Teaching Teacher Recommendation Method Based on Fuzzy Clustering and Latent Factor Model

DUNHONG YAO AND XIAOWU DENG

School of Computer Science and Engineering, Huaihua University, Huaihua 418000, China
Key Laboratory of Wuling-Mountain Health Big Data Intelligent Processing and Application in Hunan Province Universities, Huaihua 418000, China
Key Laboratory of Intelligent Control Technology for Wuling-Mountain Ecological Agriculture in Hunan Province, Huaihua 418000, China

Corresponding author: Dunhong Yao (dh_yao@hhtc.edu.cn)

ABSTRACT Colleges and universities attach great importance to the quality of undergraduate teaching. To virtually guarantee the course's teaching quality, the key lies in recommending suitable teachers for the course scientifically. It is a seemingly simple but very complicated problem. Moreover, with the development of colleges and universities, new courses are continually set up, and new teachers are introduced, which further complicates the problem. The problem has not been solved well for many years. Therefore, we propose a course teacher recommendation model (FCTR-LFM) based on fuzzy clustering and the latent factor model (LFM) to solve this problem. Firstly, under the guidance of pedagogy theories and methods, we conduct quantitative modeling for teachers and courses' relevant characteristics and combine the quantitative results with historical teaching scores to establish a large-scale sparse course teaching evaluation matrix as the recommendation dataset. Next, we adopt the improved fuzzy clustering model to realize teachers' automatic clustering according to their characteristics and use the teacher cluster to reconstruct the teaching evaluation matrix, significantly reducing the dataset's size and reducing the sparsity. Then, we used the improved LFM to predict the score items in the evaluation matrix, including the missing score items. Finally, the prediction evaluation scores are sorted according to the course, and the TOP-N recommendation of the course teachers is realized. The experimental results show that FCTR-LFM can realize the prediction and recommendation well using the optimized parameters. It effectively solves the problem that there is no scientific basis for recommending suitable teachers for the course for a long time.

INDEX TERMS Teacher characteristics, course characteristics, sparse evaluation matrix, FCTR-LFM, TOP-N recommendation, teaching quality.

I. INTRODUCTION

At present, colleges and universities still attach great importance to undergraduate teaching quality under the situation that new courses and teachers are constantly introduced. The key to improving the quality of teaching lies in whether the teacher matches the course. Studies have shown that the quality of course teaching is a complex multi-dimensional structure. Which is closely related to the teachers' educational background, degree, title, gender, age, teaching age, professional matching, knowledge structure, job burnout, and

The associate editor coordinating the review of this manuscript and approving it for publication was Juan A. Lara.

teaching style [1]–[10]. The course teaching quality is also closely related to the course's complexity, difficulty, breadth, and depth [11]. It is reflected in the results of teaching evaluation [12], [13]. An implicit relationship exists among teacher characteristics, course characteristics, and teaching quality. A way to resolve this problem is to establish an appropriate course teacher Recommendation System (RS). It can be used to excavate the implicit relationship among teacher, course, and teaching evaluation, solve cold start and data sparsity and accurately predict the teaching evaluation score between each course and each teacher. In this way, we can accurately recommend TOP-N teachers for each course to improve the teaching quality scientifically.

In recent years, studies related to education recommendation mainly focus on the recommendation system and prediction. The research on the recommendation system mainly includes the following aspects:

(i) Recommendation of teaching resources. Wu et al. [14] constructed an online education knowledge recommendation system based on an improved Neural Networks (NNs) path sorting algorithm. Oliveira et al. [15] proposed a content recommendation framework to customize content recommendations according to each learner's style. Taufik et al. [16] used the K-Nearest Neighbor (KNN) classification method to help students choose professional knowledge consistent with their abilities and interests.

(ii) Course recommendation. Xu and Zhou [17] designed a curriculum recommendation framework based on a deep learning model to extract multi-mode course features. Pan et al. [18] proposed a course recommendation model based on deep learning to generate viewpoints from different perspectives and provide students with sensible course recommendations. Wang et al. [19] proposed using attention-based convolutional NNs to predict user ratings and recommend courses with TOP-N positions.

(iii) Professional or learning object recommendation. Alghamdi et al. [20] proposed a fuzzy recommendation system to help students choose the right major. Wan et al. [21] used a Random Forest (RF) algorithm to implement a personalized professional recommendation system according to students' interests and learning ability. Sergis and Sampson [22] developed a recommendation system to uniformly support teachers' select learning objects from the existing learning object library.

Research on prediction mainly includes:

(i) Performance prediction. Huang et al. [23] designed a recommendation system based on a cross-user domain Collaborative Filtering (CF) algorithm to predict each student's performance in elective courses accurately. Zacharis [24] used classification and regression trees to analyze student course activity data to design a prediction model to predict student course performance. Imran et al. [25] proposed a student achievement prediction model based on a supervised learning Decision Tree (DT) classifier.

(ii) Student performance prediction. Data collected by Mozahem [26] from the Learning Management System were used to predict student performance in face-to-face classrooms. Chiu et al. [27] used statistical models to predict student performance based on learning behavior data. Almasri et al. [28] proposed a classifier technique based on the ensemble meta-based tree model to predict student performance.

(iii) Learn early warning and prediction. Qiu et al. [29] proposed an end-to-end dropout prediction model based on convolutional NNs to predict students' dropout problems in MOOCs. Ma et al. [30] discussed predicting students' performance before class according to the traditional classroom teaching situation. That is, testing the students at risk in each course before the beginning of each course. Zacharis [31]

analyzed the log file data in the modern learning management system and predicted the students at risk of poor performance in blended learning courses.

(iv) Other predictions. Verma et al. [32] used a machine learning classifier to predict Indian and Hungarian university students' knowledge of information and communication technology and mobile technology. Huang et al. [33] designed a novel cross-user domain CF algorithm to accurately predict each student's optional course score using the most similar advanced students' course score distribution. Verma et al. [34] predicted the student's native place based on technological awareness having various development, availability, usability, educational benefits, etc.

Hybrid recommendation systems combining several recommendation technologies are becoming critical, enhancing the accuracy and reliability of recommendations. Zhu et al. [35] proposed an off-line course recommendation model that fuses network structure features with graphical NNs and combines user interaction activities with tensor factorization to improve course students' grading accuracy not attend. Wan and Niu [36] proposed a hybrid filtering recommendation method based on the Learners' influence propagation method, self-organizing recommendation method, and sequential pattern mining, which improved the personalization diversity recommendations. Esteban et al. [37] combined CF and content-based CF to propose a hybrid recommendation system; it is used to recommend the most suitable courses to students, which improved the recommendation's reliability and performance. Nafea et al. [38] designed an effective recommendation algorithm based on the K-means clustering algorithm, cosine similarity measure, and Pearson correlation coefficient. They recommended personalized course learning objects according to students' learning styles, which improved the recommendation accuracy.

From the above research status, we can see that the current research applied to the recommendation system in higher education does not involve the critical task of recommending suitable teachers for courses. Machine learning techniques (mainly NNs, KNN, K-means, CF, DT, RF.) used in these studies can solve the above-recommended tasks well. However, the dataset used in our task is constructed from the matrix of teachers' characteristics, course characteristics, and teaching scores, which is too sparse and has cold start problems. There is a big problem using any one of the above machine learning techniques alone to achieve the course teachers' recommended tasks. Our research tries to solve these problems and improve the effectiveness and efficiency of the teacher recommendation system. We are committed to the following specific issues:

(i) The teacher's recommendation task's dataset is incredibly sparse, making it challenging to implement the recommendation using any of the above machine learning technologies alone [39]. The scarcity of data is that while there are many teachers and courses, each teacher teaches only a small number of courses. Therefore, extreme data

sparsity presents a significant challenge to User-based CF or Item-based CF recommendations [86].

(ii) Many teachers with different characteristics, especially the characteristics of teaching style and job burnout, are typical unbalanced samples. It is challenging to set a fixed number of clustering based on K-means to meet automatic clustering requirements among teachers based on characteristics.

(iii) How to express the implicit relationship among teachers' characteristics, courses' characteristics, teaching scores, and sufficient good reasonable predictive teaching scores for teachers' non-teaching courses is a considerable challenge. Suppose the missing items in the matrix were filled with the mean value using the CF-based matrix decomposition method. It is then decomposed into two matrix multiplication using Singular Value Decomposition (SVD) [40]. The calculated predicted scores could hardly express the teacher and the course's real implicit correlation.

(iv) For a high-dimensional teaching evaluation matrix, how to effectively realize dimension reduction, improve the accuracy and efficiency of prediction, and reflect the implicit relationship between objects. It is another problem that needs to be solved.

To this end, we proposed a course teacher recommendation system (FCTR-LFM) based on the fuzzy clustering and latent factor model (LFM) [41] to improve the personalization and efficiency of recommendation. The main work includes: under the guidance of pedagogy norms, a series of methods are defined to obtain quantitative results of teacher characteristics, course characteristics, and teaching performance. These results will be used to build the experimental dataset. The dataset is presented as a high-dimensional sparse evaluation matrix. A fuzzy clustering model for teachers is designed, enabling teachers to cluster automatically according to their characteristics and solve the problems of reducing the evaluation matrix's sparsity and teachers' cold starting. Based on the improved LFM, the evaluation matrix is decomposed into the product of two low-dimensional matrices containing implicit factors. It reflects the implicit relation between teacher characteristics, course characteristics, and teaching grading. Predict all missing scores in the evaluation matrix. Recommend the top N teachers according to the course. Specifically, the main contributions of this paper are as follows:

(i) Establish the experimental dataset. A series of quantitative models are defined under the pedagogical norms, which are used to quantify teachers' Educational Background, Degree, Title, Gender, Age, Teaching Age, Job Matching, Professional Matching and Teaching Style, as well as Course Difficulty and Teaching Score. Then, these are used to construct the teaching evaluation matrix as the experimental data set. In the evaluation matrix, if the teacher teaches a particular course, the corresponding evaluation value is the product of the teaching score and the course difficulty; otherwise, the value is 0. In this way, a sparse experimental data set

is established, and the correlation among the three is also reflected.

(ii) The fuzzy clustering model of teachers is established. It does not need to set the number of clusters in advance but through teachers' characteristics, reflecting flexibility. First of all, we use the teacher eigenvector to establish the teacher set and then use the standardization method to process the teacher set's data to get a feature index matrix. Then the fuzzy similarity matrix is established by the similarity coefficient method. Then the square self-synthesis method is used to obtain the transitive closure. Finally, the appropriate confidence level value is set to realize teacher clustering. In this way, we use the teacher cluster to replace the related teachers in the evaluation matrix, effectively reducing the evaluation matrix's scale and sparsity and solving the cold start problem.

(iii) Establish the prediction model of teaching evaluation. When using the teacher clusters to reduce the evaluation matrix scale, the evaluation between the teacher cluster and the course is replaced by the mean value of non-zero evaluation within the teacher cluster, which improves the LFM model. The high-dimensional sparse evaluation matrix is then decomposed into the product of two low-dimensional matrices with implicit internal relations to predict the evaluation between teacher clusters and curriculum effectively. This way, it solves the problem that other methods can not reflect the implicit relationship among teacher characteristics, curriculum characteristics, and teaching scores and improves prediction efficiency.

(iv) The TOP-N teachers are recommended according to the course. First of all, the evaluation of the course in the prediction evaluation matrix is sorted in descending order, and the top teacher cluster is selected as the recommended object. The cluster teachers are then ranked in descending order according to their evaluation in the fuzzy similarity matrix, and the top n teachers are selected for recommendation. If the number of recommended teachers in the cluster is not enough, the teachers who are in the second cluster will be selected according to the same method. In this way, the efficiency of the TOP-N recommendation is improved.

In this way, we are expected to solve course teachers' recommendation problems and find a new scientific method to ensure course teaching quality.

This study's remainder is constructed as follows: In Section II, we discuss the related work on partition clustering methods and TOP-N recommendations. In section III, we describe data collection and preprocessing. In Section IV, we describe FCTR-LFM in detail. In Section V, we provide the experimental results and evaluate them. Finally, in Section VI, we conclude the paper and point out new directions for research.

II. RELATED WORK
In this section, we primarily focus on the partition clustering and TOP-N recommendation. Hence, we discuss the related work about the partition clustering and TOP-N recommendation methods.

