• RESEARCH PAPER •

Personalized Forgetting Mechanism with Concept-Driven Knowledge Tracing

Shanshan WANG[1], Ying HU[2], Xun YANG[3], Zhongzhou ZHANG[4],
Keyang WANG[5] & Xingyi ZHANG[6]

[1,2]*Information Materials and Intelligent Sensing Laboratory of Anhui Province, Anhui University, Anhui 230601, China;*
[3]*School of Data Science, University of Science and Technology of China, Anhui 230026, China;*
[4] *College of Computer Science, Sichuan University, Chengdu 610065, China;*
[5]*zhejiang Dahua Technology Co., Ltd. Zhejiang 310053, China;*
[6]*School of Computer Science and Technology, Anhui University, Anhui 230601, China*

Abstract   Knowledge Tracing (KT) aims to trace changes in students' knowledge states throughout their entire learning process by analyzing their historical learning data and predicting their future learning performance. Existing forgetting curve theory based knowledge tracing models only consider the general forgetting caused by time intervals, ignoring the individualization of students and the causal relationship of the forgetting process. To address these problems, we propose a **C**oncept-driven **P**ersonalized **F**orgetting knowledge tracing model (CPF) which integrates hierarchical relationships between knowledge concepts and incorporates students' personalized cognitive abilities. First, we integrate the students' personalized capabilities into both the learning and forgetting processes to explicitly distinguish students' individual learning gains and forgetting rates according to their cognitive abilities. Second, we take into account the hierarchical relationships between knowledge points and design a precursor-successor knowledge concept matrix to simulate the causal relationship in the forgetting process, while also integrating the potential impact of forgetting prior knowledge points on subsequent ones. The proposed personalized forgetting mechanism can not only be applied to the learning of specifc knowledge concepts but also the life-long learning process. Extensive experimental results on three public datasets show that our CPF outperforms current forgetting curve theory based methods in predicting student performance, demonstrating CPF can better simulate changes in students' knowledge status through the personalized forgetting mechanism.

Keywords   knowledge tracing, concept-driven matrix , forgetting mechanism, ability personalization, cognitive modeling

Citation   Shan S W, Ying H, Xun Y et al. Personalized Forgetting Mechanism with Concept-Driven Knowledge Tracing.

1   Introduction

With the rapid development of online education, it has become increasingly important to accurately assess students' knowledge status. Knowledge Tracing (KT) plays a vital role in online learning, which is the task of modelling student knowledge over time so that we can accurately predict the student's performance on future interactions. Knowledge tracing [36, 39] models not only help educators better understand students' academic abilities [34] but also provide targeted educational resources and personalized learning strategies. By analyzing students' historical learning data, *i.e.*, learning behaviors and performance [8], they can accurately predict their performance and knowledge level on future learning tasks [5]. Considering the high cost and importance of data collection in online education, the use of knowledge tracing models is expected to grow in prevalence and importance within the field of education.

Existing knowledge tracing (KT) models [16, 22] have achieved tremendous success in predicting students' performance. For example, Deep Knowledge Tracing (DKT) [1, 4] utilizes Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM) networks [16, 19, 20] to predict correct answers.

