

A Personalized Matching System for Management Teaching Resources Based on Collaborative Filtering Algorithm

<https://doi.org/10.3991/ijet.v15i13.15353>

Yongqing Shi ^(✉)

Nanjing University of Aeronautics and Astronautics, Nanjing, China
Anhui Science and Technology University, Fengyang, China
shiyq@ahstu.edu.cn

Xiaojiang Yang

Nanjing University of Aeronautics and Astronautics, Nanjing, China

Abstract—To realize education informatization, it is highly necessary to recommend teaching resources to students that can enhance their learning interest and improve teaching quality. This paper develops a personalized matching system for management teaching resources based on collaborative filtering (CF) algorithm. Firstly, the authors set up a user interest model, designed the flow and algorithm for personalized matching, and improved the similarity calculation method. Next, a personalized recommendation algorithm was developed based on the CF, and a personalized matching engine was constructed with the aid of Apache Mahout. The experimental results show that the proposed CF algorithm can effectively improve the recommendation quality, and push personalized teaching resources to each user; the learners are highly satisfied with the personalized matching system. The research results shed new light on personalized recommendation of teaching resources, opening up a new way to education informatization.

Keywords—Collaborative filtering algorithm, management, teaching resources, personalized matching

1 Introduction

The rapid development of computers, information technology and network technology has changed the people's way to live and learn, and also set off a wave of education informatization in China. Meanwhile, various network teaching resources have emerged. Although people can obtain more learning opportunities and resources through the network, they still feel it more difficult to find the information they need quickly and effectively, so it is particularly important to apply recommendation technology to the education field and provide learners with suitable teaching resources.

In 1992, researchers at the XP Research Center applied the CF technology to the mail system for the first time [1]. Personalized recommendation technology has been applied earlier in e-commerce systems, e.g., Amazon has increased user purchases through personalized recommendation algorithms [2]. At the beginning of the 21st century, foreign scholars conducted research on the application of personalized recommendation systems in online learning. The Open University Neitherland constructed a personalized recommendation system and summarized the application scenarios of various personalized recommendation technologies [3]. In recent years, this technology has developed rapidly, and been used in almost every field. Domestic websites such as JD.com, Alibaba, and Dangdang are mature in the application of personalized recommendation technology [4]. Comparatively speaking, the application of personalized recommendation technology in the education field started late, without efficient promotion and high user experience. Until 2008, the research results were gradually enriched. For example, Yang Lina et al. designed a technical process of the virtual learning community based on the personalized recommendation technology [5]; Wang Yanfang proposed a personalized e-Learning system [6]. Through reviewing the relevant references at home and abroad, it's found that the current application of domestic personalized recommendation technology in the education field still has problems such as lack of intelligence in the recommendation system, relatively single recommendation strategy, and defects in the personalized recommendation mechanism [7].

In view of the above, this paper firstly gives a brief introduction to the relevant theories about the personalized matching system of teaching resources, and designs the related recommendation process of the personalized matching system based on the CF algorithm with the aim to meeting the teaching needs of management. Also, the traditional Pearson similarity algorithm was improved, and a recommendation engine was built based on Apache Mahout. Finally, the experimental analysis was conducted to verify the effectiveness and practicability of the improved algorithm and personalized matching system.

2 Related Theories of Personalized Matching System for Teaching Resources

2.1 Classification of recommendation algorithms

To reduce the users' blind search for resources, the personalized recommendation system interacts with the user, and actively recommends the information that meets the users' personalized needs and interests to them [8]. At present, there are many algorithms and technologies for personalized recommendation. The three common recommendation algorithms are analyzed as follows:

1. Rule-based recommendations. Rules can be discovered through the data mining technology of association rules or be formulated by users themselves. The quality and quantity of rules determine the quality of rule-based recommendation. Key-

words, description layer and user interface layer are the three main components of the rule-based recommendation system [9], in which the keywords are the foundation, the description layer defines the users and resources, and the user interface layer is based on the other two, providing users with personalized recommendation results. The disadvantage of this method is that the quality of the rules is difficult to guarantee and cannot be updated in real time.

