

Affection-enhanced Personalized Question Recommendation in Online Learning

Mingzi Chen¹, Xin Wei^{1,2*}, Xuguang Zhang¹, and Lei Ye¹

¹School of Communications and Information Engineering, Nanjing University of Posts and Telecommunications
Nanjing 210003, China

[e-mail: souththtc@163.com, zyg26247@126.com, yel@njupt.edu.cn]

² Institute of Higher Education, Nanjing University of Posts and Telecommunications
Nanjing 210003, China

[e-mail: xwei@njupt.edu.cn]

*Corresponding author: Xin Wei

*Received October 20, 2023; revised November 13, 2023; accepted December 7, 2023;
published December 31, 2023*

Abstract

With the popularity of online learning, intelligent tutoring systems are starting to become mainstream for assisting online question practice. Surrounded by abundant learning resources, some students struggle to select the proper questions. Personalized question recommendation is crucial for supporting students in choosing the proper questions to improve their learning performance. However, traditional question recommendation methods (i.e., collaborative filtering (CF) and cognitive diagnosis model (CDM)) cannot meet students' needs well. The CDM-based question recommendation ignores students' requirements and similarities, resulting in inaccuracies in the recommendation. Even CF examines student similarities, it disregards their knowledge proficiency and struggles when generating questions of appropriate difficulty. To solve these issues, we first design an enhanced cognitive diagnosis process that integrates students' affection into traditional CDM by employing the non-compensatory bi-dimensional item response model (NCB-IRM) to enhance the representation of individual personality. Subsequently, we propose an affection-enhanced personalized question recommendation (AE-PQR) method for online learning. It introduces NCB-IRM to CF, considering both individual and common characteristics of students' responses to maintain rationality and accuracy for personalized question recommendation. Experimental results show that our proposed method improves the accuracy of diagnosed student cognition and the appropriateness of recommended questions.

Keywords: Personalized question recommendation, affection-aware, cognitive diagnosis, collaborative filtering, intelligent tutoring systems.

1. Introduction

With the development of online learning platforms, more and more students are practicing exercises through online learning platforms. The online learning platform provides rich and diverse learning resources to meet the personalized and diversified learning needs of students, devoted to stimulating their interest and motivation in adaptive learning. However, the rich question bank that platforms provide also makes some students get lost in abundant information. Fortunately, intelligent tutoring systems use artificial intelligence technologies to analyze and assess students' learning behaviors and learning outcomes, providing them with intelligent learning assistance such as personalized diagnosis, feedback, and guidance. Personalized question recommendation, as an emerging application of ITS, is committed to adapting to the differences and variations of different students and recommending appropriate questions for them [1]. Personalized question recommendation can also save teachers' time and energy and enhance the relevance and effectiveness of teaching. Since there are a huge number of questions online, selecting the proper ones for students is still a challenge.

The first challenge is how to recommend personalized questions for students. Popular recommendation methods include content-based filtering (CBF) and collaborative filtering (CF), which have achieved excellent results in item recommendation. However, their performance in question recommendation is not quite good because excessive focus on the student's response ignores the student's other generated information. CBF recommends based on the similarity questions and their recommendation target, generating the recommendation contents [2]. When the data is limited, CBF results in a lack of variety and novelty in the recommendation. Even though CF utilized probabilistic matrix factorization (PMF) to reduce the dimensions of sparse data, it has not yet leveraged student personalized data (*e.g.*, affection, cognition), reducing the rationality of recommended questions [3]. Cognitive diagnosis-based question recommendation can recommend suitable questions for students based on their cognition. Whereas, the accuracy of cognitive diagnosis largely influences the effect of personalized question recommendation. Therefore, it is still a concern to improve the accuracy of cognitive diagnosis.

Another challenge is how to improve the accuracy of cognitive diagnosis models (CDM). It has been proven that inadequate considerations account for low cognitive diagnosis results. Currently, some research has taken more aspects from student online learning data to improve cognitive diagnosis results. [4] gathered students' forgetting features to CDM. [5] incorporated rich educational contextual features into the existing CDM to enhance the impact of the external environment on students' implicit cognition. These studies reveal that considering more aspects contributes to the improvement of cognitive diagnosis. In the question recommendation scenario, student affection is related to their preferences for questions, affecting their learning motivation and learning outcomes [6-7]. Moreover, [8-10] confirmed that applying affective information to recommender systems can effectively improve their effectiveness. Thus, it is necessary to introduce affection to CMD to model their personalized cognition, which will guide the subsequent question recommendation.

To tackle the above-mentioned challenges, we propose an affection-enhanced personalized question recommendation method (AE-PQR). By introducing CDM to CF, both students' individual and common personalities can be effectively modeled. Since personalized question recommendation heavily relies on the student's cognition, we try to improve the accuracy of cognitive diagnosis to guide effective question recommendation. Specifically, we integrate student affection as an outer cognition into the cognitive diagnosis process. Combining the predicted responses by affection-aware CDM and CF, personalized questions can be

recommended for students based on their final predicted responses. Our method selects proper difficulty questions that match students' cognition to improve the efficiency and effectiveness of students' exercise practice. The main contributions of this paper are as follows:

1) We design a process for enhanced cognitive diagnosis. By employing the non-compensatory bi-dimensional item response model (NCB-IRM), the enhanced cognition adaptively integrates students' outer cognition and inner cognition to model their individual personalities. Specifically, students' affection is treated as the outer cognition.

2) We propose an AE-PQR. It introduces NCB-IRM to CF to be rational and accurate for personalized question recommendation in online learning. Specifically, PMF is employed to utilize the students' common characteristics. Considering both individual and common characteristics, the recommended questions of AE-PQR are adjusted to the students' requirements.

Through extensive experiments, the effectiveness and rationality of the proposed method are validated. Compared to comparative methods, the AE-PQR can diagnose student cognition more accurately. Moreover, it is proved to be suitable for the students' question recommendation.

The structure of this paper is arranged as follows: Section 2 introduces existing relevant works on question recommendation. Section 3 introduces the proposed AE-PQR models. Section 4 presents the experimental results, and Section 5 concludes the work of this paper.

2. Related Work

This section first introduces the current recommendation methods from three aspects: CF, CDM, and affection.

2.1 Collaborative Filtering Recommendation

CF employs the user's historical behavioral data and similar users' behavioral data to predict the items that they may be interested in. It assumes that there is an association between each user and the item. The recommendations are categorized into the following two types: user-based recommendation [11] and item-based recommendation [12]. User-based recommendation recommends items to users with similar preferences. Item-based recommendation recommends items that are similar to those users have already purchased.

However, CF relies heavily on the similarity between users and items. It suggests that the number of user-item ratings affects the accuracy of the recommendation method. If historical data is insufficient, CF encounters data sparsity and cold-start problems. Based on the matrix factorization algorithm, PMF can solve the above-mentioned problems and deal with large-scale question recommendations. PMF maps users and items to a common low-dimensional vector space to compute the latent features separately. Thus, PMF is highly adaptive to sparse data and can also be used for large-scale student-question recommendation.

However, PMF overlooks the characteristics of skills (*e.g.*, the relationship between questions and skills) in question recommendation. Currently, more and more researchers are concerned about how to also consider the features of skills in the question recommendation. In [13, 14], researchers developed recommendation methods that combine the CDM (*i.e.*, IRT, MIRT, DINA, *etc.*) in PMF. They also suggested improving the recommendation accuracy in terms of students' personalized intensity and CDM construction [13].

