \beta \times D_2) \qquad \text{Formula(2)}$$

It should be pointed out that as more and more user resources are browsed and read, even if the user has not given a rating, the system can predict their interest in the content of the resource based on their behavior, which is completely interpretable as a substitute for zero rating items. For the convenience of calculation, the values of α and β in this article are both 0.5.

Similar Neighbor Set Calculation and Recommendation Strategy

Score matrix acquisition

User ratings are numerical and can be directly obtained from the system, forming an M×N的rating matrix, as shown in Table 2, where r_{ij} represents the rating of user i on user resource j. It should be pointed out that not every item has a rating r, and when user i does not rate resource j, r_{ij} is 0.

Table 2. User-User Resource Rating Matrix M×N

User \ Resource	$item_1$	…	$item_j$	…	$item_N$
$User_1$	r_{11}	…	r_{1j}	…	r_{1N}
…	…	…	…	…	…
$User_i$	r_{i1}	…	r_{ij}	…	r_{iN}
…	…	…	…	…	…
$User_M$	r_{m1}	…	r_{mj}	…	r_{mn}

Matrix sparsity completion

At present, a feasible solution for recommendation systems to solve the cold start problem is to use the basic information and initial project selection registered by users when they first enter the system to initialize their interests.For the sparsity problem of collaborative filtering recommendation, the sparsity of the rating matrix needs to be calculated first, and the sparsity of the matrix needs to be completed based on the analysis of user behavior and interest. The formula for calculating sparsity is:

$$\text{Sparsity}(M \times N) = 1 - \frac{\text{RNum}}{\text{UNum} \times \text{INum}} \qquad \text{Formula(3)}$$

The above formula shows the relationship between the number of ratings(RNum), the number of users(UNum), the number of user resources (INum) in the matrix and sparsity. A sparsity threshold λ is set in the recommendation system. When the sparsity is greater than this threshold, that is, sparsity>λ, it indicates that the evaluation matrix of users' resources is too sparse. Generally, the sparsity threshold λ=95% is reasonable. When the matrix sparsity is greater than 95%, the matrix is considered to be sparse. At this time, according to the characteristics of user behavior, the user's display and implicit scoring data are averaged to obtain the final score to complete the sparse scoring matrix. Therefore, the revised score is expressed as follows:

$$\hat{R}=\begin{cases}R\text{，Sparsity}<\lambda\\ \dfrac{D+R}{2}\text{，Sparsity}\geq\lambda\end{cases}\qquad\text{Formula(4)}$$

Similar neighbor set

The user based collaborative filtering method originates from the idea that users who like the same item the most are similar, and recommendations can be made between such similar groups, that is, recommending items that similar group users like to target users. The core of this method is to find a set of similar neighbors based on user similarity, and return the rating results of Top-N similar neighbors as recommendation prediction results to the user. The reason why cosine similarity is widely used is because its algorithm is simple and efficient.

(1)Cosine similarity refers to the cosine value of the vector angle formed by two things in a vector space, which is a measure of the difference between two things. The range of values is between 0 and 1. The closer the cosine value is to 1, the closer the angle is to 0, indicating that the two things are more similar. Use a vector composed of one row from the table above to represent user i is ratings for all items. If there are no ratings, use Sparsity in the equation above to replace it, which treats the user as an n-dimensional vector. The cosine similarity formula is:

$$sim(a,b)=\cos(a,b)=\frac{\sum_{j\in n}R_{a,j}\times R_{b,j}}{\sqrt{\sum_{j\in n}R_{a,j}^2}\times\sqrt{\sum_{j\in n}R_{b,j}^2}}\qquad\text{Formula(5)}$$

Among them, Sim(a,b) represents the similarity between user a and user b, $R_{a,j}$ and $R_{b,j}$ respectively represent user a and user b rating of Project j.