2. Content-based recommendations. Content-based recommendations relies on user access content through browsing, evaluation, and sharing etc., and uses various technologies to judge user behavior and needs [10]. It has the disadvantage that a lot of analysis work needs to be performed before recommendation, but a large increase in the number of web page user visits makes it difficult to analyze. It also cannot mine new interesting resources for users.
3. Collaborative filtering-based recommendations. This algorithm can discover the user's preference by mining the historical behavior data [11]. Prediction and recommendation are the main functions of CF technology, with the advantages of high degree of personalization and obvious effects. Thus, this recommendation algorithm is widely used.
4. Hybrid recommendation. To obtain the best recommendation results and take advantage of various recommendation methods in the actual application process, two or more methods are often combined to form a new hybrid recommendation method. Collaborative filtering technology is often one of the essential technologies in the hybrid recommendation technology.

2.2 Introduction of collaborative filtering technology

Currently, item-based, memory-based and user-based CF algorithms are widely used in various fields. This paper analyzes the following two CF algorithms respectively.

User-based collaborative filtering algorithm: The user-based CF algorithm can discover the potential interest of the target user and obtain more accurate recommendation result. The algorithm mainly includes the following three steps:

1. *Input user-item evaluation matrix:* The calculation formula is shown in (1), and the user's preference for the item is usually expressed by a number between 0-5 [12].

$$R(m, n) = \begin{pmatrix} R_{11} & R_{12} & \cdots & R_{1n} \\ R_{21} & R_{22} & \cdots & R_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ R_{m1} & R_{m2} & \cdots & R_{mn} \end{pmatrix} \quad (1)$$

where, R_{ij} is the rating of item j by user i , m is the user, and n is the item.

2. *Calculate user similarity and form neighbors:* Common methods for calculating user similarity are Pearson similarity algorithm and cosine similarity algorithm.

The former was used in this study. The user's similarity $\text{sim}(i,j)$ can be calculated by formula (2) [13].

$$\text{Sim}(i, j) = \frac{\sum_{k \in I_{i,j}} (R_{i,k} - \bar{R}_i)(R_{j,k} - \bar{R}_j)}{\sqrt{\sum_{k \in I_{ij}} (R_{i,k} - \bar{R}_i)^2} \sqrt{\sum_{k \in I_{ij}} (R_{j,k} - \bar{R}_j)^2}} \quad (2)$$

where, R_i and R_j are the average evaluation of items by users i and j ; $R_{i,k}$ are user i 's rating for item k ; the set of items rated jointly by users i and j is: $I_{ij} = I_i \cap I_j$.

After the similarity calculation, there are two ways to determine the neighbor set of the target user. One is to select the k users with the largest similarity and form the neighbor set; the second is to set the similarity threshold. Only the similarity over this threshold can be allowed into the neighbor set.

3. *Generate recommendations:* The user's predicted rating for an item can be calculated by formula (3). The recommendation set consists of the N highest predicted items that are not in the user's rated item set [14].

$$P_{u,i} = \bar{R}_u + \frac{\sum_{n \in S_u} \text{Sim}(u,n) (R_{u,i} - \bar{R}_n)}{\sum_{n \in S_u} \text{Sim}(u,n)} \quad (3)$$

where, $R_{u,i}$ is the score of user n on item i , $\text{Sim}(u,n)$ is the similarity between users n and i , R_u and R_n are the average evaluation of users n and i on the item.

Item-based collaborative filtering algorithm: This algorithm has the advantages such as calculation offline, and saving computing time, and the disadvantage that it cannot tap the potential interest of users. Similar to the user-based algorithm, it also needs three steps, and their first step is the same. So, only the last two steps are introduced in detail.

1. *Calculate similarity:* $\text{Sim}(i,j)$ similarity of items i and j is shown in the calculation formula (4):

$$\text{Sim}(i, j) = \frac{\sum_{k \in U_{ij}} (R_{k,i} - \bar{R}_i)(R_{k,j} - \bar{R}_j)}{\sqrt{\sum_{k \in U_{ij}} (R_{k,i} - \bar{R}_i)^2} \sqrt{\sum_{k \in U_{ij}} (R_{k,j} - \bar{R}_j)^2}} \quad (4)$$

where, the average scores of items i and j are R_i and R_j , and $R_{k,j}$ is the score of user k on item i .

2. *Generate recommendations:* The user's prediction score for the item is calculated by the method of weight summation. To ensure the predicted value within the defined range, the weight sum is divided by the similarity sum of the similar items. It's calculated as:

$$P_{u,i} = \frac{\sum_{\text{similar-items},n} (S_{i,n} * R_{u,n})}{\sum_{\text{similar-items},n} (|S_{i,n}|)} \quad (5)$$

where, $S_{i,n}$ is the degree of similarity between items i and n , and $R_{u,n}$ is the score of item n in the similar item set by the user u .

