Preference Cognitive Diagnosis for Student Performance Prediction

PEICHAO JIANG1 AND XIAODONG WANG2
1Key Laboratory of Artificial Intelligence and Personalized Learning in Education of Henan Province, Henan Normal University, Xinxiang 453007, China
2Big Data Engineering Lab of Teaching Resources & Assessment of Education Quality, Henan Province, Henan Normal University, Xinxiang 453007, China
Corresponding author: Xiaodong Wang (wxd@htu.edu.cn)

This work was supported in part by the Henan Normal University through the Cooperation Project, and in part by the Zhengzhou No. 2 Senior High School under Grant 520119160001 and Grant 5202069169001.

ABSTRACT Knowledge states modeling is a fundamental issue in online education. One of its tasks is to discover the potential knowledge capacity of students for predicting their performance (i.e., scores on exercises). Current studies either depend on cognitive diagnosis approaches or apply collaborative filtering. However, the prediction accuracy of traditional cognitive diagnosis is insufficient, and collaborative filtering has difficulty ensuring the interpretability of prediction. Actually, students usually read some auxiliary text learning materials that they are interested in, namely, preferred learning material, to consolidate what they have learned. Preference cognitive diagnosis means that the preferred learning materials can reflect the students’ knowledge states (i.e., proficiency for knowledge concepts) to some extent, which is beneficial for predicting students’ performance. Therefore, we propose a preference cognitive diagnosis method (PreferenceCD) to model students’ knowledge states. Specifically, we first design the Direct-Indirect method to acquire students’ preferred learning materials. This method mines important information from students’ reading content that can reflect their preference for learning materials to acquire those preferred learning materials directly. Moreover, it discovers preferred learning materials indirectly by analyzing the similarity of students’ learning behaviors during the reading process. Subsequently, we calculate students’ preference degree for knowledge concepts based on the acquired preferred learning materials and diagnose their proficiency for knowledge concepts by applying a cognitive diagnosis model. After that, we combine the above two aspects to model students’ knowledge states and further predict their scores on exercises. The experimental results on a real-world dataset demonstrate the effectiveness of PreferenceCD with both accuracy and interpretability. The accuracy, root mean square error (RMSE), and mean absolute error (MAE) of PreferenceCD are 0.7614, 0.4805, and 0.2386, respectively, which outperforms related works by about 2-12% in terms of these evaluation metrics.

INDEX TERMS Online education, preferred learning material, preference cognitive diagnosis, knowledge states, student performance prediction.

I. INTRODUCTION

Online education platforms, such as massive open online course (MOOC) [1], intelligent tutoring systems [2], [3], and mobile autonomous school (MAS) 1 [4], [5], provide an important means for students’ self-learning and assisted instruction. One of the crucial issues of these educational platforms is to model the knowledge states of students, which aims to discover the potential knowledge capacity of students during the learning process, such as their proficiency on specific knowledge concepts [6]. Generally, the effectiveness of knowledge states modeling can be verified by predicting students’ possible scores on a series of exercises, i.e., predicting student performance (PSP). PSP aims to evaluate whether students can respond to the corresponding exercises correctly (correct is “1” and incorrect is “0”) [6], [7], which can be further extended to some educational applications, such as exercise recommendation [8], [9] and teaching plan improvement [10].

In the literature, massive efforts have been made to solve knowledge states modeling for PSP from both educational psychology and data mining areas. In educational

1Mobile Autonomous School is an online education platform jointly developed by Zhengzhou No. 2 Senior High School, Zhengzhou, China, and Henan Normal University, Xinxiang, China. It takes the tablet computer as the learning terminal, aims to give full play to the enthusiasm and initiative of students in learning under the leading role of teachers.
psychology, cognitive diagnosis models (CDMs) employ students’ response logs on certain exercises to model their knowledge states (e.g., proficiency for knowledge concepts) [11] and predict scores with the Q-matrix (i.e., an exercise-knowledge concept matrix) [12]. In terms of data mining, collaborative filtering (CF) has been devoted to PSP [13]–[15]. Among them, matrix factorization (MF) [16] is a typical prediction technique, which decomposes students’ score matrix into latent feature vectors of students and exercises. Despite the importance of previous studies, there are still some limitations in existing methods. First, since the students’ knowledge states is somewhat concealed, CDMs may have errors in inferring only through the response logs. Thus, CDMs fail to ensure the prediction accuracy. Second, latent feature vectors decomposed by MF are difficult to understand, which makes the prediction results less interpretable, i.e., the correspondence between elements in the latent vector and specific knowledge concepts cannot be clearly described. In summary, the limitation of existing approaches is that they cannot make the PSP with both accuracy and interpretability.

Actually, students usually read some auxiliary text learning materials that they are interested in to consolidate what they have learned (e.g., making up for their lack of proficiency on certain knowledge concepts). As shown in Fig. 1, a student has some unique learning behaviors when reading the learning materials. For example, the student who is interested in certain learning materials will collect them in their favorite list. He will also read some learning materials frequently to improve their learning performance. Therefore, if students collect or often read a specific learning material, it is likely to be the students’ preferred learning material. For better illustration, Table 1 shows a learning material example.

Intuitively, the preferred learning materials can reflect students’ knowledge states, i.e., preference cognitive diagnosis. Fig. 2 displays a toy example of preference cognitive diagnosis. Before responding to exercises, a student read his preferred learning materials (e.g., L1, L2, and L3) related to the specific knowledge concepts examined in the exercises. As the student has a keen interest in these learning materials, he may spend more time and energy reading them. Accordingly, his mastery for these preferred learning materials may be higher. Furthermore, there is a strong correlation between learning materials and exercises (e.g., both are associated with knowledge concepts K1, K2, K3, K4, and K5). After reading, the student may have certain proficiency for the specific knowledge concepts to improve performance (e.g., improve scores on exercise E2 and E3). Therefore, preference cognitive diagnosis can be beneficial to PSP.

Unfortunately, it is not easy to model students’ knowledge states in combination with preference cognitive diagnosis for PSP. In this process, there are several technical challenges. First, as mentioned above, the preferred learning materials of students can reflect their knowledge states, so the acquisition of preferred learning materials is the prerequisite for preference cognitive diagnosis. Therefore, how can students’ preferred learning material be acquired accurately? Second, how can we apply the acquired preferred learning materials to model the students’ knowledge states and obtain effective prediction results?

To address these two-side challenges, in this article, we first design the Direct-Indirect method by adopting direct and indirect strategies to acquire students’ preferred learning materials. Then, we propose a preference cognitive diagnosis method (PreferenceCD) to model students’ knowledge states. Specifically, Direct-Indirect acquires those preferred learning materials directly by mining the important information from students’ reading content that can reflect their preference for learning materials. Additionally, it discovers preferred learning materials indirectly by analyzing the similarity of students’ learning behaviors during the reading process. In PreferenceCD, we first calculate students’ preference degree for knowledge concepts based on the preferred learning materials acquired by Direct-Indirect and diagnose their proficiency for knowledge concepts by using CDM. After that, we combine the above two aspects to model students’ knowledge states and predict their scores on a series of exercises. The contributions of this article can be summarized as follows.

1) For the first time, the concept of preference cognitive diagnosis is considered in predicting student performance.

2) Direct-Indirect adopts direct and indirect strategies to acquire students’ preferred learning materials, which can be effectively applied to preference cognitive diagnosis.

