

Research on Personalized Resource Recommendation Based on User Profile and Collaborative Filtering Algorithm

Dongxu Liu

College of Information Science, Zhejiang Open University, Hangzhou, China
Email: 1184010040@qq.com

How to cite this paper: Liu, D. X. (2023). Research on Personalized Resource Recommendation Based on User Profile and Collaborative Filtering Algorithm. *Advances in Applied Sociology*, 13, 841-849. <https://doi.org/10.4236/aasoci.2023.1311048>

Received: October 20, 2023

Accepted: November 19, 2023

Published: November 22, 2023

Copyright © 2023 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

With the flourishing development of online education, the problem of information overload in learning resources is becoming increasingly prominent. This study proposes a learning resource recommendation method that combines user profiling and collaborative filtering algorithms. It involves acquiring both static and dynamic user data from an online learning platform, constructing a user profile label library, conducting user group clustering using the K-means algorithm, calculating user similarity within each group and identifying the most similar users to the target user, ultimately generating the resource recommendation list based on the learning preferences of these similar users. Personalized recommendations for learning resource are of significant importance for improving learning effectiveness on online learning platforms, enhancing user satisfaction, and promoting the development of personalized education.

Keywords

Online Learning, Personalized Recommendation, Collaborative Filtering, User Profile, K-Means Algorithm

1. Introduction

In recent years, the rapid development of the internet has led to an increasing number of people acquiring knowledge through online learning. Online learning methods have eliminated the constraints of time and location, allowing learners to schedule their own study time and location. Additionally, online learning platforms offer a wealth of learning resources and flexible teaching modes. Learners can not only independently select learning materials but also enjoy a

vivid and rich learning experience through multimedia and interactive tools. Therefore, the widespread development of online learning in the field of education has provided students with more opportunities and choices, driving innovation and progress in education (Liu et al., 2020).

With the flourishing development of online education, the issue of information overload in learning resource has become increasingly prominent. Learners often find it challenging to navigate through the massive amount of learning resources. Therefore, recommendation technology has been attempted in online education systems to achieve personalized recommendations for learning resources (Xu & Guo, 2020). Personalized recommendations are of significant importance in enhancing the effectiveness of online learning platforms, increasing user satisfaction, and promoting the development of personalized teaching.

Research on personalized resource recommendations has been conducted both domestically and internationally for a considerable time. Khribi et al. (2008) initiated course recommendations based on the historical learning information of the target user and the content of the learning courses. Katuk & Ryu (2011) proposed learning path recommendations based on the emotions of learners. Salehi et al. (2013) modeled the ratings of learning resources and applied collaborative filtering algorithm to recommend learning resources to users. Experiments demonstrated an improved performance of the recommendation system. Wang (2017) created models for learners' learning interests and achieved personalized recommendations on online education platforms by enhancing collaborative filtering algorithm. Xiong (2019) proposed an enhancement scheme for collaborative filtering algorithm by merging label weights. Tan & Zhang (2021) integrated feature information from user models and resource models into collaborative filtering algorithm to accomplish personalized resource recommendations. Currently, the collaborative filtering algorithm has become the mainstream recommendation technology, but it is still affected by issues like cold start and sparse data matrices, leading to suboptimal recommendation outcomes.

Compared to resources in other domains, online learning resources exhibit strong specialization, significant clustering effects, and high resource complexity. Additionally, individual learners possess complex personal characteristics, such as learning styles, preferences, and cognitive levels, all of which significantly impact the quality of recommendations for online learning resources. Therefore, to provide more accurate recommendation services, introducing user profiles is a highly effective solution (Zhai, 2018). User profiles, as data analysis tools that depict user characteristics and understand user needs, are widely applied in the fields of intelligent recommendations and precision marketing (Li et al., 2016). Online learning platforms record students' learning behavioral data. By analyzing this data, information about students' learning interests, patterns, and motivation can be obtained, which can be used to construct user profiles (Sun & Dong, 2020). The organic integration of user profiles and learning re-

source recommendations will result in more accurate recommendations that better align with users' knowledge, skills, and interest preferences.

Liu et al. (2019) proposed a personalized learning service model that aligns with user needs from the perspective of big data profiles. Meng (2021) calculated the compatibility between users and learning resources by comparing resource features and user characteristics. Ling et al. (2022) introduced a user profile-based collaborative filtering algorithm that maintains the efficiency and accuracy of recommendations as the user base expands. Most of these studies primarily focus on analyzing learner behavior data and learning resource characteristics, lacking comprehensive analysis of learners' overall characteristics and multi-dimensional learner features analysis. Comprehensive analysis contributes to the refinement of personalized learner models. Therefore, to achieve more precise recommendation services, it is essential to focus not only on improving calculation methods but also on researching learners' personalized features.

