IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 20 March 2024, accepted 10 April 2024, date of publication 16 April 2024, date of current version 23 April 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3388718

RESEARCH ARTICLE

DKVMN-KAPS: Dynamic Key-Value Memory Networks Knowledge Tracing With Students' Knowledge-Absorption Ability and Problem-Solving Ability

WEI ZHANG[1,2], ZHONGWEI GONG[1], PEIHUA LUO[1], AND ZHIXIN LI[1]

[1]Faculty of Artificial Intelligence in Education, Central China Normal University, Wuhan 430079, China
[2]National Engineering Laboratory for Educational Big Data, Central China Normal University, Wuhan 430079, China

Corresponding author: Zhongwei Gong (gzw@mails.ccnu.edu.cn)

This work was supported by the National Natural Science Foundation of China under Grant 62377024.

ABSTRACT Knowledge tracing aims to predict students' future question-answering performance based on their historical question-answering records, but the current mainstream knowledge tracing model ignores the individual differences in different students' knowledge-absorption and problem-solving abilities, which leads to a poor prediction of students' question-answering performance by the model. To solve this, Dynamic Key-Value Memory Networks Knowledge Tracing with Students' Knowledge-Absorption Ability and Problem-Solving Ability (DKVMN-KAPS) is proposed in this paper. Firstly, a hierarchical convolutional neural network is used to consider students' knowledge mastery at multiple time steps, and then quantify students' knowledge-absorption ability, aiming to more accurately portray students' knowledge states; secondly, an autoencoder is used to dynamically update students' problem-solving ability at each time step; and finally, students' question answering performance is predicted by considering the students' knowledge state, problem-solving ability, and question features. Extensive experiments on three datasets show that the prediction performance of DKVMN-KAPS outperforms existing models and improves the prediction accuracy of deep knowledge tracing models.

INDEX TERMS Knowledge tracing, knowledge-absorption ability, problem-solving ability, individual differences.

I. INTRODUCTION

Knowledge tracing (KT) is a research hotspot in the field of personalized learning [1], which tracks changes in students' knowledge state through their history of answering questions to predict their future performance in answering questions. KT aims to provide feedback on students' weak knowledge links, assist teachers in optimizing their teaching plans, and help students adjust their learning plans, thus improving the efficiency and quality of personalized learning [2].

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

KT is mainly categorized into traditional KT models and deep learning-based KT models. Bayesian Knowledge Tracing (BKT) [3], as a classic traditional KT model, uses the Hidden Markov Model (HMM) to model each know-ledge point individually in order to predict students' mastery of specific knowledge points. Deep Knowledge Tracing (DKT) marks the first application of deep learning techniques in KT, opening up new paths of research [4]. Dynamic key-value neural network (DKVMN) [5] adds a storage module and corresponding read/write mechanism based on a traditional neural network to track students' mastery states of each knowledge point, which improves the model's

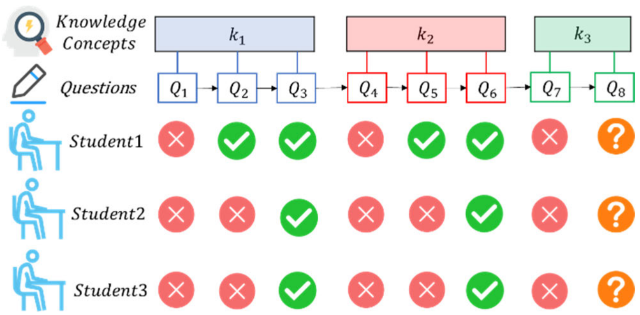

FIGURE 1. **Examples of the effect of knowledge absorption ability on how quickly students acquire knowledge.**

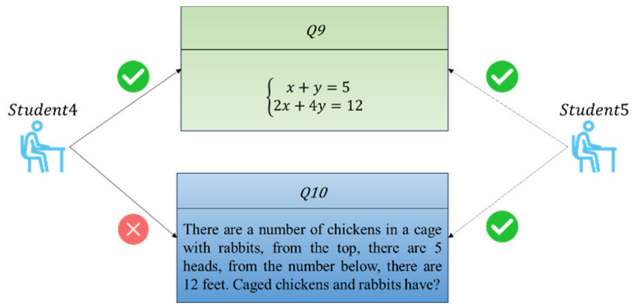

FIGURE 2. **Examples of the impact of problem-solving ability on students' answer results.**

interpretability. Compared to traditional KT models, deep KT models have powerful feature learning and modeling capabilities [6]. However, most deep KT models have the following shortcomings: (1) they do not take into account the fact that different students' knowledge-absorption ability (students' ability to absorb and understand knowledge points) is different, which leads to an inaccurate portrayal of students' knowledge states in the model; (2) they ignore the differences in students' problem-solving ability. The existing model assumes that students have the same problem-solving ability, which to some extent leads to poor model prediction performance.

Existing research suggests that students' knowledge-absorption and problem-solving abilities are key factors that influence their question-answering performance [7]. To provide insight into how individual differences (i.e., knowledge absorption ability and problem-solving ability) affect students' question-answering performance, two examples of the effects are provided. Figure 1 demonstrates the effect of knowledge absorption ability on how quickly students acquire knowledge: in answering questions Q1-Q8 involving knowledge points k1 and k2, Student1 quickly mastered these points with fewer errors, while Student2 and Student3 needed more practice to master them, which indicates that Student1 is significantly better at comprehending and assimilating knowledge than Student2 and Student 3. Figure 2 shows the effect of problem-solving ability on the answer results: when Student4 and Student5 were faced with two identical questions involving the same knowledge points, Student4 only answered the simpler Q9 correctly and answered the more complex Q10 incorrectly;on the contrary, Student5 answered both questions correctly, which shows that Student5 has stronger problem-solving ability than Student4 and is able to fully understand the key information in the questions and use the knowledge and ability he has learned to create effective methods and strategies to solve the problems. Nevertheless, the knowledge-absorption ability and problem-solving ability of different students are not given in advance, which makes it very challenging to measure them.

To solve the above problems, this paper proposes a Dynamic Key-Value Memory Networks Knowledge Tracing with Students' Knowledge-Absorption Ability and Problem-Solving Ability. First, the model uses hierarchical convolutional neural networks [9] to analyze students' mastery of knowledge points in the continuous learning process, so as to dynamically update students' knowledge-absorption ability, aiming at more accurately portraying students' knowledge states; second, the model adopts the auto-encoder (AE) [10] to capture the potential relationship that exists between students' responses to different questions, and dynamically models students' problem-solving ability at each time step, to evaluate students' problem-solving ability. Finally, considering that question features are also important factors affecting students' responses, this paper innovatively combines students' knowledge states, problem-solving ability, and question features, and uses extreme gradient boosting (XGBoost) [11] to predict the probability of students answering the next question correctly.