2.2 Cognitive Diagnosis for Question Recommendation

CDM [15] is often used to measure students' cognition by conducting a comprehensive study of their learning data over time. The input is students' answer logs, which are composed of question information and student responses. The output is the students' diagnosed knowledge proficiency or cognition of each skill.

Instead of recommending popular questions like CF, CDM recommends questions that are suitable for the students' cognition. Specifically, CDM first diagnoses the students' cognition based on their answer logs. Then, according to the students' cognition, it filters the recommended question list by a suitable difficulty. Finally, it recommends the proper questions that align with students' cognition. The CDM recommendation provides more reasonable results for the question recommendation.

However, due to CDM's limited effectiveness, wrongly assessed cognition would have a direct impact on question recommendation. In addition, CDM only models students' individual personalities but ignores the common characteristics. Therefore, through the effectiveness of CDM-based CF recommendation [13, 14], we can imply that it is necessary to introduce a comprehensive method for personalized question recommendation.

2.3 Affection-aware Recommendation

Affection plays a significant role in students' learning processes. The American humanistic psychologist Rogers believed that human cognitive activities were accompanied by certain affective factors [16]. It indicates that human affection must be closely related to the current environment and cognitive activities. The question recommendation also focuses on the environment and cognitive activities. Therefore, studying the students' affection during their learning process is crucial for personalized recommendation.

Generally, affection can be categorized into two types: positive affection (*e.g.*, concentration, joy, pride, enjoyment, *etc.*) and negative affection (*e.g.*, boredom, frustration, anxiety, shame, *etc.*) [17, 18]. However, the recommendation method largely ignores affection as a source of user context because affection is difficult to measure and easily misinterpreted.

To solve this problem, researchers developed several methods for extracting and coding affection. [10] proposed a method for automatically extracting affective context from user comments on YouTube short films using the Baker-Rodrigo protocol [19] to code students' affection. With these methods, more and more emotion-aware recommendation methods are being developed. In [8], they developed hybrid information fusion methods to design the emotion-aware recommender system. In [9], they proposed a general emotion-aware personalized music recommendation method. Based on user affective profiles, they estimated whether the items to be recommended were suitable for the user's current affective state.

All the emotion-aware recommendation methods suggest that they can effectively reflect user preferences and improve recommendation accuracy, which inspires affective-enhanced question recommendation. As a branch of multidimensional item response theory (MIRT), NCB-IRM [20] is an extension of item response theory (IRT). It models students' cognition as multiple latent traits. In NCB-IRM, "non-compensatory" means that even if a student has higher cognition in one dimension, it cannot compensate for lower cognition in another dimension. This model technically makes it possible to treat affection as one of the latent traits.

3. Enhanced Cognitive Diagnosis and Question Recommendation

3.1 Problem Formation

The goal of our formulation is to effectively 1) integrate affective features into the cognitive diagnosis process and 2) utilize both individual and common personalities for question recommendation. The personalized recommendation framework is shown in **Fig. 1**.

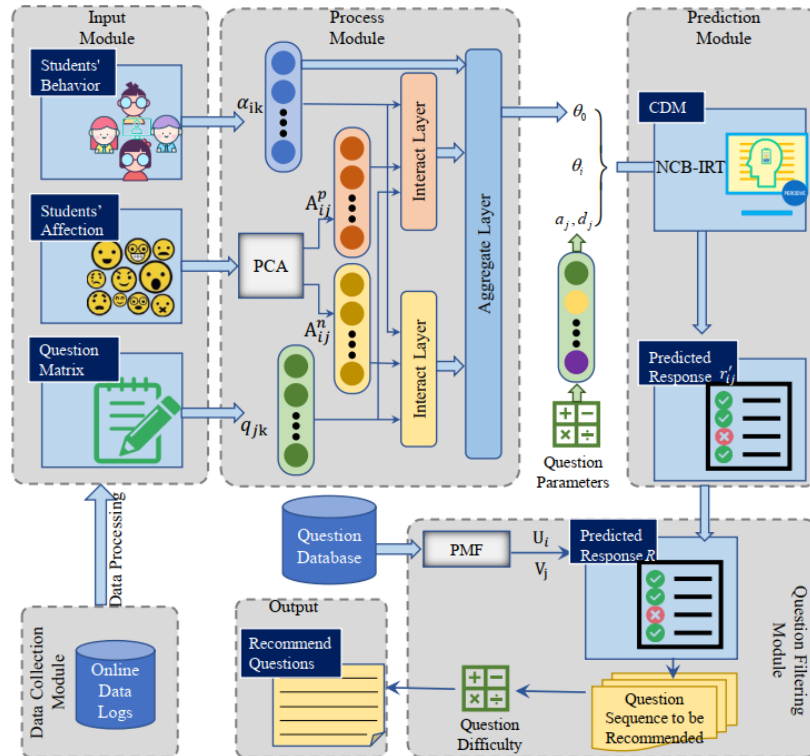


Fig. 1. The framework of personalized question recommendation in online learning.

During the learning process, students generated a series of data points. Assuming there are I students, J questions, and K skills in the online learning platform, the students' affective data is denoted as E , students' behavior data is denoted as B , and answer logs are denoted as R . The symbols and descriptions involved in this paper are shown in **Table 1**. The relationship between some of the symbols is explained as follows: from the relationship between students and questions, if $r_{ij} = 1$, it means that student i answered question j correctly; if $r_{ij} = 0$, it means that student i answered question j incorrectly. From the relationship between questions and skills, if $q_{jk} = 1$, it means that question j examines skill k ; if $q_{jk} = 0$, it means that question j does not examine skill k . The q_{jk} is denoted as the Q matrix, which is usually annotated by experts.

Table 1. Variable description

Variable	Description
i	the student i
j	the question j
k	the skill k
r_{ij}	the student i 's response to the question j
q_{jk}	the question j examined skill k
A_{ij}	the student i 's affection for the question j
α_{ik}	student i 's latent mastery in skill k

3.2 Cognitive Diagnosis with Affective Features

If we substitute NCB-IRM with CDM, the upper part of Fig. 1 is a cognitive diagnosis process, namely, the affection-enhanced cognitive diagnosis model (AE-CDM). AE-CDM consists of three parts: the input module, the processing module, and the prediction module. The input includes student behavior data \mathbf{B} , affective data \mathbf{E} , and question-skill matrix \mathbf{Q} . The processing module uses the student's latent skill mastery α_{ik} , student affection A_{ij} , and question-skill q_{jk} as input for the subsequent layer. In Fig. 2, The processing module contains one clustering layer, two interaction layers, and one aggregation layer. Specifically, the interaction and aggregation layers apply self-attention to obtain the outer cognition θ_o using the affective feature. In the prediction module, the outer cognition θ_o is fused with the inner cognition θ_i to obtain the student's enhanced cognition, $\theta = C(\theta_o, \theta_i)$.