(2)Because the above cosine similarity does not take into account the difference of user rating scale, the accuracy will be greatly deviated. Therefore, the algorithm is modified by subtracting the average rating vector of all resources from the rating vector. When calculating, the average rating vector is represented by $\overline{R}_a$ and $\overline{R}_b$ respectively, and the modified cosine similarity formula is:

$$sim(a,b)=\frac{\sum_{j\in n}(R_{a,j}-\overline{R_a})\times(R_{b,j}-\overline{R_b})}{\sqrt{\sum_{j\in n}(R_{a,j}-\overline{R_a})^2}\times\sqrt{\sum_{j\in n}(R_{b,j}-\overline{R_b})^2}}\qquad\text{Formula(6)}$$

(3)When the sparsity of the user resource evaluation matrix M × n is greater than the set value λ, the modified comprehensive score $\hat{R}$ is added to the similarity algorithm through the above method to form an improved similarity calculation method. The formula is as follows:

$$sim(a,b)=\frac{\sum_{j\in n}(\hat{R}_{a,j}-\overline{R_a})\times(\hat{R}_{b,j}-\overline{R_b})}{\sqrt{\sum_{j\in n}(\hat{R}_{a,j}-\overline{R_a})^2}\times\sqrt{\sum_{j\in n}(\hat{R}_{b,j}-\overline{R_b})^2}}\qquad\text{Formula(7)}$$

Among them, $\hat{R}_{a,j}$ represents the revised rating of user a for resource j. After the above formula is calculated, the nearest neighbor set Z={U_1, U_2, U_3, …, U_{m-1}} of users with similar interests to the target user will be obtained.

Score prediction calculation

According to the nearest neighbor set calculated in the previous step, the score of the target user can be predicted according to the score of the neighbor set, and the top-N recommendation sequence can be generated from high to low. Therefore, the calculation formula for the prediction score R' of target user a on user resource j is as follows:

$$R'_{a,j} = \overline{R_a} + \frac{\sum_{b \in Z} \left\lfloor sim(a,b)(\hat{R}_{b,j} - \overline{R_b}) \right\rfloor}{\sum_{b \in Z} \left| sim(a,b) \right|} \qquad \text{Formula(8)}$$

Where, $\overline{R_a}$ and $\overline{R_b}$ respectively represent the average score vectors of target user a and neighbor user b, $\hat{R}_{b,j}$ is the modified score of user b on resource j, sim(a,b)represents the similarity between user a and user b, and Z is the nearest neighbor set calculated in the previous step. The score prediction results calculated by this formula are sorted from high to low, and top-N resources are recommended to users.

User resource collaborative filtering recommendation strategy

Based on the above analysis, the user user resource collaborative filtering recommendation process is shown in Figure 5.The specific steps of the LBA-UCF method are described as follows:

(1) Construct a user user resource rating matrix to form an M×N matrix as shown in the table.

(2) Use formula 3 to calculate Sparsity(M×N) of the scoring matrix in step①, and compare the sparsity with the threshold λ.

(3) If the matrix is not sparse, go directly to step 8.If the matrix is sparse, calculate the comprehensive score $\hat{R}$ of users after integrating user behavior correction according to formula 5, formula 6 and formula 7.

(4) Calculate the nearest neighbor set Z={U_1,U_2,U_3,…,U_{m-1}}of the target user according to formula 7.

(5) Calculate the prediction score of the items to be recommended according to formula 8 and generate Top-n user resources to recommend to the target user.