A. PARTITION CLUSTERING METHODS

Clustering methods have many types. The partition-based clustering method with low computational complexity is frequently applied to deal with clustering large datasets. The representative algorithms include K-Means [42], mixed density clustering, graph clustering, and fuzzy clustering. K-Means algorithm is a widely used and efficient clustering method. However, this algorithm has obvious limitations, such as the quality of the clustering result depends on the selection of the initial clustering center, the sensitivity to abnormal sample points, and it can only handle numerical datasets. From the 1960s to the present, many researchers have improved the K-Means algorithm to overcome these limitations. Bradley and Fayyad [43] proposed an improvement strategy to overcome the initial center's influence. Pelleg and Moore [44] proposed a variant of the algorithm X-Means algorithm to accelerate the iterative process. Berkhin and Becher [45] extended the K-Means algorithm to the field of distributed clustering. The K-MODES algorithm proposed by Nguyen [46] overcomes the K-Means algorithm. The Means algorithm can only deal with the defects of numerical data. The hybrid density clustering algorithm transforms unsupervised learning into a supervised classification method [47]. These probability distributions are commonly used distributions, such as Gaussian distribution and t distribution. Maximum Likelihood Estimation is an essential algorithm in parameter estimation. Among these algorithms, the expectation-maximization algorithm is the most commonly used [48], [49]. The Cluster Analysis Statistical Test algorithm is also considered a clustering algorithm based on a probability model [50]. The typical graph clustering algorithm is the Clustering Identification via Connectivity Kernels algorithm based on a minimum weighted segmentation [51], [52]. This algorithm establishes a graph theory association between the sample points and the graph's vertices and sample points (edges and weights on the edges). The method based on Ding and Li [53] proposed random walks is also a graph clustering algorithm. Spectral Clustering algorithm transforms clustering into solving secondary optimization problems, can identify clusters of arbitrary shapes and converge to the optimal global solution [54], and has a wide range of applications in the field of image analysis. Lee et al. [55] and Shi et al. [56] also proposed a fast and adaptive graph-based relaxed clustering algorithm. In 1981, the fuzzy clustering method was implemented by Bezdek for the first time. The algorithm is the famous Fuzzy C-Means algorithm [57], a widely used clustering algorithm in image segmentation. This algorithm uses the degree of membership to determine the similarity of sample points and is a fuzzy clustering method based on the objective function. Subsequently, the research team of Hathaway [58] performed global optimization of the fuzzy objective function.

B. TOP-N RECOMMENDATION METHODS

The TOP-N recommendation is to recommend a list of items that may be of interest to the user through the user's implicit feedback information for their reference. The two major categories of TOP-N recommendation methods are as follows: pointwise and pairwise methods.

Pointwise methods learn the model parameters by minimizing a pointwise loss function to fit users' absolute rating values. Pan et al. [59] proposed a type of collaborative filtering (OCCF) problem and proposed two methods to solve the problem: one is based on negative sample weight. The other is based on negative sample sampling. Hu et al. [60] regard the user's rating data as an indicator of positive and negative preferences related to different confidence values. Ning and Karypis [61] proposed a Sparse Linear Method for TOP-N recommendation. This method learns the sparse aggregation coefficient matrix of the item by solving the regularization optimization problem. Kabbur et al. [62] proposed a Factorial Item Similarity Model (FISM) for the TOP-N recommendations to improve TOP-N recommendations' effectiveness. In their model, the item-item similarity matrix is decomposed into the product of two low-rank latent factor matrices.

Pairwise methods attempt to directly optimize ranking-oriented metrics based on the assumption that users prefer the rated items to the unrated items. Rendle et al. [63] are the pioneers of pairwise recommendation. They proposed a general OPTimization criterion for Bayesian personalized ranking (BPR). Pan and Chen [64] proposed a group BPR (GBPR) model by introducing rich interactions between users. Shi et al. [65] presented the Collaborative Less is More Filtering model by directly maximizing the metric MRR. The work in [66] proposed a Tensor Factorization for MAP maximization model as a context-aware TOP-N recommendation model by maximizing the MAP metric. These pairing methods can achieve better performance than point pair methods. However, these methods treat all unobserved feedback as negative feedback, thus ignoring the users' relative preferences in the unobserved items. This separation results in an extreme imbalance between positive and negative samples.

Recently, a few studies have tried to solve this problem. Zhao et al. [67] developed a model called Social BPR based on the assumption that users tend to choose the goods selected by their friends. Song and Meyer [68] proposed a Generalized Area under the Receiver Operating Characteristic Curve metric to quantify ranking performance in signature social networks. Liu et al. [69] proposed a TOP-N recommendation algorithm called Collaborative Pair Learning Ranking based on the view that users tend to prefer the items selected by neighbors. Yu et al. [70] tried to combine various types of user-item relationships into a unified pairwise ranking model to optimize map and MRR ranking indicators. Although these methods take advantage of users' preferences in the

unobserved projects, they are all based on heuristic rules, and their performance is not satisfactory.

III. DATA ACQUISITION AND PREPROCESSING

In this section, we establish quantitative models of teacher characteristics, course difficulty, and teaching scores under pedagogy theory guidance.

A. TEACHER CHARACTERISTIC DATA

Existing studies have shown that teachers' relevant characteristics significantly correlate with students' learning effects [1]–[4]. Basow and Silberg [5] found that students' evaluation of male teachers is higher than that of female teachers. Cohen [6] found a positive linear correlation between teaching age and evaluation scores by analyzing thousands of university teachers' student evaluation results. Greimel-Fuhrmann and Geyer [12] showed that the teacher's professional title was positively correlated with student evaluation. Agaoglu [7] found that the professional title is high, and the students' evaluation results are great when the teacher's age and teaching age are long. Spooren and Mortelmans [13] found that teachers' age, gender, academic qualifications, and professional titles also have a specific influence on the results of students' evaluation of teaching. However, the current research on teacher characteristics' influence on teaching evaluation results considers individual teacher characteristics such as gender, professional title, age, and educational background. Few studies consider all the background characteristics of teachers and their interaction effects. Therefore, we decided to use teacher education, degree, title, gender, age, teaching age, professional matching, and job burnout as teacher's characteristic data through serious argumentation. The specific statute methods are defined as follows.

Def 1: Educational background factor (*Ebf*): *Ebf* is set to 1, 0.7, and 0.5 for the doctor, master, and undergraduate college, respectively.

Def 2: Degree factor (*Df*): *Df* is set to 1, 0.7, 0.5, and 0 for doctor's degree, master's degree, bachelor's degree, and non-degree, respectively.

Def 3: Professional title factor (*Ptf*), *Ptf* set to 1, 0.8, 0.5, and 0.2 for the professor, associate professor, lecturer, and junior title.

Def 4: Gender factor (*Gf*), *Gf* set to 1 and 0.5 for male and female, respectively.

Def 5: Age factor (*Af*). The Age of university teachers was divided into six stages, according to [6]. *Af* was set as 0.3, 0.4, 0.6, 0.7, 0.8, and 1 for those aged 35 or below, 36~40, 41~45, 46~50, 51~55, and over 55 years old.

Def 6: Teaching age factor (*Taf*), which represents the number of years of teachers' teaching work, reflects the growth process of teachers' teaching and marks the teachers' teaching profession's growth stage. We divided the teaching experience into six sections, according to [8], [9]. The teaching experience factors less than 2 years were set as 0.1, those of 3~5 years as 0.3, those of 6~9 years as 0.5, those

of 10~15 years as 0.7, those of 16~20 years as 0.9, and those of more than 21 years as 1.

Def 7: Job burnout factor (*Jbf*). Job burnout refers to individuals' long-term reactions due to their inability to effectively cope with the work's continuous pressure. The negative mood, attitude, and bad personality characteristics caused by this factor easily affect students and lead to teachers' psychological disorders [71]. The job burnout's characteristic data come from three dimensions: Emotional Exhaustion, Depersonalization, and Reduced Personal Accomplishment. These three dimensions' data can be obtained through the survey results of the MBI-educators survey designed by Christina Maslach and verified by the Cronbach-α coefficient method. The acquisition process is shown in Figure 1.
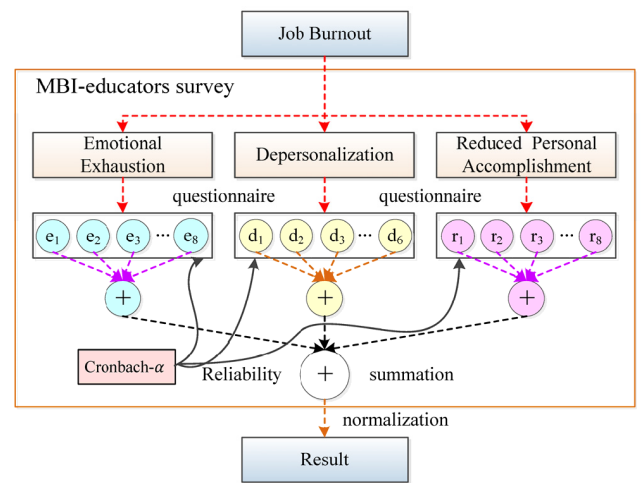

FIGURE 1. Job burnout acquisition process. The MBI-Educators questionnaire consists of 22 questions with five choices from 1 to 5 on each questionnaire. These problems are divided into three dimensions, in which e1, . . . , e8 are eight questions in the "Emotional Exhaustion" dimension, d1, . . . , d6 are six questions in the "Depersonalization" dimension, and r1, . . . , r8 are eight questions in "Reduced Personal Accomplishment" dimensions. The answers to the questions in each dimension should be summed and normalized after the Cronbach-α coefficient method is verified. Finally, credible burnout survey results can be obtained.

Then, the normalization method defined in (1) is used to obtain the burnout factor. Teacher burnout is serious when the burnout factor is large.

$$Jbf = \frac{Jb_{\text{original}} - Jb_{\text{mean}} + abs\left(floor\left((Jb_{\text{original}} - Jb_{\text{mean}})_{\text{min}}\right)\right)}{Jb_{\text{median}}}$$

(1)

where Jb_{original} means the sum of the original values of the three dimensions according to the survey results of the teaching age segment, Jb_{mean} means the average; $abs\left(floor\left((Jb_{\text{original}} - Jb_{\text{mean}})_{\text{min}}\right)\right)$ is the absolute value of the rounding (which may be negative) of the minimum value of the difference, and Jb_{median} is the median.

Def 8: Professional matching factor (*Pmf*). If one stage of the teacher's undergraduate, master, and doctoral education is matched with the taught major, then the *Pmf* is set as 0.5, 0.8 if two stages matched, and 1 if all three stages matched.

B. TEACHING STYLE DATA

The teaching style manifests teachers' preferred, habituated, charismatic, and stable teaching ideas, methods, skills, ideas, and manners formed in the long-term teaching activities and are different from those of other teachers. This style is the product of teachers' personality and teaching skills [72] and is closely related to teachers' thoughts, cultivation, and character. Teaching style has an important influence on the teaching effect. It affects the formation of students' learning style and attitude and the students' personality characteristics, interest cultivation, cooperative spirit, and learning efficiency [73].

We adopt the famous American psychologist Sternberg's teaching style classification method, which divides teachers' teaching styles into seven categories from cognitive style [74] (Table 1).

TABLE 1. Types of teaching styles.

Style	Description
Legislative (*Ls*)	People of legislative style prefer tasks that involve creativity and love to propose their ideas and plans. They make choices on their own and enjoy autonomy.
Executive (*Es*)	People of executive style work according to established instructions and rules and prefer tasks with clear structures.
Judicial (*Js*)	People of judicial style love to make evaluations about rules and procedures and enjoy judging other people's performance.
Global (*Gs*)	People of the global style pay great attention to the general picture of matters and like abstract ideas.
Local(*Locs*)	People of the local style pay great attention to concrete details.
Liberal (*Libs*)	People of liberal style love to go beyond rules and maximize challenges. They are highly tolerable to changes and ambiguities.
Conservative (*Cs*)	People of conservative style prefer to adhere to existing rules and procedures. They try to avoid changes and ambiguities.

Students are greatly familiar with their teachers' teaching styles. We first revised the teaching style evaluation scale (TSTI) [10] compiled by Grigeorenko and Sternberg to accurately obtain the teachers' teaching style. The revision's main content is to change each question's mood in the questionnaire from the teacher's perspective to the student's point of view to facilitate the questionnaire survey of students. However, 7 subscales and 49 questions remain without changing the number and meaning of the questions. In each question, seven options are given, including "completely inconsistent", "not very consistent", "slightly consistent", "basically consistent", "consistent", and "very consistent", with the corresponding score of 1–7.

Then, the collected questionnaires are sorted out and analyzed. During processing, each teaching style type' value is calculated according to (2) for the questionnaire submitted by each student.

$$\text{TS}_i = \sum_{j \in R_i} v_j \tag{2}$$

where TS_i represents the value belonging to the *i*th teaching style, $i \in \{Ls, Es, Js, Gs, Locs, Libs, Cs\}$; $j \in R_i$ represents the range of different question numbers according to the value of *i*, R_i is divided by TSTI agreement; and $v_j \in \{1, \ldots, 7\}$ represents the corresponding point value of choice in question *j*.