3) Experiments on the real-world dataset from MAS [4], [5] demonstrate that the PreferenceCD improves the accuracy, root mean square error (RMSE), and mean absolute error (MAE) by approximately 2-12% compared with existing works.

| Attribute   | Description                                                                 |
|-------------|-----------------------------------------------------------------------------|
| Type        | History                                                                     |
| Title       | World Literature and Art since the 19th Century 46942986-48B7-4c62-b8c1-e7a0a77b26fa |
| Item ID     | World Literature and Art since the 19th Century has a large period. At this stage, the content includes four aspects of literature, art, music, and television art. Literature and art have greatly enriched the cultural treasure house of mankind and are a precious cultural heritage shared by all mankind... |
| Content     |                                                                             |
The rest of the paper is organized as follows. Section II introduces the related work, and section III illustrates the problem overview. Section IV and Section V detail the proposed method of preferred learning material acquisition (i.e., Direct-Indirect) and student performance prediction (i.e., PreferenceCD), respectively. A thorough analysis of the effectiveness of Direct-Indirect and PreferenceCD as well as a comparison to existing works is given in Section VI. Finally, conclusions are drawn in Section VII.

II. RELATED WORK
Since the learning materials belong to the text type, the acquisition of preferred learning material involves text processing, which requires high extraction ability for the text information. Therefore, we first introduce text mining research and then illustrate the related works on student performance prediction.

A. TEXT MINING
Generally, existing text mining approaches can be divided into two categories: word-based approaches and the sentence-based approaches.

For the word-based approaches, the text content was numerically characterized by relevant information of words (e.g., statistical and semantic information), such as term frequency-inverse document frequency (TF-IDF) [17] and word embedding [18], [19]. Considering the correlation between words in the text, the authors in [20] proposed the latent Dirichlet allocation model (LDA), which added the Dirichlet prior distribution to the polynomial distribution of texts, topics, and words. Sentence-based approaches, such as convolutional neural networks (CNNs) [21], [22] and recurrent neural networks (RNNs) [23], [24], presented the numerical representation of text at the sentence level, which was effectively adopted in natural language processing (NLP) [25], [26]. However, the CNN and RNN input was based on the word vector generated by word embedding methods. Without the support of the sizable specific corpus for training, text mining would be affected.

The corpus of learning materials in the MAS system [4], [5] is insufficient for training the specific word vector model. However, due to the rich content and complex semantics of the corpus, it is necessary to consider the influence of the correlation between words on text mining with learning materials.

B. STUDENT PERFORMANCE PREDICTION
We illustrate existing modeling approaches for PSP from the following two aspects: educational psychology (i.e., cognitive diagnosis) and data mining (i.e., collaborative filtering).

1) COGNITIVE DIAGNOSIS
In educational psychology, cognitive diagnosis performs PSP by discovering students’ knowledge states from their response logs on exercises [11]. Generally, traditional CDMs can be divided into two categories: continuous models and discrete models. Item response theory (IRT) [27], [28] is a typical continuous model that characterizes each student as a variable, i.e., a latent trait that describes students’ comprehensive knowledge states, following a logic-like function. In comparison, discrete models, such as deterministic inputs, and the noisy-and gate model (DINA) [12], [29], represented each student as a binary vector, indicating whether students mastered the relevant knowledge concepts in the Q-matrix. Although the prediction results of traditional CDMs are more interpretable, they are usually not accurate enough. To improve the effectiveness of prediction, many researchers have developed CDMs [3], [6], [30], [31]. For example, the authors in [3] proposed NeuralCD, which applied a neural network to learn the complex interactions between students and exercises. The results tested by the two datasets were approximately 71.9% and 80.4% in accuracy. The authors in [6] designed FuzzyCDF to predict students’ scores on subjective and objective types of exercises. The fuzzy set and educational hypotheses were considered to be an effective means for measuring students’ performance. RMSE was between 0.33-0.42 under different datasets and ratios of testing exercises. However, they only explained the interpretability of the prediction results according to the students’ proficiency for specific knowledge concepts, which is not comprehensive enough.

2) COLLABORATIVE FILTERING
Recently, researchers have attempted to use collaborative filtering in data mining to predict students’ performance. It can be divided into two categories: neighbor-based approaches
and model-based approaches. Among them, neighbor-based approaches, such as K-NearestNeighbor (kNN) [13], might determine students who are similar to the target students according to their response logs on exercises for predicting target students’ scores. Model-based approaches, such as matrix factorization, have been widely applied in PSP [15], [32]–[34]. For example, the latent trait vectors of students and exercises were factored by nonnegative matrix factorization (NMF) [32] and probabilistic matrix factorization (PMF) [33] to predict students’ performance. In [34], a multirelational factorization model was proposed for PSP in an intelligent tutoring system, which exploits several possible relationships between students, exercises, and their metadata. RMSE was between 0.296-0.433 under different datasets. To capture the changes in the process of students responding to exercises, the authors in [35] considered the influence of both learning theory and Ebbinghaus forgetting curve theory and combined them into a unified probability framework. RMSE was between 0.26-0.32 under different datasets. Although the MF technique improved the prediction accuracy to some extent, it inferred that each dimension of the latent trait vector could not be associated with the specific knowledge concepts, making the prediction results less interpretable. Moreover, a state-of-the-art article [36] is recommended for readers who are interested in predicting students’ academic performance.

Our work differs from previous studies as follows. First, existing approaches mostly focus on utilizing students’ response logs on exercises for their performance prediction. We additionally consider the effect of the auxiliary learning materials that students are interested in on their performance. Second, we combine preference cognitive diagnosis with the traditional cognitive diagnosis method, aiming to model students’ knowledge states more comprehensively. Third, our work can obtain PSP results with both accuracy and interpretability.

III. PROBLEM OVERVIEW

For the specific challenges to be proposed, we formally define the problems of preferred learning material acquisition and student performance prediction. Our problem overview is shown in Fig. 3.

In the MAS system [4], [5] (an online learning platform that records students’ learning behaviors), assume there are \(|U|\) students, \(|V|\) exercises, \(|D|\) knowledge concepts, and \(|N|\) learning materials, which can be represented as \(S = \{S_1, S_2, \ldots, S_U\}\), \(E = \{E_1, E_2, \ldots, E_V\}\), \(K = \{K_1, K_2, \ldots, K_D\}\), and \(L = \{L_1, L_2, \ldots, L_N\}\), respectively. Each student performs exercises individually and their response logs \(R = \{r_{uv}\}_{U \times V}\), where \(r_{uv} = 1\) if student \(S_u\) answers exercise \(E_v\) correctly, and \(r_{uv} = 0\) otherwise. On the other hand, students may read some learning materials and their reading logs \(C_u = \{l_{un}r_{un}\}, \ldots |n \in 1, 2, \ldots, N\}\), where \(l_{un}\) represents that student \(S_u\) has read learning material \(L_n\) and \(r_{un}\) is the corresponding reading times. After reading, students may collect some learning materials in their favorite lists \(X_u = \{x_{un}, \ldots |n \in 1, 2, \ldots, N\}\), where \(x_{un}\) represents student \(S_u\) who collected learning material \(L_n\). In addition, we have Q-matrix \(Q = \{q_{vd}\}_{V \times D}\) (labeled by experts), where \(q_{vd} = 1\) if exercise \(E_v\) is related to knowledge concept \(K_d\), and \(q_{vd} = 0\) otherwise; M-matrix \(M = \{m_{nd}\}_{N \times D}\) (labeled by experts), where \(m_{nd} = 1\) if learning material \(L_n\) is related to knowledge concept \(K_d\), and \(m_{nd} = 0\) otherwise. It is noted that the M-matrix, as prior educational knowledge, shows the correlation between learning materials and knowledge concepts.