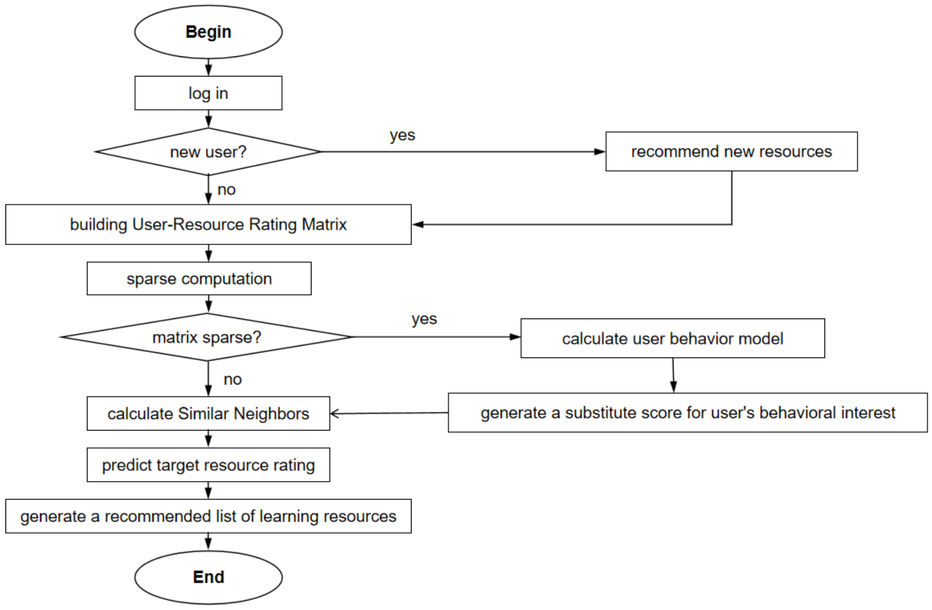

Figure 5 user resource collaborative filtering recommendation process

RESULTS

Dataset and Preprocessing

The experimental data of this study is the access data set from the online user platform of a university, which contains the complete student access records and user resource information required for this study. The data set contains basic information and user behavior data of 1345 users and 142 user resources. The information is intercepted from March 1, 2023 to June 30, 2023. The dataset is an information database. The data set used in this study contains the following four data tables.

(1) User information table: it mainly stores the user's personal information, including the basic information filled in at the time of registration and the initial interest options provided by the system and independently selected by the user. The more detailed the system can master the user's personal information, the better it can improve the accuracy of collaborative filtering recommendation.

(2) User resource table: it mainly stores the information of user resources. These user resources include courseware ppt, electronic textbooks, teaching videos, exercises and tests, which are presented in text, audio, video and other modes, and the eigenvalues are manually marked. In order to facilitate classification, use and management, each resource is marked with several labels, which are set by professional teaching staff. The linked items in the table correspond to the specific resources of the resource library.

(3) User behavior data table: it mainly stores user behavior data and is the structured presentation of behavior logs. All kinds of behavior data of users in the system are tracked and recorded by the system, such as the number of times and duration of browsing resources, as well as operations such as downloading, sharing, collecting and recommending. After the analysis of user behavior data, it can be used as an implicit score to better reflect the user's interest and preference for resources.

(4) Resource rating data table: mainly stores the rating information of users, which can be regarded as a rating matrix. This table is the main data base of collaborative filtering algorithm and is the explicit feedback of users' interests and preferences, but it is usually very sparse and needs to be handled appropriately when used.

(5) Resource rating data table: mainly stores the rating information of users, which can be regarded as a rating matrix. This table is the main data base of collaborative filtering algorithm and is the explicit feedback of users' interests and preferences, but it is usually very sparse and needs to be handled appropriately when used.

The detailed information of specific data sheets and attributes related to this experiment is described as follows.

1) User={UID，Gender，Age，profession，Region}, representing the user number, gender, age, occupation

and region respectively.

2) Item={IID，Title，Type，Size，Tag，Description}, representing resource number, name, category, capacity, label and description respectively.

3) Visit={UID，IID，StartTime，EndTime，Duration}, representing User ID, resource ID, start time, end time and duration respectively.

4) Collect={UID，IID，ClTime}, representing user ID, resource ID and collection time respectively.

5) Share={UID，IID，YesNo}, respectively indicating user ID, resource ID and whether to share.

6) Download={UID，IID，YesNo}, respectively indicating user ID, resource ID and whether to download.

7) Comment={CID，CUID，Title、Content、CmTime}, respectively indicating comment number, reviewer ID, title, content and release time.

8) Score={UID，IID，Rate、STime}, representing user ID, resource ID, score and scoring time respectively.