The values of seven teaching styles submitted by each student were calculated according to (3) to remove a small number of random questionnaire samples. The questionnaire samples with a variance value of less than or equal to 0.3 were deleted. Almost all questions are the same option because they do not seriously think about the answer.

$$\sigma^2 = \frac{1}{7} \sum_{i=Ls}^{Cs} \left(TS_i - \overline{TS} \right)^2 \tag{3}$$

where $\overline{TS}$ represents the mean of the seven teaching styles obtained from this sample.

We use the Cronbach-α coefficient [75] method shown (4) to verify that the remaining samples are credible.

$$\alpha = \frac{k}{k-1} \left(1 - \frac{\sum_{i=1}^k \varsigma_i^2}{\varsigma_t^2} \right) \tag{4}$$

where k (here 49) represents the number of problems; and $\sum_{i=1}^k \varsigma_i^2$ represents the score variation of item *i*, which is given by (5).

$$\sum_{i=1}^k \varsigma_i^2 = \sum_{i=1}^k \frac{\sum x^2}{n} - \sum_{i=1}^k \left(\frac{\sum x}{n} \right)^2 \tag{5}$$

where x in (5) is the score given by each student for each question, and n is the number of valid questionnaires.

ς_t^2 in (4) represents the variance of the total scores obtained from all valid questionnaires and is given by (6).

$$\varsigma_t^2 = \sum_{t=1}^n x_t^2 / n - \left(\sum_{t=1}^n x_t \right)^2 / n^2 \tag{6}$$

x_t in (6) is the sum of the scores given by each student for all questions.

C. TEACHING SCORING DATA

Given that some characteristic data of teachers dynamically change over time, the acquisition time of teaching scoring data should be consistent with that of teachers' characteristic data. The teaching scoring data are mainly based on the comprehensive scoring results. Such data come from the student, supervisory, and peer scoring. The teaching score was normalized by Definition 9.

Def 9: Teaching score (*Ts*) is normalized by (7) because teachers can teach more than one course at a time.

$$Ts_{t,course} = \frac{v_{t,course} - v_{\min}}{v_{\max} - v_{\min}} \times (E_{\max} - E_{\min}) + E_{\min} \tag{7}$$

where $Ts_{t,course}$ denotes the normalized value of the teaching score of the *course* taught by the teacher *t*; $v_{t,course}$ denotes

the comprehensive result of the teaching score of the *course* taught by the teacher t; v_{min} denotes the lowest value in all the teaching scores; v_{max} denotes the highest value in all the teaching scores; E_{max} denotes the highest value after normalization (generally takes one); and E_{min} denotes the lowest value after normalization, which is set to 0.1 in this study.

D. COURSE DIFFICULTY FACTOR

Course Difficulty Factor (*Cdf*) is difficult to quantify [76]. We establish the model shown in (8) to quantify the course difficulty based on the course difficulty model established by Shi *et al.* [77] and guided by the revision method proposed by Li [78].

$$Cdf = \eta \frac{Dp}{Ch} + (1 - \eta) \frac{Sp}{Ch} \quad (8)$$

where *Cdf* represents the difficulty coefficient of the course. When the *Cdf* value is large, the difficulty of the course is great. *Dp* represents the depth of the course. This factor uses the sum of the abstraction levels to express the depth of the course. The abstraction level represents the degree of abstraction related to concepts and principles and the degree of association between concepts, quantified by the sum of all knowledge chains' abstraction degrees. *Sp* is the breadth of the course, which refers to the scope and breadth of the field involved in the course content, which can be quantified by the number of course objectives that integrate the knowledge, skills, process, methods, emotional attitudes, and values of the corresponding course. *Ch* is the course's class hours, which refers to the time required to complete the course content using class hours to quantify. *Dp/Ch* represents the depth of the course per unit time, which is called comparable depth. *Sp/Ch* is the course breadth per unit time, which is called comparable breadth. η ($0 < \eta < 1$) is the weighting coefficient, reflecting the degree of emphasis on the course's comparable depth and breadth.

IV. MODEL AND ALGORITHM DESIGN

In this section, we first introduce the teacher's fuzzy clustering and latent factor model of the teaching evaluation matrix. Then, we introduce the teacher recommendation model and algorithm design of the course teaching that incorporates the teacher's fuzzy clustering.

A. TEACHER FUZZY CLUSTERING DESIGN

The clustering method's classification principle is to divide each data sample into related categories according to the specified method. The essence of fuzzy clustering analysis [79] is not to set the category in advance but to construct the fuzzy matrix by studying the object's attributes. On this basis, the classification relationship is determined. Therefore, the characteristics of uncertainty and the object of study can be objectively reflected.

We design a fuzzy clustering model for teachers to realize the automatic classification of teachers. This model uses 15 characteristics, such as teachers' education, degree, title,

gender, age, teaching age, job burnout, professional matching, and teaching style (including Legislative, Executive, Judicial, Global, Local, Liberal, and Conservative) to construct a fuzzy matrix, which can effectively determine the classification relationship. The clustering results can provide effective and accurate recommendation information for the course arrangement.

1) TEACHERS' CHARACTERISTIC DATA STANDARDIZATION

We let the set of teacher objects be $X = \{x_1, x_2, \ldots, x_m\}$. Each teacher x_i has d ($d = 15$) characteristic indicators, that is, x_i can be represented by the following d-dimension characteristic indicator vector:

$$x_i = (x_{i1}, x_{i2}, \ldots, x_{id}), \quad i = 1, 2, \ldots, m$$

where x_{ij} represents the jth characteristic index of the ith teacher. All characteristic indexes of the m teachers form a matrix, which is called $X^* = (x_{ij})_{m \times d}$, and X^* is the characteristic index matrix of X:

$$X^* = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1d} \\ x_{21} & x_{22} & \cdots & x_{2d} \\ \cdots & \cdots & \cdots & \cdots \\ x_{m1} & x_{m2} & \cdots & x_{md} \end{pmatrix} \quad (9)$$

Given that the teaching style is different from other features in dimension and order of magnitude, data standardization should be used in the feature index matrix to unify each index in a common numerical range. In this study, we use the standardization method to standardize the data to calculate the mean value and variance of the j column of the characteristic index matrix X^*. Then, we make the following transformation:

$$x'_{ij} = \frac{x_{ij} - \bar{x}_j}{\sigma_j}, \quad i = 1, 2, \ldots, m; \ j = 1, 2, \ldots, d \quad (10)$$

where, $\bar{x}_j = \frac{1}{m} \sum_{i=1}^{m} x_{ij}, \ \sigma_j^2 = \frac{1}{m} \sum_{i=1}^{m} (x_{ij} - \bar{x}_j)^2, j = 1, 2 \ldots, d.$

The characteristic index matrix after data standardization is expressed as follows:

$$X' = \begin{pmatrix} x'_{11} & x'_{12} & \cdots & x'_{1d} \\ x'_{21} & x'_{22} & \cdots & x'_{2d} \\ \cdots & \cdots & \cdots & \cdots \\ x'_{m1} & x'_{m2} & \cdots & x'_{md} \end{pmatrix} \quad (11)$$

2) CONSTRUCT THE TEACHER'S FUZZY SIMILAR MATRIX

According to the standardized data, the similarity coefficient $y_{ij} \in [0, 1]$ between teachers $x_i = \left(x'_{i1}, x'_{i2}, \cdots, x'_{id}\right)$ and $x_j = \left(x'_{j1}, x'_{j2}, \cdots, x'_{jd}\right)$ is determined by the similarity coefficient method, and the fuzzy similarity matrix $Y = (y_{ij})_{m \times m}$ is established.

$$y_{ij} = \frac{\sum_{k=1}^{d} \left|x'_{ik} - \bar{x}_i\right| \left|x'_{jk} - \bar{x}_j\right|}{\sqrt{\sum_{k=1}^{d} (x'_{ik} - \bar{x}_i)^2} \cdot \sqrt{\sum_{k=1}^{d} (x'_{jk} - \bar{x}_j)^2}} \quad (12)$$

where $\bar{x}_i = \frac{1}{d} \sum_{k=1}^{d} x'_{ik}, \bar{x}_j = \frac{1}{d} \sum_{k=1}^{d} x'_{jk}.$

If $y_{ij} < 0$, then we replace y_{ij} with $y'_{ij} = \frac{1+y_{ij}}{2}$. If the value range of y_{ij} is not in [0, 1], then $y'_{ij} = \frac{y_{ij}-w}{W-w}$ is used instead of r_{ij}, where $W = \max_{i,j}(y_{ij})$, and $w = \min_{i,j}(y_{ij})$. In this way, we establish the teacher's fuzzy similarity matrix $Y_{m \times m}$.

$$Y = \begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1m} \\ y_{21} & y_{22} & \cdots & y_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ y_{m1} & y_{m2} & \cdots & y_{mm} \end{pmatrix} \qquad (13)$$

3) FUZZY CLUSTERING

A new fuzzy equivalent matrix is constructed from fuzzy similarity matrix Y. First, the square self-synthesis method was used to work out the transitive closure $t(Y)$ of fuzzy similarity matrix Y. Specifically, Y^2, Y^4 were calculated starting from Y and repeatedly multiplying by itself. When $Y^k \cdot Y^k = Y^k$ for the first time, we obtain $t(Y) = Y^k$. Then, the appropriate confidence level value ($\lambda \in [0, 1]$) is appropriately selected to find the λ truncated matrix $t(Y)_\lambda = \left(y'_{ij}(\lambda)\right)_{m \times m}$ of $t(Y) = \left(y'_{ij}\right)_{m \times m}$, where

$$y'_{ij}(\lambda) = \begin{cases} 1, & y'_{ij} \geq \lambda \\ 0, & y'_{ij} < \lambda, \end{cases}$$

which is an equivalent Boolean matrix on X. With regard to x_i, $x_j \in X$, if $y'_{ij}(\lambda) = 1$, then teachers x_i and x_j can be classified into the same class at λ confidence level.

B. DESIGN OF THE MATRIX LATENT FACTOR MODEL OF THE TEACHING EVALUATION MOMENTS

The goal of the teacher recommendation system for courses is to analyze the historical teaching evaluation by majors of college teachers, build an appropriate model, calculate the predictive evaluation of teachers' teaching in other courses, and recommend teachers for courses by TOP-N recommendation [80] according to the predictive evaluation. We cannot simply deny teachers' incompetence in teaching other courses based on their historical teaching records. However, we need to find out the implied recommendation coefficient between teachers and courses to serve as the basis.

The SVD algorithm is a common data dimensionality reduction technology. The evaluation matrix is a large-scale multi-dimensional sparse matrix in the course teacher recommendation system due to many teachers and courses. Accordingly, this matrix requires a large amount of storage space to convert sparse matrices into dense ones. The algorithm can only be applied to matrix decomposition with low dimensions. Thus, the SVD algorithm is not an ideal teaching teacher recommendation algorithm.

