IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 28 March 2024, accepted 9 April 2024, date of publication 18 April 2024, date of current version 21 May 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3390675

RESEARCH ARTICLE

Prediction of Students' Grade by Combining Educational Knowledge Graph and Collaborative Filtering

YIWEN ZHANG [1,2], VLADIMIR Y. MARIANO [1], AND REX P. BRINGULA [3]

[1]College of Computing and Information Technologies, National University, Manila 1008, Philippines
[2]Big Data and Artificial Intelligence College, Anhui Xinhua University, Hefei 230088, China
[3]College of Computer Studies and Systems, University of the East, Manila 1008, Philippines

Corresponding author: Yiwen Zhang (yiwenzh@ustc.edu.cn)

This work was supported in part by the Natural Science Foundation of Anhui Provincial Education Department under Grant KJ2020A0784, in part by Anhui Province Teaching and Research Project 2021jyxm0616, and in part by Students of Anhui Xinhua University Innovation Project AH202012216092.

ABSTRACT Traditional collaborative filtering-based grade prediction methods overly rely on students' historical grades and overlook the content correlation between courses, resulting in lower accuracy in predicting student grades. This paper proposes a grade prediction method that combines the educational domain knowledge graph with collaborative filtering, gathering course semantic information and constructing a course knowledge graph as auxiliary information for grade prediction. Through experimentation, it has been demonstrated that the integration of the educational knowledge graph and collaborative filtering in the grade prediction method uncovers more semantic relationships between courses, thereby improving the accuracy of predicting grades for related courses. The Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) metrics show a decrease when compared to collaborative filtering and K-means algorithms.The method in this paper allows for more personalized learning and recommendation in the knowledge-rich field of education with semantic richness.

INDEX TERMS Grade predicted, knowledge graph, collaborative filtering, recommended algorithm.

I. INTRODUCTION

In recent years, with the rapid growth of data and information across various industries, the focus of research has been on how to quickly find suitable information for users from a vast amount of data. To address the problem of information overload, recommendation algorithms have been widely applied in scenarios dealing with information overload. Recommendation systems have shown excellent performance in various fields, including telecommunications, finance, agriculture, e-commerce, medicine, and education [1], [2]. In the education sector, recommendation systems have been used to provide personalized teaching for students and teachers, making it a current research focus.

Among these efforts, predicting student grades is a key aspect of educational data mining. By using prediction methods to forecast student grades, teachers can identify factors that influence student performance promptly, reduce the risk of failing and dropping out, and adjust teaching plans and methods to implement personalized teaching. Additionally, grade prediction can also be applied to knowledge tracking [3], we can track students' mastery of learned knowledge points based on their learning time series. Student profiling [4], The author analyzes students' historical grades and related knowledge points to gain insights into their mastery of the knowledge points. Online assessments [5], In order to optimize online resources and provide better personalized resource services for learners, the content of online resources and learners' learning progress are analyzed. Resource recommendations [6] is to provide learners with high-quality personalized resource recommendations through data analysis and modeling. Therefore, grade prediction plays an important role in improving teaching quality and holds significant research significance and practical value.

The associate editor coordinating the review of this manuscript and approving it for publication was Fabrizio Messina.

© 2024 The Authors. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License.
For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

Currently, there has been substantial research on student grade prediction. Burman et al. [7] collected students' psychological characteristic data and classified student grades into three categories (high, medium, and low). They used multi-class Support Vector Machines (SVM) for grade prediction. Deho et al. [8] used causal tree methods to analyze the factors influencing student grades during online teaching at a certain Australian university in 2020 during the pandemic. Zhang et al. [9] predicted student grades by analyzing Massive Open Online Course(MOOC) learners' behavioral data using neural network algorithms and clustering algorithms. Ma et al. [10] used a multi-instance multi-label approach to predict students' performance in new courses, helping teachers grasp the learning progress of each student. Bujang et al. [11] proposed the Synthetic Minority Oversampling Technique(SMOTE) to address the issue of class imbalance and overfitting in student grade prediction. Xue Mengting [12] introduced a domain-based collaborative filtering method for predicting college student grades. In general, existing student grade prediction methods can be categorized into two main types: one that utilizes regression or classification methods for grade prediction and another that treats the prediction of student grades as a rating prediction problem in recommendation systems.

As student learning behavior exhibits time-series characteristics, and recommendation systems' prediction methods rely on users' historical behavioral data, using an analogy to recommendation systems [13], [14] to predict student grades yields higher prediction accuracy and stronger interpretability compared to other statistical methods. However, when using recommendation system-based approaches for grade prediction, there is an excessive reliance on students' historical grades, and the accuracy of grade prediction can be affected by courses with low relevance in the historical data.

Education data has more knowledge relations, and knowledge graph can be used to learn more knowledge relations and have better interpretability,knowledge graph is currently widely applied in the field of artificial intelligence, such as recommendation systems, information retrieval, natural language processing, etc. [15]. It is a large-scale semantic network that effectively integrates various information and adds it to a knowledge base, revealing the semantic information between entities and entity-property relationships. KGs are typically described and stored using triplets of "entity-relation-entity." Embedding models are commonly used to map the triplets in the knowledge graph to a low-dimensional vector space [16], where each vector represents the latent semantic relationship in the knowledge graph.

To address these issues, this paper proposes a grade prediction method that combines educational knowledge graph with collaborative filtering (KN-CF), which has the following contributions:

1) Extracting course knowledge entities and course relationships to construct a knowledge graph, mapping the knowledge graph to a low-dimensional vector space.
2) Calculate the similarity between course knowledge graph vectors.
3) Integrating collaborative filtering-based course similarity and course knowledge graph similarity for student grade prediction.