The primary contributions of this paper are as follows:

- We use the knowledge-absorption ability to characterize the speed of students' knowledge mastery. Considering the differences in students' knowledge-absorption ability, this paper uses a hierarchical convolutional neural network to consider the knowledge mastery degree of students at multiple time steps and dynamically extract students' knowledge-absorption ability, which can better assess students' knowledge state by distinguishing students' knowledge-absorption ability.
- DKVMN-KAPS considers the features of students' problem-solving ability and distinguishes students' problem-solving ability by capturing the potential relationships in the process of answering questions, which can better predict students' performance in answering questions and improve the prediction performance of the model.
- Experiments on three public real-world datasets demonstrate that DKVMN-KAPS outperforms base-line methods. The ablation experiment proves the effectiveness of the model components.

The remainder of this paper is organized as follows. In Section II, research related to this study is systematically reviewed. In Section III, the KT task is defined. In Section IV, the proposed DKVMN-KAPS model is elaborated upon. In Section V, the experimental results of the DKVMN-KAPS with the baseline model are discussed on three real datasets. In Section V-D, this paper is summarized and future work is identified.

II. RELATED WORK

Mainstream KT models are mainly divided into two categories: (1) traditional KT models represented by BKT (Bayesian Knowledge Tracing) [3]; (2) deep learning-based KT models represented by DKT (Deep Knowledge Tracing) [4], DKVMN (Dynamic Key-Value Memory Networks for Knowledge Tracing) [5] as deep learning-based KT models. This section reviews the trends and shortcomings of traditional KT models and deep learning-based KT models.

BKT [3] uses HMM to model students' knowledge states. When a student answers a question, the student's mastery of each knowledge point is updated by the HMM to predict the student's performance in answering the next question. Nevertheless, BKT has some shortcomings, such as the inability to automatically mine the correlations between knowledge points, the need for manual annotation, and the lack of consideration of students' personalized features. To compensate for these shortcomings, researchers have successively improved the BKT. Baker et al. [12] enhanced the predictive performance of the model by introducing the parameters of student blunders and guessing. Yudelson et al. [13] investigated the parameters of students' personalized learning rates and obtained better predictive performance. However, all of these models ignore the effects of students' knowledge-absorption and problem-solving abilities on their performance in answering questions and need to annotate the correspondence between each question and the related know-ledge points in advance, resulting in poor prediction accuracy of the models and the need to spend a lot of labor for annotation.

To better portray the students' knowledge state as well as to avoid the high cost caused by manual labeling, Piech et al. [4] first proposed the Deep Knowledge Tracing Model (DKT), which uses a Long Short-Term Memory neural network (LSTM) to model the student learning process, breaking the limitation of the binary state assumption of the BKT, and avoiding the high cost caused by manual labeling. The DKT uses real-time feedback for modeling the user interaction that controls the transmission state through the gating mechanism of LSTM, remembering information that needs to be memorized for a long time and forgetting unimportant information. On top of DKT, researchers have proposed a number of extended models. Yeung et al. [14] proposed a method of adding three regularization terms to the loss function of the DKT algorithm to enhance the consistency of the algorithm's predictions and improve its prediction accuracy. Zhang et al. [15] added more features at the problem level and used a self-encoder to convert high-dimensional features to low-dimensional features, which further improved the DKT model. Minn et al. [16] proposed an improved DKT model based on the dynamic clustering of students, which improves the accuracy of the model in portraying students' knowledge states by classifying their learning abilities. Li et al. [17] introduced plastic weights into the DKT model, which can be used to continuously update the model parameters after training to adapt to the cognitive characteristics of the students, thus capturing their personal development and individual differences.

To distinguish students' mastery of different knowledge points and improve the model interpretability, Zhang et al. [5] proposed the DKVMN model. DKVMN is a variant of memory-augmented neural networks (MANNs), which is a kind of neural network based on the traditional neural network by adding a storage module and the corresponding read-write mechanism. model. Specifically, the DKVMN uses a key memory matrix to store all potential knowledge point information, which is used to uncover the correlation between questions and potential knowledge points; a value memory matrix is used to store and update students' mastery of each knowledge point, and at each time step, the value memory matrix is updated with the mastery of the relevant knowledge point using the students' real answers. The greater the relevance of a question to a potential knowledge point, the greater the impact of student mastery of that knowledge point on the prediction. By adding a static key memory matrix and a dynamic value memory matrix to the neural network to model the students' knowledge state, DKVMN not only effectively alleviates the long-distance dependency problem, but also improves the problem of poor interpretability of the DKT hidden state and improves the model interpretability [18]. Improvement work on DKVMN in recent years mainly includes three aspects: the improvement of model interpretability, the supplementation of learned behavioral features, and the optimization of the model's long-distance dependency limitation. Yeung et al. [19] inspired by Bayesian deep learning, fused the learning model and Item Response Theory (IRT) to improve the DKVMN model and proposed the Deep-IRT model, and Sun et al. [20] selected the Classification Regression Tree (CART) algorithm as a classification method for students' behavioral features for behavioral features not considered by DKVMN, which improves the accuracy of DKVMN's portrayal of knowledge states. Zou et al. [21] improved the DKVMN model by simulating the learning and memory process of students from the perspective of cognitive psychology and proposed the LPKT model. The above work improves the model interpretability by constructing an interpretability module to provide a basis for feature usage. Sein MINN et al. [22] used learning rate for the first time, based on which stage-specific learning ability was integrated into deep knowledge tracing for dynamic grouping of students, and Sun et al. [23] added the clustering results of students' stage-specific learning ability as features to the model, which enriched the inputs of forgetting, adding mechanism, and final prediction. Xiao et al. [24] introduced features such as question text, knowledge point difficulty, student ability, and duration, and used multiple self-attention mechanisms to combine these features to more accurately model student knowledge states. The above studies express students' knowledge states more comprehensively by integrating students' rich learning behaviors and thus obtain more accurate prediction results. Abdelrahman et al. [25] improved the model's ability to capture long-range dependencies by introducing Hop-LSTM to discover sequential dependencies between problems in a sequence.

Compared to previous work, DKVMN and its improved model can better handle the long-distance dependency problem, show students' mastery of each knowledge point, and automatically discover similar problems. However, existing DKVMN-based optimization efforts still have shortcomings, mainly including: (1) DKVMN and its improved optimization model ignored the influence of students' knowledge-absorption ability on the speed of students' knowledge mastery when supplementing the learning features, resulting in the model's inaccurate assessment of student's knowledge states to the point of low prediction accuracy; (2) Neither the DKVMN nor its improved model takes into account the effect of students' problem-solving ability on their question-answering performance. In fact, problem-solving ability is an important assessment indicator that affects students' performance in answering questions, but due to the lack of assessment of students' problem-solving ability, there is still a lot of room for the model to improve its prediction performance.

III. PROBLEM DEFINITION

The KT model takes student-question interaction sequences as input, analyzes the interaction sequences to predict students' future question-answering performance, and provides targeted learning feedback by analyzing students' knowledge state.

In the KT task, let $S = \{s_1, s_2, \cdots, s_N\}$ and $Q = \{q_1, q_2, \cdots, q_M\}$, where $s \in S$ denotes the students, N denotes the total number of students, and $q \in Q$ denotes the questions, and the students select some of the questions from Q to answer, and the log of the student's responses, $X = \{x_1, x_2, \cdots, x_T\}$, is denoted by the dichotomous group $x_t = \{q_t, r_t\}$ where $q_t \in Q$ denotes the question that the student has done at moment t and $r_t \in \{0, 1\}$. $r_t = 1$ if the student answered the question correctly, and $r_t = 0$ if the student answered the question incorrectly. the output of the model is p_t, which denotes the probability that the student answered the question correctly at moment t. TABLE 1 lists the relevant symbols and their annotations.