Simon funk proposes an improved algorithm called funk SVD algorithm or LFM, widely used in the recommendation system based on the SVD algorithm. The core idea of LFM is to connect teachers and courses through latent factors. We set the evaluation matrix of teacher and course as $R_{m \times n}$, m is the number of teachers, n is the number of courses, and R_{ij} is the product of the score (Ts_{ij}) obtained by teacher i teaching course j and the difficulty coefficient (Cdf_j) of the course taught, which is called evaluation (i.e. $R_{ij} = Ts_{ij} \times Cdf_j$). If no direct teaching relationship exists between teacher i and course j, then $R_{ij} = 0$. The numbers of teacher m and course n are relatively large, and that of courses taught by each teacher is small in the course set. Accordingly, the evaluation matrix is a high-dimensional sparse matrix. The evaluation matrix $R_{m \times n}$ can be decomposed into two low-dimensional matrices and multiplied by the implicit model to calculate the teacher's predictive evaluation on each course. Specifically, the teacher's latent factor matrix $T_{m \times f}$ and the course latent factor matrix $C_{n \times f}$ are multiplied, as shown in (14).

$$R_{m \times n} \approx \hat{R}_{m \times n} = T_{m \times F} C^T_{n \times F} \qquad (14)$$

where F is the number of latent factors, each line of matrix $T_{m \times f}$ represents a vector of the teacher's weight on the latent factor. Meanwhile, each line of matrix $C_{n \times f}$ represents the probability distribution of every course on the latent factor vector. Then, teacher i's predictive evaluation value of course j can be calculated by (15).

$$\hat{R}(i, j) = \hat{r}_{ij} = \sum_{f=1}^{F} t_{if} c_{jf} \qquad (15)$$

where $t_{if} = T(i, f)$ represents the weight of teacher i to latent factor f, and $c_{jf} = C(j, f)$ indicates the probability distribution of course j to latent factor f. About the solution of matrices T and C, the RMSE between the observed and predicted evaluation values could be used in the training set to learn the T and C matrices if a suitable T and C matrix can be found to minimize the training set. The prediction error can also minimize the prediction error of the test set. Therefore, we define the loss function as follows.

$$H(t, c) = \sum_{(i,j) \in \text{Train}} (r_{ij} - \hat{r}_{ij})^2 = \sum_{(i,j) \in \text{Train}} \left(r_{ij} - \sum_{f=1}^{F} t_{if} c_{jf}\right)^2 \qquad (16)$$

where r_{ij} is the actual evaluation score of teacher i on course j, and Train is the training set. Direct optimization of the loss function may lead to overfitting of learning. Therefore, the training model's loss function is added with a term to prevent overfitting, and (17) is obtained.

$$H(t, c) = \sum_{(i,j) \in \text{Train}} \left(r_{ij} - \sum_{f=1}^{F} t_{if} c_{jf}\right)^2 + \omega \left(\|t_i\|^2 + \|c_j\|^2\right) \qquad (17)$$

ω is a regularization parameter, which can be obtained through experiments. A stochastic gradient descent algorithm can be used to minimize the loss function. The algorithm finds the direction of rapid descent by obtaining the partial derivatives of parameters t_{if} and c_{jf} and then continuously optimizes the parameters through iteration.

Calculating the partial derivatives of the loss function according to parameters t_{if} and c_{if} yields (18) and (19):

$$\frac{\partial H}{\partial t_{if}} = -2 \left(r_{ij} - \sum_{f=1}^{F} t_{if} c_{jf} \right) c_{jf} + 2\omega t_{if} \qquad (18)$$

$$\frac{\partial H}{\partial c_{jf}} = -2 \left(r_{ij} - \sum_{f=1}^{F} t_{if} c_{jf} \right) t_{if} + 2\omega c_{jf} \qquad (19)$$

Parameters t_{if} and c_{if} are iterated along the direction where the gradient decreases fastest until the parameters converge, and the recursive equation is obtained as follows:

$$t_{if} = t_{if} + \alpha \left(\left(r_{ij} - \sum_{f=1}^{F} t_{if} c_{jf} \right) c_{jf} - \omega t_{if} \right) \qquad (20)$$

$$c_{jf} = c_{jf} + \alpha \left(\left(r_{ij} - \sum_{f=1}^{F} t_{if} c_{jf} \right) t_{if} - \omega c_{jf} \right) \qquad (21)$$

where a is the learning rate. The iteration will rapidly decline when a is high.

C. TEACHER RECOMMENDATION MODEL AND ALGORITHM DESIGN BASED ON FUZZY CLUSTERING AND LFM

The LFM model algorithm performs well in solving the low-dimensional sparse scoring matrix; however, its efficiency is low when if the matrix is large. We propose the FCTR-LFM that integrates fuzzy teacher clustering to solve the low efficiency of the LFM model algorithm when dealing with large-scale sparse matrices. The main idea of the model is first to use fuzzy clustering to cluster teachers automatically. Then, we use teacher clusters to replace the teachers in the original FLM model, thereby greatly reducing the evaluation matrix scale and improving the efficiency of the model.

When establishing the FCTR-LFM model, the LFM model must be improved, including the evaluation matrix based on teacher clustering and the prediction evaluation model. The implementation process of the TOP-N recommendation algorithm must also be improved.

1) IMPROVEMENT DESIGN OF THE OBSERVATION AND EVALUATION MODEL BETWEEN TEACHER CLUSTER AND COURSE

The evaluation matrix $R_{m \times n}$ is a matrix established according to the relationship between the teacher and the course. The value of R_{ij} is determined according to whether the teacher has taught the course and the difficulty coefficient of the course. When the teacher has taught the course, R_{ij} takes the product of Ts_{ij} and Cdf_j; otherwise, it takes a zero value. After Z teacher clusters are formed by fuzzy clustering, we use the relationship between teacher clusters and courses to establish a new smaller evaluation matrix $R_{Z \times n}$. Many teachers may be assigned to the same cluster $z \in \{1, 2, \ldots, Z\}$. In cluster z, some courses have been taught by teachers, and others have

not been taught. How to determine the observed evaluation value R_{zj}? After a careful analysis, we think that the observed evaluation value R_{zj} between teachers cluster z and course j can be determined by the mean value of nonzero evaluation between teachers and courses in the cluster shown in (22).

$$R_{zj} = \frac{1}{p} \sum_{q=1}^{p} R_{ij}, \quad \forall t_i \in z \qquad (22)$$

where p is the number of nonzero evaluation values R_{ij} between teacher i and course j in the cluster.

2) IMPROVED DESIGN OF THE PREDICTIVE EVALUATION VALUE MODEL

The predicted evaluation value $\hat{R}(z, j) = \hat{r}_{zj}$ of the teacher cluster z on the course j can be modified from (15) to (23).

$$\hat{R}(z, j) = \hat{r}_{zj} = \sum_{f=1}^{F} t_{zf} c_{jf} \qquad (23)$$

where $t_{zf} = T(z, f)$ represents the weight of teacher cluster z to latent factor f, and $c_{jf} = C(j, f)$ indicates the probability distribution of course j to latent factor f. Thus, we can obtain a new predictive evaluation matrix $\hat{R}_{Z \times n}$.

3) IMPROVED DESIGN OF THE TOP-N RECOMMENDATION METHOD

In the process of recommending N teachers for course j, first, the rows in the prediction evaluation matrix $\hat{R}_{Z \times n}$ are arranged in descending order according to $\hat{r}_{zj}$, and the top teacher cluster z_1 is selected as the recommended candidate cluster. In teacher cluster z_1, the TOP-N teachers are recommended for course j according to the similarity of teachers in the fuzzy similarity matrix Y. If the total number of teachers in z_1 cluster is less than the recommended number N, then we select the remaining recommended teachers from the second cluster z_2 from the prediction evaluation matrix $\hat{R}_{Z \times n}$. This process goes on until the TOP-N teachers are recommended.

According to the above analysis, the recommended algorithm of the FCTR-LFM model is designed as follows:

Input: number of teachers m, teacher characteristic index matrix $X^* = (x_{ij})_{m \times d}$, confidence level λ, number of courses n, number of latent factor F, number of iterations N, learning efficiency α, and regularization parameter ω

Output: list of TOP-N teachers recommended by course

---

1. //Fuzzy clustering of Teachers
2. $X' = \left(x'_{ij} \right)_{m \times d}$ ← According to (10), the teacher characteristic index matrix $X^* = (x_{ij})_{m \times d}$ is standardized by using the standardized method;
3. $Y = (y_{ij})_{m \times m}$ ← According to (12), the fuzzy similarity matrix between teachers is established by the similarity coefficient method;
4. //The transitive closure $t(Y)$ of fuzzy similar matrix Y is obtained by the square self-synthesis method
5. $t(Y) \leftarrow Y, k = 2, Y^k = Y \cdot Y$;

6. while(t(Y) $\neq$ Y^k){

7. $t(Y) \leftarrow Y^k$; $v = k$, $k = 2*k$; $Y^k \leftarrow Y^v \cdot Y^v$; }

8. Select the appropriate confidence level value λ to obtain the cut matrix of $t\,(Y) = \left(y'_{ij}\right)_{m \times m}$;

9. $Z \leftarrow$ According to $t(Y)_\lambda$, solve the number of clusters after the teacher clustering;

10. //Construction of teacher and course evaluation matrix $R_{m \times n} = (R_{ij})m \times n$

11. for($i = 1, j = 1$; $i <= m, j <= n$; $i++, j++$)

12.   if(teacher i has a teaching relationship with course j)$R_{ij} = Ts_{ij} \times Cdf_j$;

13.   else $R_{ij} = 0$;

14. //Construct teacher cluster and courses evaluation matrix $R_{Z \times n} = (\,R_{zj})_{Z \times n}$

15. for($i = 1$; $i <= m$; $i++$)

16.   if($t_i \in z$)

17.   for($j = 1$; $j <=$ n; $j++$)$R_{zj} = \frac{1}{p} \sum_{q=1}^{p} R_{ij}$;

18. // Solve the predictive evaluation matrix $\hat{R}_{Z \times n}$ between the teacher cluster and the courses

19. [T, C] = Initialize[Z, n, F] // Initializing the latent factor matrix of $T_{Z \times F}$ and $C_{F \times n}$;

20. for i in range(0, N){

21.   The teacher cluster z and its course c are extracted from the training set;

22.   Positive and negative samples were randomly selected to form *Sample*;

23.   for(c, R_{zj} in samples)

24.   error $= R_{zj} - \hat{R}\,(z, j)$; // Calculate the error according to the current parameter

25.   for f in (0, F)

26. $T_{zf}+ = \alpha \times$ (error $\times C_{fc}$ - $\omega \times$ T_{zf}); $C_{fj}+ = \alpha \times$ (error $\times$ T_{zf} - $\omega \times C_{fj}$);

27. $\alpha = \alpha*0.9$; // When the optimization reaches a certain level, the learning rate must be slowed down, and the optimal value must be gradually approached

28. }

29. // Recommend top S teachers for course j

30. The prediction evaluation matrix $\hat{R}_{Z \times n}$ is arranged in descending order according to the $\hat{r}_{zj}$ of the assigned course j;

31. s = 0, i = 1;

32. do {

33.   Select teacher cluster z_i;

34.   In z_i, perform secondary sorting according to the teacher's similarity in the fuzzy similarity matrix Y;

35.   if(The number of teachers m_i in $z_i \geqq$ S-s){s + = S-s; Recommend S-s teachers;} else{Recommend m_i teachers; s = s + m_i; i++;}

36. } while(s < S)

V. EXPERIMENTS AND RESULTS ANALYSIS
In this section, we first introduce the experimental dataset, followed by the evaluation indicators. Finally, we describe the experiments and results.

A. DATASET
The experimental data come from a certain university major, and recommending teachers for courses between different majors is meaningless. According to the definition in Chapter III, we can obtain the Teaching Recommendation Dataset (TRDs), which is a sparse matrix, and the dataset statistics are shown in Table 2.

TABLE 2. Dataset statistics.

Dataset	Teachers	Courses	Scores	Density
TRDs	132	204	865	0.032123

The characteristic data examples of teachers are shown in Table 3, and the teaching styles of all teachers in the dataset are shown in Figure 2.

Figure 2 demonstrates that teachers' teaching styles are different in the seven dimensions proposed by Sternberg.

When the weighted coefficient $\eta = 0.4$ and after the maximum and minimum normalization, the courses' difficulty factors in the dataset are shown in Figure 3.

Figure 3 shows that the method adopted can effectively and correctly distinguish the difficulty of each course.

The fivefold cross-validation was adopted in our experiments. First, the dataset TRDs is divided into five mutually exclusive subsets of similar size. Each subset TRDs$_i$ keeps the data distribution as consistent as possible. That is, it is obtained by stratified sampling from TRDs. Then, the union of four subsets is used as the training set. The remaining subset is utilized as the test set. In this way, five training and test sets can be obtained. Accordingly, five training and tests can be conducted, and the final return is the average of the five test results.

B. EVALUATION INDICES
1) EVALUATION INDEX OF THE RATING PREDICTION
The root means square error (RMSE) can be used to measure the rating prediction accuracy of the algorithm:

$$\text{RMSE} = \frac{\sqrt{\sum_{(i,j) \in \text{Test}} (r_{ij} - \hat{r}_{ij})^2}}{|Test|} \tag{24}$$

where *Test* is the test set. When the RMSE value is small, the error is also small, and the recommendation system's performance is good.

2) TOP-N RECOMMENDED EVALUATION INDEX
Previous studies on recommender systems [81], [82] have shown that the common error rate indicators in rating prediction, such as mean square error (MSE) and RMSE, cannot completely characterize the performance of the recommendation algorithm. The performance of the algorithm mainly depends on the recommendation of the TOP-N item.

TABLE 3. Example of the teacher's characteristic.