Experiments are conducted using a dataset collected from users and compared with course-based collaborative filtering (C-CF), course-based clustering (C-KM), and user-based collaborative filtering (U-CF).

II. RELATED WORK

The method proposed in this paper is inspired by the application of knowledge graphs in recommendation systems, generally speaking, the use of knowledge graphs in prediction and recommendation methods is to expand semantic information. For example, Geng [17] created behavior graphs and behavior paths based on the historical behavior sequences of learners, calculating the similarity between paths, using path-based collaborative filtering to predict user behavior and make recommendations. Wang et al. [18] Established a movie knowledge graph, utilizing the information in the knowledge graph for convolution to predict user interests and recommend movies. Knowledge Graph Convolutional Networks (KGCN) is an end-to-end framework that effectively captures the inter-item relatedness by mining their associated attributes on the knowledge graph, thereby alleviating data sparsity and cold-start issues. Due to the abundance of knowledge information in the field of education, some researchers have also introduced knowledge graphs into educational recommendations, such as, Lu et al. [19] created three types of bipartite interaction graphs: student-course graph, student-student graph, and course-course graph. Utilizing these three graph relationships to represent students and courses, calculating student similarities and course similarities, and conducting grade prediction. Wang et al. [20] established user-item knowledge interaction relationships and inputted these knowledge relationships into a Transformer model to predict ratings based on user reviews. Both of the above methods utilized the representation of bipartite graph models, but did not establish a knowledge graph with ternary relationships. Reference [21] combined collaborative filtering algorithms with reinforcement learning's reward and punishment mechanism, a time-series-based student rating optimization model was constructed. The rating results are used as a measure of students' preference for courses and for recommending courses to students. Xu et al. [22] By established a course-based knowledge graph and integrating the information from the knowledge graph into collaborative filtering recommendation algorithms, recommending resources to learners, the accuracy of collaborative filtering recommendations has been improved. The authors used knowledge graph embeddings as inputs for the similarity calculation in collaborative filtering, which improved recommendation effectiveness. However, they overlooked the importance of the collaborative filtering algorithm for learning from users' historical behaviors.

In summary, there is currently more research on course recommendation, while research on grade prediction mainly employs bipartite graph models and deep learning methods. Therefore, this paper proposes "Prediction of Students' Grade by Combining Educational Knowledge Graph and Collaborative Filtering" which utilizes the rich semantic relationships in knowledge graphs to learn user preferences, and simultaneously utilizes collaborative filtering to learn historical user preferences, train optimal parameters, linearly combine the two, and complement each other.

III. A GRADE PREDICTION METHOD COMBINING EDUCATIONAL KNOWLEDGE GRAPH AND COLLABORATIVE

A. COLLABORATIVE FILTERING RECOMMENDATION ALGORITHM

Collaborative Filtering Recommendation Algorithm includes User-Based Collaborative Filtering [23] and Item-Based Collaborative Filtering [24], which are widely used in various fields such as video, e-commerce websites, music, and education recommendations. The main idea of this algorithm is to calculate the similarity between users or items based on users' historical rating records for items. Then, it recommends items that users might be interested in or predicts users' ratings for future items. The implementation of this algorithm can be divided into three parts, as follows:

1) CONSTRUCTING THE RATING MATRIX

Constructing the user-item rating matrix is the foundation of collaborative filtering algorithms.Let $U = \{u_1, u_2, u_3, \cdots, u_m\}$ be the set of m users, $I = \{i_1, i_2, i_3, \cdots, i_n\}$ be the set of n items, and $R_{m\times n}$ represent the rating matrix for users' ratings on items, as shown in formula (1):

$$\mathrm{R_{m\times n}} \begin{bmatrix} \mathrm{R_{11}} & \mathrm{R_{12}} & \mathrm{R_{13}} & \cdots & \mathrm{R_{1n-1}} & \mathrm{R_{1n}} \\ \mathrm{R_{21}} & \mathrm{R_{22}} & \mathrm{R_{23}} & \cdots & \mathrm{R_{2n-1}} & \mathrm{R_{2n}} \\ \mathrm{R_{31}} & \mathrm{R_{32}} & \mathrm{R_{33}} & \cdots & \mathrm{R_{3n-1}} & \mathrm{R_{3n}} \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \mathrm{R_{m-11}} & \mathrm{R_{m-12}} & \mathrm{R_{m-13}} & & \mathrm{R_{m-1n-1}} & \mathrm{R_{m-1n}} \\ \mathrm{R_{m1}} & \mathrm{R_{m2}} & \mathrm{R_{m3}} & & \mathrm{R_{mn-1}} & \mathrm{R_{mn}} \end{bmatrix} \tag{1}$$

In formula (1), the rating of user u_i on item I_j is represented by R_{ij}, where i takes values in the range of [1, m] and j takes values in the range of [1, n]. A higher rating value indicates a higher level of preference that user u_i has for item I_j.

2) CALCULATING ITEM SIMILARITY

If two items have a higher similarity, it indicates that the user has a higher interest in similar items that similar to the historical items can be recommended to the target user. Usually, cosine [25], Jaccard, or Pearson algorithms are used to calculate the similarity between items. Taking cosine similarity as an example, considering all users' ratings for a certain item as an m×1 one-dimensional vector, the rating vector of item i is represented as $F_i = \{r_{1i}, r_{2i}, r_{3i}, \cdots, r_{mi}\}$, and the rating vector of item j is represented as $F_i = \{r_{1j}, r_{2j}, r_{3j}, \cdots, r_{mj}\}$. Then, calculate the similarity between item i and item j is shown in formula (2):

$$A(i, j) = \frac{F_i \cdot F_j}{\|F_i\| \cdot \|F_j\|} = \frac{\sum_{u=1}^{m} r_{u_i} \cdot r_{uj}}{\sqrt{\sum_{u=1}^{m} r_{u_i}^2} \cdot \sqrt{\sum_{u=1}^{m} r_{u_j}^2}} \tag{2}$$

In formula (2), $A(i, j)$ represents the cosine similarity value between item i and item j with a range of $[-1, 1]$.

