ICIC Express Letters ICIC International ©2020 ISSN 1881-803X
Volume **14**, Number **11**, November **2020** pp. 1083–1090

KNOWLEDGE POINT RECOMMENDATION ALGORITHM BASED ON SIMILARITY OPTIMIZATION

Zhaoyu Shou[1,2], Yanguo Wang[1], Yiru Wen[1] and Huibing Zhang[3]

[1]School of Information and Communication
[2]Key Laboratory of Cognitive Radio and Information Processing Ministry of Education
[3]Guangxi Key Laboratory of Trusted Software
Guilin University of Electronic Technology
No. 1, Jinji Road, Guilin 541004, P. R. China
guilinshou@guet.edu.cn

Received February 2020; accepted May 2020

Abstract. *Due to the complexity of the structure of knowledge points in online course videos, the relative difficulty coefficient of knowledge points will be different for learners. The recommendation algorithms came into being, and the collaborative filtering algorithms are the most widely used recommendation algorithms. However, the traditional collaborative filtering recommendation algorithms have limitations in the similarity calculation between learners. Therefore, we propose a new knowledge point recommendation algorithm. The algorithm fully considers the differences in learners' relative difficulty coefficient of knowledge points in different dimensions. Besides, we define the similarity between learners at an overall level by incorporating information from knowledge points learned by non-associated learners. We comprehensively consider different aspects to improve the performance and quality of the recommendation algorithm. Comparison experiments on the real dataset demonstrate that our proposed algorithm improves the overall performance of the recommended algorithm.*
Keywords: Similarity calculation, Difference, Collaborative filtering, Knowledge point recommendation

1. **Introduction.** Quantifiable personalized education is an important complementary form of classroom education [1-3]. Meanwhile, personalized knowledge point recommendation is the core issue of adaptive teaching [4,5], which is an important way to improve the learning effect. Among the many recommendation algorithms, the collaborative filtering algorithm is the most widely used recommendation algorithm [6]. However, the traditional collaborative filtering algorithm has the problem of poor precision in the similarity calculation. To settle the problem above, Wang and Zheng [7] improved the classic similarity calculation model by considering the proportion of users' ratings of common scoring items, but the model ignores the impact of users' non-common scoring items on similarity calculation. Besides, if the data of users' common scoring items is relatively sparse, the precision of similarity calculation will decrease. Xiang and Qiu [8] used the slope-one algorithm to calculate the predicted value of the score as a backfill value to improve the precision of the algorithm. However, the cosine similarity model used in this paper does not consider the evaluation criteria of users for the projects. Zhang et al. [9] segmented the PCC (Pearson Correlation Coefficient) algorithm to improve the results based on the number of user public items and the PCC threshold. Gao et al. [10] introduced the concept of coincidence dependency to modify the traditional similarity measure. The similarity calculation of [9] and [10] only considers the overall difference of users' ratings on the project and ignores the difference of users' ratings on each project.

DOI: 10.24507/icicel.14.11.1083

To solve the problem of collaborative filtering algorithms, we fully consider the differences of learners' relative difficulty coefficient of knowledge points in different dimensions, if the smaller the difference of learners' relative difficulty coefficient on the same knowledge point, the higher the similarity between the learners. Meanwhile, learners' evaluation criteria for the relative difficulty coefficient of knowledge points are also different. Based on these two aspects, we construct a vector calculation model of relative difficulty coefficient of knowledge points. In addition, if the cognitive levels of learners differ greatly, the similarity calculation between learners is more focused on the knowledge points that the learners have learned, besides the knowledge points that they have learned together. Therefore, in order to further improve the quality of the algorithm, we reduce the impact of the number of knowledge points that learners learn together on the similarity calculation by integrating the knowledge points learned by non-associated learners. So we comprehensively improve the performance and quality of the recommendation algorithm from different aspects.

The rest of this paper is organized as follows. Section 2 presents the related definitions. In Section 3, a new knowledge point recommendation algorithm is proposed. In Section 4, various experimental results are comparatively analyzed. Finally, Section 5 presents conclusions.

2. Related Definitions.

As knowledge points are based on the course videos, the teaching videos of the online course are lectures on the knowledge points, so the knowledge point recommendation is essentially the corresponding course video recommendation. In online courses, learners have a small number of explicit scores for the course videos, so we mine the learners' hidden information to build the learners' relative difficulty coefficient of knowledge points, so as to represent learners' preference for the scoring of knowledge points.