9) In the above description,3),4),5),6),7)corresponds to user behavior, and counts each user's browsing, collecting, sharing, downloading and commenting behavior, while the user's explicit score is obtained from 8). After data cleaning, each user has at least one online user behavior or rating data.

Experimental Results and Analysis

This study first uses user behavior data to obtain a comprehensive rating for prediction, which completes the sparse rating matrix. Then, the cosine similarity algorithm is used to obtain a similar user set, which calculates the resource prediction rating and ultimately obtains the recommendation result.To verify the effectiveness of the improved algorithm, this experiment compares the two results obtained from the original scoring matrix and the comprehensive scoring matrix. The experiment is divided into two parts.

Experiment 1: Statistics and calculations of the two indicators of the number of ratings and sparsity, comparing the original scoring matrix and the improved scoring matrix, according to different user time periods (usually in units of months), take the average of the results of five experiments as the final results of this experiment, the number of ratings and sparsity calculations are shown in Table 3.

Table 3. Comparison of evaluation scores and sparsity of the two methods

Indicator / Period	Number of Ratings		Sparsity	
	pre-improvement	post-improvement	pre-improvement(%)	post-improvement(%)
1	2652	13551	98.58	92.75
2	3147	17964	98.31	90.39
3	2749	22967	98.52	87.72
4	2525	23654	98.65	87.35
5	2963	21138	98.42	88.70
Average	2807	19855	98.50	89.38

As can be seen from Table 5, the number of ratings obtained by applying the improved composite scoring matrix calculation has increased significantly, for example, in Stage 1, the number of ratings before and after the improvement has increased by a factor of 4.1, and the average incremental number of ratings for the whole stage is more than 6 times. Meanwhile the sparsity problem of the scoring matrix is also significantly improved, e.g., in stage 3, the sparsity before and after the improvement is decreased by 11%, and the average sparsity of the whole stage is decreased by 9.3%. Overall, the method proposed in this paper effectively enhances the usability of the scoring matrix, which is mainly due to the fact that by analysing the logs of the user's behaviour, the user's implicit preferences for resources are obtained and these implicit ratings are added to the calculation of the scoring matrix, which increases the number of ratings and improves the sparsity.

Experiment 2: The recommendation “Precision” and recall “Recall” two indicators to carry out statistics and calculations, comparing the difference between the indicators of the traditional recommendation results before the improvement and after the improvement of the recommendation results, to test the recommendation performance.The calculation formula is shown below:

$$\text{Precision} = \frac{\text{Projects recommended by the system and liked by users}}{\text{Projects recommended by the system}} \qquad \text{Formula(9)}$$

$$\text{Recall} = \frac{\text{Projects recommended by the system and liked by users}}{\text{Projects user favorite}} \qquad \text{Formula(10)}$$

Through the analysis, it can be seen that precision and recall are contradictory to each other, and the increase in the number of recommended items can improve recall but reduce precision.For this reason, it is necessary to use the F1-measure evaluation metric, which is close to the smaller value of accuracy and recall, and can provide a global perspective for classification accuracy.

$$\text{F1-measure} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \qquad \text{Formula(11)}$$

In this experiment, different orders of magnitude in the recommendation list are used to do comparative analysis, and Top-10, Top-20, ..., Top-100 are selected, and the three indexes of accuracy, recall and F1-measure of the recommendation results in Top-N are calculated first, and then the results of the Improved Recommendation Method (IM) are compared with those of the Traditional Recommendation Method (TM).The specific calculation results are shown in Table 4.

