
Design of a Web Content Personalized Recommendation System Based on Collaborative Filtering Improved by Combining k-means and LightGBM

Xiaoming Li

*School of Information Engineering, Linyi Vocational College, Linyi,
276017, China
E-mail: lixiaoming202412@163.com*

Received 14 January 2025; Accepted 17 March 2025

Abstract

To improve the precision of Web content personalized recommendation, a Web content personalized recommendation system based on collaborative filtering improved by combining k-means and LightGBM is proposed. Firstly, the k-means clustering algorithm (k-means) is improved by using the Rat Swarm Optimizer (RSO) algorithm to cluster and group users and Web content. At the same time, Light Gradient Boosting Machine (LightGBM) algorithm is introduced to predict the level of interest of users in web content, and collaborative filtering recommendation method improved by combining k-means and LightGBM is proposed. Then, simulation experiments are conducted, thus verifying the recommendation method. Finally, B/S architecture is used to design and test the recommendation system. The results reveal that MAE and RMSE of the collaborative filtering recommendation method is improved by combining k-means and LightGBM for recommendation on the UserBehavior dataset are 1.08% and 2.41%, respectively, and its precision,

recall and F1 are 98.76%, 98.64% and 98.53%, respectively. Therefore, a Web content personalized recommendation system based on collaborative filtering improved by combining k-means and LightGBM has perfect functional modules, and it can meet Web content personalized recommendation, which has certain practical application value.

Keywords: Machine learning, collaborative filtering, content recommendation, personalized recommendation, k-means algorithm, LightGBM algorithm.

1 Introduction

In current Internet media, Web content is the main form of Internet media, because it contains graphics, videos and other content, and more and more websites launch personalized Web pages. However, due to the rich information and large amount of data in Web content, the difficulty for users to find Web content they are interested in is actually increased. To address this issue, personalized recommendation systems have emerged.

Chenzhong et al. proposed a knowledge learning method based on personalized recommendation. By using knowledge graph information to model and analyze users' interests and preferences, the fine-grained preferences of users could be better captured and the accuracy of personalized recommendation was improved [1]. Pelanek et al. proposed a website personalized recommendation based on a rule modular approach, which realized adaptive learning website personalized recommendation by analyzing large-scale data, which had positive significance for improving students' learning efficiency [2]. Fayçal et al. proposed a recommendation method based on deep neural collaborative filtering, and developed a recommendation system based on this recommendation method, so as to achieve personalized recommendations for e-commerce. The recommendation accuracy was 0.85, click through rate was 0.12, and recall rate was 0.78 [3]. Mishra et al. proposed a personalized recommendation method based on machine learning and collaborative filtering. By using machine learning to solve the cold-start problem of the collaborative filtering recommendation method, users' interest in Web content recommendation results was improved, and personalized Web content recommendation was realized [4]. Anitha and Kalaiarasu designed a set of collaborative filtering recommendation systems based on optimized machine learning. By adopting the k-means algorithm to optimize machine learning, improving the collaborative filtering recommendation method, and

applying it to e-commerce recommendation in websites, personalized product recommendation was realized [5]. Pengjie proposed an automated machine learning improved collaborative filtering recommendation method. By using automated machine learning to optimize collaborative filtering recommendation methods, personalized recommendation of web content was achieved [6]. Mecheri et al. combined the advantages of deep learning in the field of recommendation systems to propose a neural network-based Web content recommendation method, which guided the development of future Web service recommendation work and improved the accuracy of Web content recommendation by using neural networks to analyze Web content data [7]. In order to improve the quality of Web content recommendation, Gnanasekaran et al. proposed a Web content recommendation method based on hybrid deep learning by combining a convolutional neural network and an intubator memory neural network, and collected data from the WSDream data repository for training. Experiment showed that the method improved the accuracy of Web content recommendation, with an average absolute percentage error and an average absolute scale error of 7.5% and 2%, respectively, which was in sharp contrast to the conventional methods [8].

Comparing the machine learning based on collaborative filtering and the web content recommendation method based on deep learning, it can be seen that the web content recommendation method based on collaborative filtering has certain advantages in recommendation precision. The reason for this is that the machine learning-based Web content recommendation method can effectively solve the cold-start problem and improve the precision and personalization of the recommendation results. However, the web content recommendation method based on deep learning has a poor recommendation effect in an open and dynamic environment, and is easily affected by bias in the training data, resulting in the recommendation results not meeting the user's expectations, which is also the main reason for the low recommendation precision of the current web content recommendation method [9, 10]. Therefore, this paper adopts the machine learning method of collaborative filtering as the recommendation method of a Web content personalized recommendation system. Meanwhile, considering that collaborative filtering recommendation algorithms cannot identify similar user groups and content types when constructing web user and content rating matrix methods, an improved k-means algorithm based on RSO is adopted to cluster and group similar users and content. Then, a prediction method of the LightGBM algorithm is proposed in combination with the problem of the collaborative filtering recommendation algorithm in the user's poor prediction accuracy

of interested content. The LightGBM algorithm can effectively improve the algorithm prediction precision by optimizing the algorithm and feature selection method. For example, Messaoudi and Loukili optimized the collaborative filtering algorithm using the LightGBM algorithm to meet user preferences for content, which improved user interaction with recommended content [11]; Chunxue Zhang et al. used the LightGBM algorithm to predict user click rate, which improved the content recommendation precision of the collaborative filtering algorithm [12]. Therefore, using the LightGBM algorithm to improve the collaborative filtering algorithm is feasible to improve the accuracy of the collaborative filtering algorithm in predicting the user's interested content. Based on the above analysis, this paper proposes to combine the improved k-means algorithm and LightGBM algorithm for personalized recommendation of Web content.