During the practice process, once a student answers incorrectly, DKT argues that the student's corresponding knowledge status will decline. Dynamic Key-Value Memory Network (DKVMN) [35] introduces a dynamic key-value memory matrix based on DKT to capture changes in students' knowledge state more efficiently. Context-aware Knowledge Tracing (AKT) [17] is an attention-based model that connects a learner's future responses to assess questions with their past responses. It calculates attention weights using exponential decay and context-aware relative distance metrics, alongside question similarity. Learning Process-consistent Knowledge Tracing (LPKT) [6] monitors students' knowledge status by directly modeling the student's learning process, calculates learning gain through the difference between learning units, and simulates the forgetting process through time intervals [46]. Considering that progress rates are student-specific, extending LPKT to LPKT-S [26] differentiates individual progress rates for each student. Although the aforementioned methods have successfully modeled the forgetting process, there is still a significant limitation that they fail to consider the personalized cognitive processes of students and the causal relationships of the forgetting process. As a result, this limitation weakens the interpretability [44] of these methods, making it difficult for educators to gain deep insights and limits a comprehensive understanding of students' learning and forgetting processes. Therefore, there is a need to meticulously and comprehensively consider the cognitive structure and knowledge structure of students in the KT task. As shown in Fig. 1, the different cognitive structures of students will not only lead to different learning outcomes, but also different forgetting rates. Generally, students with stronger learning abilities can acquire knowledge more effectively in the learning process and have lower forgetting rates in the forgetting process. Conversely, students with weaker learning abilities tend to have poorer grasp of knowledge and also exhibit higher rates of forgetting. Additionally, the different knowledge structures can affect the students' mastery of the knowledge because there is a hierarchical relationship between knowledge concepts. Specifically, if one knowledge point is forgotten by the student, the related knowledge points may also be forgotten to some extent.
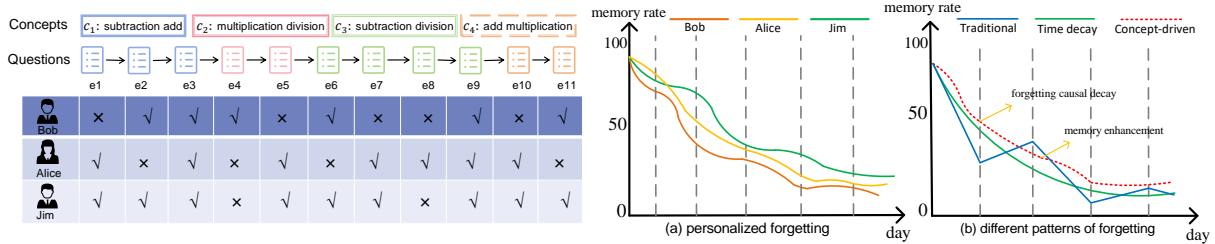

Figure 1 Examples of three learners answer a series of exercises on an online learning system. (a) indicates that the forgetting rate of various students is different. (b) represents different forgetting patterns: the traditional model considers correct answers as mastering the knowledge point, while incorrect answers indicate no impression of the knowledge point; the time decay model believes that memory of knowledge points weakens over time; and the causal forgetting model integrates the intrinsic relationship between time and knowledge points, balancing long-term and short-term memory.

To better understand the causal relationship between personalized learning and the forgetting process of students, in this paper, we propose a Concept-driven Personalized Forgetting knowledge tracing (CPF) model. Firstly, to accommodate the individual differences among students, we incorporate students' prior abilities into the modeling of both the learning and forgetting processes. Guided by cognitive theory [10] and learning efficacy theory [12], we measure students' abilities [3] through answer time, answer accuracy and question difficulty in this study. Secondly, to explore the causal relationship between knowledge concepts in the forgetting process, we design a novel causal forgetting mechanism that explicitly quantify the hierarchy of different knowledge concepts. Specifically, inspired by educational theories [28, 48], we construct a precursor-successor matrix (P-matrix) to capture the directed relations between different knowledge concepts and calculate causal forgetting weights by searching for the closest prerequisite knowledge concepts to the current exercise according to this P-matrix. Subsequently, to simulate the forgetting-review mechanism presented by students when facing the exercise with related knowledge concepts, we compute the similarity of adjacent knowledge states and adopt this similarity to update the forgetting gate, thereby modeling the forgetting process more accurately [21, 40] in the real-world scenarios. We conduct on three public datasets used in KT task. The state-of-the-art performance demonstrates the effectiveness of our CPF.

The main contributions of our paper are as follows:

- We incorporate students' personalized cognitive abilities into both the learning and forgetting pro-

cesses in the knowledge tracing, which can explicitly distinguish students' individual learning gains and forgetting rates according to their cognitive abilities.

- We design a precursor-successor knowledge concept matrix and introduce a forgetting-review mechanism, aiming to comprehensively capture the causal relationships between knowledge concepts during the forgetting process of knowledge tracing.

- By integrating the personalized cognitive processes and the causal relationships of concepts in the forgetting process, a novel Concept-driven Personalized Forgetting knowledge tracing (CPF) model is materialized in this paper. Extensive experimental results on three public datasets demonstrate the effectiveness of our CFP.