3 Personalized Matching System of Management Teaching Resources Based on Collaborative Filtering Algorithm

3.1 Establishment of user interest model

The establishment of user interest models, item matching and recommendation output are the three main stages of the personalized matching system for management teaching resources [15]. The user interest model is a formal algorithm-oriented description with a specific data structure. At this stage, the main tasks are obtaining and maintaining user habits, needs, and interests. Thus, this stage is also the foundation and core of the personalized matching system for management teaching resources. The establishment of user interest model usually includes the tasks of data collection, model representation, learning and updating.

Management teaching resources include related online courses, test paper materials, test question banks, answers to frequently asked questions, literature and cases, etc. Based on the online teaching platform, this paper takes the teacher and students as the research objects. The student's learning information includes: name, gender, major, and teaching level (undergraduate, graduate, and doctoral). Due to the wide range of students' interest, they may be also interested in non-major professional information. Therefore, the student groups were divided into management professional interest group and non-management professional interest group. The user's preference for items was expressed by the user-item scoring matrix $R(m,n)$.

3.2 Personalized matching process and algorithm design

Recommended process: As above, management teaching resources were classified into professional resources and non-professional resources, of which the latter mainly refers to the resources that students are interested in, and the former includes the resources of interest to students, boutique resources and the latest resources, as shown in Fig. 1.

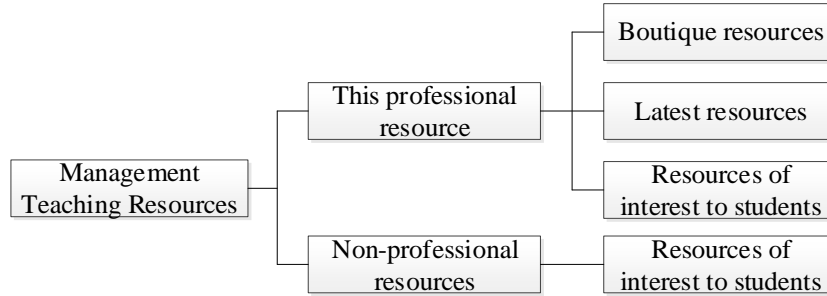


Fig. 1. Classification of teaching resources

Depending on the resources, the recommended method used will vary. Students should have the highest access rate to resources in their major. To avoid the repetition of same resources in different types of educational resources, the boutique resources were recommended first in this study. The boutique resources refer to resources with high evaluation scores. The system will compare and analyze these resources according to the user's professional codes, and then recommend the boutique resources of management to students. The system should continue to have new resources in order to ensure that users can continuously learn new knowledge. The system uses the new resources added by students within two login times and then recommends them to students. Meanwhile, to avoid excessive pushing of new resources, a threshold N will be set, so the number of pushed new resources will not exceed N .

Compared with the first two teaching resource recommendations, the resource recommendation of students' interest considers the user's personalized needs and interests, which is also the focus of this study. Fig. 2 shows the specific process: first input user-item evaluation matrix, and use the improved Pearson similarity algorithm to calculate the similarity between items; then predict the score of unrated items based on the user's evaluation of the item using the weighted sum method; among the management professional resources, select the N items with the highest ratings and that are not in the rated set of items by the user as the recommended set.