IV. THE DKVMN-KAPS MODEL

In this section, Dynamic Key-Value Memory Networks Knowledge Tracing with Students' Knowledge-Absorption Ability and Problem-Solving Ability (DKVMN-KAPS) is proposed, which achieves a more accurate prediction by integrating students' knowledge-absorption ability and problem-solving ability. The model architecture is shown in Figure 3. The DKVMN-KAPS model is mainly divided into correlation weights, knowledge-absorption ability, knowledge states update, problem-solving ability, and student performance prediction. More technical details will be presented in the following subsections.

TABLE 1. Some key notations in the DKVMN-KAPS.

Symbol	Description
S	Student sequence; $\|S\| = N$; s_i denotes the ith student
Q	Question sequence; $\|Q\| = M$; q_j denotes the jth question
X	Student answer sequence
r_t	Students' answers at time t
aw_t	Student responses to questions on different features
M^k	Key memory matrix
M^v	Value memory matrix
N	Number of memory slots in the memory matrix
k_t	Problem embedding vector at moment t
V_t	Knowledge growth after student practice
V_t^{adap}	Students' adaptive knowledge growth
ab_t	Student's problem-solving ability vector at moment t
λ	Features of knowledge-absorptive ability
w_t	Correlation weight vector
e_t	Knowledge erasure vector
a_t	Knowledge growth vector
m_t	Student mastery of current problems
f_t	Student knowledge state vector
r_t	Problem feature vector
ab_t	Probability that the student's answer is correct at time t

A. CORRELATION WEIGHT

To compute the correlation weight vector, we follow the same procedure as in the DKVMN in this paper. The correlation weights are mainly reflected in the attention weights w_t, i.e., the embedding vector k_t is used to query M^k in the model, and the query result is a weighting of the attention level of each knowledge point, which indicates the correlation between the question and each knowledge point.

First, multiply the problem label q_t with the embedding matrix $A \in R^{Q \times d_k}$ to obtain a continuous embedding vector $k_t \in R^{d_k}$, then do the inner product of the embedding vector k_t and each memory slot $M^k(i)$ in the keyed memory matrix, and finally obtain the associated weights $w_t(i) \in R^N$ by the *Softmax* function.

$$w_t(i) = Softmax(k_t^T M^k(i)) \tag{1}$$

where, $\sum_{i=1}^{N} w_t(i) = 1$, the weight w_t indicates the relevance of question q_t to each knowledge point.

B. KNOWLEDGE-ABSORPTION ABILITY

To model students' knowledge states more accurately, this paper uses convolutional neural network (CNN) [26] to extract students' knowledge-absorption ability and update students' knowledge states by knowledge-absorption ability.

CNN was originally invented for computer vision. It operates on a sequence of inputs using a fixed-size sliding window that allows for the extraction of connective links and variations between successive input elements; typically, the first layer is responsible for extracting the base features, while subsequent layers accept these outputs and recognize more complex features [27]. Thus, the multilayer convolutional structure of *CNN* can extract deep features and create a hierarchical representation of the input sequence. In this hierarchical representation, closer input elements interact at lower layers, while more distant elements interact at higher layers [28]. The hierarchical convolutional structure of the *CNN* allows us to consider the students' mastery of the knowledge points at multiple time steps and to learn from it the features of the students' knowledge-absorption ability.

This paper uses one-dimensional convolution, with $W \in R^{2d \times d_v}$ and $b \in R^{d_v}$ as the sliding window's parameters. To prevent convolution operations involving subsequent learning interactions, the second half of the sliding window is masked in this paper. Since value memory matrices store student knowledge point mastery, the sliding window takes d time-step value memory matrices as inputs, uses a *softmax* activation function, and maps them to a single output element. The number of feature maps is set to d_v. A hierarchical convolutional structure is formed by stacking L identical convolutional layers, where lower layers capture knowledge-absorption in the most recent period, and higher layers allow for monitoring farther out. The *GLU* [29] is then used as a nonlinear and a simple gating mechanism is implemented on the outputs of the convolutional layers to filter and weight the features selectively to obtain the output features. Finally, the output features are processed by Singular Value Decomposition (SVD) [30]to obtain the students' knowledge-absorption ability λ. In addition, to address the problem of vanishing gradients and network degradation, this paper adds residual connectivity between the inputs to the outputs of the convolutional layer [31]. The knowledge-absorption ability module is shown in Figure 4:

C. KNOWLEDGE STATE UPDATE

After the student completes the answer to the question, the value memory matrix is updated based on the import tuple (q_t, r_t), the associated weights w_t, and the learning efficiency λ. The knowledge growth vector $v_t \in R^{d_v}$ is obtained by multiplying the student's answer (q_t, r_t) with the embedding matrix $B \in R^{2Q \times d_v}$; and the student's adaptive knowledge growth v_t^{adap} is obtained by connecting the student's current knowledge state f_t to it.

$$v_t^{adap} = [v_t, f_t] \tag{2}$$

There exist studies that show that students who master knowledge faster forget knowledge slower instead [32], here in this paper, we consider that knowledge-absorption ability is inversely proportional to knowledge forgetting. The value memory matrix is erased by the knowledge erasure vector $e_t \in R^{d_v}$. The relevant formula is as follows:

$$e_t = Sigmoid\left(\frac{1}{\lambda} E^T v_t^{adap} + b_e\right) \tag{3}$$

$$\hat{M}_t^v(i) = M(t-1)^v(i)\,[1 - w_i(i) e_t] \tag{4}$$

Students with high knowledge-absorption ability have faster knowledge growth, which is added to the value-memory matrix via the knowledge growth vector $a_t \in R^{d_v}$. The relevant equations are shown in Eq. (5) Eq. (6)

$$a_t = Sigmoid\left(\lambda D^T v_t^{adap} + b_a\right) \tag{5}$$

$$M_t^v(i) = \widehat{M}_t^v(i)\,[1 + w_i(i) a_t] \tag{6}$$

where $\widehat{M}_t^v(i)$ denotes the value memory matrix after erasing memories and $M_t^v(i)$ denotes the value memory matrix after adding memories.