3) SELECTING NEIGHBORS

There are typically two methods for selecting neighbors. The first method involves sorting the computed item similarity values in descending order and selecting the top N items as neighbors, also known as TOP-N. The second method involves directly setting a threshold value Y and selecting all items with similarity values greater than Y as neighbor recommendations. In this paper, the TOP-N method is used to select neighbors.

B. KNOWLEDGE GRAPH AND METHOD OF REPRESENTION

The commonly used knowledge graph embedding models include matrix factorization model, bilinear model, distance-based model, single-layer neural network model, and translation model, etc [26]. Among them, the translation model, characterized by simple parameters and low algorithm complexity, is widely applied in knowledge graphs, for example, the TransE model [27]. For each given triplet (h, r, t) in the set S, where h represents the head entity, t represents the tail entity, and r represents the relationship between entities, we use E to represent the set of entities h, t, i.e., h, $t \in E$, and R to represent the set of entity relationships, i.e., r $\in R$. The TransE model represents the embedding vectors of head entity h and tail entity t as $\bar{\mathrm{h}}, \bar{\mathrm{t}}$, respectively, and the embedding vector of relationship entity r as $\bar{\mathrm{r}}$, where $\bar{\mathrm{r}}$ is also known as the translation between vectors $\bar{\mathrm{h}}$ and $\bar{\mathrm{t}}$. Through continuous training and adjustment, the model aims to make $\bar{\mathrm{h}} + \bar{\mathrm{r}}$ infinitely close to $\bar{\mathrm{t}}$, i.e., $\bar{\mathrm{h}} + \bar{r} \approx \bar{t}$. The TransE model is illustrated in Figure 1.

The loss function of the TransE model is represented by formula (3):

$$d(\bar{h}, \bar{t}, \bar{r}) = \left\|\bar{h} + \bar{r} - \bar{t}\right\|_2^2 \tag{3}$$

In the formula(3),d represents the loss function, which is also the Euclidean distance, and H denotes the L2 norm of the vectors. For all triplets in the set S, the loss function can be defined as shown in formula (4):

$$D = \sum_{(h,r,t)\in S} \sum_{(h',r,t')\in S'} \times max(0, d(\bar{h}, \bar{r}, \bar{t}) - d(\bar{h}', \bar{r}, \bar{t}') + \gamma) \tag{4}$$

In the formula(4), S represents the set of positive triplets, S' denotes the set of negative triplets obtained by randomly

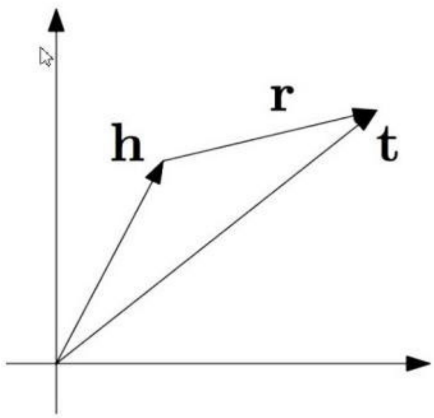

FIGURE 1. TransE Model.

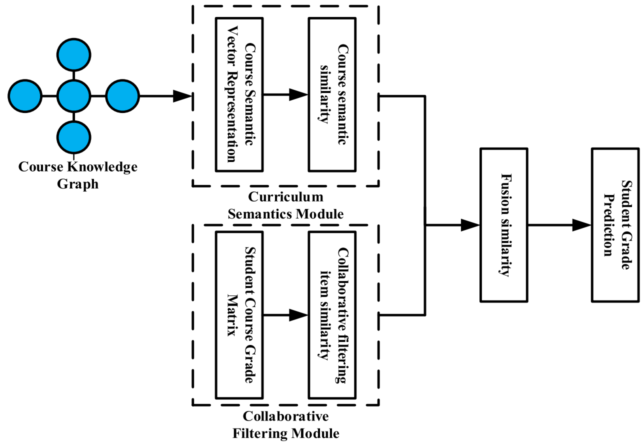

FIGURE 2. Collaborative filtering recommendation algorithm integrated with knowledge graph.

replacing the head or tail entity of positive triplets, and γ represents the distance between positive and negative triplets. During the training process of the TransE model, minimizing the value of L will make the distance $d(\overline{h}, \overline{r}, \overline{t})$ for positive samples approach 0 infinitely, and the distance $d(\overline{h}', \overline{r}, \overline{t}')$ for negative samples approach ∞.

C. TEXTRANK

TextRank is an unsupervised graph-based algorithm that can be used not only for text keyword extraction [28] but also for text summarization generation [29].First, the TextRank algorithm is used to divide the text into several parts, where each part represents a vertex, i.e., a text unit, containing words and sentences. Next, the similarity between each pair of vertices is computed, and the similarity values are assigned to the related vertex pairs. This process is iteratively performed to calculate the similarity scores between the vertices. Finally, the computed scores are stored in a matrix, and the vertex values are sorted in descending order. The keywords are then extracted from the vertices with higher values.

In this paper, the TextRank algorithm is employed for text keyword extraction, which not only takes into account the contextual context of keyword occurrences but also better reflects the semantic information of the text. As a result, the extracted keywords are more specific and highly accurate. Traditional collaborative filtering algorithms only consider the historical rating information of users on items, without considering the relationship between item semantics, and cannot deeply explore the correlation between items.This leads to a limitation in effectively uncovering the correlations among items. In this paper, we propose a method for predicting academic performance by integrating educational knowledge graphs with collaborative filtering. We incorporate the semantic content of course knowledge graphs into the collaborative filtering recommendation algorithm, addressing the deficiency of semantic information in the traditional collaborative filtering approach. This integration enhances the accuracy of performance prediction. The algorithmic process is illustrated in Figure 2.