2   Related Works

2.1   Knowledge Tracing

Knowledge tracing (KT) [7] is an important task in online learning systems, aiming to dynamically track the learner's knowledge status. Existing methods can be divided into two categories: (1) Traditional methods, such as Bayesian Knowledge Tracing (BKT) [18, 25, 30], which use binary variables to describe students' knowledge status. The Item Response Theory model (IRT) [2, 43] uses student abilities and problem characteristics to analyze student performance. Traditional KT models ignore the forgetting process. (2) Based on deep learning methods, Deep Knowledge Tracing (DKT) introduces recurrent neural networks such as RNN and LSTM for the first time. Self-Attentive Knowledge Tracing (SAKT) [59] adds an attention mechanism to the KT model for the first time [29, 31]. Context-aware attentive knowledge tracing (AKT) [17] model utilizes contextualized representations of practice and knowledge acquisition and combines attentional mechanisms with cognitive and psychometric models. Reconciling Cognitive Modeling with Knowledge Forgetting: A Continuous Time-aware Neural Network Approach (CT-NCM) [9] realistically integrates the dynamics and continuity of knowledge forgetting into the modeling of student learning processes. Learning Process-consistent Knowledge Tracing (LPKT) aims to assess the student's knowledge status by modeling the student's learning process, and considers the impact of the answer time interval and answer time on the learner's knowledge status during the learning process. Monitoring Student Progress for Learning Process-consistent Knowledge Tracing (LPKT-S) is an extension of LPKT that clearly distinguishes the individual progress rate of each student. However, the above models ignore the impact of individual differences among students on the forgetting process, and lacks exploration of the causal forgetting caused by the correlation and hierarchy between knowledge concepts.

2.2   Forgetting Curve

In pedagogy, forgetting is a complex and universal phenomenon [45], and with time, students' knowledge proficiency may decline due to forgetting factors [11]. The Ebbinghaus forgetting curve proposes [50,56,57] that students forget most quickly shortly after learning new knowledge or skills, and then the rate of forgetting gradually slows down. This also shows that when students' knowledge state tends to be stable, the proportion of forgotten knowledge will decrease [13], but this is not only due to the influence of time factor but also related to knowledge structure. Trace decay theory emphasizes [37, 49]that the knowledge students learn will gradually decay with time, and emphasizes the need for regular review and consolidation in the learning process. By actively reviewing and consolidating the learned knowledge, students can significantly delay the rate of forgetting, so that they can remember knowledge for a long time. This shows that, if there is a certain correlation between knowledge, then forgetting some knowledge will also lead to the forgetting of its related knowledge. However, based on this theory, through continuous review and consolidation, the knowledge will gradually transform into a stable knowledge state based on knowledge correlation when it reappears after forgetting.

3   Preliminaries

3.1   Relation Definition

Let E and C be the set of all different exercises and knowledge concepts respectively, where E = $\{e_1, e_2, \ldots, e_N\}$ and C = $\{c_1, c_2, \ldots, c_K\}$. Then problem sets and knowledge concept relation matrix can
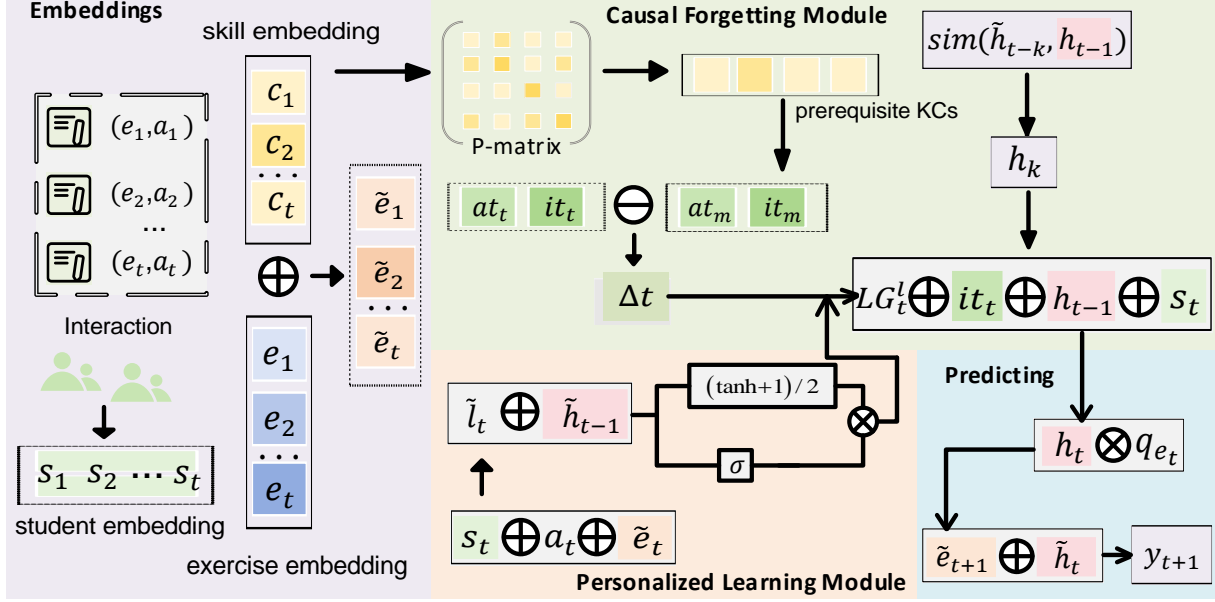
Figure 2 The main network structure of CPF model is mainly composed of learning, forgetting and prediction modules. The learning module uses the learning gate to get the personalized learning gain, uses the directed relationship of knowledge concepts to capture the causality of forgetting process, and finally predicts the future performance of students.