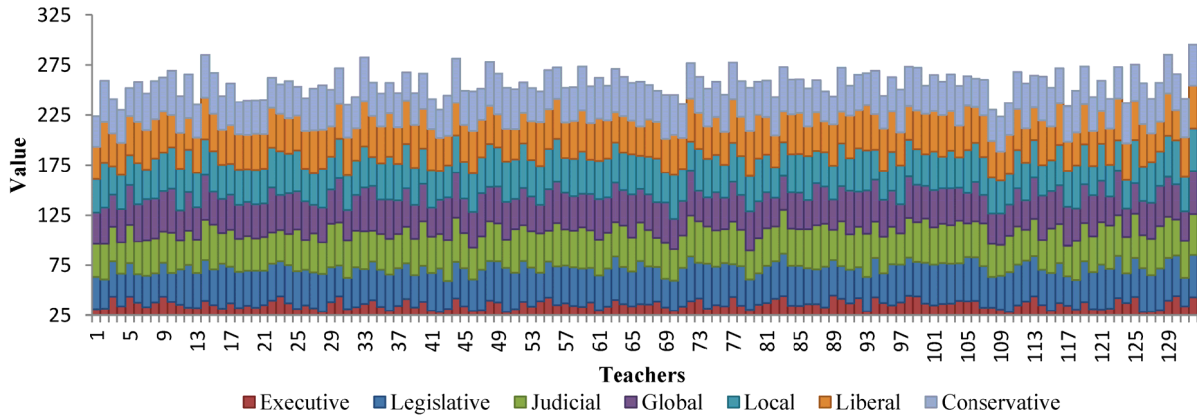| Teacher	Ebf	Df	Ptf	Gf	Af	Taf	Jbf	Pmf	Ls	Es	Js	Gs	Locs	Libs	Cs
T001	0.5	0.5	1.0	1.0	1.0	1.0	0.0614	0.5	39.4	34.8	36.4	34.8	38.8	39.1	35.7
T002	1.0	1.0	1.0	1.0	0.4	0.7	0.6571	0.5	40.0	34.3	37.0	36.3	38.7	40.9	33.8
T003	0.5	0.7	0.8	1.0	0.7	0.9	0.7314	0.5	38.6	33.9	36.5	34.7	36.5	38.6	33.3
T004	0.7	0.7	0.5	1.0	0.3	0.3	0.7321	0.8	34.8	32.4	33.5	32.1	34.5	35.1	32.8
T005	0.7	0.7	0.5	0.5	0.4	0.5	0.729	0.5	40.9	35.1	37.6	35.9	38.9	40.2	34.5

FIGURE 2. Teaching styles. The figure shows seven categories of 132 teachers' teaching styles in the experimental dataset, and the values are sorted out in the questionnaire survey of TSTI.

FIGURE 3. Course difficulty. The figure shows the difficulty coefficient of 204 courses in the data set. The larger the bubble, the more difficult the course is.

We will evaluate the hit rate through accuracy, recall, and F1-measure to fully verify the feasibility of our method:

$$P@N = \frac{\sum_{t \in T} |R(t) \cap Q(t)|}{\sum_{t \in T} |R(t)|} \qquad (25)$$

$$R@N = \frac{\sum_{t \in T} |R(t) \cap Q(t)|}{\sum_{t \in T} |Q(t)|} \qquad (26)$$

$$F1@N = \frac{2 \times \mathrm{Pr}\,ecision \times \mathrm{Re}call}{\mathrm{Pr}\,ecision + \mathrm{Re}call} \qquad (27)$$

where $R(t)$ and $Q(t)$ are the teachers' recommended list on the training and test sets.

C. EXPERIMENTAL RESULTS AND ANALYSIS

The four important parameters in the FCTR-LFM model algorithm are confidence level λ, number of latent factor F, learning rate a, and regularization parameter w. In this section, we first analyze the influence of each parameter on the algorithm. After that, we compare the algorithm with other related algorithms in terms of prediction and TOP-N recommendation accuracies. The experiment was completed on a common PC with Core i5 2.3 GHz CPU and 8 G RAM.

1) EFFECT OF THE APPROPRIATE CONFIDENCE LEVEL VALUE λ

In the FCTR-LFM algorithm, parameters F, α, N, and ω were set to 50, 0.007, 50, and 0.002, respectively. The appropriate confidence level value λ was adjusted in the interval [0, 1] by a step size of 0.1. The effects of the adjustment on the RMSE and TOP-5 of the FCTR-LFM are displayed in Figures 4 and 5, respectively.

Figures 4 and 5 demonstrate that different confidence levels λ can make teachers fuzzy cluster into different teacher clusters. This phenomenon will change the matrix's sparsity and scale when reconstructing the evaluation matrix, which significantly influences the prediction and recommendation accuracy of the algorithm. Therefore, the algorithm needs to select the appropriate confidence level λ to achieve good results.

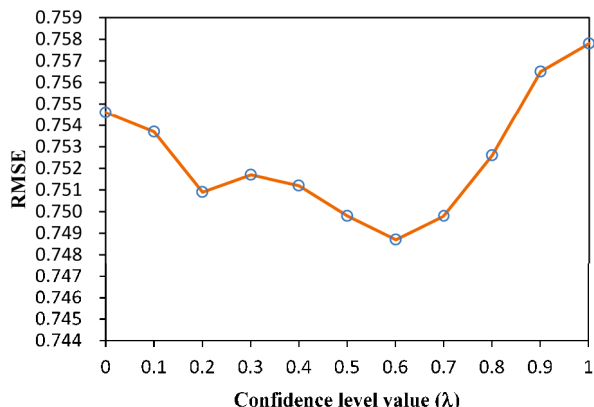**FIGURE 4.** Influence of different λ on the prediction accuracy. When other parameters are fixed, the experimental results show that the confidence level parameter λ is around 0.6, and the algorithm will obtain the optimal prediction accuracy.
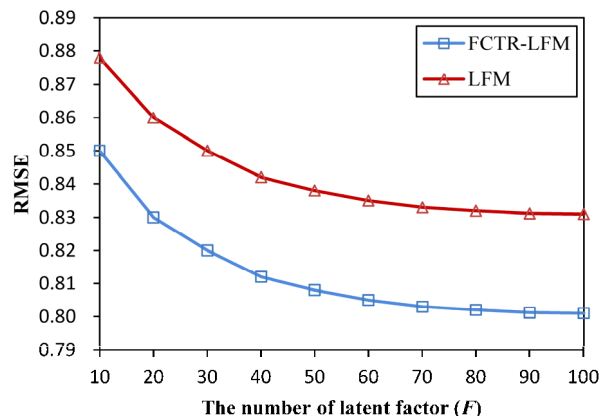

FIGURE 6. Comparison of RMSE. With the increase of F, the accuracy of both algorithms is significantly improved. The accuracy of the FCTR-LFM is slightly better than that of the LFM. When the F value exceeds 80, the accuracy of the two algorithms tends to be stable.
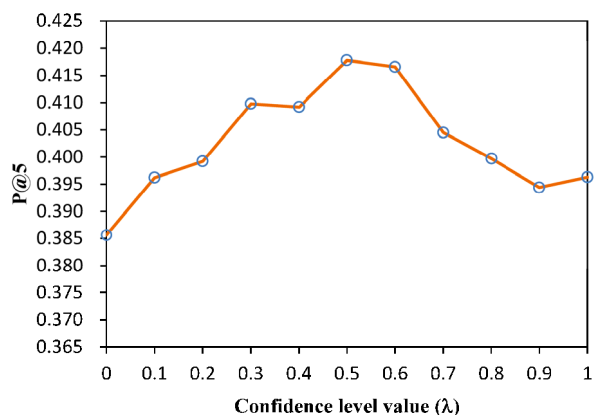

FIGURE 5. Influence of different λ on the recommendation accuracy. When other parameters are fixed, the experimental results show that the confidence level parameter λ will obtain the optimal recommendation accuracy around 0.5.

FIGURE 7. Running time comparison. In the case of small F, the running time of the FCTR-LFM model is higher than that of the LFM model. With the increase of F value, the running time of the FCTR-LFM model is significantly lower than that of the LFM model.

2) EFFECT OF THE LATENT FACTOR NUMBERS F ON THE ALGORITHM PERFORMANCE

Under the same experimental environment, parameters α, w, and N were set 0.01, 0.002, and 50. The number of latent factor F increased from 10 to 100, the FCTR-LFM algorithm ($\lambda = 0.5$) and the classic LFM algorithm on the scoring prediction accuracy and running time were compared. The results are shown in Figures 6 and 7.

Figure 6 demonstrates that when other parameters remain unchanged, the value of F has a significant influence on the overall performance of the FCTR-LFM and LFM algorithms. Within a certain range, the two algorithms' accuracy has been significantly improved with the increase of F. The score prediction accuracy of the FCTR-LFM algorithm is slightly better than that of LFM. When the value of F exceeds 80, the overall accuracy of the two algorithms is stable.

Figure 7 shows that, in terms of running time, the training time of traditional LFM is mainly derived from iterative model training; thus, the training time is closely related to the

number of iterations. The iteration time will increase with the increase in the number of latent factor F. The running time of the FCTR-LFM model is composed of fuzzy clustering and model training. Although clustering takes up part of the running time during data processing, it reduces the scoring matrix scale and the time spent in model training iteration. In the case of small F, the running time of the FCTR-LFM model is higher than that of the LFM model. Meanwhile, the running time of the FCTR-LFM model is significantly lower than that of the LFM model with the increase in F. This phenomenon shows that the FCTR-LFM algorithm has a good performance while maintaining high accuracy.

3) EFFECT OF LEARNING RATE α ON THE ALGORITHM PERFORMANCE

The learning rate controls the change of parameters in the iterative process, and its size is directly related to the final prediction results of the model. In the experiment, fixed confidence level $\lambda = 0.5$, latent factor number $F = 80$, regularization parameter $w = 0.002$, change learning rate a,

and corresponding iteration number N. The experimental results are shown in Figure 8.

FIGURE 8. Comparison of learning rate (α) with iteration times (N). As the model's learning rate goes from high to low, the convergence speed changes from fast to slow, and the number of iterations required to achieve a good recommendation effect changes from less to more.

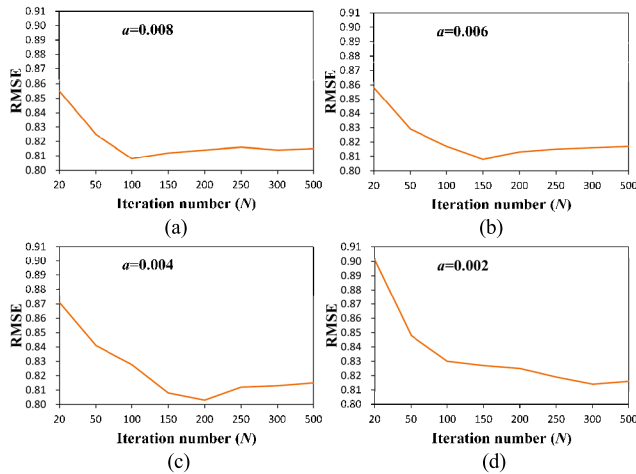Figure 8 demonstrates that the learning rate drops from 0.008 to 0.002, and the RMSE of the model changes differently. When the model's learning rate is high, the number of iterations required to achieve a good recommendation effect is relatively small, and the convergence rate of the model is relatively fast. The optimal solution will be crossed when searching for the optimal local solution along the gradient direction. When the learning rate is small, much iteration is required to achieve a good recommendation effect. The model's convergence speed is relatively slow; thus, the optimal local solution requires a large amount of time to find. Although a small learning rate can produce a low prediction error, the algorithm's convergence speed also becomes slow. Therefore, in the actual algorithm application and considering the characteristics of the model itself, factors, such as the memory space and time performance requirements of the recommended system, should also be comprehensively examined to select a good learning rate and iteration times.

4) COMPARISON WITH OTHER METHODS
The performances of several existing approaches are compared on TRDs to justify the effectiveness of our proposed method.

a: BPR
Based on implicit feedback, the BPR is extended from the TOP-N ranking recommendation algorithm by the pairing hypothesis for item ordering by using implicit feedback, Matrix Factorization, and K-Nearest Neighbors model [83].

b: FISM
The FISM alleviates the existing TOP-N recommendation algorithm's sparsity by taking the product of

two low-dimensional latent factor matrices as the similarity matrix [84].

c: FST
The Factored Similarity model with Trust (FST) introduces the mutual trust matrix and user similarity matrix into the FISM, thereby alleviating the existing TOP-N recommendation algorithm's sparsity and enhancing the accuracy of the ranking recommendation [85].

The experiments are used to observe how their recommended accuracy changes when the number of iterations increases. In the FCTR-LFM algorithm, the fixed-parameter confidence level $\lambda = 0.5$, the number of latent factor $F = 80$, the regularization parameter $w = 0.002$, and the learning rate $a = 0.006$. In the BPR, FISM, and FST algorithms, the optimal parameter settings described in the literature [61]–[63] are used. The experimental results are shown in Figures 9-11.