D. SEMANTIC SIMILARITY OF COURSE KNOWLEDGE GRAPH

The entities and relationships of the course knowledge graph are embedded into a P-dimensional semantic space using the TransE model, resulting in a semantic vector representation denoted as $C_i = (c_{1i}, c_{2i}, \ldots, c_{ni}, c_{pi})^T$,In the equation, C_i represents the semantic vector of course i and $1 \leq n \leq p$ denotes the semantic value of course i on the n-th dimension. Since the TransE model is trained using the Euclidean distance to compute errors, the semantic similarity value between two courses is also calculated using the Euclidean distance. The calculation method is illustrated in Formula (5):

$$L(C_i, C_j) = \sqrt{\sum\nolimits_{k=1}^{p} \left(c_{ni} - c_{nj}\right)^2} \tag{5}$$

where L represents the semantic distance between any two courses C_i and C_j. In order to avoid singular values in the calculation results, the distances are normalized using the specific formula shown in Equation (6):

$$G(C_i, C_j) = \frac{1}{1 + L(C_i, C_j)} \tag{6}$$

In the equation, the range of values for $G(C_i, C_j)$ is (0,1]. When the value of G approaches 1, it indicates that the semantic similarity between courses C_i and C_j is higher.

E. FUSION SIMILARITY

The course-based collaborative filtering algorithm calculates course similarity based on a matrix of students' historical course grades. The accuracy of calculating course similarity is influenced by subjective factors. For instance, different instructors assessing the same questions differently can affect course grades and subsequently impact the accuracy of course similarity. Moreover, the algorithm's accuracy is also affected by data sparsity. High data sparsity, characterized by a lack of historical student-course grade data, can result in lower prediction accuracy.On the other hand, semantic similarity between courses is calculated from the perspective of course features. The more comprehensive the course semantic features are, the more accurate the computation of course similarity becomes. This, in turn, leads to higher accuracy in predicting student grades.

The course semantic similarity obtained from the course knowledge graph and the similarity calculated using the course-based collaborative filtering algorithm are combined through a linear weighted fusion to derive the final similarity. The fusion process is represented by Formula (7):

$$W(C_i, C_j) = \beta \cdot G(C_i, C_j) + (1 - \beta) \cdot A(i, j) \tag{7}$$

In the Formula(7), $W(C_i, C_j)$ represents the fused similarity between courses C_i and C_j, $G(C_i, C_j)$ signifies the semantic similarity based on the course knowledge graph, and $A(i, j)$ denotes the similarity calculated through the course-based collaborative filtering.β is the weighting factor, with a range of values between [0,1]. When β lies within the range of (0,1), it represents the fused similarity. When β is equal to 1, it corresponds to the similarity derived from the course knowledge graph. When β is equal to 0, it corresponds to the similarity obtained from collaborative filtering.

F. COURSE GRADE PREDICTION

Predicting the course a certain student is about to study, based on Formula (7), the calculation formula for grade prediction is given by Formula (8):

$$Z_{u_i} = \frac{\sum_{j \in Y(i,N)} W(C_i, C_j) \cdot r_{uj}}{W(C_i, C_j)} \tag{8}$$

In the equation, Z_{u_i} represents the predicted course grade for student u in course i. $Y(i, N)$ signifies the N courses most similar to course i. r_{u_j} denotes the historical grades of student u in their past courses j. $W(C_i, C_j)$ stands for the fused similarity.

A parenthetical statement at the end of a sentence is punctuated outside of the closing parenthesis (like this). (A parenthetical sentence is punctuated within the parentheses.) In American English, periods and commas are within quotation marks, like "this period." Other punctuation is "outside"! Avoid contractions; for example, write "do not" instead of "don't." The serial comma is preferred: "A, B, and C" instead of "A, B and C."

IV. EXPERIMENTAL METHODOLOGY

The experimental design for this study consists of two main parts. The first part involves the integration of educational knowledge graphs and collaborative filtering for grade prediction (referred to as KN-CF). The second part encompasses comparative experiments, including: course-based collaborative filtering prediction method (referred to as C-CF), course clustering-based prediction method (referred to as C-KM), and user-based collaborative filtering prediction method (referred to as U-CF). The experimental steps are outlined as follows:

Step 1: Obtain online and offline course content and store it in TXT format, and organize student course grade data.

TABLE 1. Grade level representation.

Original Data(grade)	Transformed(grade)
$90 \le grade \le 100$; Excellent	5
$80 \le grade < 90$; Good	4
$70 \le grade < 80$; Average	3
$60 \le grade < 70$; Pass	2
$0 \le grade < 60$; Fail	1

Step 2: Extract entities and relationships from the knowledge graph, and construct the course knowledge graph.

Step 3: Calculate semantic similarity within the course knowledge graph.

Step 4: Compute similarity using course-based collaborative filtering.

Step 5: Fuse semantic similarity from the course knowledge graph and course-based collaborative filtering similarity (KN-CF).

Step 6: Utilize the fused similarity for predicting student grades.

Step 7: Compare KN-CF with C-CF, C-KM, and U-CF methods.

A. DATASET AND PREPROCESSING

The experimental data for this study was sourced from the School of Big Data and Artificial Intelligence, Software Engineering program, Anhui Xinhua University. The dataset comprises two main components: student grades and course semantic information. The course semantic information includes details about MOOC [30] courses and attributes of course materials,such as textbook name, author, publisher, publication year.