The innovation points of this paper are: first, the k-means algorithm improved by the RSO algorithm is used to cluster similar users and content, and identify similar user groups and similar content types. At the same time, the dimensionality reduction and screening of large-scale data are carried out, thus the data dimension is reduced, and the recommendation precision and effect are improved. Then, LightGBM is adopted to predict the target users and their interested content, improving the precision of web content recommendation.

The organizational structure of this paper is as follows: The first section analyzes the background and significance of the birth of the recommendation system, and points out the existing problems of recommendation systems; the second section describes the principle of the collaborative filtering recommendation method; the third section improves the collaborative filtering recommendation algorithm according to the limitations of the collaborative filtering recommendation method; the fourth section designs a Web content personalized recommendation system and tests its functions based on the improved collaborative filtering recommendation method; the fifth section summarizes research results and draws conclusions.

2 The Collaborative Filtering Recommendation Method

Collaborative filtering is a commonly used recommendation method in recommendation systems. Compared with the content-based recommendation algorithm, collaborative filtering recommendation methods can effectively recognize users who are interested in the content according to their past behavior, thereby achieving accurate and fast content recommendations

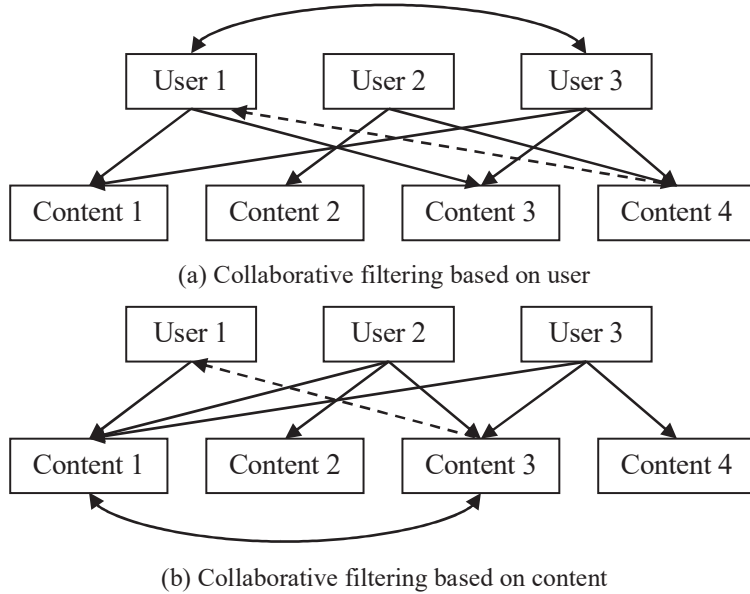


Figure 1 Schematic diagram of collaborative filtering recommendation methods.

[13–15]. Therefore, the collaborative filtering recommendation method is selected as the recommendation method.

Figure 1 shows the recommendation principle of the collaborative filtering recommendation method.

Among them, Figure 1(a) shows the principle of user-based collaborative filtering recommendation, which carries out recommendation by predicting user’s historical behavior. Figure 1(b) shows the principle of content-based collaborative filtering recommendation, which conducts recommendation on users based on the similarity of content. Because the user-based collaborative filtering recommendation method has a better recommendation effect, this recommendation method is selected for recommendation in this paper. The specific operation is as follows: First, construct a scoring matrix between users and content, compare user similarity, and construct a similar user set. Then, predict the level of interest of target users in the content. Finally, select content with higher predicted values for recommendation, so as to achieve personalized recommendation [16–18].

The construction method of a scoring matrix between users and content is shown in Equation (1), a calculation formula of similarity is shown in Equation (2), and prediction result of target users’ interest in content is shown

in Equation (3) [19–21].

$$X = \begin{bmatrix} U_1P_1 & \cdots & U_iP_1 \\ \cdots & \cdots & \cdots \\ U_1P_j & \cdots & U_iP_j \end{bmatrix} \quad (1)$$

$$sim(U_i, U_j) = \frac{V_{U_i} \cdot V_{U_j}}{|V_{U_i}| \cdot |N_{U_j}|} \quad (2)$$

$$ps(U_i, P_t) = \frac{\sum_{U_m} sim(U_i, U_m)G(U_m, P_t)}{\sum_{U_m} sim(U_i, U_m)} \quad (3)$$

where U_i represents the user; P_j indicates content; U_iP_j is the score of U_i on P_j ; $sim(U_i, U_j)$ stands for similarity between U_i and U_j ; V_{U_i} represents the scoring vector of U_i for the content; V_{U_j} is the scoring vector of U_j for the content; N_{U_j} is the content that U_j interacts with; U_m indicates the neighbor user; t is unrated content.

The recommendation process of the user-based collaborative filtering recommendation method is as follows:

First, input user behavior data, including browsing, favorites, and other content. Then, perform data cleaning and processing, and generate a user-content matrix to reflect the interaction between users and content; Next, the user-content matrix is decomposed to obtain user features and content features, and the similarity of the user feature matrix is calculated according to Equation (2). Finally, based on similarity calculation results, predict the level of interest of target users in the content and generate personalized content recommendation results [22].

The above process can be represented as a schematic diagram as shown in Figure 2.