Definition 1 (Preferred Learning Material Acquisition): Given students’ reading logs \(C\), our goal is to acquire students’ preferred learning materials, which can be applied for knowledge states modeling.

Definition 2 (Student Performance Prediction): Given students’ response logs \(R\), acquired preferred learning materials, Q-matrix \(Q\), and M-matrix \(M\), our goal is to predict students’ scores on exercises by modeling their knowledge states.

IV. PREFERRED LEARNING MATERIAL ACQUISITION

In this section, we introduce our Direct-Indirect method that can achieve the goal of preferred learning material acquisition (i.e., Definition 1). Specifically, as shown in Fig. 4, we first acquire the preferred learning materials directly from the learning material content that students have read (i.e., Direct Acquisition). This part mainly includes the formalization of learning materials and students as well as the similarity calculation between them. Subsequently, we acquire those preferred learning materials indirectly by analyzing the similarity of students’ learning behavior during the process of reading learning materials (i.e., Indirect Acquisition). In this part, the similarity of learning behavior between students is calculated (e.g., \(sm_{ij}\)). Finally, the acquired learning materials obtained by combining direct and indirect strategies are taken as the students’ preferred learning materials. For better illustration, Table 2 lists some important mathematical notations about preferred learning material acquisition.

A. DIRECT ACQUISITION

In the reading process of learning materials, students read some content that they are interested in or that is conducive to learning, which usually implies a large amount of student preference information for learning materials.
Therefore, we mine the students’ preference information from their reading content and directly acquire the preferred learning materials.

Generally, the keyword vector model can be established by using the student reading content to represent each student formally. Then, the similarity between students and learning materials is calculated to match the learning materials with a high similarity for students as their preferred learning materials. However, it is difficult to effectively mine the students’ preference information only through the keyword vector model, which will arouse the overfitting phenomenon in learning material matching. For example, if there are no learning materials similar to the students’ keyword vector, matching cannot be performed. In that case, the students’ preferred learning materials cannot be acquired accurately. Moreover, assume it only matches the learning materials similar to the keyword vector of students, the overfitting of the acquired preferred learning materials with students’ reading content may occur.

In the real scene, students may be interested in learning materials associated with their reading content, i.e., they are interested in learning materials with similar topics. For example, learning material (e.g., Ancient Chinese Economic Policy) and learning material (e.g., Ancient Chinese Business Development) do not have many similar keywords. However, they are interrelated in content. Therefore, we employ students’ reading content to represent each student formally through the two dimensions of keywords and topics. After that, we calculate students’ direct preference degree for learning materials by utilizing the cosine formula, and select the $k$ learning materials with the highest direct preference degree as the preferred student learning materials. Specifically, Direct Acquisition can be divided into the following three parts: Learning Material Formalization, Student Formalization, and Similarity Calculation.

1) LEARNING MATERIAL FORMALIZATION
To calculate the similarity between students and learning materials, we need to formalize each learning material. For learning material $L_n$, we formalize it as $L_n = \{K_n, P_n\}$, where $K_n$ and $P_n$ are the keyword vector and topic distribution vector of $L_n$. Here, we formalize the learning materials through the dimensions of keywords and topics (i.e., text mining for learning materials), which considers the correlation between words in the text.

For each learning material in $L$, we use the Jieba tool\(^2\) to gain the results of word segmentation. Then, we represent the keyword vector $K_n = \{(K_{n1}, \omega_{n1}), (K_{n2}, \omega_{n2}), \ldots\}$ by the TF-IDF algorithm\(^{[17]}\) applying the segmentation results, where $K_{nj}$ and $\omega_{nj}$ represent the keyword $j$ of learning material $L_n$ and the corresponding weight, respectively.

For each learning material in $L$, we adopt the LDA model\(^{[20]}\) for mining the latent topic distribution. LDA is a Bayesian probabilistic model with a structure of three layers (i.e., text-topic-word), which extracts the representative word list from the corpus of learning materials as a topic and eventually presents the topics of each learning material as a probability distribution. The mined topic distribution vector $P_n = \{(P_{n1}, v_{n1}), (P_{n2}, v_{n2}), \ldots\}$, where $P_{nj}$ and $v_{nj}$ represent the topic $j$ of learning material $L_n$ and the corresponding weight, respectively.

2) STUDENT FORMALIZATION
For student $S_u$, we collect the keywords and topics from the reading content and then formalize them as $S_u = \{F_u, G_u\}$.

\(^2\)https://pypi.org/project/jieba/
where $F_u$ and $G_u$ are the keyword vector and topic distribution vector of student $S_u$.

Given the reading logs $C_u = \{(L_{u1}, t_{u1}), \ldots | n \in 1, 2, \ldots, N\}$, then $L_{u1} = \{k_{u1}, P_{u1}\}$. We extract the keywords used to formalize student $S_u$ from the reading content and represent them as $F_u = \{(F_{u1}, \sigma_{u1}), (F_{u2}, \sigma_{u2}), \ldots\}$, where $F_{ui}$ and $\sigma_{ui}$ are the keyword of student $S_u$ and the corresponding weight. The $\sigma_{ui}$ is computed as follows:

$$
\sigma_{ui} = \frac{\sum_{n=1}^{\text{|C_u|}} \omega_{u_{in}}}{|P|}
$$

where $\omega_{u_{in}}$ denotes the corresponding weight of the keyword $j$ of student $S_u$ (i.e., $F_{ui}$) in the keyword vector $K_n$ of learning material $L_n$ that $S_u$ has read (i.e., $K_{un}$).

For the topics of student $S_u$, we extract from the reading content and represent them as $G_u = \{(G_{u1}, \mu_{u1}), (G_{u2}, \mu_{u2}), \ldots\}$, where $G_{ui}$ and $\mu_{ui}$ are the topic $j$ of student $S_u$ and the corresponding weight. The $\mu_{ui}$ is computed as follows:

$$
\mu_{ui} = \frac{\sum_{n=1}^{\text{|C_u|}} \nu_{u_{in}}}{\text{|C_u|}}.
$$

where $\nu_{u_{in}}$ denotes the corresponding weight of the topic $j$ of student $S_u$ (i.e., $G_{ui}$) in the topic distribution vector $P_n$ of learning material $L_n$ that $S_u$ has read (i.e., $P_{un}$).

3) SIMILARITY CALCULATION

To obtain the students’ direct preference degree for learning materials, we adopt the cosine formula to calculate the similarity between students and learning materials that both of them are formalized. The direct preference degree of student $S_u$ for learning material $L_n$ is calculated as follows:

$$
dp_{un} = \frac{\text{Sim}(S_u, L_n) + (1 - \rho)\text{Sim}(G_u, P_n)}{\sqrt{\rho^2 + (1 - \rho)^2}}
$$

where $\rho$ is the weight parameter and $\rho \in [0, 1]$, which is employed to control the proportion of the students’ keyword vector $F$ and topic distribution vector $G$.

Finally, the students’ direct preference degree for learning materials is the mined preference information. We can further select the $k$ learning materials with the highest direct preference degree as the students’ preferred learning materials.