Definition 2.1. ***Relative difficulty coefficient of knowledge points.*** *According to the definition in [11], learners' learning behaviors of knowledge points are collaboratively analyzed. The learners' learning behavior characteristics for the video learning frequencies, the video pausing and dragging frequencies, and the video learning duration are weighed as the relative difficulty coefficient of knowledge point i to the learner u, which is shown in (1).*

$$fkp_{u,i} = \alpha * v_f(u,i) + \beta * v_t(u,i) + \gamma * v_{p,d}(u,i) \tag{1}$$

where i represents the knowledge point of the video, u represents the learner, $v_f(u,i)$ represents the learning frequencies of u watching i, $v_t(u,i)$ represents the learning duration of u watching i, and $v_{p,d}(u,i)$ represents the frequency of u pausing and dragging i. In [11], in order to determine the value of (α,β,γ), Zhu et al. conducted a large number of experiments and a total of 67 different combinations (α,β,γ) were tested under the ratio of integer. Finally, it was concluded that the recommendation accuracy was the highest when the value was $(1,5,4)$.

Definition 2.2. ***Calculation of similarity between learners.*** *In the traditional similarity calculation, cosine similarity and Pearson correlation coefficient are the most widely used similarity calculation methods. The similarity between the learner u and the target learner v is calculated as follows.*

① *Cosine similarity*

$$sim(u,v) = \frac{\sum_{i=1}^{C} fkp_{u,i} \times fkp_{v,i}}{\sqrt{\sum_{i=1}^{C}(fkp_{u,i})^2}\sqrt{\sum_{i=1}^{C}(fkp_{v,i})^2}} \tag{2}$$

where $fkp_{u,i}$ represents the relative difficulty coefficient of the knowledge point i to the learner u, $fkp_{v,i}$ represents the relative difficulty coefficient of the knowledge point i to *the target learner v, and C represents the knowledge points learned by both the learner u and the target learner v.*

② *Pearson correlation coefficient*

$$sim(u,v)=\frac{\sum_{i=1}^{C}\left(fkp_{u,i}-\overline{fkp}_u\right)\left(fkp_{v,i}-\overline{fkp}_v\right)}{\sqrt{\sum_{i=1}^{C}\left(fkp_{u,i}-\overline{fkp}_u\right)^2}\sqrt{\sum_{i=1}^{C}\left(fkp_{v,i}-\overline{fkp}_v\right)^2}} \tag{3}$$

where $\overline{fkp}_u$ represents the average relative difficulty coefficient of the knowledge points learned by the learner u, and $\overline{fkp}_v$ represents the average relative difficulty coefficient of the knowledge points learned by the target learner v.

Definition 2.3. ***Predict the relative difficulty coefficient of recommended knowledge points to learners.*** *The relative difficulty coefficient of the knowledge point j that has not been learned by the target learner v can be predicted by calculating the similarity between learners through Equation (2) or (3). The prediction equation is as follows.*

$$fkp_{v,j}=\overline{fkp}_v+\frac{\sum_{u=1}^{n}\left(fkp_{u,i}-\overline{fkp}_u\right)\times sim(u,v)}{\sum_{u=1}^{n}sim(u,v)} \tag{4}$$

where n represents the number of similar learners, and $sim(u,v)$ represents the similarity between the target learner v and the learner u.

3. Knowledge Point Recommendation Algorithm Based on Similarity Optimization.

3.1. **Vector calculation based on the relative difficulty coefficient of knowledge points.** Traditional collaborative filtering algorithms have different limitations when calculating similarities between learners through similarity models.

In the similarity calculation of collaborative filtering, learners' relative difficulty coefficient of knowledge points can be regarded as an n-dimensional vector. When calculating the similarity between learners, the cosine similarity is considered by calculating the cosine value of the angle between the relative difficulty coefficient of knowledge points by learners, ignoring specific size of learners' relative difficulty coefficient of knowledge points.

On the other hand, the Pearson correlation coefficient only considers the difference in learners' relative difficulty coefficient of knowledge points as a whole, and ignores the difference of learners' relative difficulty coefficient on the same knowledge point. Therefore, we construct a new measurement standard based on learners' differences in the relative difficulty coefficient of knowledge points to better calculate the similarity between learners.