This study proposes a personalized resource recommendation method for online learning platforms. Firstly, the personalized static and dynamic user data are converted into a user label library, which is used to create user profile data samples. Next, the K-means algorithm is employed to cluster user groups, and a user-resource preference table is constructed for the target user group. Then, user similarity within the group is calculated based on the Pearson coefficient, and users with the highest similarity to the target user are selected. Finally, based on the resources preferences of these similar users, the recommendation score for each resource is generated, and suitable resources are selected to be delivered to the target use.

2. User Profile

The essence of user profiling is to describe, analyze, and present user characteristics, and abstract them into a labeled user model. For online learning platforms, the analysis of attributes and behavioral characteristics of learners is the foundation for building learner profile models. The construction of user profiles involves four steps: data collection, feature selection and label determination, label encoding, and profile presentation.

2.1. Data Collection

Data forms the foundation for portraying user profiles, and the more data collected, the more accurately the profile model can reflect a user's latent characteristics. Therefore, collecting user data comprehensively helps in constructing clear user profiles.

For online learning platforms, the data used to build user profiles can be divided into two categories: static attribute data and dynamic interaction data. Static attribute data refers to information that doesn't change over a short period of time, such as a user's gender, age, occupation, and other characteristics. Dynamic interaction data, on the other hand, includes information that change

over time as the user engages in learning activities, such as browsing history and posting records. It reflects the user's course preferences and learning habits.

During the data collection phase, static attribute data can be obtained directly from the user information unit in the database. Dynamic interaction data can be acquired through analysis of user learning logs. When building user profiles, it's important to combine static attribute features and dynamic interaction features in order to create the most authentic user profiles.

2.2. Feature Selection and Label Determination

To build user profiles, it is necessary to label and vector user feature data. Labeling involves extracting key information that represents user characteristics and converting it into discrete labels. Vectoring, on the other hand, encodes user features into numerical vector, providing computable numerical values for subsequent data analysis.

The key to data labeling is the selection of user features. For learners on online learning platforms, feature selection can be carried out from two dimensions: the first one includes static attribute features such as gender, age, occupation, and education background; the second one comprises dynamic interaction features, including learning interests, learning patterns, and forum participation. Each feature includes multiple labels, as shown in **Table 1**.

2.3. Feature encoding

Each feature in **Table 1** can be encoded by converting feature labels into numerical values or vectors. The specific methods are as follows:

- Gender feature is represented in binary form. 0 represents male, and 1 represents female.
- Age feature is represented in actual age values.
- Occupation and Education features are represented by a 4-dimensional vector. Each label is encoded as a component of the vector. For example, student label is represented as [1, 0, 0, 0], and high school label is represented as [0, 1, 0, 0].

Table 1. User features and label determination.

Features	Label
Gender	Male, Female
Age	Teenager (<20), Youth (21 - 40), Middle-aged (41 - 60), Elderly (>61)
Occupation	Student, Enterprise employee, Employees of public institutions, Freelance
Education	Junior high school, High school, Undergraduate, Graduate
Learning interest	Subject 1, subject 2, ..., subject <i>N</i>
Learning pattern	Video, Text, Exercise
Forum participation	Diving (forum posts = 0), Bubbling (1 - 10), Active (>10)

- Learning interest feature is represented by a vector, and the dimension of the vector is equal to the total number of subjects on the platform. Assuming that there are N subjects in total, if a user spends more than $1/N$ of their total learning time on a certain subject, the corresponding component of that subject is set to 1 in the vector. If a user spends more than $2/N$ of their total learning time on a certain subject, the corresponding component of that subject is set to 2, and so on.
- Learning pattern feature is represented by a 3-dimensional vector. If a user spends more than $1/3$ of their total learning time on a certain resource type, such as video resource, the corresponding component of that resource type is set to 1 in the vector. If a user spends more than $2/3$ of their total learning time on a certain resource type, the corresponding component of that resource type is set to 2, and so on.
- Forum participation feature is represented by the actual number of forum posts

Table 2 represents the sample data of user profiles after feature encoding. Each user has 7 label values, denoted as $U_j (j = 1, 2, \dots, 7)$, which correspond to the user's gender, age, occupation, education background, learning interests, learning pattern, and forum participation. Based on this sample data, various modeling and analysis can be conducted, such as similarity calculation, group clustering, resource recommendation, behavior prediction, etc., to provide data support for the next step of personalized services.