1) PROCESSING STUDENT GRADE DATA

These records contain the course scores of software engineering students enrolled in 2015, 2016 and 2017 cohorts over a period of four years, spanning 8 semesters. However, the 8th semester includes internships and thesis defense without examinable courses. Therefore, only grades from the 1st to 7th semesters are selected for the experimental dataset. Furthermore, various subjects adhere to distinct assessment standards, with some employing percentage grading and others using a letter-grade system. In order to standardize data types, a normalization process will be performed to convert scores into a range of 1 to 5 at our school, as shown in Table 1.

The processed dataset consists of a total of 14,898 grade records.

FIGURE 3. The knowledge graph of python programming course.

2) PROCESSING COURSE SEMANTIC DATA

The course semantic information primarily involves the collection of course knowledge point data, including course names, course descriptions, and course outlines. Some knowledge points come from offline teaching material information, but since the content of the offline teaching material is less or not comprehensive enough, more online information related to the course is crawled by web crawler to supplement the knowledge content.This knowledge acquisition is divided into two parts: offline course knowledge data and online course knowledge data. For offline course knowledge data, information is extracted from physical textbooks and supplementary materials. Due to the limitations or singular nature of offline resources, web scraping techniques are employed to extract course textbook outline information from the internet, serving as a complement to online course knowledge data.

B. CONSTRUCTION OF KNOWLEDGE GRAPH

In this paper, knowledge units are extracted from online and offline educational and teaching resources of different courses to form a course knowledge graph. Choose Python language programming design course and build knowledge graph as shown in Figure 3:

The steps to construct the knowledge graph are as follows:

Step 1: Utilizing the TextRank algorithm, extract knowledge points from all corresponding online and offline content for each course, representing them as entities in the knowledge graph. Given the potentially high number of entities in the course knowledge graph, to mitigate construction costs, only the top 20 most significant entities are extracted from each course. The extraction results are depicted in Table 2.

Step 2: Define relationships between entities in the knowledge graph and create triples, as illustrated in Table 3.

Table 3 shows the 8 triplet relationships established in this experiment, and the knowledge graph is created according to these 8 triplet relationships.

Based on the aforementioned rules, the dataset used in this study eventually forms 15,957 triples in the course knowledge graph. Among these triples, 80% are allocated for the training set, 10% for the validation set, and the remaining 10% for the test set.

C. EVALUATION METRICS

In this experiment, the commonly used evaluation metrics in recommendation systems,Evaluation methods in recommendation systems are divided into two types: TOP-N

TABLE 2. Extraction of course entities.

Course name	Entity
.NET Programming	Control Form Statement File Data Type Menu Array Instance Dialog Box Control Structure
C++language programming	Function Array Pointer C++ Programming Operator Statement Structure Definition Constructor
J2EE framework	Statement Implementation Data Type Object Mapping Form Application Form Dynamic Configuration
Information Security	network security computer virus encryption encryption algorithm Computer technology algorithm authentication cryptography information security

recommendation and rating prediction. TOP-N recommendation mainly measures the hit rate, commonly using metrics such as F1, recall, precision [18], [31], [32], etc. Rating prediction mainly measures the error range and commonly uses the Root Mean Square Error ($RMSE$), (Mean Absolute Error (MAE) [4], [33], are employed as evaluation criteria. These metrics have a range of $[0, +\infty]$, with values approaching 0 indicating a closer alignment between actual and predicted values, thereby making $RMSE$ and MAE closer to 0. Conversely, larger values indicate greater errors, approaching $+\infty$. The calculation formulas are illustrated in Equations (9) and (10).

$$RMSE = \sqrt{\frac{\sum_{i=1}^{x}(P_i - \hat{P}_i)^2}{x}} \tag{9}$$

In the equation:$RMSE$ represents the Root Mean Square Error (RMSE) of predicted student grades.i denotes the current prediction of course grade.P_i represents the predicted grade for the i-th course.$\hat{P}_i$ signifies the actual grade for the i-th course.x represents the total number of courses.

The specific formula for MAE, as shown in Equation (10), is as follows:

$$MAE = \frac{\sum_{i=1}^{x}\left|P_i - \hat{P}_i\right|}{x} \tag{10}$$

In the equation, MAE represents the Mean Absolute Error of predicted student grades.i denotes the current prediction of course grade.P_i represents the predicted grade for the i-th course.$\hat{P}_i$ signifies the actual grade for the i-th course.x represents the total number of courses.

The main contributions of this paper are summarized in Table 4.

V. EXPERIMENTAL RESULTS ANALYSIS

A. EXPERIMENTAL PROCEDURE

First, the construction of the course knowledge graph is performed according to the methods described in Section III-B. Then, the similarity of the knowledge graph is calculated using the Euclidean formula as shown in equation (3). Finally, the results of the knowledge graph and collaborative filtering similarity are combined, and the recommendation list is computed.

B. PARAMETER TRAINING

In accordance with Equation (7), for the fusion experiment, the weight of β is varied within the range of 0 to 1 with a step size of 0.1. The average values of the evaluation metrics for semesters 1 to 7 for $RMSE$and MAE are computed. The experimental results for different weight values of β are depicted in Figures 3 and 4.

From Figures 3 and 4, it can be observed that when β is set to 0.4, the values of $RMSE$ and MAE for the grade prediction method that combines educational knowledge graph and collaborative filtering are minimized. Therefore, in this experiment, the value of β is chosen as 0.4.

TABLE 3. Construction of course triplets.