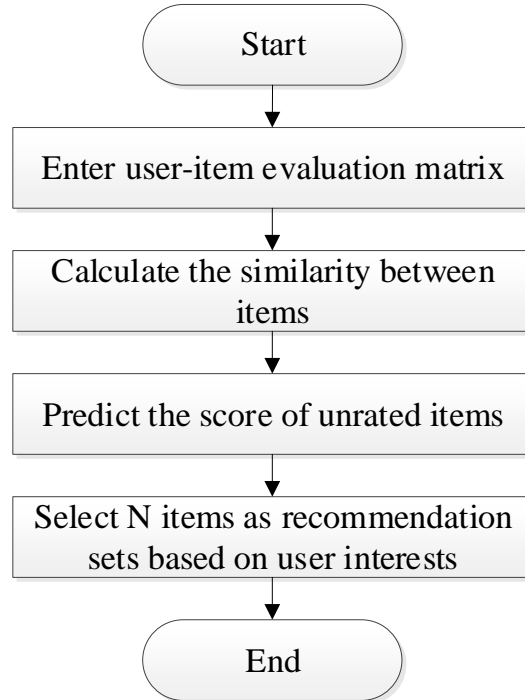


Fig. 2. Teaching resource personalized matching recommendation process

Improved similarity calculation method: When the user score data is extremely sparse, the nearest neighbor calculated by the traditional Pearson similarity algorithm is inaccurate. To effectively avoid this problem, this paper improves the traditional Pearson similarity algorithm by taking U_i and U_j as the user set evaluating i and j . When calculating the similarity, first calculate the union of the user sets scored by items i and j , as shown in formula (6):

$$U_{ij} = U_i \cup U_j \quad (6)$$

Hybrid personalized recommendation algorithm based on collaborative filtering: In order to better achieve the personalized matching of management teaching resources, this paper uses a hybrid recommendation algorithm (based on the item-user collaborative filtering algorithm), and its process is shown in Fig. 3.

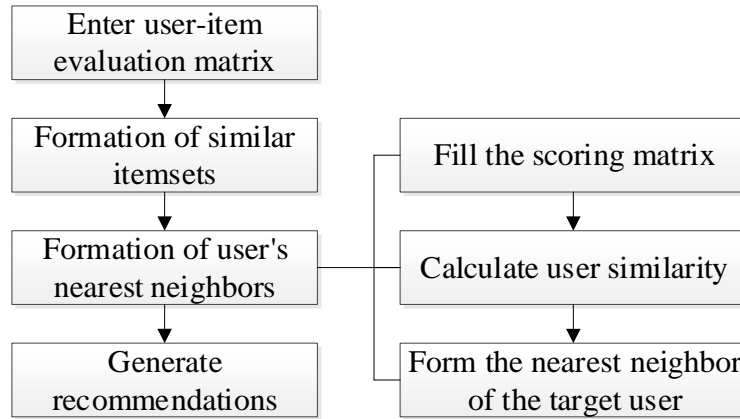


Fig. 3. Process of hybrid collaborative filtering personalized recommendation algorithm

First, enter the user-item evaluation matrix R , then use the improved Pearson similarity algorithm to calculate the similarity $Sim(i, j)$ of items i and j , and select the top N items with the maximum similarity to item k as the similarity set SI_k , which is stored in a certain database to be updated regularly.

Let the set of items rated by users u and i to be $I_{ui} = I_u \cup I_i$, I_u , and I_i , the unrated items in the set of items to be k , then user u 's rating of item k can be predicted by the similar set SI_k of items in item k , as shown in formula (7). This is repeated continuously to fill the scoring matrix. According to formula (8), the similarity between the target users u and I was calculated to select the top N users with the largest similarity to user u as their nearest neighbor set. The formula (9) was used to calculate the predicted score of item u by user u , and the top N items that were not in the set of rated items by the user with a higher score were selected as the user's recommended set.

$$P_{u,k} = \bar{R}_k + \frac{\sum_{n \in SI_k} Sim(k,n) (R_{u,n} - \bar{R})}{\sum_{n \in SI_k} Sim(k,n)} \quad (7)$$

$$Sim(u,i) = \frac{\sum_{k \in I_{ui}} (R_{i,k} - \bar{R}_i)(R_{u,k} - \bar{R}_u)}{\sqrt{\sum_{k \in I_{ui}} (R_{i,k} - \bar{R}_i)^2} \sqrt{\sum_{k \in I_{ui}} (R_{u,k} - \bar{R}_u)^2}} \quad (8)$$

$$P_{u,j} = \bar{R}_u + \frac{\sum_{n \in N_u} Sim(u,n) (R_{n,j} - \bar{R}_n)}{\sum_{n \in N_u} Sim(u,n)} \quad (9)$$

3.3 Implementation of personalized matching engine

The “taste” provided by Apache Mahout can efficiently implement the CF algorithm, and it is also convenient for users to define and implement their own recommendation algorithm, which can meet the requirements of scalability and flexibility for the recommendation engine in the personalized matching system for management teaching resources. Therefore, this paper uses Apache Mahout to build a recommendation engine. Fig. 4 shows the system login interface. Fig. 5 shows the recommendation results of teaching resources books for management majors.