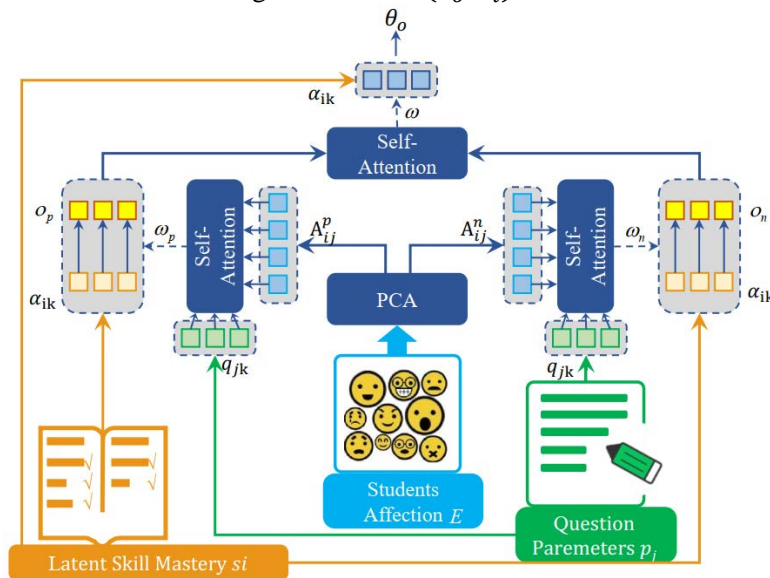


Fig. 2. Affective interaction modeling.

3.2.1 PCA-based Affection Clustering

This layer utilizes affective features by clustering complex affection. Since PCA can retain the most important features of affection data while removing noise and irrelevant features, we use PCA in this section. After applying PCA for clustering, two main types of affection are obtained: the positive affection A_{ij}^p and the negative affection A_{ij}^n .

$$\begin{aligned} A_{ij}^p &= \mathbf{E} \cdot V_1, \\ A_{ij}^n &= \mathbf{E} \cdot V_2, \end{aligned} \quad (1)$$

In (1), \mathbf{E} represents the affection data, \cdot represents the dot product operation between vectors, A_{ij}^p and A_{ij}^n represent the projection values of the affective data onto the first two principal components, and V_1, V_2 are the eigenvectors.

3.2.2 Self-attention-based Affective Interaction Layer

To address the difficulty in understanding the interaction between affective features and student behavior, two self-attention modules are used. In the interaction layer, we simulate the interaction between clustered affection A_{ij}^p, A_{ij}^n , student's latent skill mastery α_{ik} , and skill-question matrix q_{jk} . Taking the positive affective input as an example, the self-attention module uses the affective feature A_{ij}^p as query, the question representation \mathbf{p}_j as key, and the student's latent skill mastery representation \mathbf{s}_i as value. First, to assign weights to students' positive affection based on question characteristics, the attention weight ω_p is computed by the cosine similarity of query and key. Then, the interaction-weighted sum of ω_p and the student's latent skill mastery \mathbf{s}_i generates the interacted positive affective vector \mathbf{o}_p . This step adjusts for the effects of positive affection on students' knowledge proficiency.

$$\omega_p = \text{Softmax}(\text{sim}(A_{ij}^p, \mathbf{p}_j)), \mathbf{o}_p = \sum_{i=1}^I \omega_p \mathbf{s}_i, \quad (2)$$

where $\mathbf{p}_j = (q_{j1}, q_{j2}, \dots, q_{jk})$, $\mathbf{s}_i = (\alpha_{i1}, \alpha_{i2}, \dots, \alpha_{ik})$.

Similarly, the interaction layer utilizes self-attention to simulate the interaction between the student's negative affection, the student's skill mastery, and the answered questions. Thus, we can figure out the interacted negative affective vector \mathbf{o}_n .

$$\omega_n = \text{Softmax}(\text{sim}(A_{ij}^n, \mathbf{p}_j)), \mathbf{o}_n = \sum_{i=1}^I \omega_n \mathbf{s}_i, \quad (3)$$

3.2.3 Aggregation Layer Based on Different Interactive Affection

In the aggregation layer, another self-attention module is designed to aggregate the computed different interacted affective vectors. Specifically, \mathbf{o}_p and \mathbf{o}_n are used as queries and keys in the self-attention, and the student's skill mastery \mathbf{s}_i as the value. This aggregation process combines the different interacted affective vectors to obtain the outer cognition θ_o , regulating the effects of different affection on students.

$$\omega_i = \text{Softmax}(\text{sim}(\mathbf{o}_p, \mathbf{o}_n)), \theta_o = \sum_{i=1}^I \omega_i \mathbf{s}_i. \quad (4)$$

3.2.4 Enhanced Cognitive Diagnosis with Integrated Affective Features

Students' enhanced cognition θ includes the outer cognition θ_o and the inner cognition θ_i . The affection of student i on question j can be represented as:

$$A_{ij} = \begin{cases} A_{ij} & \text{if } A_{ij} \in \mathbf{E}^*; \\ \mathbf{U}_{i_A}^T \mathbf{V}_{j_A} & \text{if } A_{ij} \in \mathbf{E}^*. \end{cases} \quad (5)$$

where \mathbf{E}^* represents the affective matrix of student i on the answered question j .

The influence of affective features on outer cognition θ_o can be calculated using (1), (2), (3), (4). The inner cognition θ_i is diagnosed by CDM using the student's answer log. In order to obtain the student's enhanced cognition, NCB-IRM considers two extreme cases of the student's affective features separately:

1) Without being influenced by affective features, student i correctly answers question j . The student response function is modeled as:

$$\eta_{ij}^{\text{def}} = P(r_{ij} = 1 | A_{ij} = 0) = \frac{1}{1 + \exp\{-a_j(\theta_i - d_j)\}}, \quad (6)$$

where r_{ij} represents the response of student i on question j ; A_{ij} represents the affection of student i in answering question j ; θ_i , a_j , d_j are the student's inner cognition, question discrimination and question difficulty in the 2PL-IRT model, respectively [21].

2) Only influenced by the affective feature, student i correctly answers question j . The student response function can be modeled as:

$$\zeta_{ij}^{\text{def}} = P(r_{ij} = 1 | A_{ij} = 1) = \frac{1}{1 + \exp\{-\alpha_j(\theta_o - b_j)\}}, \quad (7)$$

Besides the same symbols as in (6), θ_o , α_j , b_j are the student's outer cognition, question discrimination, and question difficulty in the 2PL-IRT model, respectively.

Assuming that each question's response is statistically independent of the student's cognition, the responses are modeled using the Bernoulli distribution:

$$P(r_{ij} = 1 | \theta_i, a_j, d_j, \theta_o, \alpha_j, b_j, A_{ij}) = \eta_{ij}^{1-A_{ij}} \zeta_{ij}^{A_{ij}}, \quad (8)$$

where η_{ij} and ζ_{ij} represent the probabilities of student i correctly answering question j based on question practice or affective features, respectively.

As shown in Fig. 3, the response matrix is a comprehensive reflection of the prior knowledge response and the prior students' affection. According to (8), given the response \mathbf{R} , the item response function for θ_i , θ_o , a_j , α_j , b_j , d_j , is described by (9):

$$L(\theta_i, \theta_o, a_j, \alpha_j, d_j, b_j) = \prod_{i,j} (\eta_{ij}^{1-A_{ij}} \zeta_{ij}^{A_{ij}})^{r_{ij}} (1 - \eta_{ij}^{1-A_{ij}} \zeta_{ij}^{A_{ij}})^{1-r_{ij}}, \quad (9)$$

We maximize the marginal likelihood in (9) to assess students' mastery of the questions. Different from the parameter estimation for 2PL-IRT, the optimal solutions for θ_i , a_j , α_j , b_j , and d_j are obtained through a Metropolis-Hastings based Markov Chain Monte Carlo method. Specifically, all parameters are first randomized as initial values. Then, using the observed response \mathbf{R} , the conditional probability of students' inner cognition θ_i , corresponding question discrimination a_j , α_j , and corresponding question difficulty b_j , d_j are calculated. Next, the acceptance probability of the samples is computed based on the Metropolis-Hastings algorithm.