The relevance weight $w_t(i)$ of the key memory matrix input to each knowledge point is weighted and summed with the value of the corresponding slot $M_t^v(i)$ of the value memory matrix to denote the student's mastery of the current knowledge point m_t, which is calculated as follows:

$$m_t = \sum_{i=1}^{N} w_t(i) M_t^v(i) \tag{7}$$

The m_t and the embedding vector k_t are connected and passed through the fully connected layer with *Tanh* activation function, thus generating the feature vector f_t, which represents the current knowledge state of the student. The calculation is shown in Equation (8):

$$f_t = Tanh(w_f [m_t, k_t] + b_f) \tag{8}$$

D. PROBLEM-SOLVING ABILITY

Students' proficiency in answering questions accurately is contingent upon both their level of knowledge and their problem-solving abilities. The primary determinants of students' problem-solving abilities are the questions' inherent features, such as their complexity and differentiation [33]. In this paper, we fit the problem features d_t by passing the embedding vector k_t of $q_t$to a multilayer perceptual machine (MLP).

$$d_t = Tanh[MLP(k_t)] \tag{9}$$

Students' performance in responding to various questions can be more accurately predicted by differentiating their problem-solving abilities. When faced with the same problem, students with strong problem-solving abilities tend to outperform those with weaker problem-solving abilities. Even in the face of difficult problems or in the case of insufficient knowledge, students with strong problem-solving abilities can rely on their ability to achieve good performance. It is important to note that students' problem-solving ability is a dynamic process that continues to change based on their performance during problem-solving. As shown in Fig. 5, to realize the modeling of students' problem-solving ability, this paper adopts AE to process the answer results given by

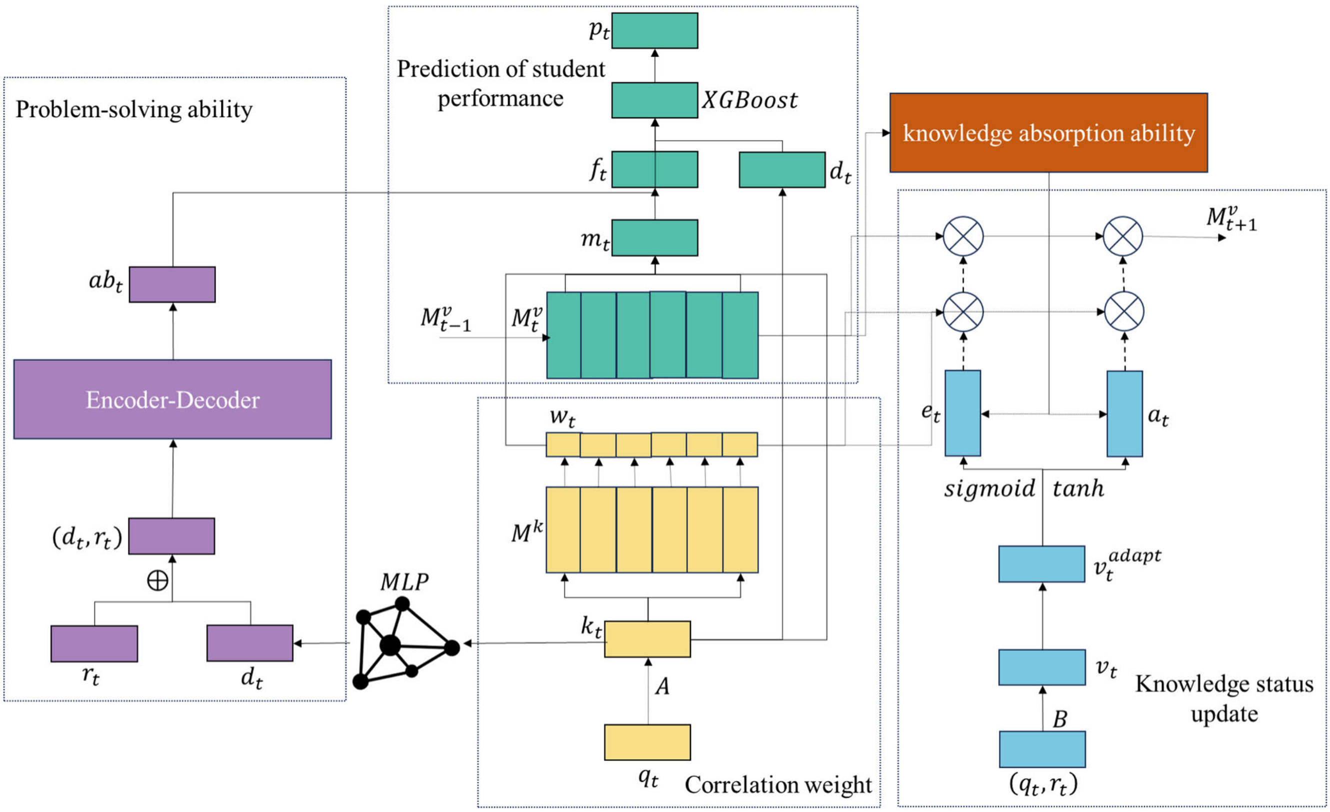

FIGURE 3. Model architecture diagram.

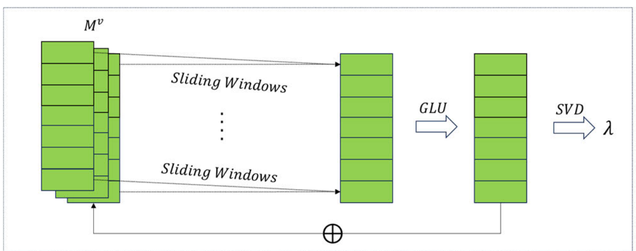

FIGURE 4. Module on knowledge-absorption ability.

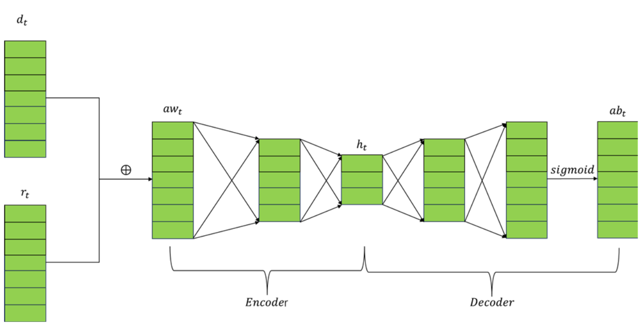

FIGURE 5. Modeling student Problem-Solving ability.

students facing questions containing different problem features to dynamically assess students' problem-solving ability.

In this paper, the question feature vector d_t and the answer vector r_t are spliced together as the student's answers to questions with different features $aw_t{=}(d_t, r_t)$, which are mapped to a lower dimensional representation after a series of hidden layers. This low-dimensional representation captures the potential relationship that exists between how students answer different questions when faced with different problems and their responses, i.e., the potential characterize-ation of a student's problem-solving ability h_t.

$$h_t = g_1(aw_t) \tag{10}$$

where g_1 denotes the encoding process of aw_t, containing two fully connected layers with *ReLu* functions and one fully connected layer without the activation function.

Subsequently, the decoder receives the output of the encoder as input and generates a vector representation of the corresponding student's problem-solving ability ab_t through a series of hidden layers that reconstruct the results as close as possible to the input questions and answers:

$$ab_t = g_2(h_t) \tag{11}$$

where g_2 denotes the decoding process of ab_t, which contains two fully connected layers with the *ReLu* function and one fully connected layer with the *Sigmoid* function.

The process of training the AE is achieved by minimizing the loss function between the input and the reconstruction result. In this paper, we use the mean square loss function to continuously adjust the parameters of the AE to minimize the reconstruction error through the backpropagation algorithm and stochastic gradient descent.