B. INDIRECT ACQUISITION

The direct strategy of acquiring the preferred learning materials may not be accurate enough by employing the students’ reading content. For example, if a student reads fewer learning materials during the learning process, i.e., the reading content is not sufficient, at this time, the Direct method may not be accurate enough to mine the students’ preference information for learning materials by utilizing insufficient reading content, thereby affecting the accuracy of learning material acquisition. Generally, students with similar learning behaviors may have the same preference for learning materials [37], [38]. Therefore, we consider analyzing the similarity of students’ learning behaviors during the reading process to discover the potential preferred learning materials.

In a pair of students, the more they read the same learning materials, the more similar their learning behavior may be. For example, students have a low proficiency for specific knowledge concepts, which may prompt them to read more of the same learning materials for consolidation. Additionally, the number of times students read the learning materials (i.e., reading times) can reflect the similarity of their learning behavior. For example, students may be interested in a certain learning material that they have read repeatedly. Therefore, we design a method for quantifying the similarity of learning behavior and thus indirectly acquire the students’ preferred learning materials.

Given the reading logs $C_u = \{L_{u1}, t_{u1}, \ldots | n \in 1, 2, \ldots, N\}$ and $C_j = \{L_{j1}, t_{j1}, \ldots | n \in 1, 2, \ldots, N\}$, the similarity of student $S_u$ to student $S_j$ on learning behavior is defined as follows:

$$
sm_{uj} = \frac{\sum_{n=1}^{\text{|C_u|}} t_{un}}{\sum_{n=1}^{\text{|C_j|}} t_{jn}} \times \mu
$$

where $t_{un}$ is the reading times of student $S_u$ for learning material $L_n$ and $In_{uj}$ is the intersection between where $S_u$ and $S_j$ have read the learning materials. $\mu$, as the weight parameter, represents the proportion of the number of elements in $In_{uj}$ and $C_j$.

The setting of $\mu$ considers the errors caused by the number of students’ reading learning materials on the similarity quantification. For example, if student $S_u$ has read fewer learning materials (i.e., $|C_u|$ is small), whereas student $S_j$ has read more learning materials (i.e., $|C_j|$ is large), $C_u = In_{uj}$ may occur (i.e., $C_u \in C_j$), if $\mu$ is not set, $sm_{uj}$ is 1. Actually, it is produced by the fact that $S_u$ reads a relatively small number of learning materials, rather than the high similarity caused by a large number of the same learning materials that
$S_u$ and $S_j$ have read together. When $S_u$ reads less, in contrast, $S_j$ reads more, $|J_{u|j}|$ may be smaller than $|C_j|$, and thus, $\mu$ is minor. As a result, $sm_{uj}$ is less than 1. Therefore, we set $\mu$ to reduce the errors caused by the number of students’ reading learning materials on the similarity quantification to some extent. In addition, (6) reveals that due to the influence of the reading times (e.g., $t_{um}$) and the parameter $\mu$, $sm_{uj}$ is not equal to $sm_{ij}$, i.e., the similarity of $S_u$ to $S_j$ is not equal to that of $S_j$ to $S_u$, which reflects the relativity of the similarity quantification between students and is more in line with the real situation.

To obtain the students’ indirect preference degree for learning materials, we first normalize the quantified similarity to between 0 and 1 and set the similarity threshold $\delta$ as 0.75. Subsequently, we group students whose similarity relative to student $S_u$ is not less than $\delta$ and set the group as $J = \{S_1, S_2, \ldots, S_j\}$. The indirect preference degree of student $S_u$ for learning material $L_n$ is computed as follows:

$$i p_{un} = \max (o_{u n 1}, o_{u n 2}, \ldots, o_{u n j})$$

$$o_{uni} = s m_{uj} \times d p_{li}, \quad i = 1, 2, \ldots, j. \quad (7)$$

where $\max (\cdot)$ denotes selecting the maximum value and $o_{uni}$ represents the indirect preference degree of student $S_u$ for learning material $L_n$ reflected by student $S_i$. After that, we can select the $k$ learning materials with the highest indirect preference degree as the preferred learning materials of students.

**V. PREFERENCE COGNITIVE DIAGNOSIS**

In this section, we introduce our preference cognitive diagnosis method (PreferenceCD) that can achieve the goal of student performance prediction (i.e., Definition 2). As shown in Fig. 5 (from top to bottom and according to different colors), our proposed method consists of three parts, which start with students’ comprehensive proficiency for knowledge concepts, then evaluates students’ actual mastery for exercises. Specifically, the students’ comprehensive proficiency for knowledge concepts (i.e., knowledge states) is calculated by their proficiency and preference degree for knowledge concepts. Subsequently, with the Q-matrix, the students’ actual mastery of exercises can be evaluated according to their knowledge states. Finally, by setting a threshold (i.e., $T_{HR}$), the students’ exercise scores are predicted based on their actual mastery for exercises. For better illustration, Table 3 displays some important mathematical notations and each step of PreferenceCD is specified in the following subsections.

**A. COMPREHENSIVE PROFICIENCY CALCULATION**

In this subsection, we illustrate the method for obtaining the comprehensive proficiency of students for knowledge concepts (i.e., knowledge states modeling). Specifically, we calculate the students’ comprehensive proficiency for knowledge concepts (e.g., $\gamma_{uv}$) by combining their proficiency (e.g., $\alpha_{ud}$) and preference degree (e.g., $\beta_{ud}$) for knowledge concepts. Then, $\gamma_{uv}$ is defined as follows:

$$\gamma_{uv} = \lambda \alpha_{ud} + (1 - \lambda) \beta_{ud}. \quad (8)$$

where $\lambda$ is the weight parameter and $\lambda \in [0, 1]$, which is applied for controlling the proportion of students’ proficiency and preference degree for knowledge concepts. Next, we introduce the calculation of $\alpha_{ud}$ and $\beta_{ud}$, respectively.

1) PROFICIENCY DIAGNOSIS

CDMs can better model students’ knowledge states. Since the IRT model [27], [28] expresses students as a variable, it is difficult to characterize the students’ proficiency for specific knowledge concepts. The DINA model [12] is simple with strong interpretability of parameters (i.e., slip and guess) for modeling students’ knowledge states from different knowledge concepts. For this, we apply the DINA model to diagnose students’ proficiency in knowledge concepts.

The DINA model describes student $S_u$ as a binary vector $\alpha_u = \{\alpha_{u1}, \alpha_{u2}, \ldots, \alpha_{uD}\}$, where $\alpha_{ud}$ represents the proficiency of $S_u$ for knowledge concept $k_d$ and $\alpha_{ud} = 1$ indicates that $S_u$ has mastered $k_d$, and $\alpha_{ud} = 0$ otherwise.

Given $\alpha_u$, the mastery of $S_u$ for exercise $E_v$ is given by the following:

$$\eta_{uv} = \prod_{d=1}^{D} \alpha_{ud} \cdot (9)$$

where $\eta_{uv} = 1$ indicates that $S_u$ has given a right response for $E_v$, and $\eta_{uv} = 0$ otherwise. In addition, DINA takes the slip and guess factors to simulate students’ responses to exercises. The probability of student $S_u$ responding to exercise $E_v$ correctly is defined as follows:

$$P_v(\alpha_u) = P(r_{uv} = 1|\alpha_u) = g_v^{1-\eta_{uv}} (1 - s_v) \eta_{uv}. \quad (10)$$
where \( s_v \) is slip, which represents the probability that students who have mastered all the knowledge concepts related to \( E_v \) but have not responded correctly. \( g_v \) is a guess, which represents the probability that students who have not mastered all the knowledge concepts related to \( E_v \) have responded to it correctly.