3.3 Affection-enhanced Personalized Question Recommendation

In item recommendation, PMF uses latent feature vectors to capture the implicit relationship between users and items, improving the accuracy and diversity of recommendations. In our personalized question recommendation scenario, the NCB-IRM framework integrates students' outer and inner cognition, enriching the information dimension of cognitive diagnosis. As shown in the upper part of Fig. 3, by introducing PMF to NCB-IRM, AE-PQR integrates enhanced cognition and latent feature vectors to enhance the personalization and adaptability of recommendations. The output of AE-PQR is the predicted students' responses.

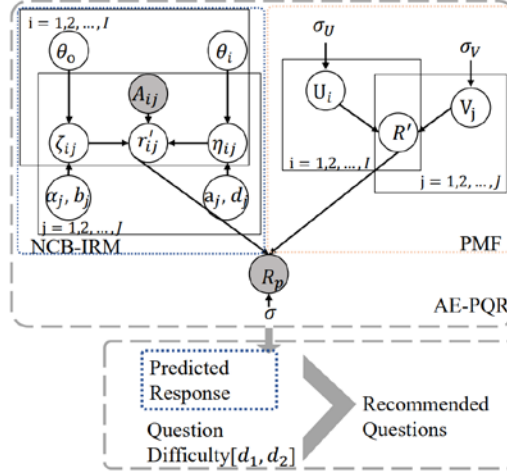


Fig. 3. The framework for personalized question recommendation.

$$\mathbf{R}_p = \mu + \rho r_{ij} + (1 - \rho) \mathbf{R}' = \mu + \rho r_{ij} + (1 - \rho) \mathbf{U}_i^T \mathbf{V}_j, \quad (10)$$

where \mathbf{R}_p is the predicted student response, μ is the average score of all students, r_{ij} represents the potential response of the students' unanswered questions calculated by AE-CDM, \mathbf{R}' represents the potential response of the students' unanswered questions evaluated by PMF, and ρ controls the proportion between the individuality and commonality of predicted students' responses.

Each part of the AE-PQR can explain a certain attribute of the observed responses. In order to make the prediction close to the ground truth, the optimization objective turns into minimizing the function in (11).

$$E = \frac{1}{2} \sum_{i=1}^I \sum_{j=1}^J I_{ij} (\mathbf{R} - \mathbf{R}_p)^2 + \frac{\lambda_U}{2} \sum_{i=1}^I \|\mathbf{U}_i\|_{Fro}^2 + \frac{\lambda_V}{2} \sum_{j=1}^J \|\mathbf{V}_j\|_{Fro}^2, \quad (11)$$

where I_{ij} equals to 1 when student i answer the question j , $\lambda_U = \frac{\sigma^2}{\sigma_U^2}$, $\lambda_V = \frac{\sigma^2}{\sigma_V^2}$ is the regularization parameter, \mathbf{U}_i and \mathbf{V}_j represent the latent feature vectors of students and questions in D dimensions, respectively, and $\|\cdot\|_{Fro}^2$ denotes the matrix normalization. Take the derivatives with respect to $\mathbf{U}_i, \mathbf{V}_j$:

$$\begin{aligned} \frac{\partial E}{\partial \mathbf{U}_i} &= (\mathbf{R} - \mathbf{U}_i^T \mathbf{V}_j) \mathbf{V}_j - \lambda_U \mathbf{U}_i, \\ \frac{\partial E}{\partial \mathbf{V}_j} &= (\mathbf{R} - \mathbf{U}_i^T \mathbf{V}_j) \mathbf{U}_i - \lambda_V \mathbf{V}_j, \end{aligned} \quad (12)$$

Then update using the stochastic gradient descent method until convergence or reaching the maximum number of iterations.

$$\begin{aligned} \mathbf{U}_i &= \mathbf{U}_i + \alpha \frac{\partial E}{\partial \mathbf{U}_i}, \\ \mathbf{V}_j &= \mathbf{V}_j + \alpha \frac{\partial E}{\partial \mathbf{V}_j}. \end{aligned} \quad (13)$$

Unlike traditional interest-based recommendations, the difficulty of recommended questions should be within an appropriate range. Particularly, we need to recommend questions with a proper difficulty that aligns with the student's cognition. The difficulty is

defined relative to the student's cognition. For student i and question j , the difficulty is defined as d_j . Similarly, the probability of student i correctly answering question j is equal to $1 - d_j$. Based on the student's cognition, AE-PQR recommends questions with difficulty in $[d_1, d_2]$ to student i from the J questions. The recommended question set is evaluated to be answered correctly between $[1 - d_2, 1 - d_1]$.

4. Experimental Analysis

4.1 Dataset

ASSISTments is a general cognitive diagnosis public dataset that records students' answer logs. We use two online datasets from the ASSISTments platform, which include student data from the years 2009–2010 and 2017. ASSIST2009 and ASSIST2017 are the benchmark datasets for modeling students' cognition and question recommendation. The ASSIST2017 dataset incorporates the Baker Rodrigo Ocumpaugh Monitoring Protocol for encoding the students' affection [19]. Adding affection data to the answer logs, ASSIST2017 records 942816 interactions from 686 students. After data processing, the datasets are split into train, test, and valid sets in a ratio of 7:2:1. **Table 2** summarizes the basic statistical data of these two datasets.

Table 2. Description of the datasets

Dataset	ASSIST2009	ASSIST2017
Train	186049	299946
Test	55760	85699
Valid	25606	42849
Max_studentid	4163	7783
Max_questionid	17746	3163
Number of skills	123	102

Part of the question-skill matrices Q from ASSIST2009 and ASSIST2017 are visualized in **Fig. 4**. Each row of the subplots represents a question set, and each column represents a skill. White cells indicate that a question tests a particular skill, while yellow cells indicate that a question does not test a certain skill. For example, in **Fig. 4 (a)**, in the first column, question Q6 tests skill K15. Most of the exercise questions test two skills or fewer, indicating that the Q matrix is very sparse.

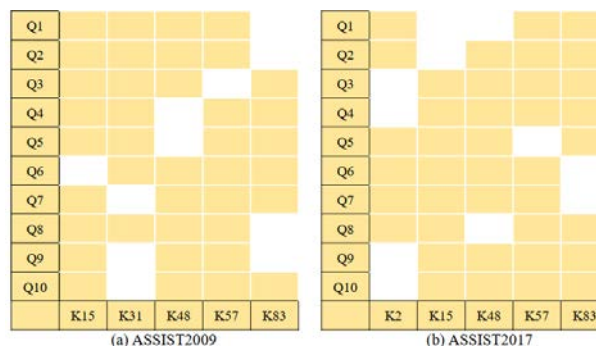


Fig. 4. Q matrix of the two datasets. (Blanks indicate that the question tests the skill).