E. PREDICTION OF STUDENT PERFORMANCE

To predict students' future question-answering performance more accurately, this paper adopts *XGBoost* to predict students' question-answering performance. *XGBoost* utilizes the integrated learning idea of gradient boosting tree, which associates students' input features with labels by means of iterative optimization and constructing regression trees and predicts students' question-answering performance by weighted average [34]. Compared with a single learner, *XGBoost*'s integrated approach enables more accurate predictions and produces better results [35]. In addition, *XGBoost* further improves the performance of the model by introducing regularization terms and flexible hyperparameter tuning to prevent overfitting. Therefore, in this paper, the *XGBoost* algorithm is applied to student performance prediction to improve the model prediction accuracy.

In this paper, students' knowledge state f_t, problem solving ability ab_t, and problem feature d_t are taken as *XGBoost* input features R=(f_t, ab_t, d_t), and students' answer results r_t are taken as labels. When *XGBoost* reaches the maximum number of iterations or satisfies the iteration termination condition, the *XGBoost* model $F(SR)$ is output, and the probability p_t of students answering the next question q_t correctly is predicted by the trained *XGBoost* model F(SR).

To optimize the model and update the problem embedding matrix A, the problem response embedding matrix B, the key memory matrix M^k, and the weights and biases of the neural network, this paper chooses the cross-entropy loss function to train the model and uses the Adam optimizer to minimize the objective function [36]. The cross-entropy loss function is shown in Equation (12):

$$L = -\sum\nolimits_t (r_t logp_t + (1 - r_t)\log(1 - p_t)) \tag{12}$$

where $\boldsymbol{p_t}$ is the predicted probability and $\boldsymbol{r_t}$ is the true label.

V. EXPERIMENTS AND ANALYSIS

In this section, several experiments are conducted to evaluate the performance of the proposed DKVMN-KAPS model. Specifically, the dataset is first introduced and the experimental setup, comparison models, and evaluation metrics of this paper are described in detail. Secondly, an extended study is conducted to demonstrate the effectiveness of the DKVMN-KAPS model by answering the following questions:

- RQ1: How does DKVMN-KAPS perform in terms of prediction accuracy compared to state-of-the-art KT models?
- RQ2: How do the key modules of the DKVMN-KAPS affect performance?
- RQ3: How well does the DKVMN-KAPS portray the state of student knowledge?

TABLE 2. Overview of the datasets.

Summary	ASSIST2009	ASSIST2015	STATICS2011
Questions	26,684	NA	1223
Students	4,151	19840	335
Skills	123	100	156
Interactions	325,637	683,801	189,297
Error rates	34.16%	26.82%	23.46%

A. DATASETS

To verify the effectiveness of the proposed DKVMN-KAPS model, this paper conducts many experiments on three public education datasets. For each dataset, this paper is divided into training sets and testing sets according to a certain proportion. The basic information of each dataset is shown in TABLE 2:

ASSIST2009 [37]: This dataset, from the ASSISTments Intelligent Tutoring System, is the most classic and widely used dataset in KT research and contains 325,637 interactions from 4,151 students covering 26,688 questions and 123 knowledge points, with an error rate of 34.16%. The error rate is the percentage of questions answered incorrectly for all interactions contained in the dataset.

ASSIST2015 [38]: This dataset also from the ASSISTments Intelligent Tutoring System, contains 683,801 interactions from 19,840 students covering 100 knowledge points with an error rate of 26.82%. Compared to the ASSIST2009 dataset, which further clarifies the dataset ASSISTment collected in 2009 by collapsing the number of concepts to exactly 100 and introducing a larger number of students, but with a slightly reduced average student interaction record.

STATICS2011 [39]: The dataset is from an engineering mechanics course at a university and contains 189,927 interactions from 333 students covering 1,223 problems and 156 knowledge points with an error rate of 23.46%. Compared to the previous two datasets, this dataset contains the fewest students, but the average number of interactions per student is the highest.

B. EXPERIMENTAL SETUP

Five-fold cross-validation is employed in this paper. Each fold randomly divides the dataset into 80% training data and 20% test data, and the training and test sets do not contain the same students. Next response of a student is predicted by using current and previous response sequence in chronological order. The input problem data is presented to the neural network using "embedding" input vectors.

For the hierarchical convolutional neural network, the convolution kernel size is 6, the number of layers in the hierarchical convolutional layer is set to 3, and the number of residual blocks is also 3. For DKVMN, the learning rate and the number of iterations were set to 0.001 and 100, respectively. The initial values of the key memory matrix and the value memory matrix were learned during training. For

XGBoost, the parameters used are those provided by default in the Python toolbox. To speed up the training process, the batch size is set to 256; to prevent overfitting, the dropout coefficient is adjusted and takes the value of 0.2. In addition, the stochastic gradient descent mechanism is applied to minimize the loss function.

C. BASELINES AND EVALUATION METRICS

To assess the validity of the DKVMN-KAPS model, BKT, DKT, DKVMN, SKVMN, KTMFF, and DKT-LCIRT are selected as baseline models in this paper.

BKT [3]: BKT is a classical KT model utilizing HMM, which assumes a student's knowledge state as a set of binary variables and updates them according to Bayes' rule.

DKT [4]: DKT is the pioneering deep KT model that utilizes LSTM to model student knowledge states.

DKVMN [5]: DKVMN uses MANNs to track the knowledge state of students, using key memory matrix and value memory matrix to store the representation of underlying knowledge points and update the knowledge state.

SKVMN [25]: SKVMN uses an improved LSTM with skip connections in its sequence modeling, enhancing the ability of DKVMN to capture long-term dependency relationships in the problem sequence.

KTMFF [24]: KTMFF introduced features such as problem text, knowledge point difficulty, student ability, and question-answering time in DKVMN. It combines these features using a multi-head self-attention mechanism to comprehensively predict the probability of students answering questions correctly.

DKT-LCIRT [40]: DKT-LCIRT not only models student learning ability but also introduces IRT to improve the interpretability of the model improving the DKVMN model.

In addition, this paper also uses the average AUC metric to evaluate the prediction performance of each model. AUC is the area surrounded by the lower axis and the ROC curve, and the AUC calculation formula contains Equation (13) and Equation (14):

$$TPR = \frac{TP}{TP + FN} \tag{13}$$

$$FPR = \frac{FP}{FP + TN} \tag{14}$$

In Equation (13), TP denotes true cases, i.e., the number of samples that the model correctly predicts as positive cases; TN denotes true negative cases, i.e., the number of samples that the model correctly predicts as negative cases; FP denotes false positive cases, i.e., the number of samples in which the model predicts a negative case as a positive one; and FN denotes false negative cases, i.e., the number of samples in which the model predicts a positive case as a negative one.

In Equation (14), FPR denotes the horizontal coordinate of the ROC curve and TPR denotes the vertical coordinate of the ROC curve. The value of AUC ranges from 0 to 1. AUC=0.5 indicates that the predictive performance of the model is equivalent to random guessing, and the larger value indicates that the predictive performance of the model is better. The predictive performance of the model is positively correlated with the value of AUC.