2.4. User Profile Presentation

User profiles can be presented in various ways, such as profile summaries, profile radar charts, profile word clouds, and so on. These presentation methods can be selected and combined based on specific needs and application scenarios, aiming to intuitively convey the core information of user profiles and help better understand and utilize user profile data.

3. Personalized Resource Recommendation Based on User Profiles and Collaborative Filtering

3.1. Collaborative Filtering

The basic idea of collaborative filtering is to predict potential preferences of target

Table 2. User profile sample.

	Gender	Age	Occupation	Education	Learning interest	Learning pattern	Forum participation
User1	0	35	[0, 1, 0, 0]	[0, 1, 0, 0]	[0, 2, 1, ..., 1]	[1, 1, 0]	6
User2	1	22	[1, 0, 0, 0]	[0, 0, 0, 1]	[0, 5, 1, ..., 0]	[1, 1, 0]	2
User3	1	44	[0, 0, 1, 0]	[0, 0, 1, 0]	[1, 0, 7, ..., 0]	[1, 0, 1]	0
...

users through mining historical behavior data and then recommend items. There are two frequently-used collaborative filtering algorithms: the first is user-based collaborative filtering, which recommends items preferred by users with similar interests to the target user. The second is item-based collaborative filtering, which recommends items similar to those previously liked by the user.

For online learning platforms, there is an obvious clustering effect among learners, meaning that students in the same major tend to take similar types of courses. Moreover, users' demand for learning resources does not change abruptly. Therefore, it is suitable to use user-based collaborative filtering algorithms to implement resource recommendation.

Before recommending resources, it is necessary to find a group of users who have similar interests to the target user. This involves clustering all users on the online learning platform based on their similarities, and partitioning them into different clusters. Group profiles can also be created for each cluster.

3.2. Group Clustering

The K-means clustering algorithm is used to segment learner groups, ensuring that learners within each group have similarity while different groups exhibit significant differences. K-means clustering is an unsupervised learning algorithm that partitions samples into different categories based on the similarity between them. The steps for performing group segmentation using the K-means algorithm are as follows.

- 1) Input the dataset and set the number of clusters to K .
- 2) Randomly select K users from the dataset as initial cluster centroids.
- 3) Calculate the distance between users in the dataset and the cluster centroids, and assign each user to the cluster of the nearest cluster centroid. Here, the Euclidean distance is used as the measurement function for assessing the distance between users and cluster centroids, with the following formula:

$$d(U, C) = \sqrt{\sum_{j=1}^7 (U_j - C_j)^2} \quad (1)$$

$d(U, C)$ represents the Euclidean distance between user U and cluster centroid C . U_j represents the numerical values of the 7 feature labels for user U , and C_j represents the numerical values of the 7 feature labels for the cluster centroid.

- 4) Recalculate the mean for each cluster and use it as the new cluster centroid.
- 5) Repeat steps 3 and 4 until the clustering results no longer change or the maximum number of iterations is reached.

3.3. Resource Recommendation

After clustering and dividing the users into groups, the specific steps for resource recommendation using collaborative filtering algorithm are as follows:

- 1) Identify the group where the target user belongs and construct a table of subject preferences for users within that group.

Taking the target user's group as the retrieval space, the learning time of users in the group for each subject are calculated to construct a table of subject preferences, as shown in **Table 3**. With a total of N subjects, if a user's study time for subject i exceeds $1/N$ of their total study time, the preference value Li for that subject is 1. If a user's study time for subject i exceeds $2/N$ of their total study time, the preference value Li for that subject is 2, and so on. The higher the Li value, the higher the user's preference for that subject.

2) Calculate the similarity between users within the group and find the top M users with the highest similarity.

Similarity calculation is a key step in the collaborative filtering algorithm. The conventional Pearson correlation coefficient is adopted to calculate user similarity, as shown below:

$$r(U_1, U_2) = \frac{\sum_{i=1}^N (x1_i - \bar{x1})(x2_i - \bar{x2})}{\sqrt{\sum_{i=1}^N (x1_i - \bar{x1})^2} \sqrt{\sum_{i=1}^N (x2_i - \bar{x2})^2}} \quad (2)$$

$r(U_1, U_2)$ is the user similarity between user U_1 and user U_2 . $x1_i$ represents the preference value of user U_1 for subject i , $\bar{x1}$ represents the average preference value of user U_1 for all subjects, $x2_i$ represents the preference value of user U_2 for subject i , and $\bar{x2}$ represent the average preference value of user U_2 for all subjects. N is the total number of subjects.

Sorting the similarity values between the target user and other users, the top M users with the highest similarity to the target user can be selected. These similarity values are denoted as $r_m (m = 1, 2, \dots, M)$.