DINA uses the expectation maximization algorithm (EM) to maximize the marginal likelihood of (10), and thus, \( s_v \) and \( g_v \) can be estimated. By maximizing the posterior probability of exercise scores of student \( S_u \) (i.e., \( r_u \)), \( \alpha_u \) can be determined as follows:

\[
\hat{\alpha}_u = \arg \max_{\alpha} P(\alpha | r_u) = \arg \max_{\alpha} L(r_u | \alpha, \hat{s}_v, \hat{g}_v) P(\alpha) = \arg \max_{\alpha} \prod_{v=1}^{V} P(r_{uv} | \alpha, \hat{s}_v, \hat{g}_v). \tag{11}
\]

Because the DINA model can only obtain students’ discrete proficiency for knowledge concepts (i.e., mastered is “1” and not mastered is “0”), we diagnose the students’ probabilistic proficiency for knowledge concepts according to all possible posterior probabilities of \( \alpha_u \) and simulate \( \alpha_{ud} \) as a variable between 0 and 1. We redefine \( \alpha_u \), and \( \alpha_{ud} \) is diagnosed as follows:

\[
\hat{\alpha}_{ud} = P(\alpha_{ud} = 1 | r_u) = \frac{\sum_{\alpha_{ud}=1}^{2D} P(\alpha | r_u)}{\sum_{\alpha_{ud}=1}^{2D} P(\alpha | r_u)} = \frac{\sum_{\alpha_{ud}=1}^{2D} L(r_u | \alpha, \hat{s}_v, \hat{g}_v) P(\alpha)}{\sum_{\alpha_{ud}=1}^{2D} \prod_{v=1}^{V} L(r_{uv} | \alpha, \hat{s}_v, \hat{g}_v) P(\alpha)} \tag{12}
\]

2) PREFERENCE DEGREE CALCULATION
Notably, students’ preferred learning materials are usually related to relevant knowledge concepts, which can express students’ interests or demands for different knowledge concepts and thus reflect students’ knowledge states. For example, when reading a certain preferred learning material, students may spend more time and energy. As a result, the student may have a higher mastery of the learning material and thus may have a higher proficiency for the specific knowledge concepts. Therefore, we acquire students’ preferred learning materials and employ them as a part of knowledge states modeling.

Here, we achieve the way to transform the acquired preferred learning materials into the students’ preference degree for knowledge concepts by adopting prior education knowledge (i.e., M-matrix). Specifically, we first extract the learning materials of students acquired by the Direct-Indirect method in Section IV and formalize them as follows:

\[
DP_u = \{ (L_{un}, dp_{un}) \} | n \in \{ 1, 2, \ldots, N \} \]
\[
IP_u = \{ (L_{un}, ip_{un}) \} | n \in \{ 1, 2, \ldots, N \} \]
\[
|DP_u| = |IP_u| = k. \tag{13}
\]

where \( DP_u \) and \( IP_u \) are the preferred learning materials of student \( S_u \) acquired by the Direct Acquisition and Indirect Acquisition.

To unify the dimensions of the corresponding preference degree (i.e., the direct and indirect preference degree of students for learning materials, such as \( dp_{un} \) and \( ip_{un} \)), we normalize them to between 0 and 1. Then, the preferred learning materials of student \( S_u \) can be further formalized as follows:

\[
SP_u = \{ (DP_u, IP_u) \} = \{ (L_{un}, sp_{un}) \} | n \in \{ 1, 2, \ldots, I \}
|SP_u| = 2k = I. \tag{14}
\]

where \( L_{un} \) and \( sp_{un} \) represent the preferred learning material \( L_n \) of student \( S_u \) and the corresponding preference degree, respectively.

The preferred learning materials are related to the specific knowledge concepts. Therefore, according to the M-matrix \( M \), the preference degree of student \( S_u \) for knowledge concept \( k_d \) is calculated as follows:

\[
\beta_{ud} = \frac{\sum_{I \in SP_u, m_{ud}=1}^{I} sp_{un}}{|O|} \\
O = \{ L_{un} | L_{un} \in SP_u, m_{ud}=1, \ n = 1, 2, \ldots, I \}. \tag{15}
\]

where \( I \) is the number of acquired preferred learning materials.

B. ACTUAL MASTERY EVALUATION AND SCORE PREDICTION
With the comprehensive proficiency of knowledge concepts obtained in subsection A, we can further evaluate the students’ actual mastery for exercises (i.e., the probability of being able to solve the exercises) and predict their scores.

Following the method in [9], we apply the observed students’ response logs and the exercise parameters (i.e., slip and guess) as priori to evaluate the student’s actual mastery for exercises. Specifically, according to the Q-matrix \( Q \), we first adopt the geometric mean method to evaluate the students’ average mastery for exercises. The details are as follows:

\[
\eta_{avgw} = \frac{1}{\sum_{d=1}^{2D}} \left[ \prod_{d=1}^{D} \text{ins}_{udw} \right] \tag{16}
\]

\[
\text{ins}_{udw} = \begin{cases} 1, & q_{ud} = 0 \ 
\gamma_{ud} \times q_{ud}, & q_{ud} = 1. \end{cases} \tag{17}
\]

where, \( \eta_{avgw} \) is the average mastery of student \( S_u \) for exercise \( E_v \). Then, we take the students’ response logs as well as the
slip and guess estimated by DINA model (e.g., $s_v$ and $g_v$) as the priori parameters to evaluate the students’ actual mastery for exercises:

$$\eta_{av} = \begin{cases} 
(1 - s_v)\eta_{avg} & \text{if } r_{av} = 1 \\
(1 - s_v)\eta_{avg} + g_v(1 - \eta_{avg}) & \text{if } r_{av} = 0.
\end{cases} \quad (18)$$

Eventually, according to the evaluated actual mastery for exercises, we can predict students’ observable scores on exercises:

$$\hat{r}_{av} = \begin{cases} 
1, & \text{if } \eta_{av} > T_{HR} \\
0, & \text{otherwise}.
\end{cases} \quad (19)$$

where $T_{HR}$ is the threshold set in advance, and we let it be 0.5.

C. SUMMARY
We introduced the details of the PreferenceCD method. It is worth noting that the knowledge states modeling of students consists of two aspects. The first is to diagnose students’ proficiency for knowledge concepts by the DINA model. The second is to apply the preferred learning materials acquired by the Direct-Indirect method and prior education knowledge (i.e., M-matrix), calculating students’ preference degree for knowledge concepts. After that, the students’ comprehensive mastery of knowledge concepts can be obtained by combining the above two aspects as the final results of knowledge states modeling. Compared with traditional CDMs, PreferenceCD considers the influence of students’ preference for learning materials on their performance. Thus, it incorporates additional preference factors (i.e., preference degree for knowledge concepts) in knowledge states modeling.

VI. EXPERIMENTS AND ANALYSIS
In this section, we first evaluate the effect of Direct-Indirect on preferred learning material acquisition. Then, we compare the performance of PreferenceCD against the baseline approaches on the PSP task. Next, we investigate the influence of the parameters ($\lambda$, $\mu$, and $I$) on the prediction results. Finally, we conduct a case study to explore the interpretability of PreferenceCD.