D. EXPERIMENTAL RESULTS AND ANALYSIS

1) ANALYSIS OF MODEL VALIDITY (RQ1)

To answer RQ1, this paper compares the DKVMN-KAPS model with the BKT, DKT, DKVMN, SKVMN, KTMFF, and DKT-LCIRT models. The best performance model and second-best performance model results are shown in bold and italics, respectively. TABLE 3 shows the comparative results of the average AUC of the seven models tested on the three publicly available datasets.

As shown in TABLE 3, it can first be seen that DKVMN-KAPS significantly outperforms the state-of-the-art DKT-LCIRT model on ASSIST2009, with an AUC improvement of 2.24%. However, the performance improvement of DKVMN-KAPS on STATICS2011 and ASSIST2015 is more limited, with only 0.62% and 0.83% improvement in AUC compared to the next best model. This paper suggests that the reason for this is that the more information about incorrect responses (i.e., the higher the percentage of incorrect responses), the richer the information about students' ability to do the question is captured by the model, and thus the more accurate the predictions are. Second, the traditional model BKT is inferior to all other models in predicting students' question answering, which may be caused by the unreasonable portrayal of students' knowledge state in BKT, while the present model adopts the deep learning method and considers the students' knowledge-absorption ability, which effectively improves the portrayal of the students' knowledge state and achieves the best results in comparison with all baseline models. Finally, DKVMN-KAPS outperforms the other models on all datasets, which is attributed to the fact that the model takes into account students' knowledge-absorption and problem-solving ability, not only realizes the accurate portrayal of students' knowledge states, but also takes into account students' knowledge states, problem-solving ability, and question features to predict whether students can answer the questions correctly, which is a more comprehensive prediction of students' answering performance and improves the prediction accuracy.

Overall, the AUC of DKVMN-KAPS performs better than other models on all datasets, and compared to advanced models such as SKVMN, KTMFF, and DKT-LCIRT, the present model improves the AUC by an average of 2.31%, 4.16%, and 2.18%, and compared to the baseline models such as DKT, DKVMN, and so forth, it improves the AUC by an average of 4.13%, 4.38%. This suggests that updating students' knowledge states by considering their knowledge acquisition at multiple time steps extracting features of students' knowledge-absorption ability from them, and modeling students' problem-solving ability by their responses to different questions are necessary to obtain better performance. So far, RQ1 has been answered in detail.

TABLE 3. Model performance comparison.

Model	ASSIST2009	ASSIST2015	STATICS2011
	AUC	AUC	AUC
BKT	0.6311	0.6423	0.7301
DKT	0.8219	0.7303	0.8265
DKVMN	0.8157	0.7268	0.8284
SKVMN	0.8363	0.7484	*0.8485*
KTMFF	0.8176	0.7305	0.8297
DKT-LCIRT	*0.8527*	*0.7645*	0.8198
DKVMN-KAPS	**0.8751**	**0.7728**	**0.8547**

2) ABLATION EXPERIMENTS(RQ2)

To answer RQ2, this paper conducts an ablation study on all three real datasets to further validate the effectiveness of each component in the model proposed in this paper. Three comparison settings are set up in this paper, details of which are given below:

DKVMN-X: Predicting student answer results via *XGBoost* considering only student knowledge state and question features.

DKVMN-PS: DKVMN-KAPS without considering students' knowledge-absorption ability.

DKVMN-KA: DKVMN-KAPS without considering students' problem-solving abilities.

To be fair, the previously mentioned variants of the model and the rest of the experimental setup remained consistent except for the changes mentioned above. TABLE 4 and Figure 6 show the comparative results of the average AUC of the three models tested on the three publicly available datasets.

Firstly, it can be observed that the knowledge-absorption ability contributes the least to the performance in the whole model, but it is not conducive to accurately portraying the students' knowledge states if the students' knowledge-absorption ability is ignored. Second, it can be observed that the student problem-solving ability module plays a pivotal role in the model predictions, and the largest decrease in the predicted results is observed if the student problem-solving ability is not considered, which implies that modeling student problem-solving ability can better predict the results of student responses. In addition, the relatively low enhancement of problem-solving ability on the model's predictive performance on the ASSIST2015 dataset may be because the average number of student interactions with problems in this dataset is low, resulting in the model not learning enough about students' problem-solving ability. Overall, the prediction accuracy of DKVMN-PS with the addition of student problem-solving ability and DKVMN-KA with the addition of knowledge absorption ability were both improved compared to DKVMN-X with the addition of question features only. By adding the student problem-solving ability module, the model prediction accuracy improved by 2.21% on average compared to the DKVMN-X model. The DKVMN-KA model with knowledge absorption ability improved the prediction accuracy by 0.34% on average compared to the DKVMN-X due to the more accurate portrayal of the student's knowledge state. The DKVMN-KAPS improves the prediction accuracy by an average of 2.46% compared to DKVMN-X with the combined effect of students' problem-solving ability and knowledge absorption ability, which indicates that DKVMN-KAPS, which considers both students' knowledge absorption ability and problem-solving ability, outperforms any individual model. Therefore, it is necessary and effective to consider students' knowledge absorption ability and problem-solving ability together. So far, this paper completes the answer to RQ2.

TABLE 4. Comparative results of ablation experiments.

Model	ASSIST2009	ASSIST2015	Statics2011
	AUC	AUC	AUC
DKVMN-X	0.8325	0.7611	0.8353
DKVMN-PS	0.8738	0.7695	0.852
DKVMN-KA	0.8369	0.764	0.8361
DKVMN-KAPS	0.8751	0.7728	0.8547

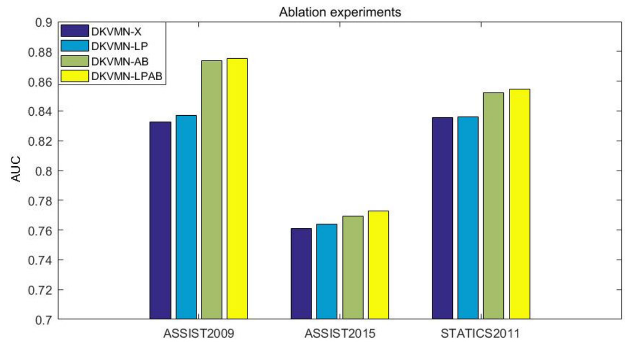

FIGURE 6. The ablation experiment result in figure.

3) EVOLUTION OF KNOWLEDGE STATES (RQ3)

To demonstrate more intuitively the accuracy of this paper's portrayal of students' knowledge states, this paper visualizes students' knowledge states on the Assistments09 dataset. By extracting the relevant weight vector for each question, the association between questions and knowledge points can be obtained. According to the knowledge state extraction method introduced in subsection IV-C, the knowledge state of students after completing each specific question can be output to get the knowledge state of students after completing each specific question, and the knowledge state of students can be visualized by plotting the knowledge state of students in the order of doing the questions into a heat map. The knowledge state visualization can not only be used for personalized instruction of students but also can verify the actual effect of the KT model to a certain extent.