be expressed as $Q_{ij} = \{q_{e_i c_j}\}_{N \times K}$. If the exercise e_i contains the knowledge concept c_j, $q_{e_i c_j}$ is set to 1; otherwise, $q_{e_i c_j} = 0$. In addition, inspired by cognitive diagnosis and based on the directed and undirected relations between knowledge concepts, we construct a knowledge concept prerequisite relation matrix similar to Q-matrix [47], that is P-matrix, representing the directed relations between different knowledge concepts. This can be expressed as $P_{ij} = \{p_{c_i c_j}\}_{K \times K}$. If there is a leading relation between the knowledge concepts c_i and c_j, then $p_{c_i c_j} = 1$; otherwise, $p_{c_i c_j} = 0$.

3.2 Problem Statement

For students, the learning process for each time step is expressed as an interaction process, which is represented by (e_t, a_t), where e_t represents the index of the exercise undertaken at time step t, and $a_t = 1$ indicates a correct response, while $a_t = 0$ indicates an incorrect response. Knowledge tracing (KT) aims to forecast a student's response to a given exercise based on their interaction history. In a formal sense, given the interaction record in the previous T time steps denoted by R=(e_1, a_1),..., (e_t, a_t),..., (e_T, a_T), the KT model seeks to assess the mastery level of the knowledge concept and predict the student's response to the exercise at time step T+1 (e_{T+1}).

4 Methodology

In this section, we introduce the proposed Concept-driven Personalized Forgetting knowledge tracing method (CPF) in detail. First, we briefly review the baseline,*i.e.*, LPKT, of our CPF. Then, we present the framework of our CPF model in detail. As shown in Fig. 2, CPF utilizes personalized knowledge acquisition and forgetting mechanisms to effectively update students' knowledge status. Simultaneously, by modeling the hierarchical relationship between knowledge concepts, it reveals the causal relationship of the forgetting process.

4.1 Overview of LPKT

LPKT consists of three modules at each learning step, including learning module, forgetting module and prediction module. Specifically, the learning phase mainly calculates the positive impact of learning gain by calculating the difference between the current interaction and an adjacent interaction, and models the student's knowledge status. In response to the negative impact of forgetting, the forgetting stage uses a forgetting gate to determine the degree of decay of the knowledge state over time. Then the knowledge

mastery level in the learning stage and forgetting stage is used to update the knowledge status, and finally the proposed prediction module is used to predict the student's performance in the next exercise.

