Personalized Course Recommendation System Fusing with Knowledge Graph and Collaborative Filtering

Gongwen Xu, Guangyu Jia, Lin Shi, and Zhijun Zhang

1School of Business, Shandong Jianzhu University, Jinan 250101, China
2School of Computer Science and Technology, Shandong Jianzhu University, Jinan 250101, China

Correspondence should be addressed to Gongwen Xu; xugongwen@sdjzu.edu.cn

Received 31 August 2021; Accepted 16 September 2021; Published 26 September 2021

1. Introduction

Recently, with the integration of Internet and education, various online education platforms have emerged and developed rapidly. These platforms have accumulated many users with their high-quality and massive resources. Especially affected by the COVID-19, the online education all over the world has developed rapidly. The daily active users of Coursera, edX, XuetangX, and other online education platforms have reached more than 10 million. The explosive growth of online education has changed the traditional teaching mode and made it a reality for people to learn anytime and anywhere. Online education has become an important educational model and technical way for learners to acquire knowledge and expand skills and academic education [1].

However, nowadays a wide range of learning directions and all kinds of courses have sprung up on the online education platform. The massive courses also bring confusion to learners and have some disadvantages for learners. Firstly, it is difficult for learners to find suitable courses in a short time; secondly, the learning route that learners are interested in is complex, while the courses in the school training program tend to be single and repetitive, so the online learners’ satisfaction is low and cannot achieve the expected teaching effect; thirdly, most online education platforms do not provide the same effective learning guidance and curriculum planning for learners as the traditional education. Learners are usually short of an in-depth understanding of the overall knowledge structure, while the number of Internet learning resources is miscellaneous, so learners fall into a large number of course choices, resulting in information overload and even the low course passing rate [2, 3].

How to provide personalized content for learners in massive courses resources is a problem worthy of research. Mining the interests of learners on the online education platform would help to better understand the needs of learners and help the platform provide personalized teaching services for learners. As the solution to this problem, the traditional interest mining methods either take the courses that learners are interested in as the interest point, or take the topic of the discussion area as the interest point. So many
2. Overview of the Course

2.1. The Course Recommendation Algorithm

Course personalized recommendation technology is one of the research hotspots in the field of online education and big data education. Many research teams all over the world have put forward the concept of personalized online education and studied personalized course recommendation algorithms in order to reduce the dropout rate of online learning, stimulate and mobilize learners’ enthusiasm for active learning, and give full play to the learning personality of different learners.

Bhaskaran and Santhi mined learners’ browsing logs according to learners’ behaviour and preferences and used the hybrid recommendation strategy proposed by ApriorAll algorithm to realize personalized recommendation for learners [8]. Based on the traditional data mining methods, Obeidat et al. [9] and others used collaborative filtering method and comparison of association rules to recommend courses for learners. The experimental results showed that grouping and clustering learners can significantly improve the recommendation effect. Huang et al. used reinforcement learning method and Markov decision process [10] to recommend exercises for learners. When recommending, they comprehensively considered the smoothness of exercise difficulty, review and preview, and the degree of learners’ participation. Liu et al. proposed using the method based on neural network [11] to track learners’ knowledge level, so as to provide learners with personalized learning path recommendation. The MCRs proposed by Zhang et al. are based on the distributed association rule mining algorithm [12], which makes the recommendation information transmitted more timely and improves the user’s course retrieval efficiency.

Chen et al. carried out a knowledge recommendation algorithm based on learners’ existing knowledge and learning materials [13], which modelled the recommendation process as a Markov decision problem. Aguilar et al. analysed the differences and similarities between the course selection system and the e-commerce platform, improved knowledge, and discovered association rules through biological heuristic algorithms, while most of the research in recent years were to explore the behaviour characteristics of learners to represent the characteristics of learners, so as to produce recommendation results [14].

Pang et al. extracted the behaviour characteristics of each learner and converted them into vectors of the same dimension, and dispersed them among similar users, so these learners would have more same courses [15]. They proposed multilayer bucket recommendation and MapReduce technology extension, so as to increase the satisfaction of the traditional collaborative recommendation algorithm.

To sum up, there have been a lot of works in learner personalized course recommendation recently years.