A. DATASET PREPARATION
We collect the dataset from the history course data during the interaction between students of Zhengzhou No. 2 Senior High School in Henan Province and the MAS system [4], [5]. The dataset includes students’ response logs $R$ on historical objective exercises, reading logs $C$ of relevant learning materials, and favorite lists $X$. In addition, it also includes the text content of each learning material as well as the Q-matrix and M-matrix marked by the front-line teachers. We denote the dataset as History. For better illustration, the brief summary of History is shown in Table 4. Furthermore, Fig. 6 displays the Q-matrix and M-matrix. In this article, we conduct experiments on the history course because the historical data have richer text information of learning materials.

Since History contains noise data, it cannot be applied in the experiments directly. To alleviate the influence of noise data on the experimental results, we process the dataset as follows. 1) For the text of each learning material, we delete some special symbols (e.g., “$” and “ ●”), letters, and numbers. As a result, it only retains the text part. 2) To extract the keywords and topics of relevant learning materials accurately, we add some extra words based on the Chinese stopword list, including the common function words, punctuation, and HTML placeholders (3,974 items totally). 3) Applying the LDA model to mine topics, we set the number of the topic as 19 (i.e., in (5), $H = 19$), to correspond to the number of knowledge concepts in History.

B. PREFERRED LEARNING MATERIAL ACQUISITION
1) EVALUATION METRICS
To demonstrate the effectiveness of Direct-Indirect for acquiring preferred learning materials, we apply precision, recall, and F-score as evaluation metrics. Their definitions are as follows:

$$\text{precision} = \frac{\sum_{u=1}^{U} |\{L_{ui} | L_{ui} \in X_u, i = 1, 2, \ldots, I\}|}{I \times U} \quad (20)$$

$$\text{recall} = \frac{\sum_{u=1}^{U} |\{L_{ui} | L_{ui} \in X_u, i = 1, 2, \ldots, I\}|}{\sum_{u=1}^{U} |X_u|} \quad (21)$$

$$F - \text{score} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}}. \quad (22)$$

where $I$ is the number of acquired preferred learning materials and $X_u$ is the favorite list of student $S_u$.  

TABLE 4. Statistics of the dataset.

| Statistics                                              | Value |
|--------------------------------------------------------|-------|
| Total number of students                                | 207   |
| Total number of exercises                               | 104   |
| Total number of learning materials                       | 146   |
| Total number of knowledge concepts                      | 19    |
| Total number of reading logs of students                | 2577  |
| Total number of learning materials in students’ favorite lists | 1677  |
2) BASELINES

In this article, we select the following baselines for comparison with Direct-Indirect.

1) Keywords-F. We let $\rho$ equal 1 (in (3)) and select the $I$ learning materials with the highest direct preference degree (i.e., $dp$) as the preferred learning materials of students.

2) Topics-G. We let $\rho$ be 0 and select the $I$ learning materials with the highest direct preference degree as the preferred learning materials of students.

3) Direct. We let $\rho$ be 0.6 and select the $I$ learning materials with the highest direct preference degree as the students’ preferred learning materials.

4) Indirect. We let $\rho$ be 0.6 and select the $I$ learning materials with the highest indirect preference degree (i.e., $ip$) as the students’ preferred learning materials by the calculation of (7).

3) EXPERIMENTAL RESULTS

The experimental results of all the methods under different numbers of acquired preferred learning materials are shown in Fig. 7. In addition, we set the weight parameter $\rho$ as 0.6 to make the experimental results of Direct and Indirect optimum. More information about the two methods is shown in Table 5 and Table 6. It is worth noting that the setting of $\rho$ is the same in Direct and Indirect. This might be because the calculation of the students’ indirect preference degree in (7) (i.e., $ip$) is affected by their direct preference degree for learning materials (i.e., $dp$). Therefore, when Direct works best (i.e., $\rho = 0.6$), Indirect also achieves the best.

From Fig. 7, when $I$ is 4, the results of Direct-Indirect are not optimum. With the increase in $I$, all the metrics of Direct-Indirect improve steadily. When $I$ is 10, the acquisition results are the best. With the further increase in $I$, the F-score of Direct-Indirect decreases, which may because the number of learning materials collected in students’ favorite lists is small, thus influencing the acquisition results. When $I$ is 4, the results of Direct-Indirect are not optimal. However, the acquisition of preferred learning materials ultimately needs to serve the knowledge states modeling of students. Assume we set $I$ to 4 as the number of acquired preferred learning materials. In that case, the smaller $I$ may cause fewer knowledge concepts related to the acquired learning materials, resulting in the inaccurate calculation of the students’ preference degree for knowledge concepts. As a result, student knowledge states modeling is also affected. Therefore, considering the actual situation and experimental results, Direct-Indirect performs better than other baselines. In particular, the better performance of Direct-Indirect over Direct and Indirect proves that the strategies of combining direct and indirect learning for acquiring preferred learning materials are superior in applying only one of the two. Therefore, the Indirect method is useful for acquiring learning materials as a supplement to the preferred student learning materials, which further improves the acquisition accuracy.

Considering the actual situation, we mainly focus on the experimental results with larger $I$. When $I$ is greater than 8,
the performance of Direct beats Keywords-F and Topics-G. The reason for better returns may be that students are interested in learning materials with similar topics. Therefore, in Direct, we extract keywords and topics from the reading content to formalize students, which can mine students’ preference information effectively and directly acquire preferred learning materials. Moreover, when \( I \) is less than 8, Keywords-F is better than Direct, which may be because students are also obsessed with learning materials that are more similar to their reading content. Actually, Keywords-F extracts the specific keywords from students’ reading content to achieve the acquisition of preferred learning materials. Compared with Direct, the preferred learning materials acquired by Keywords-F are more similar to students’ reading content. Therefore, Keywords-F has better results when \( I \) is small.

### C. STUDENT PERFORMANCE PREDICTION

#### 1) EXPERIMENTAL SETUP AND EVALUATION METRICS

To demonstrate the effectiveness of PreferenceCD, we conduct experiments on the PSP task. Specifically, in History, we verify from the two dimensions of students and exercises. For students, we randomly select 20% of them for testing, and the rest are for training. For exercises, K-fold cross-validation is adopted, which \( K = 5 \) to evaluate the stability of PreferenceCD.

For example, in the experiment with students on 20% of the testing exercises, we first diagnose students’ proficiency for knowledge concepts (e.g., \( \alpha_{ud} \)) by the response logs of all students on 80% of the training exercises. Then, we evaluate the exercise parameters (e.g., \( s_v \) and \( g_v \)) by the response logs of 80% of the training students on all exercises. After that, we further calculate the students’ comprehensive proficiency for knowledge concepts (e.g., \( \gamma_{ud} \)) and predict their scores on testing exercises. Such a cross-validation partition ensures that we do not take any parameter training.

We adopt precision, RMSE, and MAE as the evaluation metrics. Their definitions are as follows.

\[
\text{precision} = \frac{\sum_{u=1}^{U} |\hat{r}_{ui} - r_{ui}, i = 1, 2, \ldots, V|}{U \times V} \quad (23)
\]

\[
\text{RMSE} = \sqrt{\frac{\sum_{u=1}^{U} \sum_{v=1}^{V} (\hat{r}_{uv} - r_{uv})^2}{U \times V}} \quad (24)
\]

\[
\text{MAE} = \frac{\sum_{u=1}^{U} \sum_{v=1}^{V} |\hat{r}_{uv} - r_{uv}|}{U \times V} \quad (25)
\]

#### 2) BASELINES

In this article, we consider the following baselines for comparison with PreferenceCD.