In the learning stage, LPKT can model the learning gains by connecting the students' previous learning embedding l_{t-1} and the current learning embedding l_t as the basic input element of LPKT. The current learning embedding l_t consists of exercise e_t, answer a_t, and answer time at_t. LPKT considers two factors impacting learning outcomes: the interval time and the student's prior knowledge. To focus on the state of the knowledge concept related to the current exercise, LPKT multiplies the knowledge state h_{t-1} with the knowledge concept vector q_{e_t} of the current exercise to get the related knowledge state $\widetilde{h}_{t-1}$.

$$l_t = \mathbf{W}_1^T[e_t \oplus a_t \oplus at_t] + b_1, \tag{1}$$

$$\tilde{h}_{t-1} = q_{e_t} \cdot h_{t-1}, \tag{2}$$

Considering that not all learning gains directly contribute to the increasing of students' knowledge, the learning gate is designed to control students' ability to absorb knowledge and ensure a positive learning gain through linear transform.

$$lg_t = \tanh(\mathbf{W}_2^T[l_{t-1} \oplus it_t \oplus l_t \oplus \widetilde{h}_{t-1}] + b_2), \tag{3}$$

$$\Gamma_t^l = \sigma(\mathbf{W}_3^T[l_{t-1} \oplus it_t \oplus l_t \oplus \widetilde{h}_{t-1}] + b_3), \tag{4}$$

$$LG_t = \Gamma_t^l \cdot ((lg_t + 1)/2), \tag{5}$$

According to the forgetting curve theory, the memory of learned material declines exponentially over time. LPKT employs a forgetting gate to simulate the forgetting effect, primarily determined by the students' prior knowledge, learning intervals, and learning gains. Then, the output of the forgetting gate Γ_t^f is multiplied by h_{t-1} to eliminate the influence of forgetting, which can obtain the required updated knowledge state h_t of students following the t-th learning interaction.

$$\Gamma_t^f = \sigma(\mathbf{W}_4^T[h_{t-1} \oplus LG_t \oplus it_t] + b_4), \tag{6}$$

$$h_t = \widetilde{LG_t} + \Gamma_t^f \cdot h_{t-1}. \tag{7}$$

4.2 Proposed Method

4.2.1 Personalized Learning Module

In LPKT, there is no clear distinction between students' individual abilities, but the learning process of each student is different. LPKT only distinguishes different students' responses to the same exercises through the answer time, ignoring the modeling of students' abilities. In order to address this limitation, we explicitly introduce a personalized student ability s_t in our methods. Specifically, as show in Eq.8, we derive this personalized student ability $s_t \in \mathbb{R}^{d_k}$ by combining the problem difficulty $df_t \in \mathbb{R}^{d_k}$, answer time $at_t \in \mathbb{R}^{d_k}$, correct accuracy $ac_t \in \mathbb{R}^{d_k}$. Simultaneously, we combine the exercise embedding e_t with knowledge concepts to get the concept-perceived exercise embedding $\tilde{e}_t$. Then we concatenate the $\tilde{e}_t$, a_t, s_t together and apply a perceptron to extract the personalized learning emdedding $\tilde{l}_t$.

$$s_t = \alpha \otimes df_t + \beta \otimes ac_t + \mu \otimes at_t, \tag{8}$$

$$\tilde{e}_t = e_t \oplus c_t, \tag{9}$$

$$\widetilde{l}_t = \mathbf{W}_1^T[\tilde{e}_t \oplus a_t \oplus s_t] + b_1, \tag{10}$$

where $\mathbf{W}_1 \in \mathbb{R}^{(d_e+d_k+d_a) \times d_k}$ and $b_1 \in \mathbb{R}^{d_k}$ denote the weight matrix and its corresponding bias term. α, β, μ are hyper-parameters. $\otimes$ and $\oplus$ denote the element-wise multiplication and concatenation operation. d_k represents the dimension.

Considering the learning gain in LPKT requires combining the learning unit of the previous moment and the time interval for answering the exercise, we simplify this steps and further combine the student's ability to achieve personalized learning gain as follows:

$$\widetilde{lg}_t = tanh(\mathbf{W}_2^T[\widetilde{l}_t \oplus \widetilde{h}_{t-1}] + b_2), \tag{11}$$

where $\mathbf{W}_2 \in \mathbb{R}^{(2d_k) \times d_k}$ denotes the weight matrix, $b_2 \in \mathbb{R}^{d_k}$ denotes the bias term, $tanh$ denotes the non-linear activation function. Then we apply a perceptron to generate a personalized learning gain gate as:

$$\widetilde{\Gamma}_t^l = \sigma(\mathbf{W}_3^T[\widetilde{l}_t \oplus \widetilde{h}_{t-1}] + b_3) \tag{12}$$

where $\mathbf{W}_3 \in \mathbb{R}^{(2d_k) \times d_k}$ denotes the weight matrix, $b_3 \in \mathbb{R}^{d_k}$ is the bias term, σ denotes the non-linear *sigmoid* activation function.