We perform data preprocessing in this subsection. First, some empty values in the dataset are removed. Then, self-correlated features with student responses are excluded. Finally, we

select the top twenty features most correlated with student responses based on the Spearman correlation coefficient.

4.2 Baseline Models

In personalized question recommendation, CF-based recommendation and CDM-based recommendation are the two mainstream recommendation methods. Here, we apply the relatively mature and widely used recommendation models as the baseline models.

To validate AE-CDM, this subsection proposes three CDM frameworks (*i.e.*, IRT, MIRT, DINA) [7] that integrate traditional cognition with affective features for conducting enhanced cognitive diagnosis. An attention-based framework E-CDM, an affection-based framework A-CDM, and an enhanced affection-attention-based framework AE-CDM are proposed. The detailed CDM baseline models are introduced as follows:

IRT provides interpretable parameters (*i.e.*, student skill mastery, question discrimination, and difficulty). It uses a logistic function to describe students' cognition and analyze their performance.

MIRT is a continuous multi-dimensional CDM. It extends the logistic item response function of IRT.

Deterministic input, noisy "and" gate (DINA) [22] introduces a question knowledge matrix Q and considers the impact of students' slipping and guessing behaviors.

To validate the effectiveness of AE-PQR in question recommendation, we conducted comparative experiments with the following baseline models:

PMF [23] factorizes the student and question matrices into low-dimensional latent feature vectors using the student response matrix. Then the latent feature vectors can predict the absent student responses. Recommendations are made based on the predicted student responses.

CDM models students' cognition using various models (IRT, MIRT, DINA) and further predicts students' responses. It recommends challenging or easy questions for students separately.

PMF-DINA is a question recommendation method proposed by [13]. It first combines PMF and DINA to predict student responses and then recommends questions of corresponding difficulty.

4.3 Evaluation Metrics and Experimental Settings

4.3.1 Evaluation Metrics

1) AE-CDM

Since it is difficult to accurately obtain the true values of students' cognition, the performance of cognitive diagnosis is difficult to evaluate directly. Therefore, based on existing work, R is utilized to supervise the performance of AE-CDM. The performance is indirectly assessed through both regression and classification, depending on whether the prediction result is a score or a response (0 or 1).

When treating the problem as a regression task, the root mean square error (RMSE) and mean absolute error (MAE) [24] are used to quantify the distance between predicted responses and actual responses. Smaller values of RMSE and MAE indicate better predictive performance of the model.

When treating the problem as a classification task, the predicted results (1, 0) represent positive and negative instances, and evaluation metrics commonly used are the area under the ROC curve (AUC) and the prediction accuracy (ACC) [25]. The values of AUC and ACC range from 0 to 1, with values closer to 1 indicating better predictive results.

2) AE-PQR

In traditional recommendation methods, precision, recall, and F1 score are commonly used metrics. Typically, they recommend top-N items to the users. However, question recommendation is not about recommending popular or easy questions to students but rather attempting to recommend questions that align with their cognition. Therefore, the evaluation metric is set as the rate of correctly answered recommended questions, shown as (14). The higher the RACC, the closer the recommended questions are to the student's cognition.

$$\text{RACC} = \frac{\text{right}}{\text{num of recommended problems}}, \quad (14)$$

4.3.2 Experimental Setting

During the training process, Xavier initialization is used to initialize the parameters. Specifically, these parameters are filled with random values sampled from $N(0, \text{std}^2)$, where $\text{std} = \sqrt{\frac{2}{n_i + n_{i+1}}}$, n_i represents the input dimension of the neural network, and n_{i+1} represents the output dimension of the neural network. In addition, the epoch is set to 5 and the learning rate is set to 0.001. In the CDM, the question discrimination and question difficulty in IRT are set to $a_j = 4$ and $d_j = 0$ respectively. In MIRT, $a_j = 0$ and $d_j = 0$. For DINA, the slip, guess, and step size are set to $\text{max_slip} = 0.4$, $\text{max_guess} = 0.4$, and $\text{max_step} = 1000$. All models are implemented in Python using PyTorch. The hardware configuration for the experiment is a 2.3 GHz Dual-Core Intel Core i5 with 8GB of RAM, and the operating system is macOS Big Sur 11.3.1.

4.4 Experimental Result Analysis

4.4.1 Cognitive Diagnosis

1) Attention Module

The ASSIST2009 dataset only includes the student answer logs. We first validated the effectiveness of E-CDM on this dataset by applying self-attention to important features. The prediction results of CDM and E-CDM in ASSIST2009 are shown in [Table 3](#).

Table 3. Cognitive prediction results in ASSIST2009

	ACC	MAE
IRT	0.644	0.396
E-IRT	0.659	0.375
MIRT	0.576	0.429
E-MIRT	0.575	0.430
DINA	0.537	0.472
E-DINA	0.644	0.404

Intuitively, all E-CDM models show an increase in ACC compared to CDM and a decrease in MAE. It indicates that the self-attention module can effectively capture the internal correlations among student answer logs. One notable point in [Table 3](#) is that E-IRT (ACC = 0.659, MAE = 0.375) performs the best among these models, suggesting that IRT has a good fit for the ASSIST2009 dataset. Another notable point is that from the perspective of model improvement, E-DINA shows the largest improvement over the DINA model (ACC improvement of 0.107, MAE

decrease of 0.079). However, the prediction results of E-MIRT do not show improvement compared to MIRT and even show a decrease in ACC (0.01) and an increase in MAE (0.01). This implies that MIRT may already have expanded the model dimensions based on IRT, and introducing the self-attention module will cause overfitting of the model.

2) Affection-enhanced Module

To explore whether incorporating affective features will enhance the model's performance, experiments are conducted on the ASSIST2017 dataset. The student cognitive prediction results for ASSIST2017 are shown in **Table 4**. Firstly, by incorporating affective features as input, A-CDM shows significant performance improvements compared to CDM, which focuses only on student answer logs. This indicates that affective features have an important impact on the model's prediction. Moreover, AE-CDM shows improvement compared to A-CDM in AUC, ACC, and MAE.

Table 4. Cognitive prediction results in ASSIST2017

	AUC	ACC	RMSE	MAE
IRT	0.535	0.470	0.575	0.515
E-IRT	0.510	0.464	0.587	0.521
A-IRT	0.701	0.653	0.465	0.431
AE-IRT	0.741	0.682	0.451	0.412
MIRT	0.551	0.535	0.633	0.466
E-MIRT	0.551	0.538	0.634	0.466
A-MIRT	0.608	0.598	0.533	0.435
AE-MIRT	0.642	0.613	0.553	0.401
DINA	0.508	0.524	0.572	0.480
E-DINA	0.512	0.431	0.631	0.552
A-DINA	0.543	0.534	0.564	0.467
AE-DINA	0.566	0.561	0.551	0.450

3) Modeling Time

Fig. 5 compares the runtime of AE-CDM on the ASSIST2017 dataset. After data preprocessing and deduplication, the runtime of all three models (including training and prediction time) is within 150 seconds. Among the three frameworks of AE-CDM, AE-IRT requires less time with its simple model structure. In ASSIST2017, AE-IRT is not only more accurate but also less time-consuming. The results exhibit that complex multidimensional CDM structures (*i.e.*, MIRT and DINA) do not necessarily lead to better predictions.