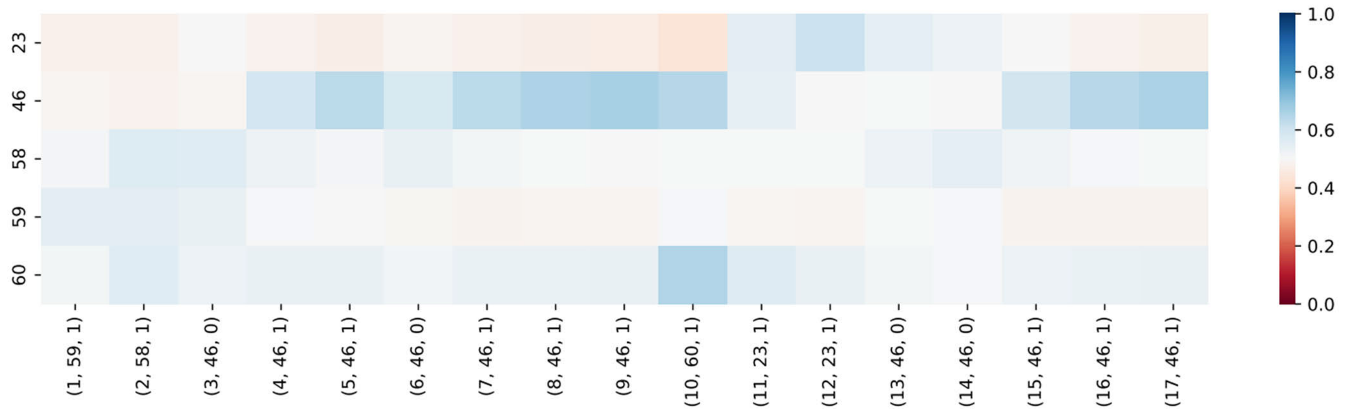

FIGURE 7. Knowledge state evolution diagram.

To track changes in a student's mastery of specific knowledge points as they learned, we randomly intercepted the student's interaction record from the Assistments2009 dataset. The horizontal coordinates in Figure 7 are the student's interaction sequence (t, kq, r), which represent the moments, the knowledge points corresponding to question q, and the student's answers, respectively, and the vertical coordinates represent five different knowledge points. The results of the experiment showed that after a student answered a question involving knowledge point 46 and answered it correctly at moments 4 and 7, the student's mastery level of the knowledge point increased at the next moment. When the student answered the question about Knowledge Point 46 incorrectly at moment 6, the student's mastery level of the knowledge point decreased at the next moment. After the student correctly answered the questions involving Knowledge Point 46 at moments 16, and 17, the student's mastery level of Knowledge Point 46 hardly changed because the student practiced Knowledge Point 46 several times, and therefore, the student should be given more time to learn other knowledge points. This shows that DKVMN-KAPS can better capture the changes in the knowledge level of the corresponding knowledge points from the student's responses. Moreover, teachers can give students practice suggestions according to their knowledge states. So far, this paper has provided a sufficient answer to RQ3.

VI. CONCLUSION

In this paper, we propose a Dynamic Key-Value Memory Networks Knowledge Tracing with Students' Knowledge-Absorption Ability and Problem-Solving Ability (DKVMN-KAPS), which considers the influence of students' knowledge-absorption ability and problem-solving ability on students' performance and obtains better prediction results than the existing models. In addition to focusing on the state of knowledge, DKVMN-KAPS introduces problem-solving ability as a supplement, and more accurately portrays students' knowledge state by considering the knowledge-absorption ability of students. Finally, we consider students' knowledge states, problem-solving ability, and question features to predict students' performance in answering questions. We have conducted extensive experiments on three public datasets, and the results show that DKVMN-KAPS can track the changes in students' knowledge states more closely to the real situation and has better performance in predicting the results.

In the future research direction, we can use more appropriate methods to construct features according to the textual features of the problem (e.g., teaching terminology, subject-specific terms, etc.) and use pedagogical theories to model students' knowledge absorption and problem-solving abilities and to better simulate the changes in students' knowledge states during the learning process.

ACKNOWLEDGMENT

This paper was partially supported by National Natural Science Foundation of China (NO. 62377024).

REFERENCES

[1] F. Ke, W. Wang, W. Tan, L. Du, Y. Jin, Y. Huang, and H. Yin, "HiTSKT: A hierarchical transformer model for session-aware knowledge tracing," *Knowl.-Based Syst.*, vol. 284, Jan. 2024, Art. no. 111300.

[2] T. Huang, S. Hu, H. Yang, J. Geng, Z. Li, Z. Xu, and X. Ou, "Response speed enhanced fine-grained knowledge tracing: A multi-task learning perspective," *Expert Syst. Appl.*, vol. 238, Mar. 2024, Art. no. 122107.

[3] A. T. Corbett and J. R. Anderson, "Knowledge tracking: Modeling the acquisition of procedural knowledge," *User Model. User-Adapted. Interact.*, vol. 4, no. 4, pp. 253–278, 1995.

[4] C. Piech, J. Spencer, J. Huang, S. Ganguli, M. Sahami, L. Guibas, and J. Sohl-Dickstein, "Deep knowledge tracking," in *Proc. 28th Int. Conf. Neural Inf. Proccess. Syst. (NIPS)*. Cambridge, MA, USA: MIT Press, 2015, pp. 505–513.

[5] J. Zhang, X. Shi, I. King, and D.-Y. Yeung, "Dynamic key-value memory networks for knowledge tracing," in *Proc. 26th Int. Conf. World Wide Web*. Perth, WA, Australia: International World Wide Web Conferences Steering Committee, Apr. 2017, pp. 765–774.

[6] Y. Zhao, H. Ma, W. Wang, W. Gao, F. Yang, and X. He, "Exploiting multiple question factors for knowledge tracing," *Expert Syst. Appl.*, vol. 223, Aug. 2023, Art. no. 119786.

[7] S. Liu, J. Yu, Q. Li, R. Liang, Y. Zhang, X. Shen, and J. Sun, "Ability boosted knowledge tracing," *Inf. Sci.*, vol. 596, pp. 567–587, Jun. 2022.

[8] S. Shen, Q. Liu, E. Chen, H. Wu, Z. Huang, W. Zhao, Y. Su, H. Ma, and S. Wang, "Convolutional knowledge tracing: Modeling individualization in student learning process," in *Proc. 43rd Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, Jul. 2020, pp. 1857–1860.

[9] M. Pavlovski, M. Alqudah, T. Dokic, A. A. Hai, M. Kezunovic, and Z. Obradovic, "Hierarchical convolutional neural networks for event classification on PMU measurements," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–13, 2021.

[10] D. P. Kingma and M. Welling, "An introduction to variational autoencoders," *Found. Trends Mach. Learn.*, vol. 12, no. 4, pp. 307–392, 2019.

[11] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*. San Francisco, CA, USA: ACM, Aug. 2016, pp. 785–794.

[12] R. S. J. D. Baker, A. T. Corbett, and V. Aleven, "More accurate student modeling through contextual estimation of slip and guess probabilities in Bayesian knowledge tracing," in *Intelligent Tutoring Systems*, vol. 5091. Berlin, Germany: Springer, 2008, pp. 406–415.