2.2. The Problems in the Traditional Course Recommendation Algorithm

At present, the existing research works mainly focused on using collaborative recommendation or data mining methods to improve the accuracy of course recommendation, but there are some problems in these methods, such as the cold start of recommendation algorithm, building a recommendation model without good performance, and carrying out an algorithm which cannot accurately recommend without initial data [16]. These problems are described as follows:

1. Accuracy of course recommendation: the recommendation algorithm using highly accurate rate can provide learners with more suitable courses and enhance their satisfaction. The accuracy of recommendation results is an important index of recommendation algorithm. If the recommendation result generated by a recommendation algorithm does not bring out good recommendation results, the recommendation results will be meaningless in some certain [17].

2. The problem of data sparsity: when using the traditional collaborative filtering method, the similarity calculation mainly depends on the learners’ rating matrix of the course. In the actual process, the number of the course rating score is very small; in some cases, the scoring users account for only 2% of the total users. It makes the course scoring matrix so sparse, and it is difficult for courses and learners to find similar neighbors, resulting in the low quality of courses recommendation [18].

3. Cold start problem: now the course platform would constantly update the content under the needs of new learners and new courses, but there is no record of new learners or new course content before updating the course platform, resulting in the failure of the recommendation algorithm to make timely and effective recommendations [19].
3. Related Research

3.1. Recommendation Based on Collaborative Filtering. Collaborative filtering algorithm is widely used in recommendation field. Traditional recommendation algorithms ignore the characteristic data of users or items in advance and only rely on the historical behaviour data of users to build models and recommend items to users. In most collaborative filtering algorithms, nearest neighbor technology is used to calculate the distance between users with the help of historical preference information [20]. The algorithm uses the weighted score of the nearest neighbor of the target user to predict the target user’s preference for specific courses, but it often faces problems such as sparse data and mining interpretation of recommendation results.

Collaborative filtering has been successfully applied in different researching fields [21]. The collaborative filtering recommendation system is based on the idea that similar users may have similar preferences or similar items may be preferred by similar users [22]. Collaborative filtering is generally divided into two categories:

(1) User-based algorithm [23]: according to the users’ behaviour record, it would find the user set which is similar in the interest and preference of the target user, then find the items liked by each user in this user set, filter out the items that the target user has already selected, and finally recommend the remaining items to the target user. In this method, the user similarity matrix calculated by the similarity formula is used to predict the target user’s score on the items liked by similar users.

(2) Item-based algorithm [24]: according to the users’ behaviour records, it calculates the similarity between items and filters out the item set with high score among the item sets generated by the target user. Then, according to the item similarity matrix, the other items, most similar to each item in the collection, would be found out, sorting and filtering the items that the target user has selected at the same time, and then finally the rest items are recommended to the target user.

The collaborative filtering algorithm uses the weighted score of the nearest neighbor of the target user to predict the target user’s preference for specific courses, but sometimes it would face the problems of sparse data and unexplained recommendation results.

3.2. Recommendation Method Based on Knowledge Graph. Google put forward the concept of knowledge graph, which uses the “entity-relation-entity” 3-tuple to describe the semantic relationship between different entities in the real world and forms a network knowledge structure through the relationship [25]. Representation learning aims to transform the objects into the low-dimensional space [26], and the aim of knowledge graph representation learning is to map the entities and relationships in the knowledge graph. The resulting vector can effectively represent the semantic relationship between entities and relationships [27].

Recently, the representation learning methods of deep learning have attracted extensive attention [28–34]. The common models of knowledge graph representation learning include distance model, energy model, matrix decomposition model, bilinear model, translation model, and so on [35]. The translation model represented by Ebisu and Ichise [36] has fewer parameters and is more popular in the research fields. For each 3-tuple \((h, r, t)\), where \(h\) and \(t\) represent head entities and tail entities, respectively, \(r\) is the relationship between them, TransE represents \(h, r, t\) as the embedded vectors \(v_h\), \(v_r\), and \(v_t\), respectively, and \(v_h\) is the translation from vectors \(v_h\) to \(v_t\), and the relationship among the three elements is expressed as follows:

\[
v_h + v_r \approx v_t. \tag{1}\]