For the learner u and the target learner v, the difference in the relative difficulty coefficient of the same knowledge point i can be expressed as the difference of the vector $|fkp_{u,i}-fkp_{v,i}|$. At the same time, for different learners u and v, there are differences in the relative difficulty coefficient of knowledge points in different dimensions, that is, learners' evaluation criteria for the relative difficulty coefficient of knowledge points are different. If learners have a good grasp of knowledge points, the relative difficulty coefficient of knowledge points is generally low; on the contrary, the relative difficulty coefficient of knowledge points is higher.

Inspired by vector computing, the difference of relative difficulty coefficient of knowledge points for different learners can be seen as the difference between different vectors. Its definition is as follows:

$$\varphi(u,v)=1-\frac{\sqrt{\sum_{i=1}^{C}|fkp_{u,i}-fkp_{v,i}|\cdot\left|\overline{fkp_{u,c}}-\overline{fkp_{v,c}}\right|}}{\max-\min} \tag{5}$$

where C represents the knowledge points learned by both the learner u and the target learner v, $\overline{fkp_{u,c}}$ represents the average relative difficulty coefficient of the knowledge point set C learned by the learner u, $\overline{fkp_{v,c}}$ represents the average relative difficulty coefficient of the knowledge point set C learned by the target learner v, max represents the maximum relative difficulty coefficient of all knowledge points for all learners, and min represents the minimum relative difficulty coefficient of all knowledge points for all learners.

We measure the differences between learners at different aspects, fully consider the differences in the relative difficulty coefficient of knowledge points in different dimensions, and calculate the similarities between learners more accurately.

3.2. **Calculation of correction factors based on similarity of non-associated learners.** The traditional similarity calculation depends on the number of knowledge points that learners learn together. If the number of knowledge points that different learners learn together is relatively sparse, the similarity calculation between learners has the defect of poor accuracy.

To solve this problem, we modify the vector calculation model in Section 3.1. If the cognitive levels of learners u and v differ greatly, the similarity calculation between learners is more focused on the knowledge points that the learners have learned, besides the knowledge points that they have learned together. Therefore, we consider the influence of non-associated learners on similarity calculation between learners. The definition is as follows.

$$\theta = \frac{C}{|K_u| \cdot |K_v|} \tag{6}$$

where K_u represents the knowledge points learned by learner u, and K_v represents the knowledge points learned by learner v.

3.3. **Knowledge point recommendation algorithm based on similarity optimization.** By combining the above correction factor θ and the difference $\varphi(u,v)$, the similarity between learners is calculated as follows.

$$sim(u,v) = \varphi(u,v) \cdot \theta \tag{7}$$

The pseudo code of the proposed recommendation algorithm is as follows.

Algorithm: The knowledge point recommendation algorithm based on similarity optimization

Input: U: Learners,
    D: Learners' learning behavior data,
    K: Knowledge point set
Output: Recommended list of knowledge points for target learners
1. The real dataset is divided into a training set and a test set, with the training set accounting for 80% and the test set accounting for 20%.
2. for each $u \in U$ in D do
3.   for each $u \in U$ in K do
4.   Calculate the learner's relative difficulty coefficient of the knowledge point $fkp_{u,i}$ using Equation (1)
5. Calculate learners' similarity $sim(u,v)$ using Equation (7)
6. Sort similarities between learners
7. Calculate predicted relative difficulty coefficient of the knowledge point $fkp_{v,j}$ using Equation (4)
8.   end for
9. end for
10. Generate recommended list of knowledge points for target learners

4. **Experimental Evaluation.** To evaluate the performance and quality of our proposed algorithm (named New), we compare the collaborative filtering algorithm based on Cos similarity (named Cos), the collaborative filtering algorithm based on Pearson correlation coefficient (named Person), improved collaborative filtering algorithm based on user-similarity [7] (named ICF), and a new similarity calculation method in collaborative filtering-based recommendation systems [12] (named JacRA by Wu et al.). The experiment will analyze and compare the performance of the five algorithms on the evaluation index, and verify the effectiveness and quality of our proposed algorithm.