[13] M. V. Yudelson, K. R. Koedinger, and G. J. Gordon, "Individualized Bayesian knowledge tracing models," in *Artificial Intelligence in Education*, vol. 7926. Berlin, Germany: Springer, 2013, pp. 171–180.

[14] C.-K. Yeung and D.-Y. Yeung, "Addressing two problems in deep knowledge tracing via prediction-consistent regularization," in *Proc. 5th Annu. ACM Conf. Learn. Scale*. London, U.K.: ACM, Jun. 2018, pp. 1–10.

[15] L. Zhang, X. Xiong, S. Zhao, A. Botelho, and N. T. Heffernan, "Incorporating rich features into deep knowledge tracing," in *Proc. 4th ACM Conf. Learn. Scale*. Cambridge, MA, USA: ACM, Apr. 2017, pp. 169–172.

[16] S. Minn, Y. Yu, M. C. Desmarais, F. Zhu, and J.-J. Vie, "Deep knowledge tracing and dynamic student classification for knowledge tracing," in *Proc. IEEE Int. Conf. Data Mining (ICDM)*. Singapore: IEEE, Nov. 2018, pp. 1182–1187.

[17] Z. Li, S. Yu, Y. Lu, and P. Chen, "Plastic gating network: Adapting to personal development and individual differences in knowledge tracing," *Inf. Sci.*, vol. 624, pp. 761–776, May 2023.

[18] H. Liu, T. Zhang, F. Li, Y. Gu, and G. Yu, "Tracking knowledge structures and proficiencies of students with learning transfer," *IEEE Access*, vol. 9, pp. 55413–55421, 2021.

[19] C.-K. Yeung, "Deep-IRT: Make deep learning based knowledge tracing explainable using item response theory," 2019, *arXiv:1904.11738*.

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

[21] Y. Zou, X. Yan, and W. Li, "Knowledge tracking model based on learning process," *J. Comput. Commun.*, vol. 8, no. 10, pp. 7–17, 2020.

[22] S. Minn, M. C. Desmarais, F. Zhu, J. Xiao, and J. Wang, "Dynamic student classiffication on memory networks for knowledge tracing," in *Advances in Knowledge Discovery and Data Mining*, vol. 11440. Cham, Switzerland: Springer, 2019, pp. 163–174.

[23] X. Sun, X. Zhao, B. Li, Y. Ma, R. Sutcliffe, and J. Feng, "Dynamic key-value memory networks with rich features for knowledge tracing," *IEEE Trans. Cybern.*, vol. 52, no. 8, pp. 8239–8245, Aug. 2022.

[24] Y. Xiao, R. Xiao, N. Huang, Y. Hu, H. Li, and B. Sun, "Knowledge tracing based on multi-feature fusion," *Neural Comput. Appl.*, vol. 35, no. 2, pp. 1819–1833, Jan. 2023.

[25] G. Abdelrahman and Q. Wang, "Knowledge tracing with sequential key-value memory networks," in *Proc. 42nd Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, Jul. 2019, pp. 175–184.

[26] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[27] A. Bello, S.-C. Ng, and M.-F. Leung, "A BERT framework to sentiment analysis of tweets," *Sensors*, vol. 23, no. 1, p. 506, Jan. 2023.

[28] H. Pei, B. Wei, K. C.-C. Chang, Y. Lei, and B. Yang, "Geom-GCN: Geometric graph convolutional networks," 2020, *arXiv:2002.05287*.

[29] Y. N. Dauphin, A. Fan, M. Auli, and D. Grangier, "Language modeling with gated convolutional networks," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 933–941.

[30] Y. Huo, D. F. Wong, L. M. Ni, L. S. Chao, and J. Zhang, "HeTROPY: Explainable learning diagnostics via heterogeneous maximum-entropy and multi-spatial knowledge representation," *Knowl.-Based Syst.*, vol. 207, Nov. 2020, Art. no. 106389.

[31] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*. Las Vegas, NV, USA: IEEE, Jun. 2016, pp. 770–778.

[32] F. I. M. Craik and R. S. Lockhart, "Levels of processing: A framework for memory research," *J. Verbal Learn. Verbal Behav.*, vol. 11, no. 6, pp. 671–684, Dec. 1972.

[33] J. Sun, M. Wei, J. Feng, F. Yu, Q. Li, and R. Zou, "Progressive knowledge tracing: Modeling learning process from abstract to concrete," *Expert Syst. Appl.*, vol. 238, Mar. 2024, Art. no. 122280.

[34] W. Su, F. Jiang, C. Shi, D. Wu, L. Liu, S. Li, Y. Yuan, and J. Shi, "An XGBoost-based knowledge tracing model," *Int. J. Comput. Intell. Syst.*, vol. 16, no. 1, p. 13, Feb. 2023.

[35] B. Chakravarthi, S.-C. Ng, M. R. Ezilarasan, and M.-F. Leung, "EEG-based emotion recognition using hybrid CNN and LSTM classification," *Frontiers Comput. Neurosci.*, vol. 16, Oct. 2022, Art. no. 1019776.

[36] W. Zhang, S. Hu, and K. Qu, "Graph attention neural network model with behavior features for knowledge tracking," *IEEE Access*, vol. 11, pp. 88329–88338, 2023.

[37] M. Feng, N. Heffernan, and K. Koedinger, "Addressing the assessment challenge with an online system that tutors as it assesses," *User Model. User-Adapted Interact.*, vol. 19, no. 3, pp. 243–266, Aug. 2009.

[38] X. Xiong, S. Zhao, V. Inwegen, and J. E. Beck, "Going deeper with deep knowledge tracing," in *Proc. Int. Educ. Data Mining Soc. Int. Educ. Data Mining Soc.*, Worcester, MA, USA, 2016, pp. 545–550.

[39] K. R. Koedinger, R. S. Baker, K. Cunningham, A. Skogsholm, B. Leber, and J. Stamper, "A data repository for the EDM community: The PSLC datashop," in *Handbook of Educational Data Mining*, vol. 43. Boca Raton, FL, USA: CRC Press, 2010, pp. 43–56.

[40] G. Li, J. Shuai, Y. Hu, Y. Zhang, Y. Wang, T. Yang, and N. Xiong, "DKT-LCIRT: A deep knowledge tracking model integrating learning capability and item response theory," *Electronics*, vol. 11, no. 20, p. 3364, Oct. 2022.

WEI ZHANG received the Ph.D. degree from Huazhong University of Science and Technology. He is currently a Professor with the Faculty of Artificial Intelligence in Education, Central China Normal University. He has published more than 60 articles in the academic journals, including 55 papers indexed by SCI, SSCI, EI, CPCI, and CSSCI. His research interests include computer applications, big data analysis, data mining, and the application of information technology in education.

ZHONGWEI GONG is currently pursuing the master's degree in software engineering with Central China Normal University. His research interest includes knowledge tracing.

PEIHUA LUO is currently pursuing the master's degree with the Faculty of Artificial Intelligence in Education, Central China Normal University. Her research interests include data mining, deep learning, and knowledge tracing.

ZHIXIN LI is currently pursuing the master's degree in software engineering with the Faculty of Artificial Intelligence in Education, Central China Normal University. His main research interests include deep learning and knowledge tracing.