In the TransE model, it aims to train vectors \(v_h\), \(v_r\), and \(v_t\) effectively to make formula (1) infinitely close to be equal. The smaller the error between them, the more likely there is a relationship \(r\) between the head and tail. Therefore, the loss function is shown as follows:

\[
f(v_h, v_r, v_t) = \|v_h + v_r - v_t\|^2_2, \tag{2}\]

where \(\| \cdot \|_2\) is the 2-norm of the vector, that is, the Euclidean distance. The total cost function for all 3-tuple samples is expressed as follows:

\[
L = \sum_{(h,r,t) \in S} \sum_{(h',r',t') \in S'} \max(0, f(v_h, v_r, v_t) - f(v_{h'}, v_{r'}, v_{t'}) + \delta), \tag{3}\]

where \(S\) is positive sample and is the set of all 3-tuple in the knowledge graph; \(S'\) is the negative samples of \(S\), that is, each existing 3-tuple in \(S\) is randomly replaced with its head entity or tail entity to obtain a new 3-tuple, and the 3-tuple does not belong to \(S\); then the set of such 3-tuples is called negative sample; and \(\delta\) is the distance between \(S\) and \(S'\). In order to minimize the cost function \(L\) during training process, it is necessary to make the loss function \(f(v_h, v_r, v_t)\) of positive samples tend to 0 and the loss function \(f(v_{h'}, v_{r'}, v_{t'})\) of negative samples tend to infinity.

4. Recommendation Algorithm Fusing with Knowledge Graph and Collaborative Filtering Technology

4.1. The Algorithm Model. At present, collaborative filtering recommendation algorithms usually use users’ historical evaluation data without considering the semantic relationship between recommendation items. To solve this problem, based on previous studies, a courses resource recommendation algorithm fusing knowledge graph and collaborative filtering technology (FKGCF) was proposed. This algorithm directly integrates the semantic similarity of the recommended object into the similarity calculation of the collaborative filtering recommendation algorithm and makes up for the deficiency that this algorithm ignores the connotation characteristics of
the item itself from the semantic perspective. The algorithm model is shown in Figure 1.

4.2. Composition of Course Knowledge Graph. The interaction information between learners and courses in the existing data sets is sparse or even missing, which will lead to the cold start problem. This problem can be alleviated by introducing other information, namely, side information, into the recommendation algorithm. Side information introduced in common recommendations such as movies and commodities includes social networks, user/item attributes, multimedia, and contexts [4]. The side information used in this paper is the knowledge graph, which is represented by 3-tuple \((C_1, R, C_2)\). The knowledge graph of the course is \(G\):

\[
G = \{(c_1, r, c_2) | c_1, c_2 \in C, \quad r \in R\},
\]

where \(c_1, c_2 \in C, C\) is the set of all the courses, and \(r \in R\); there are 5 relations in \(R\): same.instructor, same.subject, hour.low, hour.mid, and hour.high. When a learner chooses some certain courses, these courses can be connected to other events in the knowledge graph, connected to many other nonitem entities, and then connected to other course entities from these nonitem entities. For example, if there are computer courses in the selected courses, it can be related to other computer courses belonging to the same subject. Then, a 3-tuple knowledge graph is built.

In the traditional collaborative filtering algorithm, the connection is established by other learners’ historical learning records and other interactive processes. The essence of the knowledge graph constructed by the FKGCF algorithm is to establish a connection between the courses which learners have interacted and the courses which they have not interacted. These connections are not obtained from the interaction history of other learners, but through non-item entities. The method of constructing knowledge graph provides additional information sources connected between courses and a more accurate calculation method of item similarity in the algorithm, so as to improve the accuracy of recommended courses [37].

4.3. The Semantic Similarity. The entities and relationships in the knowledge graph are embedded into a d-Dimensional semantic space by using the TransE algorithm, and the item semantic vector is expressed as follows:

\[
I_i = (E_{i1}, E_{i2}, \ldots, E_{id})^T,
\]

where \(I_i\) presents the semantic vector of item \(i\) and \(E_{ik}\) is the value of the \(k\)th dimension semantic vector, where \(1 < k < d\). The TransE algorithm trains the loss function based on Euclidean distance. The similarity of item semantics is also measured by this distance, and the calculation equation is as follows:

\[
d(I_i, I_j) = \sqrt{\sum_{k=1}^{d} (E_{ik} - E_{jk})^2}.
\]

In order to unify the value of distance within the range of \((0, 1)\), the following calculation method is carried out:

\[
sim_{sem}(i, j) = \frac{1}{1 + d(I_i, I_j)} \tag{7}
\]

The greater the value of \(\sim_{sem}(i, j)\) is, the more similar the semantics of items \(i\) and \(j\) is.