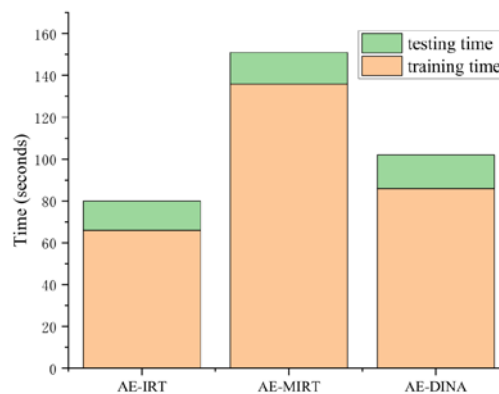


Fig. 5. Time consumption on the ASSIST2017 dataset.

4.4.2 Analysis of Model Structure

1) Ablation Experiment

This section conducts an ablation experiment to validate the effectiveness of each module in AE-CDM. We analyze the impact of clustering, interaction, and fusion layers on model performance. Specifically, each layer is replaced with a regular aggregation layer that only computes the average of the inputs. The parameters of the other three layers are kept unchanged.

Table 5 records the results of the ablation experiment of AE-CDM on the ASSIST2017 dataset. Regardless of which layer is replaced, there is a certain degree of performance decline. This indicates that each layer contributes to the final prediction performance of the model. Moreover, replacing the interaction layer has the largest impact on the final model performance. It suggests the crucial role of the interaction between student affection, student behavior, and the questions during the modeling process.

Table 5. Results of the ablation experiment

	AE-IRT				AE-MIRT				AE-DINA			
	AUC	ACC	RMSE	MAE	AUC	ACC	RMSE	MAE	AUC	ACC	RMSE	MAE
AE-CDM	0.741	0.682	0.451	0.412	0.642	0.613	0.553	0.401	0.566	0.561	0.551	0.450
- cluster	0.722	0.666	0.460	0.425	0.629	0.605	0.562	0.413	0.548	0.557	0.561	0.470
- interact (negative)	0.584	0.627	0.488	0.506	0.583	0.548	0.634	0.479	0.535	0.507	0.621	0.546
- interact (positive)	0.565	0.604	0.513	0.517	0.570	0.538	0.632	0.466	0.514	0.431	0.627	0.528
- aggregate	0.701	0.653	0.465	0.431	0.608	0.598	0.533	0.435	0.543	0.534	0.564	0.467

2) Parameter Analysis

Affection-enhanced cognition can be represented as $\theta = C(\theta_o, \theta_i) = t\theta_o + (1 - t)\theta_i$. To explore the effect of affective feature's weight t on the cognitive diagnosis outcome of AE-CDM, AUC is used as an indicator to visualize the parameter t 's impact on the AE-CDM outcomes.

As shown in **Fig. 6**, experiments are conducted on the three CDM frameworks in ASSIST2017 dataset. The model achieves the highest AUC value when the weight t is close to 0.5. Parameter t represents the weight of outer cognition for affective feature modeling, indicating that student affective features should not be overlooked in cognitive diagnosis.

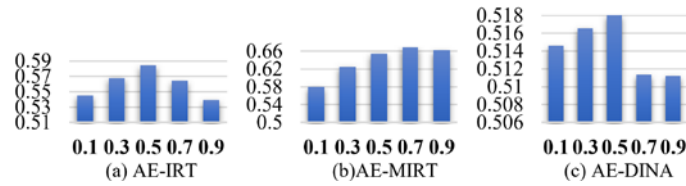


Fig. 6. Distribution of the weight t for outer cognition in the ASSIST2017 dataset.

3) Students' Cognitive Visualization

In **Fig. 7**, students' cognition is visualized on six skills (multiplication, square root, area, probability, equation solving, and pattern finding). Firstly, the students' affection is clustered into two groups in the ASSIST2017 dataset: the positive and the negative, for 205105 and 32272 interactions separately. They are denoted as n_{pos} and n_{neg} . The average cognition of students in different affection clusters is calculated using $\theta_{avg} = \frac{\sum_{i=1}^n \theta}{n}$, where θ represents the enhanced cognition and n represents the number of students in each category.

From **Fig. 7**, the following observations can be made: Firstly, students in different affection clusters exhibit constancy in their mastery of skills. Generally, students have a higher mastery level in multiplication, square-root, and pattern-finding, while their mastery level is lower in area, probability, and equation-solving. It is obvious that the prerequisite skills (*i.e.*, multiplication, square-root) of area are easier for students to master. While the difficulty of the skills (*i.e.*, area, equation-solving) increases, students' cognition decreases, which conforms to the laws of cognition [26].

Secondly, skills with higher cognition (*i.e.*, multiplication, square-root, and equation-solving) are better mastered by students who show positive affection. For instance, the cognition of positive students in multiplication, square-root is 0.01 higher than that of negative students. Similarly, skills with lower cognition (*i.e.*, area, pattern-finding, and probability) are better mastered by students who show negative affection. Through further analysis, we observe that these students mostly show confused and frustrated affection. It indicates that these skills are too difficult for these students. They have made excessive efforts to resolve the questions, consequently experiencing a negative emotional impact in the process. Overall, positive affection is correlated with higher cognition. This finding is consistent with the study in [27]. Therefore, actively mobilizing students' positive affection during the learning process may contribute to higher cognition. In addition, the students' cognitive visualization can further guide the question recommendation.

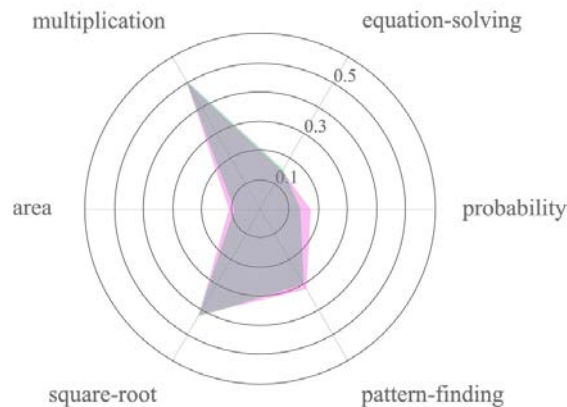


Fig. 7. Average students' cognition in different affection (positive in green, negative in pink, overlap in gray).

4.4.3 Analysis of Recommendation Effectiveness

In our proposed recommendation method, questions with difficulty levels ranging from 0 to 1 are recommended to students according to their cognition. The evaluation metric RACC represents the ratio of correctly answered recommended questions by students. In the ASSIST2009 and ASSIST2017 datasets, the relationship between the difficulty of recommended questions and RACC is shown in Fig. 8.

As the difficulty of recommended questions increases, the RACC of students for these questions decreases continuously. It indicates that these models are capable of adaptively aligning the difficulty of recommended questions with students' cognition. In other words, the recommended questions are of an appropriate difficulty level based on students' needs. On the one hand, the proposed AE-PQR, where we choose AE-PMFDINA here for intuitive comparison, outperforms PMF-DINA in terms of RACC. AE-PQR, after incorporating students' affective features, can model students' cognition more accurately. Furthermore, it can predict their responses more precisely and recommend proper questions within the set difficulty range for each individual student. On the one hand, by taking advantage of CDM and CF in recommending questions, our proposed method considers both students' individual personalities and the common relationship of questions. Our proposed model devotes effort to the reliability and interpretability of recommended questions.