3 Web Content Personalized Recommendation Based on Collaborative Filtering Improved by Combining k-means and LightGBM

3.1 Improvement of the Collaborative Filtering Recommendation Method

The recommendation method based on collaborative filtering has excellent recommendation performance. However, due to the limitations of constructing user and content scoring matrices and predicting the level of interest

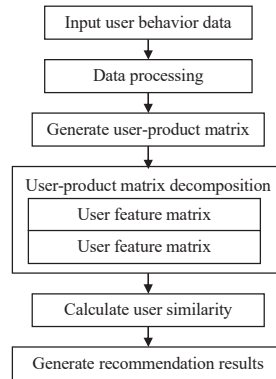


Figure 2 Recommendation process of the user-based collaborative filtering recommendation method.

of target users in content, the efficiency, accuracy, and personalization of recommendation results are not high. To solve the above problems, this paper improves the recommendation method of collaborative filtering.

3.1.1 Introduction of the k-means algorithm

The k-means algorithm is a clustering analysis method that is often used in the field of recommendation. It can provide users with more accurate personalized content recommendation results by dividing users and content into different groups and different types according to user behavior data. Therefore, in the user and content scoring matrix generated by the collaborative filtering recommendation method, this paper introduces the k-means algorithm to group users and content, so as to identify similar user groups and similar content types.

The k-means algorithm can group users and content, and recommend personalized web content to users that they are interested in. However, because the k-means algorithm is sensitive to initial clustering center, it is easy to fall into a local optimal solution, which further affects the personalization and accuracy of the recommendation results of the k-means algorithm. Therefore, to solve this problem, the RSO algorithm is adopted to carry out improvement.

The principle of the RSO algorithm to optimize the recommendation results of the k-means algorithm is to improve the accuracy and personalization of recommendation results by iteratively optimizing matching degree between users and content, so as to achieve more efficient

and accurate personalized recommendation of Web content. The specific operation process is as follows.

First, the number and location of the rat population of the RSO algorithm and cluster center search range of the k-means algorithm are initialized. Then, the social intelligence of the rat population is utilized, data are classified into the cluster with the shortest Euclidean distance, and an iterative calculation is started. Next, all possible solutions are evaluated based on their fitness values, and the solution corresponding to the best fitness value is selected to adjust the position of the rat population, thus searching for prey and attacking. Finally, when maximum iterative times are reached, output results are used as an initial clustering center of the k-means algorithm.

The optimization process of the k-means algorithm using the RSO algorithm is shown in Figure 3.

3.1.2 Introduction of the LightGBM algorithm

The LightGBM algorithm is a machine learning algorithm based on decision tree, which has the advantages of good training effect and difficult fitting. Moreover, it is often used in click rate prediction and target interest degree prediction. Therefore, this paper introduces the LightGBM algorithm

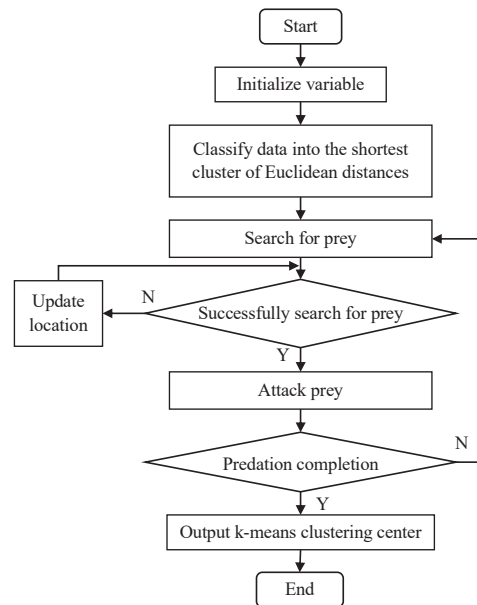


Figure 3 Process of optimizing the k-means algorithm using the RSO algorithm.

to improve the prediction process of the user-based collaborative filtering recommendation method, thereby improving its recommendation accuracy.

The principle and process of the LightGBM algorithm are as follows:

First, initialize $F_0(x)$ learner based on a priori information, as in Equation (4):

$$F_0(x) = \lg \frac{P(y = 1|x)}{1 - P(y = 1|x)} \quad (4)$$

where $P(y = 1|x)$ represents the proportion of sample $y = 1$ in training samples.

Then, M regression trees are established, and the residual value of the m th tree is calculated according to Equation (5):

$$r_{m,i} = y_i - \frac{1}{1 + e^{-F(x_i)}} \quad (5)$$

where $F(x_i)$ and y_i are fitting data and $r_{m,i}$ stands for leaf node region.

Finally, strong learner $F_M(x)$ is calculated according to Equation (6):

$$F_M(x) = F_0(x) + \sum_{m=1}^M \sum_{j=1}^{J_m} c_{m,j} I, \quad x \in r_{m,j} \quad (6)$$

where j represents the j th tree; $c_{m,j}I$ is the residual value corresponding to the j th tree; I is a random number.

3.2 Collaborative Filtering Recommendation Process Based on k-means and LightGBM Improvement

Web content recommendation includes shopping platform recommendation, personalized website recommendation, news hot spot recommendation and other content. For convenience of experiment, this paper takes shopping platform recommendation as an example. This study adopts the collaborative filtering recommendation method improved by combining the k-means algorithm and the LightGBM algorithm to conduct personalized recommendation on shopping platforms in Web content. The specific recommendation process is as follows:

- (1) Collect and sort out behavioral data on websites such as users recommended by shopping platforms and their historical shopping platform click-through rates, and page views, then normalize and standardize these data.