Table 4. Results of the evaluation of the three indicators of the recommended methodology

Indicators / Number	Precision		Recall		F1-measure	
	TM	IM	TM	IM	TM	IM
Top-10	1.000	1.000	0.200	0.200	0.333	0.333
Top-20	0.952	0.985	0.340	0.400	0.501	0.569
Top-30	0.918	0.943	0.460	0.520	0.613	0.670
Top-40	0.866	0.876	0.550	0.610	0.673	0.719
Top-50	0.890	0.886	0.600	0.660	0.717	0.756
Top-60	0.856	0.871	0.680	0.760	0.758	0.812
Top-70	0.845	0.858	0.750	0.820	0.795	0.839
Top-80	0.832	0.845	0.820	0.870	0.826	0.857
Top-90	0.821	0.833	0.880	0.920	0.849	0.874
Top-100	0.811	0.820	0.940	0.982	0.871	0.894

In order to better observe the changes in the data, the data in Table 4 were plotted as a graph, as shown in Figure 6, to observe the trend of the curve.

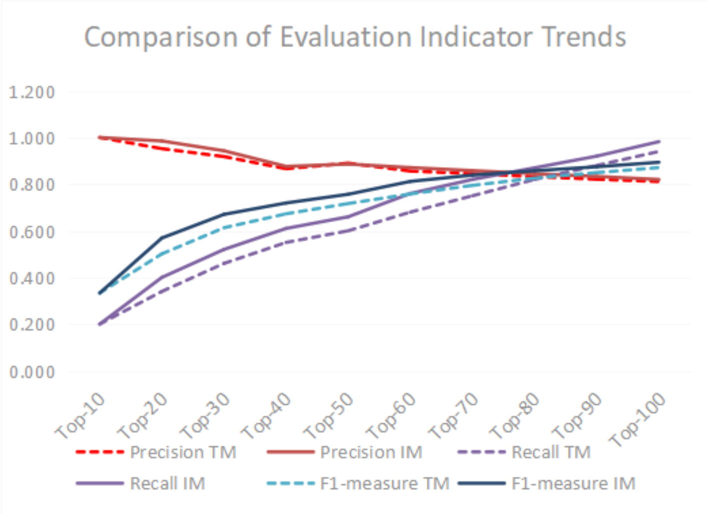

Figure 6. Comparison of evaluation data for the three indicators of the recommended methodology

In Figure 6, we can see that the traditional collaborative filtering recommendation method TM, in which only explicit ratings are involved, is compared with the improved recommendation method IM, which combines explicit ratings and implicit ratings proposed in this paper. With the increasing number of recommendations, for the same indicators, the improved recommendation method IM has higher accuracy, recall and F1-measure values compared to the traditional recommendation method TM; at the same time, for different indicators, the accuracy precision of the recommendation results has a slightly decreasing trend, with an average decrease ratio of 1.4%, while the recall and F1-measure metrics both increase smoothly, with an average improvement rate of 8.4% and 5.6%, respectively. The enhancement ratios of recall and F1-measure are significantly larger than the decline ratio of accuracy, indicating that it is worthwhile and reasonable to enhance recall and F1-measure at the expense of accuracy. This result is mainly due to the comprehensive user interest score obtained by combining the implicit predictive score generated by the user's user behaviour with the user's display score, which improves the traditional collaborative filtering recommendation algorithm on the basis of appropriately supplementing the user-resource matrix, enabling users to obtain relatively more compliant user-resource recommendations even if they are not used to scoring or maliciously scoring. Through the analysis of this experiment, it is effectively verified that the LBA-UCF method proposed in this chapter for the user-user-resource knowledge graph can improve the recommendation quality.

It is emphasised here that the user-interest model as well as the user-resource scoring model are based on the user ontology and the user-resource knowledge graph, the reason being that knowledge graph-like graph databases are more conducive to the discovery and computation of complex relationships between entities, and provide technical support for educational knowledge services.

DISCUSSION

Based on the conceptual architecture of the ontology, this study incorporates user-based collaborative filtering recommendation algorithms into an online user system to achieve personalised user recommendation. Aiming at the sparsity of the user-resource scoring matrix and the problem of severe user scoring in user collaborative filtering recommendation, starting from the implicit preferences contained in user behaviour, we apply explicit scoring in combination with implicit scoring generated by browsing and clicking behaviour to complement the scoring matrix; at the same time, aiming at the usability of similar user recommendation, we propose a knowledge graph-based online user resource collaborative The method can effectively improve the traditional collaborative user recommendation. This method can effectively improve the problems of cold start and sparsity as well as over-reliance on user ratings in traditional user collaborative filtering algorithms, and the experimental evaluation results show that the personalized recommendation method based on knowledge graph proposed in this study has a better recommendation quality in the online user system. The shortcoming lies in the problem of fusing the rating vectors generated under the knowledge graph-based architecture with the collaborative recommendation algorithm, which is a difficulty we will overcome in the next research.