Entity	Relation	Entity
Student	Course Name	Score
course name	Include	Keyword Entity
course name	Textbook	Textbook Name
course name	Author	Author Name
course name	publishing house	Concrete publisher
course name	Time	Specific publication time
course name	Course type	Elective
course name	Belong	Specific Category

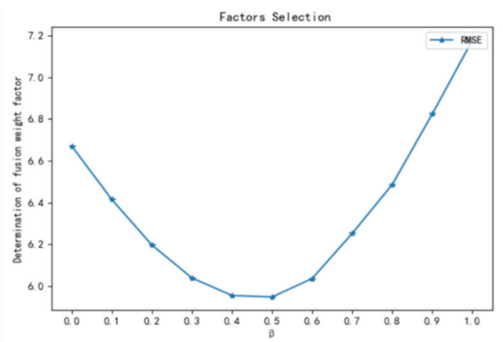

FIGURE 4. RMSE results under different weight values.

C. EXPERIMENTAL RESULTS AND ANALYSIS

The experiments are conducted in four groups, comparing course-based collaborative filtering (C-CF), course clustering-based (C-KM), user-based collaborative filtering (U-CF), and the proposed grade prediction method that combines educational knowledge graph and collaborative filtering (KN-CF).

1) C-CF

This is a course-based collaborative filtering prediction method. This method calculates the similarity between courses based on a user's historical courses and ratings, and predicts grades.

2) C-KM

Rating prediction method based on course clustering. This method establishes a course knowledge graph, clusters courses, and predicts grades based on the similarity of courses within the same cluster.

3) U-CF

User-based collaborative filtering prediction method. This method calculates similar users to the target user and predicts course exam grades based on the similarity between users.

4) KN-CF

The method proposed in this paper, based on combining educational knowledge graph and collaborative filtering, is compared with the methods mentioned above. In the experiment,with β set to 0.4, the comparative results of the four groups in terms of RMSE and MAE metrics are illustrated in Figures 5 and 6, respectively.

From Figures 5 and 6, it can be observed that, except for the 2nd semester, the fused method of course knowledge graph and collaborative filtering generally outperforms course-based collaborative filtering, user-based collaborative filtering, and K-means-based methods in terms of RMSE

TABLE 4. Contribution of Paper.

Method	Contribution
Build a knowledge graph	Learn more knowledge and expand the modle's information
Computing KG similarity	Computing course similarity from a semantic perspective
Fused KG-CF	Enhancing collaborative filtering recommendation methods

TABLE 5. Comparison result of RMSE.

	Term-2	Term-3	Term-4	Term-5	Term-6	Term-7
	RMSE	RMSE	RMSE	RMSE	RMSE	RMSE
C-CF	1.3615	1.4439	1.1539	1.1310	1.0885	1.0376
C-KM	**1.1239**	1.0725	1.0396	1.0611	0.9720	0.9980
U-CF	1.1602	1.1686	1.1775	1.2133	1.0276	1.0923
KN-CF	1.2068	**0.9516**	**0.9802**	**0.9823**	**0.8929**	**0.9408**

TABLE 6. Comparison result of MAE.

	Term-2	Term-3	Term-4	Term-5	Term-6	Term-7
	MAE	MAE	MAE	MAE	MAE	MAE
C-CF	1.0702	1.1684	0.8435	0.8414	0.7962	0.7670
C-KM	**0.8464**	0.8124	0.8052	0.8068	0.7238	0.7435
U-CF	0.8521	0.8774	0.8545	0.9068	0.7264	0.8246
KN-CF	0.9280	**0.7046**	**0.7298**	**0.7382**	**0.6418**	**0.7035**

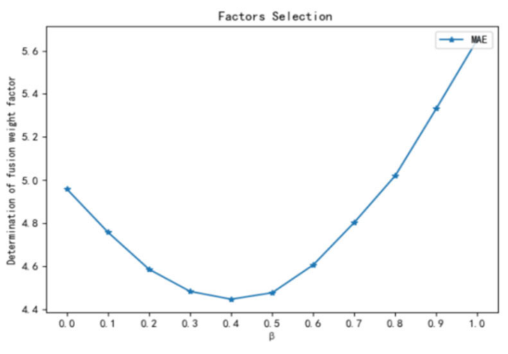

FIGURE 5. MAE results under different weight values.

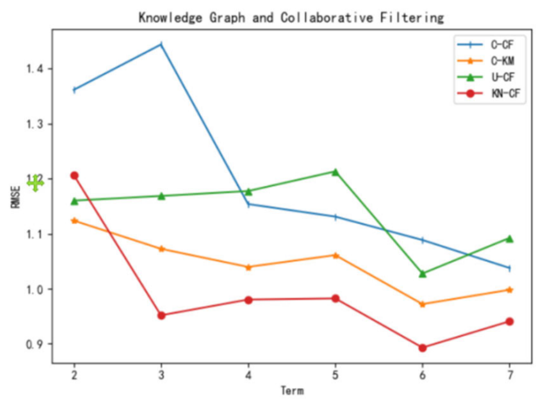

FIGURE 6. RMSE comparison results.

and MAE metrics. This indicates that the course knowledge content learned by the knowledge graph contributes to the discovery of knowledge relevance, enhancing the accuracy of similarity computation and reducing prediction errors. The relatively higher errors in the 2nd semester are likely due to the fact that the courses undertaken by students during that period are primarily common courses. As specialized courses increase in subsequent semesters, the content becomes more interrelated, and the knowledge learned by the graph becomes more accurate in describing the courses. Thus, overall, the

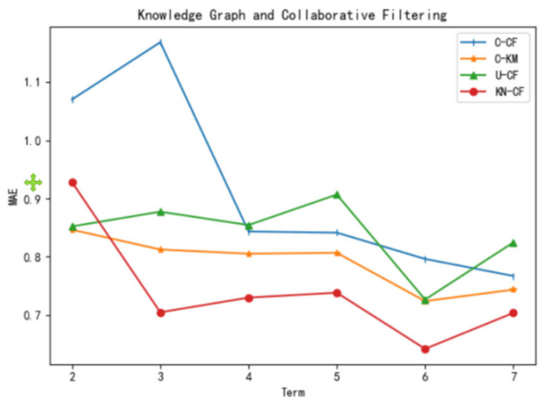

FIGURE 7. MAE Comparison Results.

fused approach of course knowledge graph and collaborative filtering exhibits lower RMSE and MAE errors compared to other methods.