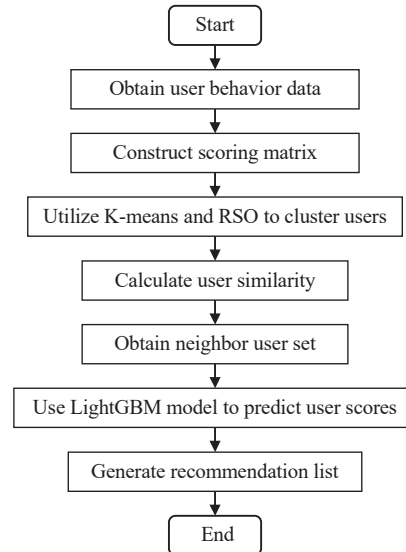


Figure 4 Recommendation process of collaborative filtering improved by k-means and LightGBM.

- (2) Utilize the k-means algorithm improved by the RSO algorithm, cluster users and their behavioral data separately, then construct the user and content scoring matrices.
- (3) According to the similarity calculation formula of the k-means algorithm, calculate the similarity of users, then obtain the set of neighboring users.
- (4) Adopt the LightGBM algorithm to predict users' level of interest in recommended shopping platforms and generate personalized recommendation lists for shopping platforms.
- (5) Select the top five shopping platforms recommended as the final recommendation results and recommend them to users, that is, realize the personalized recommendation of web content.

The above process can be represented by a diagram, as in Figure 4.

4 Simulation Experiment

4.1 Experimental Environment Construction

This paper constructs a recommendation method of collaborative filtering improved by combining k-means and LightGBM based on python language

Table 1 Examples of the experimental data

UserID	CategoryID	ItemID	Behavior	Datetime	Timestamps	Dates	Hours
1	15632	5216	Pv	13:11	1532644512	2021/11/25	2
2	15647	5789	Pv	12:00	1563245973	2021/11/28	5
3	16539	5326	Pv	18:36	1536549852	2021/11/30	6
...

and the pytorch framework, and runs it on an Ubuntu20.04 operating system. For a system hardware environment, this paper selects NVIDIA GeForce RTX 3080 GPU and Intel Core i7-11800H CPU as hardware devices, with a memory size of 32 GB.

4.2 Data Source and Preprocessing

In this paper, the IRIS dataset is used to verify the performance of the RSO algorithm against the k-means algorithm. The IRIS dataset contains 150 data samples, which are divided into Iris Setosa, Iris Versicolour and Iris Virginica, with 50 samples for each category. Moreover, it contains four attributes, namely Sepal.Length, Sepal.Width, Petal.Length, and Petal.Width.

This study uses the Taobao User Behavior dataset provided by Alibaba's Tianchi data platform as the experimental dataset to verify the recommendation effectiveness of the recommendation method of collaborative filtering improved by combining k-means and LightGBM. This dataset contains a total of 1 billion pieces of data from 25 November 2021 to 3 December 2021, including user IDs, content IDs, content category IDs, behavior types, and timestamps on the Taobao platform.

Due to the large amount of UserBehavior data and garbage data, missing values, duplicate values and outliers in data are deleted before performing the experiment. At the same time, to filter out valid data, the FILTER function is called to screen data, thereby obtaining 20,000 pieces of data that can be used for this experiment. Table 1 illustrates examples of the experimental data.

4.3 Evaluation Indicators

In this paper, the mean absolute error (MAE), root mean square error (RMSE), precision, recall and F1 score were selected as evaluation indicators. Their calculation formulas are:

$$MAE = \frac{\sum_{i=1}^n |y_i - y'_i|}{n} \quad (7)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y'_i - y_i)^2} \quad (8)$$

$$precision = \frac{TP}{FP + TP} \quad (9)$$

$$recall = \frac{TP}{TP + FN} \quad (10)$$

$$F1 = \frac{precision \times recall}{precision + recall} \times 2 \quad (11)$$

where n represents sample size; y_i indicates measured value; y'_i stands for predicted value; TP and FP represent the number of positive samples that are correctly determined to be positive classes and the number of negative samples that are incorrectly determined to be positive samples, respectively; FN is the number of positive samples that are incorrectly determined to be negative samples.

4.4 Parameter Settings

In this paper, the population size and iteration times of the RSO algorithm were set to 50 and 300, respectively, and the number of clusters was set to 3.

4.5 Results and Analysis

1. Improvement verification of the k-means algorithm. Based on the IRIS dataset, this study validated the effectiveness of improvements carried out on the k-means algorithm. Table 2 shows the verification results of the improvement effect.

As seen in Table 2, the clustering standard deviation of the k-means algorithm is 1.58, the silhouette coefficient is 0.38, and its Calinski–Harabasz index is 320.12. The clustering standard deviation of the k-means algorithm improved by the RSO algorithm is 0.01, the silhouette coefficient is 0.71, and its Calinski–Harabasz index is 562.23. Therefore, the k-means algorithm

Table 2 Improvement verification results of the k-means algorithm

Algorithm	Standard Deviation	Silhouette Coefficient	Calinski–Harabasz Index
k-means	1.58	0.38	320.12
RSO improved k-means	0.01	0.71	562.23

Table 3 Comparison of recommendation effect evaluation indicators of the k-means algorithm before and after improvement

Recommendation Method	Mean Absolute Error (MAE)/%	Root Mean Square Error (RMSE)/%	Precision/%	Recall/%	F1 Score/%
k-means improved based on RSO+collaborative filtering	5.68	6.89	90.12	90.28	90.10
Based on k-means+collaborative filtering	7.12	8.43	87.15	86.23	86.21

improved by the RSO algorithm has smaller clustering standard deviation, larger silhouette coefficient and Calinski-Harabasz index on the IRIS dataset, indicating it has stronger global optimization ability and better optimization results.