FIGURE 9. P@N comparison. It can be seen from the figure that with the increase of iteration number (N), the recommendation accuracy of the four algorithms decreases. However, FCTR-LFM is relatively slightly higher than other algorithms.

FIGURE 10. R@N comparison. As the iteration number (N) increases, the four algorithms' recall rates are all increase.

The above figures show that we combine fuzzy clustering with the LFM model to improve the FCTR-LFM model's

FIGURE 11. F1@N comparison. The value of F1 is the harmonic average of precision and recall.

design. In the TRDs, we compare the existing related models after the optimal parameters are selected. Some advantages can be observed in the recommendation accuracy of course teachers.

5) COMPARISON WITH USER-BASED CF, ITEM-BASED CF, AND PERSONALRANK

The CF algorithm refers to a recommendation algorithm designed based on user behavior data, which mainly includes: neighborhood-based algorithms, latent factor model, and graph-based random walk algorithm. Typical neighborhood-based algorithms are user-based collaborative filtering algorithms (User-based CF) and item-based collaborative filtering algorithms (Item-based CF) [86]. In this paper, the user is the teacher, and the item is the course. The typical representative of the latent factor model is LFM. The FCTR-LFM is based on LFM. The typical representative of the graph-based random walk algorithm is PersonalRank [87].

In the experiment, we fixed the parameters α, w, N, and F in the FCTR-LFM algorithm as 0.006, 0.002, 80, and 100, respectively. We fixed the parameter probability p in PersonalRank as 0.6. We compare the P@5 of these four algorithms on different sparse evaluation matrices. To obtain evaluation matrices with different sparsity degrees, we can consider adjusting the confidence level value λ to increase from 0 to 1 during teacher fuzzy clustering, with a step size of 0.1. After using different teacher clusters to reconstruct the evaluation matrix, the evaluation matrix's sparsity will be different. The results are shown in Figure 12.

As shown in Figure 12, with the change of confidence level (λ), the number of teacher clusters will differ. The sparsity degree of the reconstructed evaluation matrix will be different. When user-based CF and item-based CF are used to recommend the sparse matrix, their recommendation accuracy is low. However, the FCTR-LFM algorithm can give more accurate prediction values for all items in the sparse matrix, making the matrix dense, so the recommendation

accuracy is high. The PersonalRank algorithm will randomly walk according to the probability p, establish the edges between the teachers and the courses, calculate the PR value between all courses that the teachers can be walked. And then, it recommends according to the size of the PR value, so its recommendation accuracy is higher than User-based CF and Item-based CF but lower than our designed FCTR-LFM.

FIGURE 12. Compare the P@5 of FCTR-LFM, PersonalRank, User-based CF, and Item-based CF at different confidence level values (λ).

6) RECOMMENDED CASE

In our dataset TRDs, take the example of recommending five teachers for a course numbered C015 ($Cdf = 0.644316$). The course's difficulty co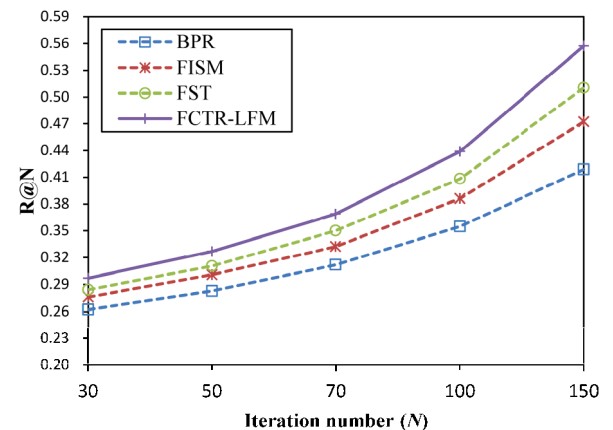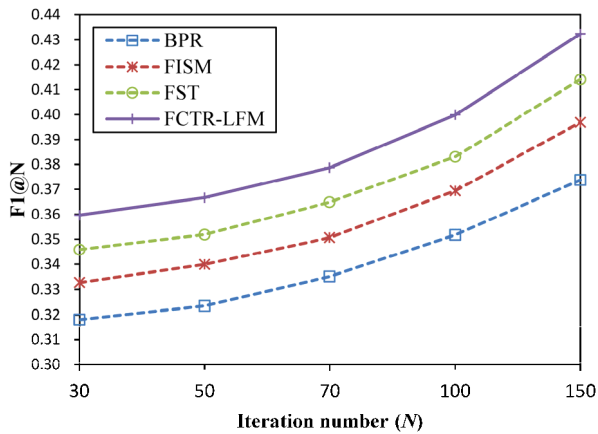efficient, the characteristics of the teachers who taught the course, and the course evaluation scores obtained by these teachers are shown in Table 4.

As shown from the table, these four teachers have different characteristics and get different teaching scores (Ts) for the course C015. The remaining teachers who are not listed have not taught the course. In the experiment, we fixed the parameters λ, F, N, a, and w as 0.9, 80, 50, 0.006, and 0.002. The operation process is as follows:

Step 1: realize the fuzzy clustering of teachers.

(i) The teacher characteristics matrix X^* is standardized to obtain the feature index matrix X', as shown at the bottom of the next page.

(ii) Construct the teacher's fuzzy similar matrix Y

$$Y = \begin{bmatrix} 1 & 0.66 & 0.62 & \cdots & 0.63 & 0.76 & 0.70 \\ 0.66 & 1 & 0.61 & \cdots & 0.69 & 0.69 & 0.71 \\ 0.62 & 0.61 & 1 & \cdots & 0.90 & 0.48 & 0.79 \\ \cdots & \cdots & \cdots & \ddots & \cdots & \cdots & \cdots \\ 0.63 & 0.69 & 0.90 & \cdots & 1 & 0.52 & 0.81 \\ 0.76 & 0.69 & 0.48 & \cdots & 0.52 & 1 & 0.64 \\ 0.70 & 0.71 & 0.79 & \cdots & 0.81 & 0.64 & 1 \end{bmatrix}$$

(iii) Fuzzy clustering

First, the square self-synthesis method was used to work out the transitive closure $t(Y)$ of fuzzy similarity matrix Y. Then, the appropriate confidence level value λ is appropriately selected to find the λ truncated matrix $t(Y)_{\lambda}$. The results

TABLE 4. Current teaching teachers and teaching scores of course C015 (*Cdf* = 0.644316).

Teacher	Ebf	Df	Ptf	Gf	Af	Taf	Jbf	Pmf	Ls	Es	Js	Gs	Locs	Libs	Cs	Ts
T004	0.7	0.7	0.8	1	0.3	0.7	0.39	0.8	32.5	34	31	33.6	34	32.1	33	0.77
T013	0.7	0.7	0.5	1	0.3	0.5	0.23	0.8	34.8	35.4	33.5	35.1	34.5	35.1	34.8	**0.92**
T053	0.5	0.7	0.8	1	0.7	0.9	0.28	0.5	38.6	33.9	36.5	34.7	36.5	38.6	33.3	0.87
T058	1	1	0.8	1	0.6	0.7	0.39	0.8	33.7	34.3	36.8	32.3	34.5	35.5	34.6	0.79

are as follows.

$$
t(Y) = \begin{bmatrix}
1 & 0.86 & 0.90 & \cdots & 0.91 & 0.85 & 0.91 \\
0.86 & 1 & 0.86 & \cdots & 0.86 & 0.85 & 0.86 \\
0.90 & 0.86 & 1 & \cdots & 0.90 & 0.85 & 0.90 \\
\cdots & \cdots & \cdots & \ddots & \cdots & \cdots & \cdots \\
0.90 & 0.86 & 0.90 & \cdots & 1 & 0.85 & 0.91 \\
0.85 & 0.85 & 0.85 & \cdots & 0.85 & 1 & 0.85 \\
0.91 & 0.86 & 0.91 & \cdots & 0.91 & 0.85 & 1
\end{bmatrix}
$$

$$
t(Y)_{0.9} = \begin{bmatrix}
1 & 0 & 1 & \cdots & 1 & 0 & 1 \\
0 & 1 & 0 & \cdots & 0 & 0 & 0 \\
1 & 0 & 1 & \cdots & 1 & 0 & 1 \\
\cdots & \cdots & \cdots & \ddots & \cdots & \cdots & \cdots \\
1 & 0 & 1 & \cdots & 1 & 0 & 1 \\
0 & 0 & 0 & \cdots & 0 & 1 & 0 \\
1 & 0 & 1 & \cdots & 1 & 0 & 1
\end{bmatrix}
$$

We can get 19 teacher clusters from $t(Y)_{0.9}$, which are denoted as $Z = \{z_1, z_2, \ldots, z_{19}\}$. Among them, T013, T053, and t058 are in the same cluster $z_7 = \{$T013, T039, T058, T053, T011, T089, T070, T095, T037$\}$, and T004 is assigned to another cluster $z_{13} = \{$T042, T004, T108, T001, T109$\}$. The order of teachers in each cluster has been sorted by similarity.

Step 2: use LFM to predict the evaluation matrix.

(i) Teachers' clustering is used to reconstruct the evaluation matrix.

We use the teacher cluster to reconstruct the evaluation matrix. According to (22), we first calculate the teaching score between the teacher cluster and the corresponding course and then multiply it by the course difficulty coefficient

as the evaluation score. In this way, the original matrix size of 132 × 204 is reduced to 19 × 204, which greatly reduces the evaluation matrix scale, and obtains the following evaluation matrix $R_{19 \times 204}$

$$
R_{19 \times 204} = \begin{bmatrix}
 & & & 0.4961 & & \cdots \\
 & & & & 0.5395 & \cdots \\
 & & 0.7481 & & & \cdots \\
0.5376 & & & 0.5541 & & \cdots \\
\cdots & \cdots & \cdots & \cdots & \cdots & \ddots & \cdots \\
 & & & & & \cdots & 0.6437
\end{bmatrix}
$$

In R, the evaluation score of the z_7 cluster on the course C015 is 0.5541, and that of the z_{13} cluster is 0.4961.

(ii) Use LFM to predict the evaluation matrix $\hat{R}_{19 \times 204}$.

$$
\hat{R}_{19 \times 204} = \begin{bmatrix}
0.4102 & 0.3901 & 0.3317 & 0.4974 & 0.4098 & \cdots & 0.3074 \\
0.3856 & 0.5293 & 0.2901 & 0.0236 & 0.5401 & \cdots & 0.0683 \\
0.1476 & 0.7514 & 0.4534 & 0.2143 & 0.2563 & \cdots & 0.4817 \\
0.5382 & 0.0341 & 0.3844 & 0.5583 & 0.5012 & \cdots & 0.2911 \\
\cdots & \cdots & \cdots & \cdots & \cdots & \ddots & \cdots \\
0.2367 & 0.5461 & 0.2367 & 0.4591 & 0.3714 & \cdots & 0.6523
\end{bmatrix}
$$

It can be seen from $\hat{R}_{19 \times 204}$ that the teacher cluster z_7 has the highest score of 0.5583 for the predictive evaluation of the course C015.

Step 3: recommend the TOP-5 teachers for the C015 course.

Therefore, we recommend the TOP-5 teachers in cluster z_7 as the teachers of the course C015. Among them, T039 and

$$
X^* = \begin{bmatrix}
0.7 & 0.7 & 1.0 & 1.0 & 0.7 & 0.9 & 0.29 & 0.8 & 32.29 & 31.00 & 32.71 & 31.86 & 33.43 & 32.14 & 30.57 \\
1.0 & 1.0 & 0.8 & 1.0 & 1.0 & 1.0 & 0.93 & 1.0 & 29.23 & 31.59 & 35.23 & 36.40 & 45.00 & 40.63 & 41.43 \\
0.7 & 0.7 & 0.5 & 0.5 & 0.4 & 0.5 & 0.25 & 0.8 & 35.49 & 43.61 & 34.62 & 31.85 & 28.41 & 32.75 & 34.47 \\
0.7 & 0.7 & 0.8 & 1.0 & 0.3 & 0.7 & 0.39 & 0.8 & 32.50 & 34.00 & 31.00 & 33.63 & 34.00 & 32.13 & 33.00 \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
0.7 & 0.7 & 0.5 & 1.0 & 0.8 & 0.9 & 0.29 & 0.5 & 42.33 & 43.00 & 41.00 & 42.67 & 42.67 & 43.00 & 40.67
\end{bmatrix}
$$

$$\Downarrow$$

$$
X' = \begin{bmatrix}
-0.28 & -0.49 & 2.32 & 0.70 & 0.69 & 0.80 & -0.52 & 0.09 & -1.06 & -1.10 & -0.98 & -0.97 & -0.77 & -1.12 & -1.22 \\
1.94 & 1.85 & 1.17 & 0.70 & 2.19 & 1.27 & 2.28 & 1.75 & -1.77 & -0.97 & -0.39 & 0.04 & 1.85 & 0.82 & 1.11 \\
-0.28 & -0.49 & -0.55 & -1.44 & -0.80 & -1.07 & -0.70 & 0.09 & -0.31 & 1.69 & -0.54 & -0.97 & -1.91 & -0.99 & -0.39 \\
-0.28 & -0.49 & 1.17 & 0.70 & -1.30 & -0.13 & -0.09 & 0.09 & -1.01 & -0.43 & -1.39 & -0.58 & -0.64 & -1.13 & -0.70 \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \vdots \\
-0.28 & -0.49 & -0.55 & 0.70 & 1.19 & 0.80 & -0.52 & -2.39 & 1.28 & 1.56 & 0.96 & 1.42 & 1.32 & 1.37 & 0.95
\end{bmatrix}
$$

T011 are the teachers who have not taught the course before, while the original teacher T004 is no longer recommended.