2) Generate a subject recommendation list based on the subject preference values and similarity values of the selected users.

Based on the similarity values and subject preference values of the selected M users, the recommendation score for each subject on the online learning platform can be calculated. The formula is as follows:

$$G_i = \sum_{m=1}^M r_m Li_m \quad (3)$$

G_i is the recommendation score for subject i . r_m represents the similarity values between the user m and the target user, Li_m represents the preference values of user m for subject i . The recommendation scores for each subject can be sorted in descending order to obtain the recommended list for the target user.

Table 3. Subject preference values of users in group.

	Preference value of Subject 1 ($L1$)	Preference value of Subject 2 ($L2$)	...	Preference value of Subject i (Li)	...	Preference value of Subject N (LN)
User1	0	2	...	1	...	1
User2	0	0	...	8	...	1
User3	6	0	...	3	...	0
...

4. Conclusion

This study analyzes the static attribute data and dynamic behavioral data of users on an online learning platform. With this information, a user profile label library is constructed. The K-means algorithm is employed to segment users into groups based on the user similarity, and the most similar users to the target user are identified. Finally, a resource recommendation list for the target user is generated by considering the subject preferences and similarity values of these similar users. This is a personalized resource recommendation method that integrates user profiling and collaborative filtering. It helps analyze comprehensively the learning needs and preferences of learners on the online learning platform, thereby providing more personalized resource recommendations and more accurate services.

Acknowledgements

The research is funded by Zhejiang Open University 2023 First Class Curriculum Construction Project (Grant No. XYLKC202309).

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

- Katuk, N., & Ryu, H. (2011). Does a Longer Usage Mean Flow Experience? An Evaluation of Learning Experience with Curriculum Sequencing Systems (CSS). In *Sixth IEEE International Symposium on Electronic Design* (pp. 13-18). IEEE. <https://doi.org/10.1109/DELTA.2011.12>
- Khribi, M. K., Jemni, M., & Nasraoui, O. (2008). Automatic Recommendations for ELearning Personalization Based on Web Usage Mining Techniques and Information Retrieval. In *Eighth IEEE International Conference on Advanced Learning Technologies* (pp. 241-245). IEEE. <https://doi.org/10.1109/ICALT.2008.198>
- Li, B., Wang, Y., & Liu, Y. X. (2016). Application of User Portrait and Intelligent Recommendation Based on Big Data Technology and K-Means. *Modern Computer*, 16, 11-15.
- Ling, K., Jiang, J. L., & Li, S. Q. (2022). Collaborative Filtering Recommendation Algorithm Based on Improved User Profile. *Computer Simulation*, 39, 534-541.
- Liu, H. O., Liu, X., Yao, S. M. et al. (2019). Precision Service Model for Individual Learning Based on In-Depth Portrait of Big Data. *Research on Library Science*, 15, 68-74.
- Liu, Z. Y., Lomovtseva, N., & Korobeynikova, E. (2020). Online Learning Platforms: Reconstructing Modern Higher Education. *International Journal of Emerging Technologies in Learning*, 15, 4-21.
- Meng, J. Y. (2021). *Design and Implementation of Learning Resource Recommendation Platform by Integrating User Profile*. Master's Thesis, Xidian University.
- Salehi, M., KamaLabadi, I. N., & Ghouschi, M. G. (2013). An Effective Recommendation Framework for Personal Learning Environments Using a Learner Preference Tree and a GA. *IEEE Transactions on Learning Technologies*, 6, 350-363. <https://doi.org/10.1109/TLT.2013.28>

-
- Sun, F. Q., & Dong, W. C. (2020). Research on Online Learning User Portrait based on Learning Analysis. *Modern Educational Technology*, 4, 5-11.
- Tan, Z. T., & Zhang, M. J. (2021). Research on Learning Resource Recommendation Model Based on Collaborative Filtering Algorithm. *Computer Technology and Development*, 31, 31-35.
- Wang, L. L. (2017). Research on Personalized Recommendation Technology of Online Learning Based on Collaborative Filtering. *Microcomputer Applications*, 33, 49-51.
- Xiong, C. P. (2019). *Personalized Collaborative Filtering Recommendation Algorithm Based on Label Weight*. Master's Thesis, Xinjiang University.
- Xu, Y. J., & Guo, J. (2020). Recommendation of Personalized Learning Resources on K12 Learning Platform. *Computer Systems & Application*, 29, 217-221.
- Zhai, X. F. (2018). Study on the Intelligent Recommendation System of Personalized Resources Based on Personas. *Journal of Library and Information Science*, 3, 17-21.