To further verify the effectiveness of improvement conducted on the k-means algorithm, k-means+collaborative filtering and RSO+k-means+collaborative filtering were used for recommendation on the UserBehavior experimental dataset. Table 3 depicts the recommendation results.

According to Table 3, based on RSO+k-means+collaborative filtering, the average absolute error is 5.68%, which is 1.44% lower than that of k-means+collaborative filtering. Moreover, the precision, recall, and F1 score have been improved to varying degrees. Therefore, RSO+k-means+collaborative filtering has a more accurate recommendation effect than k-means+collaborative filtering, which proves that using RSO algorithm to improve k-means algorithm is effective. The reason for this is that RSO algorithm improves the accuracy of initial clustering center of k-means algorithm, which makes the clustering effect of the k-means algorithm on users and content better, thus improving the precision of the collaborative filtering recommendation method.

2. Validation of the recommendation method of collaborative filtering improved by combining k-means and LightGBM. To verify the effectiveness of the recommendation method of collaborative filtering improved by combining k-means and LightGBM, RSO+k-means+collaborative filtering, LightGBM+collaborative filtering, and RSO+k-means+LightGBM+collaborative filtering were used for recommendation on the UserBehavior

Table 4 Evaluation indicators of different recommendation methods

Recommendation Method	Mean	Root	Precision/ %	Recall/ %	F1 Score/ %
	Absolute Error (MAE)/%	Mean Square Error (RMSE)/%			
Based on RSO+ k-means+ collaborative filtering	5.68	6.89	90.12	90.28	90.10
Based on LightGBM+ collaborative filtering	6.23	7.14	90.35	90.32	90.12
RSO+k-means+ LightGBM+ collaborative filtering	1.08	2.41	98.76	98.64	98.53

dataset. Table 4 shows evaluation indicators of the recommendation results of different recommendation methods.

As seen in Table 4, on the experimental dataset, the RSO+k-means+LightGBM+collaborative filtering recommendation method has the lowest MAE and RMSE, which are 1.08% and 2.41%, respectively, and it has the highest precision, recall and F1 score, which are 98.76%, 98.64% and 98.53%, respectively. The RSO+k-means+LightGBM+collaborative filtering recommendation method has a more accurate recommendation effect. This shows that the RSO+k-means+LightGBM+collaborative filtering recommendation method has excellent results in Web content personalized recommendation.

To further verify the superiority of the collaborative filtering recommendation method improved by combining k-means and LightGBM, the proposed method, the recommendation method based on Tb-Bgat, Tinybert and Bigru [23], and the recommendation method based on partial order set [24] were compared on the UserBehavior data set. Table 5 illustrates the comparison results.

Analyzing Table 5 reveals that compared with the existing recommendation methods, the proposed method has obvious recommendation advantages over the UserBehavior experimental dataset, its MAE and RMSE are lower than those of recommendation method based on Tb-Bgat, Tinybert and Bigru and the recommendation method based on partial order set, while its

Table 5 Evaluation indicators for recommendation results of different recommendation methods

Recommendation Method	Mean	Root	Precision/ %	Recall/ %	F1 Score/ %
	Absolute Error (MAE)/%	Mean Square Error (RMSE)/%			
Recommendation method [23] based on Tb-Bgat with Tinybert and Bigru	6.12	5.89	87.65	88.34	89.26
Recommendation method [24] based on partial order set	6.45	6.43	89.23	89.20	89.12
RSO+k-means+LightGBM+collaborative filtering	1.08	2.41	98.76	98.64	98.53

precision, recall and F1 score are higher than those of the recommendation method based on Tb-Bgat, Tinybert and Bigru and the recommendation method partial order set; it also exhibits better recommended performance. Thus, the RSO+k-means+LightGBM+collaborative filtering recommendation method has advantages over Web content personalized recommendation. The reason for this is that this method combines the advantages of the k-means algorithm and the LightGBM algorithm, which can effectively mine users' preferences and potential interests in content attributes, and better meet users' individual needs, so it has a better recommendation effect.

4.6 Design of a Web Content Personalized Recommendation System of Collaborative Filtering Improved by Combining k-means and LightGBM

4.6.1 Overall framework

The purpose of a Web content personalized recommendation system is to provide users with personalized web content that they are interested in according to their historical and current behavior, so as to meet their preferences [25, 26]. Therefore, the system should have functions such as data analysis and content recommendation. Based on the above analysis, a B/S architecture is adopted to design the overall framework of this system, as shown in Figure 5.