The results of grade prediction for four groups from the 2nd semester to the 7th semester in terms of RMSE and MAE metrics are shown in Table 5,6:

From TABLE 5 and TABLE 6 it can be shown that the method proposed in this paper performs poorly in terms of MAE and RMSE metrics in the second semester. The reason for this is that the first two semesters mainly consist of general education courses such as English and mathematics. It is only after the second semester that more professional courses are offered, where the semantic relationships in the knowledge graph can be better manifested. Therefore, starting from the second semester, the effectiveness of the model proposed in this paper is optimal. This is because through the knowledge graph, a large number of semantic relationships between courses are learned, improving the accuracy of similarity calculation between courses. Additionally, by integrating collaborative filtering methods, the model learns from students' historical learning behaviors, thus enhancing the accuracy of the model's predictions.

VI. CONCLUSION, RECOMMENDATIONS, IMPLICATIONS, AND FUTURE WORK

This study addresses the limitations of collaborative filtering methods, which heavily rely on historical student grades and often overlook semantic information of courses as well as the influence of subjective factors on course grades. A novel approach is proposed for grade prediction, combining educational knowledge graphs and collaborative filtering. This approach bridges the gap in computing course similarity within collaborative filtering algorithms by incorporating knowledge semantics. Experimental results demonstrate that the proposed method achieves better results in terms of RMSE and MAE metrics compared to other methods.However, the dataset used in this experiment is relatively small, potentially introducing some errors. The algorithm that combines knowledge graph and collaborative filtering can learn more knowledge relationships from knowledge-rich educational datasets, leading to improved prediction accuracy.

The research findings can provide a basis for relevant studies in the knowledge-rich field of education. For example, they can contribute to student grade prediction, personalized course customization, course recommendations, and other related areas.

Therefore, in future research, the next step would involve applying the grade prediction method that combines educational knowledge graph and collaborative filtering to larger public online education datasets for experimental verification.

REFERENCES

[1] H. Ko et al., "A survey of recommendation systems: Recommendation models, techniques, and application fields," *Electronics*, vol. 11, no. 1, p. 141, Jan. 2002.

[2] Z. Yang, "Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions," *J. Thoracic Disease*, vol. 12, no. 3, pp. 165–174, Mar. 2020.

[3] X. Sun, X. Zhao, Y. Ma, X. Yuan, F. He, and J. Feng, "Muti-behavior features based knowledge tracking using decision tree improved DKVMN," in *Proc. ACM Turing Celebration Conf. China*, May 2019, pp. 1–6.

[4] C. L. Ren and X. L. Wang, "Research on students' performance portrait method based on multidimensional data," *J. Phys., Conf. Ser.*, vol. 1237, no. 3, pp. 32–43, 2019.

[5] C. Zhu, W. Sun, Q. Liu, Y. Yuan, C. Fang, and Y. Huang, "HomoTR: Online test recommendation system based on homologous code matching," in *Proc. 35th IEEE/ACM Int. Conf. Automated Softw. Eng. (ASE)*, Sep. 2020, pp. 1302–1306.

[6] H. Li, H. Li, S. Zhang, Z. Zhong, and J. Cheng, "Intelligent learning system based on personalized recommendation technology," *Neural Comput. Appl.*, vol. 31, no. 9, pp. 4455–4462, Sep. 2019.

[7] I. Burman and S. Som, "Predicting students academic performance using support vector machine," in *Proc. Amity Int. Conf. Artif. Intell. (AICAI)*, Feb. 2019, pp. 756–759.

[8] O. B. Deho, L. Liu, S. Joksimovic, J. Li, C. Zhan, and J. Liu, "Assessing the causal impact of online instruction due to COVID-19 on students' grades and its aftermath on grade prediction models," in *Proc. ACM Conf. Inf. Technol. for Social Good*, Sep. 2022, pp. 32–38.

[9] Y. Zhang and W. Jiang, "Score prediction model of MOOCs learners based on neural network," *Int. J. Emerg. Technol. Learn. (iJET)*, vol. 13, no. 10, p. 171, Oct. 2018.

[10] Y. Ma, C. Cui, X. Nie, G. Yang, K. Shaheed, and Y. Yin, "Pre-course student performance prediction with multi-instance multi-label learning," *Sci. China Inf. Sci.*, vol. 62, no. 2, pp. 1–3, Feb. 2019.

[11] S. D. A. Bujang, A. Selamat, R. Ibrahim, O. Krejcar, E. Herrera-Viedma, H. Fujita, and N. A. Md. Ghani, "Multiclass prediction model for student grade prediction using machine learning," *IEEE Access*, vol. 9, pp. 95608–95621, 2021.

[12] M. T. Xue, "Research on prediction of college students' achievement based on neighborhood collaborative filtering," *Modern Comput.*, vol. 11, no. 5, pp. 26–29, 2020.

[13] L. Huang, C.-D. Wang, H.-Y. Chao, J.-H. Lai, and P. S. Yu, "A score prediction approach for optional course recommendation via cross-user-domain collaborative filtering," *IEEE Access*, vol. 7, pp. 19550–19563, 2019.