4.1. **Dataset of the experiment.** Our experiment is carried out based on the learning behavior data of students on the online learning platform. The course we choose is the data structure and algorithm. There are 1,198 learners and 207 video knowledge points in the dataset, and it contains learning behavior logs for 207 knowledge points. For the purpose of experimental evaluation, the dataset is divided into a training set and a test set, with the training set accounting for 80% and the test set accounting for 20%. All experiments are conducted in Matlab R2017.

4.2. **Experimental results and analysis.**

4.2.1. *Influence of neighbor size on prediction accuracy.* The size of neighbors (the number of similar learners) affects the prediction accuracy (MAE) of the recommendation algorithm, which further affects the quality of the recommendation algorithm. MAE measures the average deviation between the predicted relative difficulty coefficient of knowledge points and the actual relative difficulty coefficient of knowledge points. The lower the value of MAE is, the higher the prediction accuracy is. The calculation formula of MAE is shown in (8).

$$\mathrm{MAE} = \frac{1}{e}\sum_{i=1}^{e}\left|fkp_{v,j} - fkp_{v,i}\right| \tag{8}$$

where e represents the number of knowledge points in the test set, $fkp_{v,j}$ represents the predicted relative difficulty coefficient of the knowledge point of the learner v, and $fkp_{v,i}$ represents the actual relative difficulty coefficient of the knowledge point of the learner v.

As shown in Figure 1, by adjusting the size of neighbors, we compare the performance of the proposed algorithm and the comparison algorithms in terms of prediction accuracy. As can be seen from Figure 1, the proposed algorithm has better prediction accuracy than the comparison algorithms for all sizes of neighbors.

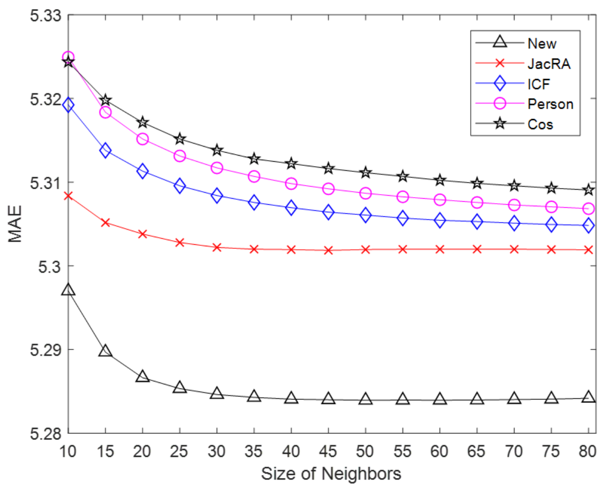

Figure 1. Influence of neighbor size on prediction accuracy

After the size of neighbors reaches 30, the improvement of prediction accuracy of the proposed algorithm is no longer significant. The downtrend of MAE tends to converge, which indicates that our proposed algorithm has better stability and can achieve better prediction accuracy of relative difficulty coefficient of knowledge points without too many neighboring learners, this is good for reducing the negative impact of sparse data on the recommendation effect. Therefore, the optimal size of the neighbors we chose in the following experiments is 30. At the same time, as the size of neighbors increases, the prediction accuracy of our proposed algorithm and comparison algorithms gradually decreases, indicating that the more the number of neighbors, the higher the prediction accuracy.

4.2.2. *Performance measure.* The task of the recommendation algorithm is to recommend useful knowledge points to learners, and the number of recommended knowledge points will directly affect the quality of the recommended algorithm. We select 9 different sizes of Top-N recommended knowledge points with N varying from 5 to 23 as a measure. In order to verify the performance of the proposed algorithm, we measure the performance of the recommendation algorithm with precision, recall and F1, and analyze how the effect of the recommendation changes with the number of recommended knowledge points N. The calculation formulas for precision P_v and recall R_v are shown in (9).

$$\begin{cases} \text{Precision} = \dfrac{|rec(v) \cap real(v)|}{|rec(v)|} \\ \text{Recall} = \dfrac{|rec(v) \cap real(v)|}{|real(v)|} \end{cases} \tag{9}$$

where $rec(v)$ represents the list of knowledge points recommended to the learner v, and $real(v)$ represents the list of knowledge points actually learned by the learner v.