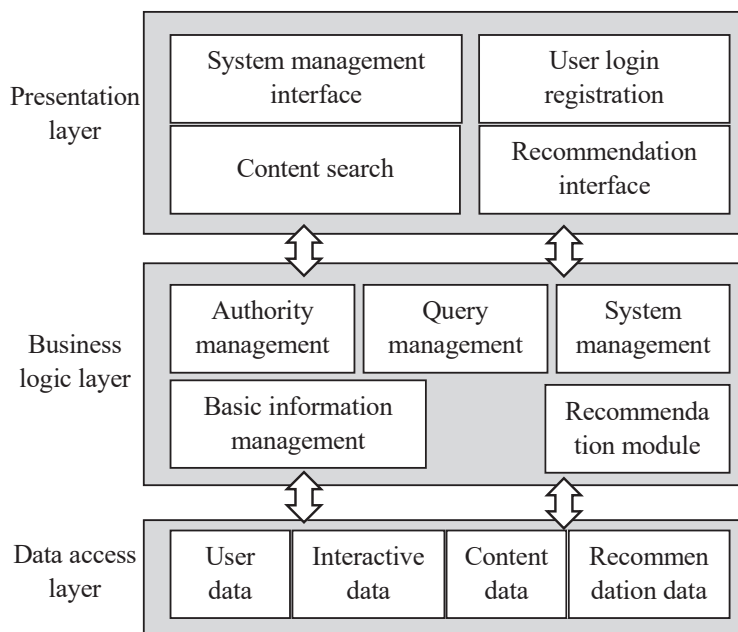


Figure 5 Overall framework of the system.

The system consists of a data access layer, business logic layer, and presentation layer [27, 28]. Among them, the data access layer is used to collect user data, user current behavior and historical behavior content data, analyze, process and store data, and generate recommendation data and interactive data. The business logic layer is used for front-end and back-end data interaction, including data query, information management, system management, authority management, and recommendation modules. Furthermore, the presentation layer is responsible for displaying recommendation results and achieving information interaction with users.

4.6.2 Design of system function modules

According to the functional requirements of a Web content personalized recommendation system, it is necessary to meet the functions of user search, content recommendation, information management, and system management. Therefore, this study designs the function modules of this system as shown in Figure 6.

The system function modules contain three modules: user management, content recommendation and background management [29, 30]. Among

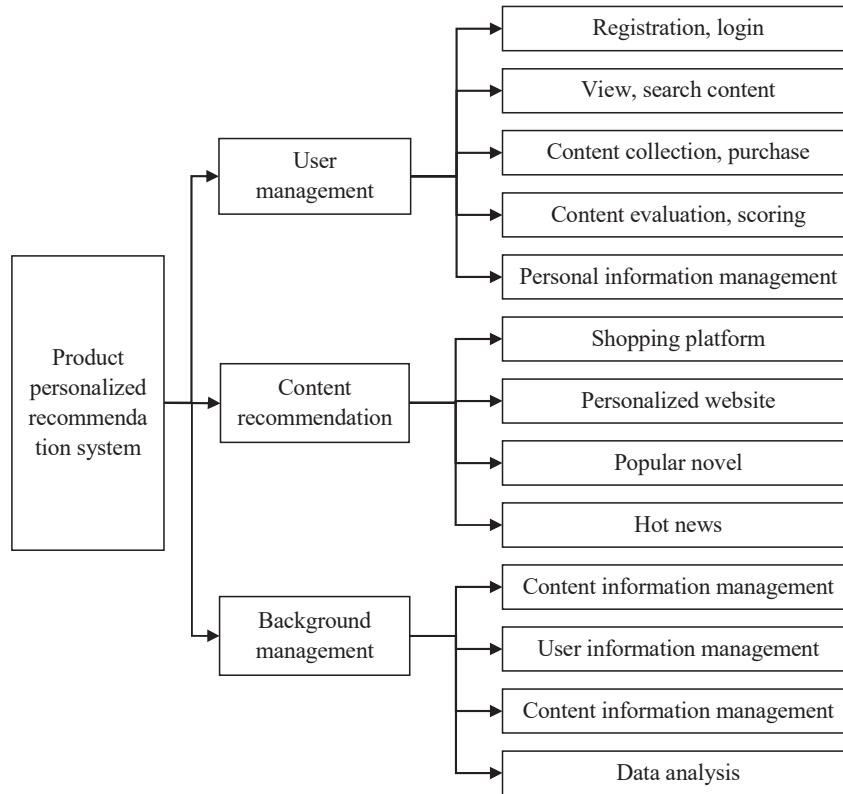


Figure 6 Structure diagram of the system function modules.

them, the user management module includes five sub-modules, namely registration and login, view and search, content collection, content evaluation, and personal information management. The content recommendation module includes four sub-modules, namely shopping platform, personalized website, hot news, and popular novel; the background management module includes four sub-modules, namely content information management, user information management, content and data analysis [31, 32].

4.6.3 System test

1. Construction of a test platform. In this study, the Java language and SpringBoot framework were used to develop the backend of the system, and MySQL was used as the database. In addition, Vue was used to realize the functional interaction of the front-end pages, and processbuilder class was

Table 6 Test cases and test results

Function Module	Test Case	Test Result
User management module	<ol style="list-style-type: none"> 1. Users enter an incorrect account or password 2. Users enter the correct account and password 3. Users enter home page of the system and clicks to view the content they are interested in 	<ol style="list-style-type: none"> 1. Users fail to log in to the system and the login failure is displayed 2. When users enter the system, the system automatically switches to the home page 3. Users can click to view any content they are interested in in the system, and automatically jump to the details page of content they are interested in
Content recommendation module	<ol style="list-style-type: none"> 1. Users click to enter content recommendation module, check whether the system can recommend the content they are interested in 	<ol style="list-style-type: none"> 1. Users can view the content of content recommendation module
Background management module	<ol style="list-style-type: none"> 1. Background administrator enters the system and deletes, adds, and modifies system users and content 2. Background administrators analyze user behavior and other data to mine user content preferences 	<ol style="list-style-type: none"> 1. Background administrators can delete, add, and modify users and content 2. Background administrators can perform operations such as data analysis

used to call the recommendation method of collaborative filtering improved by machine learning in Java environment, so as to construct a test platform for a Web content personalized recommendation system based on machine learning improved collaborative filtering [33, 34].