1) **IRT** [27]. IRT is a cognitive diagnosis model that models students’ latent traits as well as the parameters of exercises such as difficulty and discrimination.

2) **DINA** [26]. DINA is a cognitive diagnosis model that models students’ proficiency for knowledge concepts as well as the slip and guess factors of exercises with the Q-matrix.

3) **PMF** [33]. PMF refers to probabilistic matrix factorization, which is a latent factor model projecting students and exercises into a low-dimensional space.

4) **NMF** [32]. NMF refers to nonnegative matrix factorization, which is a latent nonnegative factor model.

5) **KNN** [13]. The similarity between students can be calculated by applying the cosine formula according to students’ response logs. Then the students who are most similar to the target students are found, and their scores are taken as the target students’ performance.

6) **PMF-CD** [9]. PMF-CD combines students’ knowledge states (adopted by DINA) and their common characteristics in learning (adopted by PMF) to realize PSP.

#### 3) EXPERIMENTAL RESULTS

The PSP results of all the methods on History are shown in Table 7. Here, we consider two implementations of matrix factorization, PMF and NMF with 5 latent factors. In PreferenceCD, we set the number of acquired preferred learning materials (i.e., \( I \)) to 10 and the weight parameter \( \lambda \) (in (8)) to 0.65, making the experimental results the best.

| Model   | Evaluation Metrics |
|---------|---------------------|
|         | Precision | RMSE | MAE    |
| IRT [27] | 0.6423 | 0.5896 | 0.3579 |
| DINA [26] | 0.6533 | 0.5602 | 0.3465 |
| PMF [33] | 0.7215 | 0.5127 | 0.2783 |
| NMF [32] | 0.7011 | 0.5303 | 0.2988 |
| KNN [13] | 0.7329 | 0.5002 | 0.2625 |
| PMF-CD [9] | 0.7564 | 0.5012 | 0.2662 |
| PreferenceCD | 0.7614 | 0.4805 | 0.2386 |

From the table, we can observe that PreferenceCD possesses the best performance, and its accuracy, RMSE, and MAE are 0.7614, 0.4805, and 0.2386, respectively. Specifically, combining the preference degree and proficiency of knowledge concepts to model students’ knowledge states beats the traditional CDMs (i.e., IRT and DINA). Compared with the CF method (i.e., PMF, NMF, and KNN), PreferenceCD further improves the prediction accuracy. It is also superior to the method of combining the students’ knowledge states and their common characteristics (i.e., PMF-CD). Therefore, PreferenceCD can model students’ knowledge states more accurately. It also shows that students read those learning materials that they are interested in (i.e., preferred learning material) reflects their knowledge states to some extent, demonstrating the rationality of preference cognitive diagnosis.

### D. PARAMETER SETTING

In PreferenceCD, parameter \( \lambda \) (in (8)) is applied for adjusting the students’ comprehensive proficiency for knowledge
concepts (i.e., the results of knowledge states modeling), and \( \lambda \in [0, 1] \). The larger (smaller) the \( \lambda \) is, the more dependent students’ proficiency (preference degree) for knowledge concepts is. Because the change in \( \lambda \) may affect the results of knowledge states modeling, the PSP results of PreferenceCD will change. In addition, as described in Section V, we adopt the preferred learning materials to calculate the preference degree of students for knowledge concepts. The number of acquired preferred learning materials (i.e., \( I \)) may vary the calculated preference degree of students’ knowledge concepts, which will influence their knowledge states. Therefore, we take the precision metric as an example and set different \( \lambda \) and \( I \) to observe the experimental results. The details are shown in Fig. 8.

In the figure, we can observe that when \( \lambda \) is 0.65 and \( I \) is set to 10, PreferenceCD performs the best. Moreover, the setting of \( \lambda \) means that the preference degree of knowledge concepts possesses a lower proportion in the knowledge states modeling of students. As a result, the performance of students depends more on their proficiency in knowledge concepts.

**Case Study**

Here, we present an example of the modeling results of a student’s knowledge states by applying PreferenceCD in Fig. 9. The upper part of Fig. 9 shows the Q-matrix of three exercises on four knowledge concepts and the student’s response logs. The bars in the underneath subfigure represent the student’s three knowledge states: proficiency, preference degree, and comprehensive proficiency for the specific knowledge concepts. The black lines and marks represent the knowledge states that we imagine the student should at least achieve if they want to do the exercises correctly. For example, exercise \( E_2 \) is related to knowledge concept \( K_3 \), so we believe that the student’s proficiency for \( K_3 \) needs to be above 0.5 to respond to \( E_2 \) correctly.

From the figure, we can observe several phenomena. First, the student responds to \( E_1 \) incorrectly, which may be due to the low preference degree and proficiency of the student for \( K_2 \) (i.e., lower than 0.5). Second, the student responds to \( E_2 \) correctly, which may occur because of a high comprehensive proficiency for \( K_3 \) (i.e., higher than 0.5). Third, the proficiency of the student for \( K_4 \) is low, so the student should not be able to respond to \( E_3 \) correctly. However, the fact is the opposite. This case may result since the student’s high preference degree for \( K_4 \) makes the student’s comprehensive proficiency for \( K_4 \) higher so that the student can perform \( E_3 \) correctly. Therefore, students’ preference degree for knowledge concepts can reflect their knowledge states to a certain extent, which is conducive to the PSP task. Thus, analyzing the PSP results of PreferenceCD in combination with the preference degree of students for knowledge concepts (i.e., preference cognitive diagnosis) has strong interpretability.

**Conclusion**

In this article, we proposed a preference cognitive diagnosis method, PreferenceCD, to model the knowledge states for predicting students’ performance. Specifically, we first designed the Direct-Indirect method to obtain students’ preferred learning materials and thus calculated their preference degree for knowledge concepts. Then, we diagnosed the proficiency of knowledge concepts by applying the DINA model and combining the preference degree with proficiency to calculate students’ comprehensive proficiency for knowledge concepts. Based on this, we evaluated students’ actual mastery of exercises and predicted their scores on exercises. The experimental results on the real-world dataset revealed the effectiveness of our method with both accuracy and interpretability.

However, there is still some room for improvement. First, due to the narrow audience of the MAS system, as well as the strong correlation between data (i.e., learning materials, knowledge concepts, and exercises), it is difficult for us to collect large-scale datasets for better verification. Second, we can test more CDMs to diagnose the proficiency of students for knowledge concepts. Third, there may be some other exercise types that should be considered, such as subjective exercises, to make predictions. The above problems are our future research directions.
REFERENCES

[1] M. Aparicio, T. Oliveira, F. Bacao, and M. Painho, “Gamification: A key determinant of massive open online course (MOOC) success,” Inf. Manage., vol. 56, no. 1, pp. 39–54, Jan. 2019.

[2] A. Joshi, D. Allesio, J. Magee, J. Whitehill, I. Arroyo, B. Woolf, S. Schraff, and M. Betke, “Affect-driven learning outcomes prediction in intelligent tutoring systems,” in Proc. 14th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG), May 2019, pp. 1–5.

[3] F. Wang, Q. Liu, E. Chen, Z. Huang, Y. Chen, Y. Yin, Z. Huang, and S. Wang, “Neural cognitive diagnosis for intelligent education systems,” in Proc. AAAI, 2020, pp. 6153–6161.