Fig. 4. System login interface

System menu		Welcome to use this system			
Personal center	Home Page	Book management			
Announcement management	School resources				
Teaching resource management	Search by title:	Management	Search		
Courseware management	Name	Publishing house	Author	Views	
Book management	<input type="checkbox"/> Fundamentals of Management	Guangdong Higher Education Press	Cai Muecheng Huang Qiuwen	734	
Problem Management	<input type="checkbox"/> Management	China Social Sciences Press	Wu Zhaoyun et al	613	
	<input type="checkbox"/> Principles of Management	China Renmin University Press	Fang Zhenbang Bao Chunlei	584	
	Off-campus resources that may be of interest				
	Management	China Renmin University Press	Stephen P Robbins	635	
	Management and Life	China Renmin University Press	Paula Caproni	437	
Job management					
Feedback					

Fig. 5. Recommended results of teaching resources books for management majors

3.4 Evaluation of the algorithm

Experimental data set and evaluation standard: This study selects 300 students of management majors in a college, records their learning process in detail, and collects 1,375 evaluation data of 2,500 teaching resources from users. The experimental data were then divided into two parts: training set and test set. The initial value of the number was set to 2, increasing by 2 each time to test the change of the average absolute error (MAE). In this way, the recommendation quality of the personalized matching system of teaching resources was tested.

The MAE was selected as the evaluation standard. The smaller the MAE value, the higher the recommended quality. It's calculated as:

$$MAE = \frac{\sum_{i=1}^m MAE_i}{m}$$

$$MAE_i = \frac{\sum_{i=1}^n |p_i - q_i|}{n}$$
(10)

where, m is the number of all users, pi and qi are the predicted score and the actual user score respectively.

Analysis of experimental results: Fig. 6 compares the MAE values of the two item-based CF algorithms. It can be seen from the figure that the improved CF algorithm had a significantly lower MAE than the traditional CF algorithm; MAE of the traditional algorithm gradually increased with the number of nearest neighbors, while that of the improved CF algorithm gradually decreased, indicating that it can effectively improve the recommendation quality.

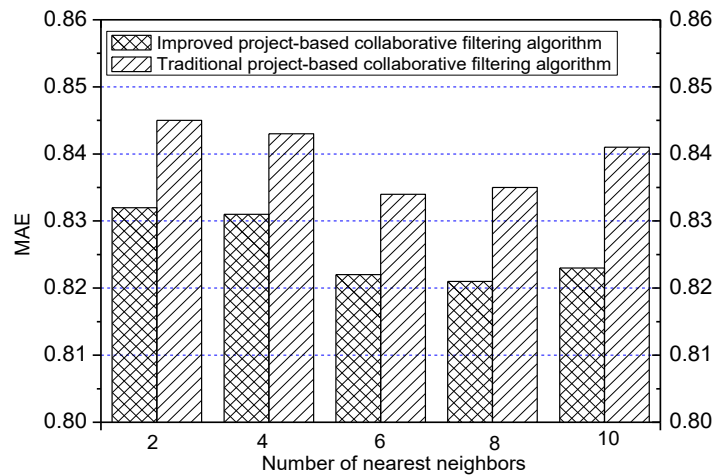


Fig. 6. Comparison of MAE values of two item-based collaborative filtering algorithms

Fig. 7 compares the MAE values of the three CF algorithms. It can be seen from the figure that the MAE value of the hybrid CF algorithm gradually decreased with the increase of the number of nearest neighbors, while that of the other two algorithms gradually decreased first and then increased; with the same number of nearest neighbors, the MAE value of the hybrid algorithm was smaller than the other two algorithms. The experimental results showed that the quality of the hybrid collaborative filtering is better than the item-based and user-based CF algorithm.