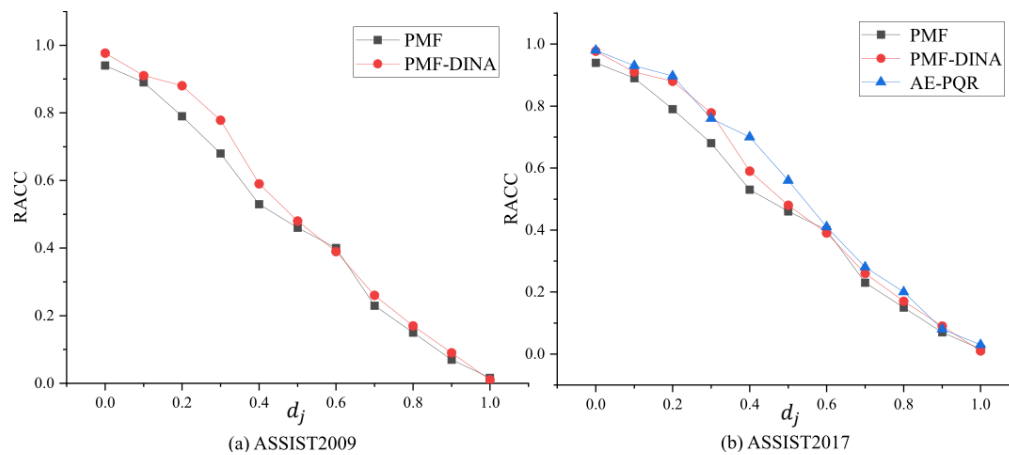


Fig. 8. The impact of different question difficulty levels d_j on students' RACC.

According to the zone of proximal development (ZPD) theory, the difficulty level of questions should align with students' current cognitive levels [28].

A case study is presented to show how AE-PQR recommends questions for two students. Fig. 9 shows the cognition of students A and B on six skills in ASSIST2017. It can be seen that Student A has a great mastery of area and pattern-finding but a poor mastery of square-root and probability. Student B has a good mastery of square-root and pattern-finding but a bad mastery of area and probability. When selecting the questions with d_j between 0.3 and 0.5 for recommendation, AE-PQR recommends questions 894, 2401, and 1597 to student A, and questions 4558, 2115, and 1624 to student B. The correlation between questions and skills is shown in Table 6.

Among them, it can be seen from Fig. 9 that student A's mastery of the skills square-root and probability is weak. The recommended question 894 tests square-root and recommended question 2401 tests probability. Student B is weak in mastering area and probability. The recommended question 4558 tests the skills of area, and the recommended question 2115 tests

the skills of probability. Due to Student A's high mastery of area, the student is not recommended for this skill-relevant question. This case study indicates that our proposed AE-PQR can recommend the corresponding personalized questions for each student. When recommending difficult questions, the results of the recommended questions are highly interpretable.

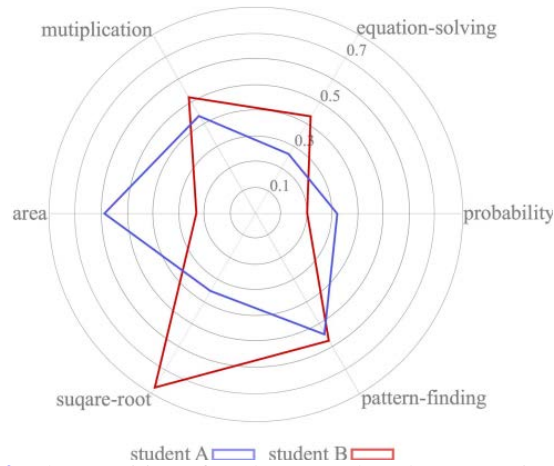


Fig. 9. The cognition of student A and student B on six skills.

Table 6. Recommended questions in a case study

	Square-root	area	probability	noskill
Student A	894	\	2401	1597
Student B	\	4558	2115	\

5. Conclusion

To recommend reasonable questions, we proposed an AE-PQR that incorporated CDM into CF. To model affective features appropriately, it used PCA and a hierarchical self-attention module to simulate the interaction between students' different affection and their outer cognition. Subsequently, it adaptively combined students' outer cognition with their inner cognition using the NCB-IRM framework, improving the accuracy of predicting individual personalities. Moreover, the common characteristics were utilized by employing PMF. Combining individual and common characteristics, our AE-PQR was able to predict students' responses to unanswered questions. It guided the regulation of the difficulty of recommendation questions, ensuring that the recommended questions were suitable for students' cognition. Experiment results on two public datasets confirmed the accuracy of cognitive diagnosis and the rationality of the recommended question using AE-PQR. It indicated that affective features significantly influence the results of cognitive diagnosis and question recommendation, highlighting the importance of considering affective factors in students' online behavioral data.

This paper attempted to simulate the interaction between students' affective features and other students' answering behaviors through a hierarchical self-attention module. In future research, as students' answer information becomes more diverse, it would be beneficial to consider the interaction of fine-grained time series learning data using neural network architectures like the Transformer, which is based on attention mechanism.

Acknowledgement

This work is partly supported by the National Natural Science Foundation of China (Grant No. 62277032), Education Scientific Planning Project of Jiangsu Province (Grant No. B/2022/01/150), Higher Education Reform Foundation of Jiangsu Province (Grant No. 2023JSJG021), China Postdoctoral Science Foundation (Grant No. 2022M721694), Teaching Reform Research Project of Nanjing University of Posts and Telecommunications (Grant No. JG00223JX01, JG00215JX01), Jiangsu Provincial Qinglan Project, and Priority Academic Program Development of Jiangsu Higher Education Institutions.