In the knowledge graph, these pieces of feature information form a “entity-relationship-entity” 3-tuple. The more similar the semantic information is, the closer the course vector is, and the more similar the users’ love for it. Using these rich semantic data, the cold start problem of recommendation system is solved to a certain extent.

4.4. Fusing Similarity. After the item semantic similarity \(sim_{sem}\) was obtained by knowledge graph representation algorithm, the similarity in the collaborative filtering recommendation algorithm was calculated, as shown in

\[
sim_{CF}(i, j) = \cos(S_i, S_j) = \frac{S_i \cdot S_j}{\|S_i\| \cdot \|S_j\|} = \frac{\sum_{k=1}^{m} r_{ki} \cdot r_{kj}}{\sqrt{\sum_{k=1}^{m} r_{ki}^2} \cdot \sqrt{\sum_{k=1}^{m} r_{kj}^2}} \tag{8}
\]

where \(\sim_{CF}(i, j)\) is the cosine similarity of collaborative filtering of items \(i\) and \(j\), and \(S_i\) and \(S_j\) are the user’s scoring vectors for items \(i\) and \(j\), respectively.

Next, we fuse two similarities, \(\sim_{CF}(i, j)\) and \(sim_{sem}(i, j)\), as shown in

\[
sim(i, j) = \alpha \cdot sim_{sem}(i, j) + (1 - \alpha) \cdot sim_{CF}(i, j), \tag{9}
\]

When \(\alpha = 1\), it is equivalent to only considering the collaborative filtering; when \(\alpha = 0\), it is equivalent to only considering the semantic similarity.
where \( \text{sim}(i, j) \) represents fusing similarity of item \( i \) and \( j \), \( \text{sim}_{\text{sem}}(i, j) \) is the items semantic similarity, \( \text{sim}_{\text{CF}}(i, j) \) is the collaborative filtering item similarity, and \( \alpha \) is the weighted factor, whose value range is \([0, 1]\). When \( \alpha \) is 0, the algorithm is based on items. When \( \alpha \) is 1, the algorithm is based on semantic. When \( \alpha \in (0, 1) \), it is the fusing recommendation algorithm. The collaborative filtering recommendation algorithm is to calculate the similarity of items indirectly from the perspective of users’ rating of items.

4.5. Rating Prediction and Generation of Recommendation List. According to the fusing similarity, the user’s score on an item that has not been evaluated before is predicted. The calculation equation of prediction score is as follows:

\[
P_{ui} = \frac{\sum_{j \in (N(u) \cap S(i,k))} \left( \text{sim}(i, j) \times r_{uj} \right)}{\sum_{j \in (N(u) \cap S(i,k))} \text{sim}(i, j)},
\]

where \( p_{ui} \) represents the prediction score that user \( u \) gives to item \( i \). \( N(u) \) refers to all item sets evaluated by user \( u \), \( S(i,k) \) refers to \( k \) adjacent items with the greatest similarity to item \( i \), \( N(u) \cap S(i,k) \) refers to the intersection of two item sets, and the intersection result is the reference for prediction score; \( \text{sim}(i, j) \) is the fusing similarity of items \( i \) and \( j \); \( r_{uj} \) is the user \( u \)’s rating of item \( j \).

The higher the prediction score is, the more interested the user is. The prediction score of all items is calculated by using equation (10). And then they are sorted according to the scoring results, giving priority to the top-\( N \) items recommended to the user.

5. Experimental Process

5.1. Experimental Environment and Data Set Processing. The experimental data comes from XuetangX. 1000 courses and 10000 members are collected for experiments. The precision, recall, and F1 value are used as evaluation indexes to measure the performance of each algorithm. The precision refers to the probability that the user is interested in the recommended course list, and recall refers to the probability that the course that the user is interested in appears in the recommendation list. F1 value is the harmonic average of precision and recall [38]. The calculating methods are shown as follows:

\[
\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}},
\]

\[
\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}},
\]

\[
F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}.
\]

In order to measure the performance of the algorithm more accurately, this paper uses \( k \)-cross verification and the \( k \) value is 5; that is, the experimental data are randomly grouped in 5 parts, one of which is selected as the testing set and the other 4 as the training set. A total of 5 tests are carried out, and the average result of 5 tests is used as the final evaluation of this algorithm.