Figure 2 shows the performance of the five recommendation algorithms in terms of precision and recall for different numbers of recommended knowledge points. As shown in Figure 2, we constantly adjust the number of recommended knowledge points N and compare the performance of the five algorithms in precision and recall.

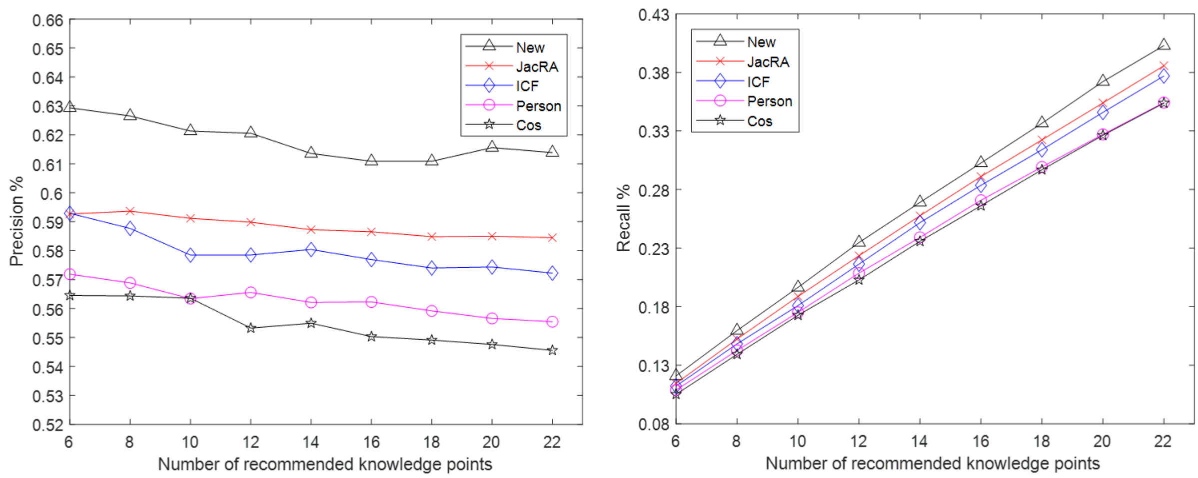

FIGURE 2. Performance of the five recommendation algorithms in precision and recall

It can be seen from Figure 2 that our proposed recommendation algorithm performs better than the other four recommendation algorithms in terms of precision and recall for any number of recommended knowledge points. We analyze the reasons: the Cos algorithm ignores the specific size of learners' relative difficulty coefficient of knowledge points, while the Person algorithm does not consider the difference of learners' relative difficulty coefficient on the same knowledge point, as for the ICF algorithm, it ignores the knowledge points learned by learners respectively, while the JacRA algorithm does not integrate the evaluation standard of learners' relative difficulty coefficient of knowledge points. The dimensions considered by our proposed algorithm are more comprehensive, which shows that the similarity calculation model that we proposed has better advantages than the model of the comparison algorithms.

In addition, it can be seen from Figure 2 that as the number of recommended knowledge points increases, the overall trend of the precision of the proposed algorithm and the comparison algorithms will decrease, and the recall will increase. The reason for the opposite trend of recall and precision is that in personalized recommendation, recall and precision are often contrary to each other, and the improvement of precision is usually at the expense of recall. Therefore, we consider both the precision and the recall through the F1 value, so as to evaluate the performance of the algorithms more comprehensively.

F1 is the weighted harmonic average of precision and recall, which is convenient to comprehensively compare the precision and recall of the algorithm. The calculation formula of F1 is shown in (10).

$$\mathrm{F1} = \frac{2 \times P_v \times R_v}{P_v + R_v} \tag{10}$$

As can be seen from Figure 3, our proposed algorithm outperforms the other four algorithms in terms of F1 for all numbers of recommended knowledge points. In addition, it can be observed that as the number of recommended knowledge points increases, the F1 of the five algorithms also gradually increases indicating that the more the number of recommended knowledge points, the higher the F1 value. It further confirms the feasibility of the proposed algorithm and improves the quality of the recommendation algorithm.