VI. CONCLUSION

It is an important means to improve colleges' and universities' teaching quality to arrange suitable teaching teachers scientifically. It is also the goal pursued by the teaching management of colleges and universities. However, to achieve this goal, there are still greater difficulties. The difficulties mainly lie in quantifying teachers' characteristics, course difficulty, and teaching performance effectively, establishing an implicit correlation model among teachers, courses, and teaching performance, and how to use algorithms to make predictions and recommendations efficiently and accurately. These difficulties have not been well solved. Therefore, under the guidance of pedagogy theories and methods, we scientifically established the quantitative models of teacher characteristics and course difficulty. Combined it with historical evaluation data and established a high-dimensional large-scale sparse course teacher recommendation evaluation matrix as the data set, which implied the correlation among the three. On this basis, we construct the FCTR-LFM algorithm based on fuzzy clustering and LFM. The algorithm firstly adopts the improved fuzzy clustering to realize automatic clustering according to the characteristics of teachers. It then uses the clustered teacher cluster to reconstruct the dataset to reduce the data set's size and sparsity. Then the improved LFM is used to predict the teaching evaluation score between the teacher cluster and each course, which improves the prediction efficiency. Finally, the improved TOP-N recommendation method is used to sort the teacher clusters according to the course and the predicted evaluation. Then the TOP-N recommendation is realized according to the teacher similarity degree within the cluster.

Experimental results show that this algorithm can effectively implement the recommendation of course teachers and has the following characteristics:

(i) According to teachers' characteristics, the improved fuzzy clustering method can realize the unsupervised automatic clustering of teachers, which reflects the flexibility. It solves the problem of cold start and avoids the disadvantage of using other clustering methods to force teachers to be divided into the specified number of categories.

(ii) The teacher cluster after clustering is used to reconstruct the data set, which greatly reduces the size of the data set, reduces the sparsity, and avoids the influence of extremely sparse data on the recommendation results.

(iii) Using the improved LFM model, the recommended matrix is decomposed into the product of two low-order matrices with an implicit factor association, which improves the prediction efficiency and accuracy.

(iv) The improved TOP-N recommendation algorithm is used to sort the teacher clusters according to the course. The teacher clusters are then sorted according to the teacher similarity, which realizes the TOP-N recommendation and improves the recommendation's efficiency and accuracy.

Although this method solves the problem of scientifically recommending suitable teachers to the course, it also has some limitations:

(i) The workload of data collection is heavy. Teachers' characteristics are conducive to solving problems not only depend on the methods and results of previous research on teacher characteristics but also on whether the methods of sorting out the results of questionnaire data are properly used. Besides, the course difficulty coefficient's acquisition also depends on the knowledge structure in the course syllabus.

(ii) The parameter optimization of the algorithm is excessively dependent on the data set. The algorithm can obtain different optimization parameters on different data sets, which need to be determined by many experiments.

To sum up, our future research focuses include: (i) Further excavate the influencing factors of recommending suitable teachers for the course. (ii) Build an effective method to simplify the data acquisition process. (iii) Find more effective methods to improve the accuracy of prediction and recommendation.

REFERENCES

[1] R. Chetty, J. N. Friedman, and J. E. Rockoff, "Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood," *Amer. Econ. Rev.*, vol. 104, no. 9, pp. 2633–2679, 2014.

[2] C. T. Clotfelter, H. F. Ladd, and J. L. Vigdor, "Teacher credentials and student achievement: Longitudinal analysis with student fixed effects," *Econ. Edu. Rev.*, vol. 26, no. 6, pp. 673–682, Dec. 2007.

[3] D. N. Harris and T. R. Sass, "Teacher training, teacher quality and student achievement," *J. Public Econ.*, vol. 95, nos. 7–8, pp. 798–812, Aug. 2011.

[4] M. Gao and Q. Liu, "Effective teacher attributes: A comparison of American and Chinese secondary teacher candidates' perceptions," *J. Int. Soc. Teacher Educ.*, vol. 16, no. 2, pp. 85–95, 2012.

[5] S. A. Basow and N. T. Silberg, "Student evaluations of college professors: Are female and male professors rated differently?" *J. Educ. Psychol.*, vol. 79, no. 3, pp. 308–314, 1987.

[6] P. A. Cohen, "Student ratings of instruction and student achievement: A meta-analysis of multisection validity studies," *Rev. Educ. Res.*, vol. 51, no. 3, pp. 281–309, Sep. 1981.

[7] M. Agaoglu, "Predicting instructor performance using data mining techniques in higher education," *IEEE Access*, vol. 4, pp. 2379–2387, 2016.

[8] R. T. Lee and B. E. Ashforth, "A meta-analytic examination of the correlates of the three dimensions of job burnout," *J. Appl. Psychol.*, vol. 81, no. 2, p. 123, 1996.

[9] S. M. Azeem and N. A. Nazir, "A study of job burnout among university teachers," *Psychol. Developing Societies*, vol. 20, no. 1, pp. 51–64, Jan. 2008.

[10] R. J. Sternberg and E. L. Grigorenko, "Styles of thinking in the school," *Eur. J. High Ability*, vol. 6, no. 2, pp. 201–219, 1995.

[11] L. Sun, K. Duan, J. Zhou, and J. Chen, "Data analysis of university curriculum knowledge points based on multidimensional structure," *J. Jinzhong Univ.*, vol. 37, no. 3, pp. 59–62, 2020.

[12] B. Greimel-Fuhrmann and A. Geyer, "Students' evaluation of teachers and instructional quality-analysis of relevant factors based on empirical evaluation research," *Assessment Eval. Higher Educ.*, vol. 28, no. 3, pp. 229–238, Jun. 2003.

[13] P. Spooren and D. Mortelmans, "Teacher professionalism and student evaluation of teaching: Will better teachers receive higher ratings and will better students give higher ratings?" *Educ. Stud.*, vol. 32, no. 2, pp. 201–214, Jun. 2006.

[14] J. Wu and Q. Feng, "Recommendation system design for college network education based on deep learning and fuzzy uncertainty," *J. Intell. Fuzzy Syst.*, vol. 38, no. 6, pp. 7083–7094, Jun. 2020.

[15] A. Oliveira, M. Teixeira, and C. Neto, "Recommendation of educational content to improve student performance: An approach based on learning styles," in *Proc. 12th Int. Conf. Comput. Supported Edu.*, Prague, Czech Republic, 2020, pp. 359–365.

[16] I. Taufik, Y. A. Gerhana, A. I. Ramdani, and M. Irfan, "Implementation K-nearest neighbor for student expertise recommendation system," *J. Phys., Conf. Ser.*, vol. 1402, no. 7, pp. 1–5, 2019.

[17] W. Xu and Y. Zhou, "Course video recommendation with multimodal information in online learning platforms: A deep learning framework," *Brit. J. Educ. Technol.*, vol. 51, no. 5, pp. 1734–1747, Sep. 2020.

[18] X. Pan, X. Li, and M. Lu, "A MultiView courses recommendation system based on deep learning," in *Proc. Int. Conf. Big Data Informatization Educ. (ICBDIE)*, Zhangjiajie, China, Apr. 2020, pp. 502–506.

[19] J. Wang, H. Xie, O. T. S. Au, D. Zou, and F. L. Wang, "Attention-based CNN for personalized course recommendations for MOOC learners," in *Proc. Int. Symp. Educ. Technol. (ISET)*, Bangkok, Thailand, Aug. 2020, pp. 180–184.

[20] S. Alghamdi, N. Alzhrani, and H. Algethami, "Fuzzy-based recommendation system for university major selection," in *Proc. 11th Int. Joint Conf. Comput. Intell.*, Vienna, Austria, 2019, pp. 317–324.

[21] J. Wan, Q. Sun, X. Li, J. Ding, and Q. Zhu, "Personalized professional recommendation system based on undergraduate questionnaires," in *Proc. 17th Int. Symp. Distrib. Comput. Appl. Bus. Eng. Sci. (DCABES)*, Wuxi, China, Oct. 2018, pp. 140–143.

[22] S. Sergis and D. G. Sampson, "Learning object recommendations for teachers based on elicited ICT competence profiles," *IEEE Trans. Learn. Technol.*, vol. 9, no. 1, pp. 67–80, Jan. 2016.

[23] 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.

[24] N. Z. Zacharis, "Classification and regression trees (CART) for predictive modeling in blended learning," *Int. J. Intell. Syst. Appl.*, vol. 10, no. 3, pp. 1–9, Mar. 2018.

[25] M. Imran, S. Latif, D. Mehmood, and M. S. Shah, "Student academic performance prediction using supervised learning techniques," *Int. J. Emerg. Technol. Learn.*, vol. 14, no. 14, pp. 92–104, 2019.

[26] N. A. Mozahem, "Using learning management system activity data to predict student performance in face-to-face courses," *Int. J. Mobile Blended Learn.*, vol. 12, no. 3, pp. 20–31, Jul. 2020.

[27] Y. C. Chiu, H. J. Hsu, J. Wu, and D. L. Yang, "Predicting student performance in MOOCs using learning activity data," *J. Inf. Sci. Eng.*, vol. 34, no. 5, pp. 1223–1235, 2018.

[28] A. Almasri, E. Celebi, and R. S. Alkhawaldeh, "EMT: Ensemble meta-based tree model for predicting student performance," *Sci. Program.*, vol. 2019, pp. 1–13, Feb. 2019.

[29] L. Qiu, Y. Liu, Q. Hu, and Y. Liu, "Student dropout prediction in massive open online courses by convolutional neural networks," *Soft Comput.*, vol. 23, no. 20, pp. 10287–10301, Oct. 2019.

[30] Y. Ma, C. Cui, J. Yu, J. Guo, G. Yang, and Y. Yin, "Multi-task MIML learning for pre-course student performance prediction," *Frontiers Comput. Sci.*, vol. 14, no. 5, pp. 1–10, Oct. 2020.

[31] N. Z. Zacharis, "A multivariate approach to predicting student outcomes in Web-enabled blended learning courses," *Internet Higher Educ.*, vol. 27, pp. 44–53, Oct. 2015.

[32] C. Verma, V. Stoffová, and Z. Illés, "Prediction of students' awareness level towards ICT and mobile technology in Indian and Hungarian University for the real-time: Preliminary results," *Heliyon*, vol. 5, no. 6, pp. 1–9, 2019.

[33] 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.

[34] C. Verma, V. Stoffova, Z. Illes, S. Tanwar, and N. Kumar, "Machine learning-based student's native place identification for real-time," *IEEE Access*, vol. 8, pp. 130840–130854, 2020.

[35] Y. Zhu, H. Lu, P. Qiu, K. Shi, J. Chambua, and Z. Niu, "Heterogeneous teaching evaluation network based offline course recommendation with graph learning and tensor factorization," *Neurocomputing*, vol. 415, pp. 84–95, Nov. 2020.

[36] S. Wan and Z. Niu, "A hybrid E-learning recommendation approach based on Learners' influence propagation," *IEEE Trans. Knowl. Data Eng.*, vol. 32, no. 5, pp. 827–840, May 2020.

[37] A. Esteban, A. Zafra, and C. Romero, "Helping university students to choose elective courses by using a hybrid multi-criteria recommendation system with genetic optimization," *Knowl.-Based Syst.*, vol. 194, pp. 1–14, Apr. 2020.

[38] S. M. Nafea, F. Siewe, and Y. He, "A novel algorithm for course learning object recommendation based on student learning styles," in *Proc. Int. Conf. Innov. Trends Comput. Eng. (ITCE)*, Aswan, Egypt, Feb. 2019, pp. 192–201.