2. Testing process. After the users enter the system by entering the correct account and password, they will check whether the system can display the various modules properly, including the user management module, the content recommendation module and the backend management module. Then, users enter the different modules and perform the corresponding operations. When entering the user management module, users add, delete and modify the user information, and then users check if the operation can be performed correctly. After entering the content recommendation module, users check if the system can recommend content of interest. When entering the backend

management module, users check if user behavior data analysis and other operations can be performed, so as to verify the effectiveness of the system.

3. Test cases and test results. Table 6 shows specific test cases and test results.

Table 6 depicts that each function module of the system can perform corresponding operations and view the personalized recommendation results of Web content of interest. It shows that this system has a certain degree of accuracy and personalization in Web content personalized recommendation, which meets the needs of practical applications.

5 Conclusion

In conclusion, adopting the RSO algorithm to improve the k-means algorithm improves the accuracy and efficiency of the k-means algorithm for user and content clustering. Combining the improved k-means and LightGBM to improve the collaborative filtering recommendation method can improve the recommendation accuracy of the collaborative filtering recommendation method. Furthermore, for the Web content personalized recommendation system constructed, each function module runs normally, which has certain accuracy and personalization, and meets the application needs of Web content personalized recommendation. According to the experimental results, the clustering standard deviation, silhouette coefficient and Calinski–Harabasz index of the RSO algorithm improved k-means algorithm on IRIS dataset are 0.01, 0.71 and 562.23, respectively, showing excellent global optimization ability and optimization results. On the UserBehavior dataset, MAE and RMSE of the collaborative filtering recommendation method improved by combining k-means and LightGBM are 1.08% and 2.41%, respectively, and its accuracy, recall, and F1 score are 98.76%, 98.64%, and 98.53%, respectively, demonstrating excellent recommendation performance. The reason for this is that this method combines the advantages of the k-means algorithm and the LightGBM algorithm, which can effectively mine users' preferences and potential interests in content attributes, and better meet the personalized needs of users. Overall, system function modules are perfect, which can achieve the personalized recommendation of Web content. However, due to the constraints of conditions, there are still some shortcomings in this paper that need to be improved. For example, in terms of personalized recommendation research, deep learning-based personalized recommendation has also achieved excellent recommendation results, such

as deep residual networks and attention mechanism networks. These deep learning methods improve the accuracy of recommendation results by analyzing complex data. Therefore, the next step of this research will consider combining the advantages of deep learning to further optimize system ability of analyzing and processing data, so as to make system recommendation results more accurate.

References

- [1] Chenzhong B, Wenqiang L, Hantao D, et al., 'A multi-view heterogeneous knowledge learning method for personalized POI recommendation', *Journal of Intelligent & Fuzzy Systems*, 2024, 46(4): 7763–7777.
- [2] Pelánek R, Effenberger T, Jarušek P, 'Personalized recommendations for learning activities in online environments: a modular rule-based approach', *User Modeling and User-Adapted Interaction*, 2024, 34(4): 1399–1430.
- [3] Fayçal M, Manal L, 'E-commerce personalized recommendations: a deep neural collaborative filtering approach', *Operations Research Forum*, 2024, 5(1):56–62.
- [4] Mishra N K, Mishra A, Barwal N P, et al., 'Natural language processing and machine learning-based solution of cold start problem using collaborative filtering approach', *Electronics*, 2024, 13(21): 4331–4331.
- [5] Anitha J, Kalaiarasu M, 'Retraction note to: Optimized machine learning based collaborative filtering (OMLCF) recommendation system in e-commerce', *Journal of Ambient Intelligence and Humanized Computing*, 2022, 14(Suppl 1): 331–331.
- [6] Pengjie L, Fucheng P, Xiaofeng Z, et al., 'CF-DAML: Distributed automated machine learning based on collaborative filtering', *Applied Intelligence*, 2022, 52(15): 17145–17169.
- [7] Mecheri K, Klai S, Souici-Meslati L, 'Deep learning based web service recommendation methods: A survey', *Journal of Intelligent & Fuzzy Systems: Applications in Engineering and Technology*, 2023, 44(6): 9879–9899.
- [8] Gnanasekaran A, Chinnasamy A A, Parasuraman E, 'Analyzing the QoS prediction for web service recommendation using time series forecasting with deep learning techniques', *Concurrency and computation: practice and experience*, 2022, 34(28): e7356.1–e7356.17.