References

- [1] J. Yun and T. Park, "An Analysis of University Students' Needs for learning support functions of learning management system augmented with artificial intelligence technology," *KSII Trans. Internet Inf. Syst.*, vol. 17, no. 1, pp. 1-15, Jan. 2023. [Article \(CrossRef Link\)](#)
- [2] C. Channarong, C. Paosirikul, S. Maneeroj, and A. Takasu, "HybridBERT4Rec: a hybrid (content-based filtering and collaborative filtering) recommender system based on BERT," *IEEE Access*, vol. 10, pp. 56193-56206, May. 2022. [Article \(CrossRef Link\)](#)
- [3] F. Chen, C. Lu, Y. Cui, and Y. Gao, "Learning outcome modeling in computer-based assessments for learning: A sequential deep collaborative filtering approach," *IEEE Trans. Learn. Technol.*, vol. 16, no. 2, pp. 243-255, Apr. 2022. [Article \(CrossRef Link\)](#)
- [4] 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 Transactions on Information Systems (TOIS)*, vol. 38, no. 2, pp. 1-33, Apr. 2020. [Article \(CrossRef Link\)](#)
- [5] Y. Zhou, Q. Liu, J. Wu, F. Wang, Z. Huang, W. Tong, H. Xiong, E. Chen, and J. Ma, "Modeling context-aware features for cognitive diagnosis in student learning," in *Proc. of KDD*, Virtual Event, Singapore, pp. 2420-2428, 2021.
- [6] H. Tian, C. Gao, X. Xiao, H. Liu, B. He, H. Wu, H. Wang, and F. Wu, "SKEP: Sentiment knowledge enhanced pre-training for sentiment analysis," in *Proc. of the 58th Annual Meeting of the ACL*, Online, pp.4067-4076, 2022.
- [7] D. Ghosal, D. Hazarika, A. Roy, N. Majumder, R. Mihalcea, and S. Poria, "Kingdom: Knowledge-guided domain adaptation for sentiment analysis," in *Proc. of the 58th Annual Meeting of the ACL*, Online, pp.3098-3210, 2022.
- [8] Y. Qian, Y. Zhang, X. Ma, H. Yu, and L. Peng, "Ears: Emotion-aware recommender system based on hybrid information fusion," *Inf. Fusion*, vol. 46, pp. 141-146, Mar. 2019. [Article \(CrossRef Link\)](#)
- [9] A. Abdul, J. Chen, H.-Y. Liao, and S.-H. Chang, "An emotion-aware personalized music recommendation system using a convolutional neural networks approach," *Appl. Sci.*, vol. 8, no. 7, pp. 1103-1112, Jul. 2018. [Article \(CrossRef Link\)](#)
- [10] C. Orellana-Rodriguez, E. Diaz-Aviles, and W. Nejdl, "Mining affective context in short films for emotion-aware recommendation," in *Proc. of the 26th ACM Conference on Hypertext & Social Media*, Guzelyurt, Northern Cyprus, pp. 185-194, 2015.
- [11] Z.-D. Zhao and M.-S. Shang, "User-based collaborative-filtering recommendation algorithms on hadoop," in *Proc. of WKDD*, NW Washington, DC, USA, pp. 478-481, 2010.
- [12] Y.-L. Lin and N.-D. Ding, "Competitive gamification in crowdsourcingbased contextual-aware recommender systems," *Int. J. Man-Mach. Stud.*, vol. 177, pp. 103083-103092, Sept. 2023. [Article \(CrossRef Link\)](#)
- [13] Q. Li, X. Liu, X. Xu, and S. Lin, "Personalized test question recommendation method based on unified probabilistic matrix factorization," *Journal of Computer Applications*, vol. 38, no. 3, pp. 639-643, Aug. 2018. [Article \(CrossRef Link\)](#)

- [14] Z. Wu, M. Li, Y. Tang, and Q. Liang, "Exercise recommendation based on knowledge concept prediction," *Knowledge-Based Syst.*, vol. 210, pp. 106481-106490, Dec. 2020. [Article \(CrossRef Link\)](#)
- [15] T. Xin, C. Wang, P. Chen, and Y. Liu, "Cognitive diagnostic models: Methods for practical applications," *Front. Psychol.*, vol. 13, pp. 895399-895408, Apr. 2022. [Article \(CrossRef Link\)](#)
- [16] M. Chen, X. Wei, L. Zhou, "Integrated media platform-based virtual office hours implementation for online teaching in post-COVID-19 pandemic era," *KSII Trans. Internet Inf. Syst.*, vol. 15, no. 8, pp. 2732-2748, Aug. 2021. [Article \(CrossRef Link\)](#)
- [17] M.-T. Cheng, W.-Y. Huang, and M.-E. Hsu, "Does emotion matter? an investigation into the relationship between emotions and science learning outcomes in a game-based learning environment," *Br. J. Educ. Technol.*, vol. 51, no. 6, pp. 2233-2251, Nov. 2020. [Article \(CrossRef Link\)](#)
- [18] A. A. AlDahdouh, "Emotions among students engaging in connectivist learning experiences," *Int. Rev. Res. Open Distrib. Learn.*, vol. 21, no. 2, pp. 98-117, Apr. 2020. [Article \(CrossRef Link\)](#)
- [19] Y. Wang, N. T. Heffernan, and C. Heffernan, "Towards better affect detectors: Effect of missing skills, class features and common wrong answers," in *Proc. of LAK*, Poughkeepsie, NY, USA, pp. 31-35, 2015.
- [20] J.-P. Fox, R. K. Entink, and M. Avetisyan, "Compensatory and noncompensatory multidimensional randomized item response models," *Br. J. Math. Stat. Psychol.*, vol. 67, no. 1, pp. 133-152, Feb. 2014. [Article \(CrossRef Link\)](#)
- [21] A. Comotti, "Assessing psychometric scales through irt-based modelling with application to covid-19 data," Ph.D. dissertation, Statistica E Finanza Matematica, Università degli Studi di Milano-Bicocca, Milan, Italy, 2022.
- [22] Y. Gu, "Generic identifiability of the dina model and blessing of latent dependence," *Psychometrika*, vol. 88, no. 1, pp. 117-131, Mar. 2023. [Article \(CrossRef Link\)](#)
- [23] N. Fusi, R. Sheth, and M. Elibol, "Probabilistic matrix factorization for automated machine learning," in *Proc. of NIPS*, Red Hook, NY, USA, pp. 3352-3361, 2018.
- [24] H. Pei, B. Yang, J. Liu, and L. Dong, "Group sparse bayesian learning for active surveillance on epidemic dynamics," in *Proc. of AAAI*, New Orleans, LA, USA, pp.800-807, 2018.
- [25] A. P. Bradley, "The use of the area under the roc curve in the evaluation of machine learning algorithms," *Pattern Recognit.*, vol. 30, no. 7, pp. 1145-1159, Jul. 1997. [Article \(CrossRef Link\)](#)
- [26] N. Chater and G. D. Brown, "From universal laws of cognition to specific cognitive models," *Cogn. Sci.*, vol. 32, no. 1, pp. 36- 67, Jan. 2008. [Article \(CrossRef Link\)](#)
- [27] Z. A. Pardos, R. S. Baker, M. O. San Pedro, S. M. Gowda, and S. M. Gowda, "Affective states and state tests: Investigating how affect and engagement during the school year predict end-of-year learning outcomes," *J. Learn. Anal.*, vol. 1, no. 1, pp. 107-128, May. 2014. [Article \(CrossRef Link\)](#)
- [28] F. Liu, L. Zhao, J. Zhao, Q. Dai, C. Fan, and J. Shen, "Educational process mining for discovering students' problem-solving ability in computer programming education," *IEEE Trans. Learn. Technol.*, vol. 15, no. 6, pp. 709-719, Dec. 2022. [Article \(CrossRef Link\)](#)



Mingzi Chen received her B.S. degree at Computer and Information Science from New York Institute of Technology, New York, U.S.A and Nanjing University of Posts and Telecommunications, Nanjing, China in 2018. She is currently a Ph.D. candidate in Nanjing University of Posts and Telecommunications. Her current research interests include educational information science and technology, multimedia communications.



Xin Wei received the Ph.D. degree major in information and communication engineering from Southeast University, Nanjing, China, in 2009. He is currently a Professor with the Nanjing University of Posts and Telecommunications, Nanjing. His research interests include the areas of multimedia communications, educational technology.



Xuguang Zhang received the Ph.D. degree in information and communication engineering from the Nanjing University of Posts and Telecommunications, China, in 2022, where he is currently a Postdoctoral Researcher. His research interests are in the area of multimedia communications and computing, resource allocation, and signal processing in wireless networks.



Lei Ye received the Ph.D. degree major in information and communication engineering from Nanjing University of Posts and Telecommunications, Nanjing, China, in 2014. She is currently an Associate Professor with the Nanjing University of Posts and Telecommunications, Nanjing. Her research interests include the areas of speech signal processing, educational technology.