ACKNOWLEDGMENTS

This article is supported by (Fund projects) :

(1) The Science and Technology Research Program of Hubei Provincial Department of Education (Project No. B2021189).

(2) The Youth Teacher Research Fund Project of Wuhan Sport University (Project No. 2021330).

(3) The Teaching Research Project of Wuhan Sport University (Project No. 202128).

REFRENCES

[1] Gong S．A collaborative filtering recommendation algorithm based on user clustering and item clustering [J].Journal of Software,2010,5(7) : 745-752．

[2] Salehi M， Kamalabadi N．Attribute-based recommender system for learning resource by learner preference tree[C]．Proceedings of the 2nd International E-conference on Computer and Knowledge Engi- neering，IEEE，2012: 133-138．

[3] Quan Z．Collaborative filtering recommendation based on user per- sonality[C]．International Conference on Information Management，Innovation Management and Industrial Engineering，IEEE， 2013: 307- 310．

[4] Yao Jin-bo，Yu Yi-cheng，Yu Zhuo-er，et al．Improvement on col- laborative filtering algorithm based on PCA default-values[J]．Journal of Jilin University(Information Science Edition)，2011，29 (5) : 494- 497．

[5] Zhao Wei，Lin Nan，Han Ying，et al．User-based collaborative filte- ring recommendation algorithm based on improved K-means clus- tering[J]．Journal of Anhui University(Natural Science Edition)，2016，40(2) : 32-36．

[6] Wang Xue-rong，Wan Nian-hong．Cloud pattern collaborative filtering recommender algorithm using user behavior correlation clustering[J]．Journal of Computer Applications，2011，31(9) ：2421-2425

[7] Chen Aixia, Li Ning Regression recommendation algorithm based on high-speed user machine and nearest neighbor [J]. Journal of Hebei University (Natural Science Edition), 2017, 37 (6).

[8] Xie Guomin, Zhang Tingting, Liu Ming, etc.Collaborative Theme Regression Recommendation Model Based on SDAE Feature Representation [J]. Computer Engineering and Science, 2019,41 (5).

[9] Xiang Xiaodong,Huang Piao,Qiu Zixian. Collaborative filtering algorithm based on comprehensive similarity weighted SlopeOne algorithm [J]. Statistics and Decision, 2020, (24)

[10] Hammou B, Lahcen A, Mouline S. An Effective Distributed Predictive Model With Matrix Factorization and Random Forest for Big Data Recommendation Systems [J].Expert Systems With Applications,2019,(137).

[11] Natarajan S,Vairavasundaram S,Natarajan S, et al. Resolving Data Sparsity and Cold Start Problem in Collaborative Filtering Recommender System Using Linked Open Data[J]. Expert Systems With Applications,2020,(149).

[12] Wang Jianren, Ma Xin, Duan Ganglong Improved K-means clustering k-value selection algorithm [J]. Computer Engineering and Applications, 2019, 55 (8)

[13] Ma Xin, Duan Ganglong, Wang Jianren, etc Research on Airline Customer Grouping Based on Improved Contour Coefficient Method [J]. Operations Research and Management, 2021, 30 (1)

[14] Zhang Wen, Cui Yangbo, Li Jian, etc Research on Collaborative Filtering Recommendation Based on Clustering Matrix Approximation [J]. Operations Research and Management, 2020, 29 (4)

[15] Tang Zekun, Huang Bingqing, Li Lian Collaborative filtering recommendation algorithm based on improved Canopy clustering [J]. Computer Application Research, 2020, 37 (9)