- [9] Zhang Y, 'Study on improved personalised music recommendation method based on label information and recurrent neural network', *International Journal of Information and Communication Technology*, 2024, 24(1): 48–59.
- [10] Chen S, Ke X, Zhang X, 'A study of personalised recommendation methods for multimedia ELT online course', *International Journal of Computational Systems Engineering*, 2024, 8(1–2): 96–106.
- [11] Messaoudi Fayçal, Loukili Manal, 'E-commerce Personalized Recommendations: a Deep Neural Collaborative Filtering Approach', *SN Operations Research Forum*, 2024, 5(1): 5.1–5.25.
- [12] Chunxue Zhang, Liqing Qiu, Caixia Jing, Chengai Sun, 'TIAE-DSIN: A time interval aware deep session interest network for click-through rate prediction', *Expert Systems with Application*, 2024, 249(sepaaptaa): 123531.1–123531.12.
- [13] Shi J, Deng Y, 'A personalised push method of English mobile reading resources based on tag similarity', *International Journal of Business Intelligence and Data Mining*, 2024, 24(3–4): 266–277.
- [14] Anitha J, Kalaiarasu M, 'Optimized machine learning based collaborative filtering (OMLCF) recommendation system in e-commerce', *Journal of Ambient Intelligence and Humanized Computing*, 2020, 12(6): 1–12.
- [15] Shuaikang H, Lifang P, Xinyin T, et al., 'Do platform recommendations in the fund market work? Evidence from a quasi-experimental study', *Industrial Management & Data Systems*, 2024, 124(12): 3274–3297.
- [16] Das S G, Nayak B, 'Financial inclusion: An application of machine learning in collaborative filtering recommender systems', *International Journal of Recent Technology and Engineering (IJRTE)*, 2020, 8(6): 4243–4247.
- [17] Li R, Shu Y, Cao Y, et al., 'Federated cross-view e-commerce recommendation based on feature rescaling', *Scientific Reports*, 2024, 14(1): 29926–29929.
- [18] Wang Y, 'Research on the recommendation strategy of dual-channel manufacturers for hybrid e-commerce platforms', *Frontiers in Physics*, 2024, 1(2): 1455165–1455169.
- [19] Bellar O, Baina A, Ballafkih M, 'Sentiment analysis: Predicting product reviews for E-commerce recommendations using deep learning and transformers', *Mathematics*, 2024, 12(15): 2403–2406.

- [20] Jin Z, Ye F, Nedjah N, et al., 'A comparative study of various recommendation algorithms based on E-commerce big data', *Electronic Commerce Research and Applications*, 2024, 6(8): 101461–101463.
- [21] Yanbin N, 'Consumer psychology in the digital age: How online environments shape purchasing habits', *Proceedings of Business and Economic Studies*, 2024, 7(5): 20–29.
- [22] Wang P, Cao H, Li P, et al., 'Quantum representation based preference evolution network for E-commerce recommendation', *Physica A: Statistical Mechanics and its Applications*, 2024, 65(4): 130155–130158.
- [23] Jing C, Weiyu Y, 'TB-BGAT with TinyBERT and BiGRU in personalized course recommendations', *International Journal of Information and Communication Technology Education (IJICTE)*, 2024, 20(1): 1–15.
- [24] Lizhu Y, Qian W, 'Personalized multi-attribute recommendation method based on partial order set', *Journal of Intelligent & Fuzzy Systems*, 2024, 46(4): 8741–8754.
- [25] Arias V A, Bedoya U H, Ruiz G D J, et al., 'Artificial intelligence and recommender systems in e-commerce. Trends and research agenda', *Intelligent Systems with Applications*, 2024, 2(4):200435–200438.
- [26] Zhu X, Xia X, Wu Y, et al., 'Enhancing explainable recommendations: Integrating reason generation and rating prediction through multi-task learning', *Applied Sciences*, 2024, 14(18): 8303–8305.
- [27] Wang X, Zhang C, Xu Z, 'A product recommendation model based on online reviews: Improving PageRank algorithm considering attribute weights', *Journal of Retailing and Consumer Services*, 2024, 8(1): 104052–104052.
- [28] Park M, Oh J, 'Enhancing E-commerce recommendation systems with multiple item purchase data: A bidirectional encoder representations from transformers-based approach', *Applied Sciences*, 2024, 14(16): 7255–7255.
- [29] Meng W, Chen L, Dong Z, 'The development and application of a novel E-commerce recommendation system used in electric power B2B sector', *Frontiers in Big Data*, 2024, 7(1): 1374980–1374980.
- [30] Juan D, Wenkuan C, 'Enhancing community E-commerce repurchase prediction through information entropy analysis', *International Journal of e-Collaboration (IJeC)*, 2024, 20(1): 1–18.
- [31] Lv S, Wang J, Deng F, et al., 'A hybrid recommendation algorithm based on user nearest neighbor model', *Scientific Reports*, 2024, 14(1): 17119–17119.

- [32] Liu Y, Chang X, Yang S, et al., “‘Pets make you spend more!’ Impact of pet ownership on consumer purchase decisions”, *Journal of Business Research*, 2024, 18(3): 114838–114839.
- [33] Li L, Helin L, Xun L, et al., ‘Graph agent transformer network with contrast learning for cross-domain recommendation of E-commerce’, *Journal of Cases on Information Technology (JCIT)*, 2024, 26(1): 1–16.
- [34] Wang H, Jiang G, Hong M, et al., ‘Graph-based bootstrapped latent recommendation model’, *Electronic Commerce Research and Applications*, 2024, 6(8): 101446–101448.

Biography



Xiaoming Li was born in Shandong, China in 1982. From 2000 to 2004, she studied at Shandong Economic University and received her bachelor’s degree in 2004. From 2007 to 2009, she studied at Dongbei University of Finance and Economics and received her Master’s degree in 2009. She has been working at Linyi Vocational College since 2008 and is currently an associate professor. She has worked in the field of computer application teaching and campus informatization construction for nearly 20 years, and her research interests are included computer application, artificial intelligence and big data.