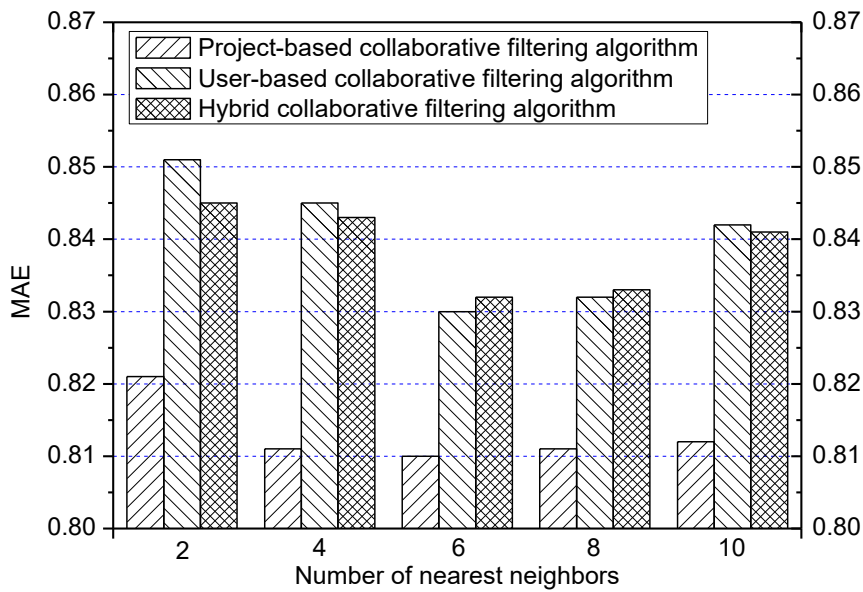


Fig. 7. Comparison of MAE values of three collaborative filtering algorithms

Fig. 8 shows the results of the learners' satisfaction survey. It can be seen that most learners are satisfied with the personalized matching system of management teaching resources based on CF algorithm, indicating that the system can meet the learners' personalized learning to a certain extent, and have a certain recommendation effect, but a small number of learners are still dissatisfied with the system, indicating that the system still needs to be continuously improved according to the needs of learners in the future.

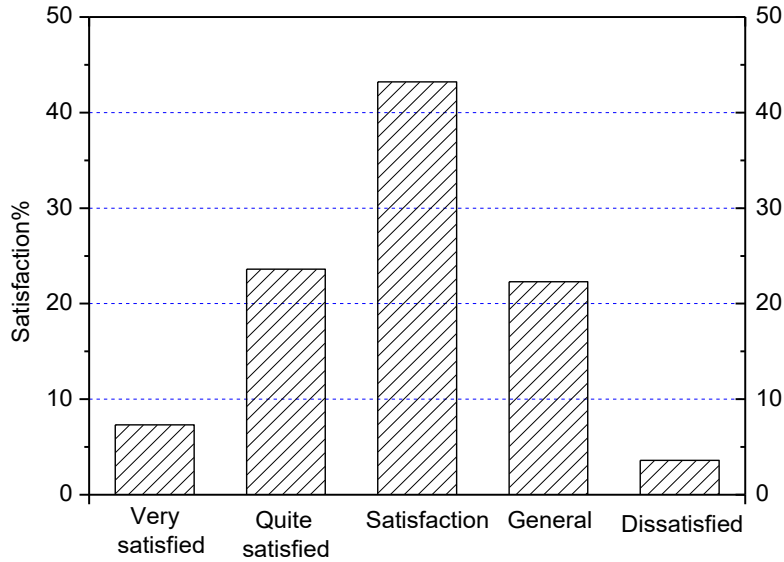


Fig. 8. Learner satisfaction

4 Conclusion

This paper applies the recommendation technology to the management teaching, and studies the personalized matching system of management teaching resources based on the CF algorithm. The specific conclusions are as follows:

1. According to the teaching needs of management, the recommendation process of the personalized matching system for management teaching resources based on CF algorithm was designed in detail, and the traditional Pearson similarity algorithm was improved.
2. In order to realize the functions of the personalized matching system, a recommendation engine based on Apache Mahout was constructed.
3. The experiment was conducted to analyze the recommendation quality of the personalized matching system for management teaching resources. It's concluded that the improved CF algorithm and the personalized matching system proposed in this paper can achieve good results in the personalized recommendation for users.

5 Acknowledgement

This work is supported by the funded program on excellent talent cultivation in colleges and universities of Anhui Province (gxyq2018047), the teaching research project of Anhui Province (2018jyxm0675), the innovation development research program in Social Science of Anhui Province (2019CX082) and the Quality Project of Anhui Science and Technology University (XJ201946).