[39] H. Drachsler, K. Verbert, O. C. Santos, and N. Manouselis, "Panorama of recommender systems to support learning," in *Recommender Systems Handbook*. Berlin, Germany: Springer, 2015, pp. 421–451.

[40] D. Ben-Shimon, L. Rokach, and B. Shapira, "An ensemble method for top-N recommendations from the SVD," *Expert Syst. Appl.*, vol. 64, pp. 84–92, Dec. 2016.

[41] W. Zhang, J. Wang, and W. Feng, "Combining latent factor model with location features for event-based group recommendation," in *Proc. 19th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD)*, New York, NY, USA, 2013, pp. 910–918.

[42] K. Krishna and M. N. Murty, "Genetic K-means algorithm," *IEEE Trans. Syst. Man, Cybern. B, Cybern.*, vol. 29, no. 3, pp. 433–439, Jun. 1999.

[43] S.-Q. Wang and D.-M. Zhu, "Research on selecting initial points for k-means clustering," in *Proc. Int. Conf. Mach. Learn. Cybern.*, Madison, WI, USA, Jul. 2008, pp. 91–99.

[44] D. Pelleg and A. W. Moore, "X-means: Extending k-means with efficient estimation of the number of clusters," in *Proc. ICML*, San Francisco, CA, USA, vol. 1, 2000, pp. 727–734.

[45] P. Berkhin and J. D. Becher, "Learning simple relations: Theory and applications," in *Proc. SIAM Int. Conf. Data Mining*, Arlington, VA, USA, Apr. 2002, pp. 420–436.

[46] H. H. Nguyen, "Privacy-preserving mechanisms for k-modes clustering," *Comput. Secur.*, vol. 78, pp. 60–75, Sep. 2018.

[47] C. Fraley and A. E. Raftery, "Model-based clustering, discriminant analysis, and density estimation," *J. Amer. Stat. Assoc.*, vol. 97, no. 458, pp. 611–631, Jun. 2002.

[48] J. E. Gentle, "The EM algorithm and extensions," *Biometrics*, vol. 54, no. 1, p. 395, 1998.

[49] Y. Zhou, S. Xu, C. Jin, and Z. Guo, "Multiple point sets registration based on expectation maximization algorithm," *Comput. Electr. Eng.*, vol. 70, pp. 1–11, Aug. 2018.

[50] A. Ben-Dor, R. Shamir, and Z. Yakhini, "Clustering gene expression patterns," *J. Comput. Biol.*, vol. 6, nos. 3–4, pp. 281–297, 1999.

[51] R. Sharan and R. Shamir, "CLICK: A clustering algorithm with applications to gene expression analysis," in *Proc. ISMB*, San Diego, CA, USA, vol. 8, 2000, pp. 307–316.

[52] M. N. Nguyen, A. Y. L. Sim, Y. Wan, M. S. Madhusudhan, and C. Verma, "Topology independent comparison of RNA 3D structures using the CLICK algorithm," *Nucleic Acids Res.*, vol. 45, no. 1, p. e5, Jan. 2017.

[53] C. Ding and K. Li, "Centrality ranking in multiplex networks using topologically biased random walks," *Neurocomputing*, vol. 312, pp. 263–275, Oct. 2018.

[54] R. Shang, Z. Zhang, L. Jiao, W. Wang, and S. Yang, "Global discriminative-based nonnegative spectral clustering," *Pattern Recognit.*, vol. 55, pp. 172–182, Jul. 2016.

[55] C. Lee, O. Zaiane, H. Park, J. Huang, and R. Greiner, "Clustering high dimensional data: A graph-based relaxed optimization approach," *Inf. Sci.*, vol. 178, no. 23, pp. 4501–4511, Dec. 2008.

[56] D. Shi, J. Wang, D. Cheng, and J. Gao, "A global-local affinity matrix model via EigenGap for graph-based subspace clustering," *Pattern Recognit. Lett.*, vol. 89, pp. 67–72, Apr. 2017.

[57] J. C. Bezdek, R. Ehrlich, and W. Full, "FCM: The fuzzy C-means clustering algorithm," *Comput. Geosci.*, vol. 10, nos. 2–3, pp. 191–203, Jan. 1984.

[58] R. J. Hathaway and J. C. Bezdek, "Fuzzy C-means clustering of incomplete data," *IEEE Trans. Syst., Man Cybern., B, Cybern.*, vol. 31, no. 5, pp. 735–744, 2001.

[59] R. Pan, Y. Zhou, B. Cao, N. N. Liu, R. Lukose, M. Scholz, and Q. Yang, "One-class collaborative filtering," in *Proc. 8th IEEE Int. Conf. Data Mining*, Pisa, Italy, Dec. 2008, pp. 502–511.

[60] Y. Hu, Y. Koren, and C. Volinsky, "Collaborative filtering for implicit feedback datasets," in *Proc. 8th IEEE Int. Conf. Data Mining*, Pisa, Italy, Dec. 2008, pp. 263–272.

[61] X. Ning and G. Karypis, "SLIM: Sparse linear methods for top-N recommender systems," in *Proc. IEEE 11th Int. Conf. Data Mining*, Vancouver, BC, Canada, Dec. 2011, pp. 497–506.

[62] S. Kabbur, X. Ning, and G. Karypis, "FISM: Factored item similarity models for top-N recommender systems," in *Proc. 19th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD)*, Chicago, IL, USA, 2013, pp. 659–667.

[63] S. Rendle, C. Freudenthaler, Z. Gantner, and L. Schmidt-Thieme, "BPR: Bayesian personalized ranking from implicit feedback," in *Proc. UAI*, Montreal, QC, Canada, 2009, pp. 452–461.

[64] W. Pan and L. Chen, "GBPR: Group preference based Bayesian personalized ranking for one-class collaborative filtering," in *Proc. IJCAI*, Beijing, China, 2013, pp. 2691–2697.

[65] Y. Shi, A. Karatzoglou, L. Baltrunas, M. Larson, N. Oliver, and A. Hanjalic, "CLiMF: Collaborative less-is-more filtering," in *Proc. IJCAI*, Beijing, China, 2013, pp. 3077–3081.

[66] Y. Shi, A. Karatzoglou, L. Baltrunas, M. Larson, A. Hanjalic, and N. Oliver, "TFMAP: Optimizing MAP for top-n context-aware recommendation," in *Proc. 35th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr. (SIGIR)*, Portland, OR, USA, 2012, pp. 155–164.

[67] T. Zhao, J. McAuley, and I. King, "Leveraging social connections to improve personalized ranking for collaborative filtering," in *Proc. 23rd ACM Int. Conf. Conf. Inf. Knowl. Manage. (CIKM)*, Shanghai, China, 2014, pp. 261–270.

[68] D. Song and D. A. Meyer, "Recommending positive links in signed social networks by optimizing a generalized AUC," in *Proc. AAAI*, Austin, TX, USA, 2015, pp. 290–296.

[69] H. Liu, Z. Wu, and X. Zhang, "CPLR: Collaborative pairwise learning to rank for personalized recommendation," *Knowl.-Based Syst.*, vol. 148, pp. 31–40, May 2018.

[70] L. Yu, C. Zhang, S. Pei, G. Sun, and X. Zhang, "Walkranker: A unified pairwise ranking model with multiple relations for item recommendation," in *Proc. AAAI*, New Orleans, LA, USA, 2018, pp. 2596–2603.

[71] A. Gulliver, L. Farrer, K. Bennett, and K. M. Griffiths, "University staff mental health literacy, stigma and their experience of students with mental health problems," *J. Further Higher Educ.*, vol. 43, no. 3, pp. 434–442, Mar. 2019.

[72] C. H. D. Larenas, A. V. R. Moran, and K. J. P. Rivera, "Comparing teaching styles and personality types of EFL instructors in the public and private sectors," *Profile Issues Teachers Prof. Develop.*, vol. 13, no. 1, pp. 111–127, 2011.

[73] N. Michel, J. J. Cater, and O. Varela, "Active versus passive teaching styles: An empirical study of student learning outcomes," *Hum. Resour. Develop. Quart.*, vol. 20, no. 4, pp. 397–418, Sep. 2009.

[74] W. Zhao, "Research on matching college teachers' teaching style with students' learning style," *J. Beijing Polytech. College*, vol. 16, no. 1, pp. 90–93, 2017.

[75] K. S. Taber, "The use of Cronbach's alpha when developing and reporting research instruments in science education," *Res. Sci. Educ.*, vol. 48, no. 6, pp. 1273–1296, Dec. 2018.

[76] M. Novak and D. Oreški, "Fuzzy knowledge-based system for calculating course difficulty based on student perception," *Comput. Appl. Eng. Educ.*, vol. 24, no. 2, pp. 225–233, Mar. 2016.

[77] N. Z. Shi, F. Z. Kong, and S. W. Li, "A model of degree of difficulty of course: Contrast of difficulty degree of geometry course in China's compulsory education," *J. Northeast Normal Univ. (Social Sci.)*, vol. 6, pp. 151–155, 2005.

[78] G. F. Li, "Deviation and correction in the application of course difficulty model–discussion with professor Shi Ningzhong," (in Chinese), *Shanghai Res. Educ.*, no. 3, pp. 50–53, 2010, doi: 10.3969/j.issn.1007-2020.2010.03.013.

[79] D. L. Davies and D. W. Bouldin, "A cluster separation measure," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-1, no. 2, pp. 224–227, Apr. 1979.

[80] H. Steck, "Evaluation of recommendations: Rating-prediction and ranking," in *Proc. 7th ACM Conf. Recommender Syst. (RecSys)*, Hong Kong, 2013, pp. 213–220.

[81] Q. Wang and J. H. Wang, "Collaborative filtering recommendation algorithm combining trust mechanism with user preferences," *Comput. Eng. Appl.*, vol. 51, no. 10, pp. 261–265, 2015.

[82] Y. Pan, F. He, and H. Yu, "Social recommendation algorithm using implicit similarity in trust," *Chin. J. Comput.*, vol. 41, no. 1, pp. 65–81, 2018.

[83] A. Krohn-Grimberghe, L. Drumond, C. Freudenthaler, and L. Schmidt-Thieme, "Multi-relational matrix factorization using Bayesian personalized ranking for social network data," in *Proc. 5th ACM Int. Conf. Web Search Data Mining (WSDM)*, Seattle, WA, USA, 2012, pp. 173–182.

[84] S. Kabbur, X. Ning, and G. Karypis, "FISM: Factored item similarity models for top-N recommender systems," in *Proc. 19th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD)*, New York, NY, USA, 2013, pp. 659–667.

[85] G. Guo, J. Zhang, F. Zhu, and X. Wang, "Factored similarity models with social trust for top-N item recommendation," *Knowl.-Based Syst.*, vol. 122, pp. 17–25, Apr. 2017.

[86] M. Papagelis and D. Plexousakis, "Qualitative analysis of user-based and item-based prediction algorithms for recommendation agents," *Eng. Appl. Artif. Intell.*, vol. 18, no. 7, pp. 781–789, Oct. 2005.

[87] J. Hu, L. Liu, C. Zhang, J. He, and C. Hu, "Hybrid recommendation algorithm based on latent factor model and PersonalRank," *J. Internet Technol.*, vol. 19, no. 3, pp. 919–926, 2018.

DUNHONG YAO was born in 1972. He received the M.S. degree in software engineering from Chongqing University, Chongqing, China, in 2008.

He is currently a Senior Research Professional with the Key Laboratory of Wuling-Mountain Health Big Data Intelligent Processing and Application in Hunan Province Universities, a Research Assistant with the Key Laboratory of Intelligent Control Technology for Wuling-Mountain Ecological Agriculture in Hunan Province, and an Associate Professor with the School of Computer Science and Engineering, Huaihua University. His research interests include machine learning, education data mining, prediction and recommendation, and artificial intelligence techniques.

XIAOWU DENG was born in 1973. He received the M.S. degree in control theory and control engineering from the Shandong University of Science and Technology, Shandong, China, in 2007, and the Ph.D. degree in control science and engineering from the Automation College, Hangzhou Dianzi University, Hangzhou, China.

He is currently a Senior Research Professional with the Key Laboratory of Wuling-Mountain Health Big Data Intelligent Processing and Application in Hunan Province Universities, a Research Assistant with the Key Laboratory of Intelligent Control Technology for Wuling-Mountain Ecological Agriculture in Hunan Province, and an Associate Professor with the School of Computer Science and Engineering, Huaihua University. His research interests include tensor learning, intelligent fault diagnostics, and big data analysis.