5.2. Experimental Process and Result Analysis

5.2.1. Experiment of Different Embedding Dimensions. When knowledge graph representation learning is carried on, different embedding dimensions will have different performance. The experimental results are shown in Figure 2. It can be seen that the precision and recall change with the increasing of embedding dimensions. Therefore, when the embedding dimension is 150, the recommendation algorithm performs best.

5.2.2. Experiment of Different \( \alpha \) Values. \( \alpha \) represents the fusion factor of recommendation algorithm, its value range is \([0, 1]\). The experimental results are shown in Figure 3. It can be seen that when \( \alpha \) equals to 0.4, the precision and recall of the recommendation algorithm perform best.

5.2.3. Experiment of Different \( k \) Values. The \( k \) value represents the number of adjacent items with the greatest similarity to items when calculating similarity. The experimental results are shown in Figure 4. It can be seen that when \( k \) equals to 5, the precision and recall of the recommendation algorithm perform best.
results in the experiment. Seven dimensions, 50, 75, 100, 125, 150, 175, and 200, are selected for the experimental comparison. The similarity fusing weight factor \( \alpha \) is set to 0.6, and the maximum number of neighbors \( k \) of similarity is set to 30. The final experimental comparison results are shown in Figure 2.

From Figure 2, it can be found that there is no significant difference among the seven different embedding dimensions. But the precision and recall value are better when the dimension is 100.

5.2.2. Experiment of Different Fusing Weight Factor Ratio.
Fusing weight factor \( \alpha \) controls the proportion of semantic similarity and collaborative filtering similarity in the final fusing similarity, and it is a key factor of this algorithm. The value range of \( \alpha \) is [0, 1], and in the experiment \( \alpha \) is assigned from 0 to 1 with a step of 0.1, so there are 11 steps altogether.

The embedding dimension of knowledge graph presentation learning is set to 100, and the maximum number of neighbors \( k \) of similarity is set to 30. The results can be seen in Figure 3.

Experiments show that when the fusion reaches a certain proportion, the precision and recall of the proposed algorithm are higher than the single semantic content-based recommendation algorithm (\( \alpha = 1 \)) and the item-based algorithm (\( \alpha = 0 \)). When the fusion weight \( \alpha \) is 0.6, the precision and recall reach the best value.

5.2.3. Comparison of Experimental Results. To verify the performance of the modified collaborative filtering recommendation algorithm proposed in this paper, experimental comparisons are made with traditional algorithms, user-basedCF and item-basedCF. The embedding dimension of knowledge graph representation learning in this algorithm is
set to 100. The fusing weight $\alpha$ is set to 0.6 and the recommended number top-$N$ is set to 10. The experimental results are shown in Figure 4. The $N$ value refers to the number of neighbors in the nearest neighbor algorithm KNN. The figure illustrates that the precision of all comparative experiments increases with the increase of $n$ value while the recall decreases, and the F1 value increases first and then decreases.

Through experimental comparison, it is obvious that the proposed FKGCF algorithm has higher precision, recall, and F1 value than the traditional algorithms.

6. Conclusions

Aiming at the problem that the collaborative filtering recommendation algorithm only uses the users' historical evaluation data without considering the semantic relationship between the recommended objects, a modified FKGCF algorithm was proposed. The algorithm not only uses the users' evaluation information of the course, but also uses the internal semantic information of the item itself. The results show that this algorithm improves the precision and recall and can realize the courses recommendation more accurately and efficiently. However, the algorithm in this paper also has some limitations, which does not consider the problem of user interest drift, because the users' interest is likely to change with the development of time, and the previous historical data is time-effective. Therefore, according to the time dynamics of historical data, establishing relevant time series recommendation models is a direction worthy of further research.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported in part by the Shandong Education Department Teaching Reform Project (Z2016M014, Z2016M016, and Z2016Z013).

References