6 References

- [1] Deng, X.Y., Wang, C. (2018). A hybrid collaborative filtering model with context and folksonomy for social recommendation. *Ingénierie des Systèmes d'Information*, 23(5): 139-157. [https://doi.org/10.3166/ ISI.23.5.139-157](https://doi.org/10.3166/ISI.23.5.139-157)
- [2] Bendahmane, M., El Falaki, B., Benattou, M. (2019). Toward a personalized learning path through a services-oriented approach, *International Journal of Emerging Technologies in Learning*, 14(15): 52-66. <https://doi.org/10.3991/ijet.v14i15.10951>
- [3] Wang, J., Clements, M., Yang, J., Vries, A.P.D., Reinders, M.J.T. (2010). Personalization of tagging systems. *Information Processing & Management*, 46(1): 58-70. <https://doi.org/10.1016/j.ipm.2009.06.002>
- [4] Anderson, M., Leflore, J.L., Anderson, J.D.M. (2013). Evaluating videotaped role-modeling to teach crisis resource management principles. *Clinical Simulation in Nursing*, 9(9): e343-e354. <https://doi.org/10.1016/j.ecns.2012.05.007>
- [5] Tarn, K.Y., Ho, S.Y. (2005). Web personalization as a persuasion strategy: An elaboration likelihood model perspective. *Information Systems Research*, 16(3): 271-291. <https://doi.org/10.1287/isre.1050.0058>
- [6] Choi, I.Y., Oh, M.G., Kim, J.K., Ryu, Y.U. (2016). Collaborative filtering with facial expressions for online video recommendation. *International Journal of Information Management*, 36(3): 397-402. <https://doi.org/10.1016/j.ijinfomgt.2016.01.005>
- [7] Chung, T.S., Rust, R.T., Wedel, M. (2009). My mobile music: An adaptive personalization system for digital audio players. *Marketing Science*, 28(1): 52-68. <https://doi.org/10.1287/mksc.1080.0371>
- [8] Tawafak, R.M., Romli, A., Malik, S.I., Shakir, M., Al Farsi, G. (2019). A systematic review of personalized learning: Comparison between e-learning and learning by coursework program in Oman, *International Journal of Emerging Technologies in Learning*, 14(9): 93-104. <https://doi.org/10.3991/IJET.V14I09.10421>
- [9] Tesone, D.V. (2004). Development of a sustainable tourism hospitality human resources management module: a template for teaching sustainability across the curriculum. *International Journal of Hospitality Management*, 23(3): 207-237. <https://doi.org/10.1016/j.ijhm.2003.10.003>
- [10] Lin, H., Yang, X., Wang, W., Luo, J. (2014). A performance weighted collaborative filtering algorithm for personalized radiology education. *Journal of Biomedical Informatics*, 51: 107-113. <https://doi.org/10.1016/j.jbi.2014.04.015>
- [11] Lin, H., Yang, X., Wang, W. (2014). A content-boosted collaborative filtering algorithm for personalized training in interpretation of radiological imaging. *Journal of Digital Imaging*, 27(4): 449-456. <https://doi.org/10.1007/s10278-014-9678-z>
- [12] Kahn, C.E. (2005). Collaborative filtering to improve navigation of large radiology knowledge resources. *Journal of Digital Imaging*, 18(2): 131-137. <https://doi.org/10.1007/s10278-004-1910-9>
- [13] Sun, H., Zheng, Z., Chen, J., Lyu, M.R. (2013). Personalized web service recommendation via normal recovery collaborative filtering. *IEEE Transactions on Services Computing*, 6(4): 573-579. <https://doi.org/10.1109/tsc.2012.31>
- [14] Lei, W., Qing, F., Zhou, J. (2016). Improved personalized recommendation based on causal association rule and collaborative filtering. *International journal of distance education technologies*, 14(3): 21-33. <https://doi.org/10.4018/ijdet.2016070102>
- [15] Zhao, M.J. (2014). A knowledge-based teaching resources recommend model for primary and secondary school oriented distance-education teaching platform. *lecture notes in electrical engineering*, 269: 511-521. https://doi.org/10.1007/978-94-007-7618-0_50

7 Authors

Yongqing Shi was born in Nantong, Jiangsu, Lecturer, Graduate for Ph.D. degree of College of Economics and Management, Nanjing University of Aeronautics and Astronautics, her research interests include Management Science and Engineering and Theory and practice of higher education (shiyq@ahstu.edu.cn).

Xiaojiang Yang was born in Nantong, Jiangsu, Professor of College of Economics and Management, Nanjing University of Aeronautics and Astronautics, his research interests include Management Science and Engineering and Engineering and Theory and practice of higher education (ahkjxyjk@126.com).

Article submitted 2020-05-01. Resubmitted 2020-06-04. Final acceptance 2020-06-04. Final version published as submitted by the authors.