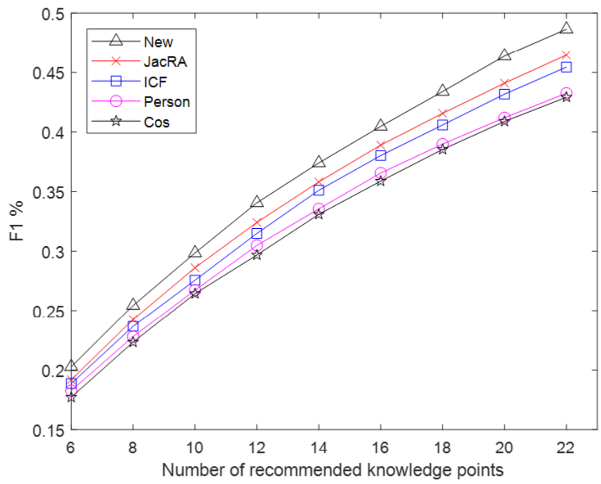

Figure 3. F1 measure of the five algorithms against different recommended number of knowledge points

From the experimental results, we can see that the overall performance and effect of our proposed algorithm are better than that of the comparative algorithms, and the evaluation index of the experiment has been greatly improved.

5. **Conclusions.** Aiming at the limitation of traditional collaborative filtering on similarity calculation, we propose a knowledge point recommendation algorithm based on similarity optimization. It fully considers the differences in learners' relative difficulty coefficient of knowledge points in different dimensions, and at the same time incorporates information about knowledge points learned by non-associated learners, improving the performance and quality of the recommendation algorithm. The future work will focus

on artificial intelligence and optimize the performance indicators of the recommended algorithm with the combination of relevant knowledge of deep learning.

Acknowledgments. This work was supported by the National Natural Science Foundation of China (61967005, 61662013, and U1501252), Innovation Project of GUET Graduate Education (2020YCXS022), the Key Laboratory of Cognitive Radio and Information Processing Ministry of Education (CRKL190107), Natural Science Foundation of Guangxi province (2016GXNSFAA380149), the Key Lab of Trusted Software (kx201511).

REFERENCES

[1] H.-J. Niu and C.-J. Chang, A novel method guiding IC manufacturing R&D direction: Perspective from knowledge integration innovation, *International Journal of Innovative Computing, Information and Control*, vol.14, no.4, pp.1371-1388, 2018.

[2] J. K. Tarus, Z. Niu and D. Kalui, A hybrid recommender system for e-learning based on context awareness and sequential pattern mining, *Soft Computing*, vol.22, no.8, pp.2449-2461, 2018

[3] H. Zhu, F. Tian, K. Wu et al., A multi-constraint learning path recommendation algorithm based on knowledge map, *Knowledge-Based Systems*, vol.143, pp.102-114, 2018.

[4] L. Wang, Z. Zeng, R. Li et al., Cross-domain personalized learning resources recommendation method, *Mathematical Problems in Engineering*, pp.1-6, 2013.

[5] X. Ma and L. Ye, Career goal-based e-learning recommendation using enhanced collaborative filtering and PrefixSpan, *International Journal of Mobile and Blended Learning*, vol.10, no.3, pp.23-37, 2018.

[6] X. Liang, *Recommended System Practice*, People's Post and Telecommunications Publishers, Beijing, 2012.

[7] W. Wang and J. Zheng, Improved collaborative filtering algorithm based on user-similarity, *Journal of East China Normal University (Natural Science)*, no.3, pp.60-66, 2016.

[8] X. Xiang and Z. Qiu, Research on collaborative filtering algorithm based on slope-one algorithm to improve score matrix filling, *Application Research of Computers*, vol.36, no.4, pp.1064-1067, 2019.

[9] L. Zhang, W. Duan and G. Ye, Collaborative filtering algorithm based on improved multi-segmented PCC, *Computer Engineering and Applications*, vol.54, no.9, pp.116-120, 2018.

[10] S. Gao, W. Liu, Y. Cui et al., Collaborative filtering integrating multiple user behaviors, *Computer Science*, vol.43, no.9, pp.227-231, 2016.

[11] H. Zhu, Y. Liu, F. Tian et al., A cross-curriculum video recommendation algorithm based on a video-associated knowledge map, *IEEE Access*, vol.6, pp.57562-57571, 2018.

[12] X. Wu, Y. Huang and S. Wang, A new similarity computation method in collaborative filtering based recommendation system, *2017 IEEE the 86th Vehicular Technology Conference (VTC-Fall)*, pp.1-5, 2017.