[4] R. Wang, “A probe into the teaching model based on mobile autonomous school taking Zhengzhou no. 2 Senior high School as an example,” (in Chinese), J. Chin. Soc. Educ., vol. 7, pp. 59–62, Aug. 2015.

[6] R. Wang, “Research on mobile classroom teaching mode in information environment,” (in Chinese), J. Chin. Soc. Educ., vol. 12, pp. 9–12, Feb. 2015.

[7] Q. Liu, Z. Huang, Y. Yin, E. Chen, H. Xiong, Y. Su, and G. Hu, “EKT: Experience-aware knowledge tracing for student performance prediction,” IEEE Trans. Knowl. Data Eng., early access, Jun. 24, 2020, doi: 10.1109/TKDE.2019.2924374.

[8] Y. Huo, D. F. Wong, L. M. Ni, L. S. Chao, and J. Zhang, “Knowledge modeling via contextualized representations for LSTM-based personalized exercise recommendation,” Inf. Sci., vol. 523, pp. 266–278, Jun. 2020.

[9] Z. Chen, T. Huang, X. Liu, R. Wu, H. Chen, and G. Hu, “Cognitive diagnosis based personalized question recommendation,” (in Chinese), Chin. J. Comput., vol. 40, no. 1, pp. 176–191, 2017.

[10] Y. Liu, Q. Liu, R. Wu, E. Chen, Y. Su, Z. Chen, and G. Hu, “Collaborative learning team formation: A cognitive modeling perspective,” in Proc. Int. Conf. Database Syst. Adv. Appl. Springer, 2016, pp. 383–400.

[11] L. V. DiBello, L. A. Roussos, and W. Stout, “31a review of cognitively diagnostic assessment and a summary of psychometric models,” Handb. Statist., vol. 26, pp. 979–1030, Aug. 2006.

[12] J. de la Torre, “DINA model and parameter estimation: A didactic,” J. Educ. Behav. Statist., vol. 34, no. 1, pp. 69–73, Dec. 2014.

[13] J. Ramos, “Using tf-idf to determine word relevance in document queries,” in Proc. 7th World Eng. Edu. Forum Appl. Sci., vol. 7, no. 16, p. 5601, Aug. 2020.

[14] E. Jembere, R. Rawatlal, and A. W. Pillay, “Factorization for predicting student performance,” in Proc. 7th World Eng. Educ. Forum (WEEF), Nov. 2017, pp. 513–518.

[15] P. Jiang, X. Wang, Preference Cognitive Diagnosis for Student Performance Prediction.

[16] C. Yan, Q. Su, and J. Wang, “MoGCN: Mixture of gated convolutional neural network for named entity recognition of chinese historical texts,” IEEE Access, vol. 8, pp. 181629–181661, 2020.

[17] S. Amin, M. I. Uddin, S. Hassan, A. Khan, N. Nasser, A. Alharr, and H. Alyami, “Recurrent neural networks with TF-IDF embedding technique for detection and classification in tweets of dengue disease,” IEEE Access, vol. 8, pp. 131522–131533, 2020.

[18] D. Mishra, B. Naik, R. M. Sahoo, and J. Nayak, “Deep recurrent neural network (Deep-RNN) for classification of nonlinear data,” in Proc. Comput. Intell. Pattern Recognit. Springer, 2020, pp. 207–215.

[19] F. Wang, Q. Liu, E. Chen, Z. Huang, Y. Chen, Y. Yin, Z. Huang, and A. Dick, “What value do explicit high level concepts have in vision to language problems?” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 203–212.

[20] Z. Huang, Q. Liu, E. Chen, H. Zhao, M. Gao, S. Wei, Y. Su, and G. Hu, “Question difficulty prediction for READING problems in standard tests,” in Proc. AAAI, 2017, pp. 1352–1359.

[21] G. Rasch, “On general laws and the meaning of measurement in psychology,” in Proc. 4th Berkeley Symp. Math. Statist. Probabl., vol. 4, 1961, pp. 321–333.

[22] S. Cheng, Q. Liu, E. Chen, Z. Huang, Z. Huang, Y. Chen, H. Ma, and G. Hu, “DIRT: Deep learning enhanced item response theory for cognitive diagnosis,” in Proc. 28th ACM Int. Conf. Inf. Knowl. Manage., Nov. 2019, pp. 2397–2400.

[23] J. de la Torre and N. D. Minchin, “The G-DINA model framework,” in Handbook Diagnostic Classification Models. Springer, 2019, pp. 155–169.

[24] Z. Huang, Q. Liu, Y. Chen, L. Wu, K. Xiao, E. Chen, H. Ma, and G. Hu, “Learning or forgetting? A dynamic approach for tracking the knowledge proficiency of students,” ACM Trans. Inf. Syst., vol. 38, no. 2, pp. 1–33, Mar. 2020.

[25] K. Nagatani, Q. Zhang, M. Sato, Y.-Y. Chen, F. Chen, and T. Okhuma, “Augmenting knowledge tracing by considering forgetting behavior,” in Proc. World Wide Web Conf., May 2019, pp. 3101–3107.

[26] D. D. Lee and H. S. Seung, “Algorithms for non-negative matrix factorization,” in Proc. NIPS, 2001, pp. 556–562.

[27] N. Thai-Nghe, T. Horváth, and L. Schmidt-Thieme, “Factorization models for forecasting student performance,” in Proc. Educ. Data Mining, 2010, pp. 1–5.

[28] N. Thai-Nghe and L. Schmidt-Thieme, “Multi-relational factorization models for student modeling in intelligent tutoring systems,” in Proc. 7th Int. Conf. Knowl. Syst. Eng. (KSE), Oct. 2015, pp. 61–66.

[29] Y. Chen, Q. Liu, Z. Huang, L. Wu, E. Chen, R. Wu, Y. Su, and G. Hu, “Tracking knowledge proficiency of students with educational priors,” in Proc. ACM Conf. Inf. Knowl. Manage., Nov. 2017, pp. 989–998.

[30] K. T. Chui, R. W. Liu, M. Zhao, and P. O. D. Pablos, “Predicting students’ performance with school and family tutoring using generative adversarial network-based deep support vector machine,” IEEE Access, vol. 8, pp. 86745–86752, 2020.

[31] D. W. Liu, J. Wu, and G. Zhang, “A fuzzy tree matching-based personalized E-Learning recommender system,” IEEE Trans. Fuzzy Syst., vol. 23, no. 6, pp. 2412–2426, Dec. 2015.

[32] P. Dwivedi and K. K. Bharadwaj, “E-learning recommender system for a group of learners based on the unified learner profile approach,” Expert Syst., vol. 32, no. 2, pp. 264–276, Apr. 2015.

PEICHAO JIANG received the B.E. degree in network engineering from Henan Polytechnic University, Jiaozuo, China, in 2018. He is currently pursuing the M.S. degree in computer science with Henan Normal University, Xinxiang, China. His current research interests include data mining and intelligent education. In addition to student performance prediction, he has applied machine learning to personalized exercise recommendation.

XIAODONG WANG received the Ph.D. degree in educational technology from East China Normal University, Shanghai, China. He is currently a Professor and the Dean of the College of Computer and Information Engineering, Henan Normal University. He has published lots of articles on refereed journals and conferences, including the Journal of Computer Research and Development and Computer Science. His current research interests include ontological engineering, data mining, and machine learning.