[14] M. Ceyhan, S. Okyay, Y. Kartal, and N. Adar, "The prediction of student grades using collaborative filtering in a course recommender system," in *Proc. 5th Int. Symp. Multidisciplinary Stud. Innov. Technol. (ISMSIT)*, Oct. 2021, pp. 177–181.

[15] M. Wang, L. Qiu, and X. Wang, "A survey on knowledge graph embeddings for link prediction," *Symmetry*, vol. 13, no. 3, p. 485, Mar. 2021.

[16] Q. Wang, Z. Mao, B. Wang, and L. Guo, "Knowledge graph embedding: A survey of approaches and applications," *IEEE Trans. Knowl. Data Eng.*, vol. 29, no. 12, pp. 2724–2743, Dec. 2017.

[17] L. Geng, ''The recommendation system of innovation and entrepreneurship education resources in universities based on improved collaborative filtering model,'' *Comput. Intell. Neurosci.*, vol. 2022, pp. 1–9, Jun. 2022.

[18] H. Wang, M. Zhao, X. Xie, W. Li, and M. Guo, ''Knowledge graph convolutional networks for recommender systems,'' in *Proc. World Wide Web Conf.*, May 2019, pp. 3307–3313.

[19] X. Lu, Y. Zhu, Y. Xu, and J. Yu, ''Learning from multiple dynamic graphs of student and course interactions for student grade predictions,'' *Neurocomputing*, vol. 431, pp. 23–33, Mar. 2021.

[20] Q. Wang, X. Cao, J. Wang, and W. Zhang, ''Knowledge-aware collaborative filtering with pre-trained language model for personalized review-based rating prediction,'' *IEEE Trans. Knowl. Data Eng.*, vol. 36, no. 3, pp. 1170–1182, Mar. 2024.

[21] K. K. Jena et al., ''e-Learning course recommender system using collaborative filtering models,'' *Electronics*, vol. 12, no. 1, p. 157, Jan. 2023.

[22] G. Xu, G. Jia, L. Shi, and Z. Zhang, ''Personalized course recommendation system fusing with knowledge graph and collaborative filtering,'' *Comput. Intell. Neurosci.*, vol. 2021, pp. 1–8, Sep. 2021.

[23] R. Lavanya and B. Bharathi, ''Movie recommendation system to solve data sparsity using collaborative filtering approach,'' *ACM Trans. Asian Low-Resource Lang. Inf. Process.*, vol. 20, no. 5, pp. 1–14, Sep. 2021.

[24] Q.-T. Truong, A. Salah, and H. W. Lauw, ''Bilateral variational autoencoder for collaborative filtering,'' in *Proc. 14th ACM Int. Conf. Web Search Data Mining*, Mar. 2021, pp. 292–300.

[25] W. Q. Xiao, S. J. Yao, and S. M. Wu, ''An improved Top-N collaborative filtering recommendation algorithm,'' *Comput. Appl. Res.*, vol. 35, no. 1, pp. 105–112, 2018.

[26] A. Rossi, D. Barbosa, D. Firmani, A. Matinata, and P. Merialdo, ''Knowledge graph embedding for link prediction: A comparative analysis,'' *ACM Trans. Knowl. Discovery Data*, vol. 15, no. 2, pp. 1–49, Apr. 2021.

[27] A. Bordes, N. Usunier, and A. Garcia-Duran, ''Translating embeddings for modeling multi-relational data,'' in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 26, 2013, pp. 1–9.

[28] X. Tian, ''Study on keyword extraction using word position weighted TextRank,'' *Data Anal. Knowl. Discovery*, vol. 29, no. 9, pp. 30–34, 2013.

[29] H. T. Huynh, N. Duong-Trung, D. Q. Truong, and H. X. Huynh, ''Vietnamese text classification with TextRank and Jaccard similarity coefficient,'' *Adv. Sci., Technol. Eng. Syst. J.*, vol. 5, no. 6, pp. 363–369, 2020.

[30] Y. Luo, G. Zhou, J. Li, and X. Xiao, ''Study on MOOC scoring algorithm based on Chinese university MOOC learning behavior data,'' *Heliyon*, vol. 4, no. 11, Nov. 2018, Art. no. e00960.

[31] B. Ray, A. Garain, and R. Sarkar, ''An ensemble-based hotel recommender system using sentiment analysis and aspect categorization of hotel reviews,'' *Appl. Soft Comput.*, vol. 98, Jan. 2021, Art. no. 106935.

[32] Z. Z. Darban and M. H. Valipour, ''GHRS: Graph-based hybrid recommendation system with application to movie recommendation,'' *Expert Syst. Appl.*, vol. 200, Aug. 2022, Art. no. 116850.

[33] S. S. Choudhury, S. N. Mohanty, and A. K. Jagadev, ''Multimodal trust based recommender system with machine learning approaches for movie recommendation,'' *Int. J. Inf. Technol.*, vol. 13, no. 2, pp. 475–482, Apr. 2021.

YIWEN ZHANG was born in Fuyang, Anhui, China, in 1980. She received the B.S. and M.S. degrees from the University of Science and Technology of China, in 2003 and 2009, respectively.

Since 2019, she has been a Professor with the Big Data and Artificial Intelligence Department, Texas A&M University, College Station. She is the author of two books, more than 30 articles, and more than ten inventions. Her research interests include recommendation algorithm and recommendation systems, educational recommendation method, and application of prediction method in weather environment.

VLADIMIR Y. MARIANO is currently a Full Professor with the College of Computing and Information Technologies, National University, Manila, Philippines. His research interests include image processing and pattern recognition and intelligent information processing.

REX P. BRINGULA is currently a Full Professor with the College of Computer Studies and Systems, University of the East (UE). He has published more than 100 articles. His research interests include affective computing, cyber-behavior, environmental issues, human–computer interaction, IT education, internet studies, mobile learning, and web usability.