Furthermore, we think that the mastery of the knowledge concepts contained in a certain exercise may have a positive impact on the mastery of their related knowledge concepts in the learning process. In order to capture this impact between the relevant knowledge concepts, we directly concat the knowledge state h_{t-1} with the relevant knowledge concept vector p_{e_t} to obtain the degree of mastery of relevant knowledge concepts $\tilde{h}_{t-1}^c$ as:

$$\tilde{h}_{t-1}^c = p_{e_t} \oplus h_{t-1}, \tag{13}$$

$p_{e_t} \in \mathbb{R}^{d_k}$ represents the relevant knowledge concept vector, which is obtained by matching the related knowledge concepts of the current knowledge concept in P-matrix defined in section 3.1. By embedding the relevant knowledge concepts $\tilde{h}_{t-1}^c$ into the learning interaction, we can get the final knowledge acquisition $\widetilde{LG}_t^c$ as follows:

$$LG_t^c = \widetilde{\Gamma}_t^l \cdot [(\widetilde{lg}_t + 1)/2] + ||\sigma(\tilde{h}_t^c)||_2 \tag{14}$$

$$\widetilde{LG}_t^c = q_{e_t} \cdot LG_t^c \tag{15}$$

4.2.2 *Causal Forgetting Module*

The student forgetting process is similar to the learning process. In the learning process, mastering a specifc knowledge point may have a positive impact on related knowledge concepts and different students always get personalized knowledge gain in each interaction. Conversely, in the forgetting process, forgetting a knowledge point can negatively influence related knowledge points. Simultaneously, students with different cognitive abilities may also represent individual forgetting rate. Therefore, we design a personalized and causal forgetting module to capture this influence and monitor the personalized forgetting situation in this section.

Causal forgetting mechanism. In order to capture the causal relationship between the knowledge concept contained in the current exercise and its related knowledge concept in the forgetting process. we design a causal forgetting mechanism in this study. Specifically, we first obtain the response time at_t and the interval time it_t for the current exercise, and identify the knowledge concept contained in the current exercise at the same time. Then, we use the P-matrix defined in section 3.1 to match all prerequisite knowledge concepts corresponding to this knowledge concept, and further obtain the response time at_t and interval time it_t for exercises corresponding to these prerequisite knowledge concepts. It should be noted that there is a one-to-one correspondence between exercises and knowledge concepts, and a one-to-many relationship between knowledge concepts. So we determine the closest prerequisite knowledge concept to the current knowledge concept by considering the time step intervals between the exercises corresponding to these prerequisite knowledge concepts and the current exercise. For example, if the exercise e_m contains the prerequisite knowledge concept c_j of the knowledge concept c_i in current exercise e_t and the exercise e_m is closest to the current exercise e_t, we will select knowledge concepts in the exercise e_m as the closest prerequisite knowledge concept to the current knowledge concept. Finally, as defined in Eq.16 and Eq.17, we construct a forgetting weight by calculating the time difference between the exercises corresponding to the nearest prerequisite knowledge concept and the current exercise:

$$\triangle t_{tm} = |(at_t + it_t) - (at_m + it_m)| \quad \text{where } t > m, \quad 0 \leqslant t, m \leqslant T - 1 \tag{16}$$

$$w_{tm} = \delta \left(1 + (exp(\triangle t_{tm} + \lambda))\right)^{-1} \tag{17}$$

The hyperparameter λ is used to adjust the offset of the forgetting weight. If the knowledge concept c_i in the current exercise successfully matches a prerequisite knowledge concept c_j in the P-matrix, then $p_{ij} \neq 0$; otherwise, $p_{ij} = 0$. Furthermore, we believe that forgetting a knowledge concept not only affects the mastery of that specifc concept but also impacts the mastery of its related concepts. Therefore,

we multiply the forgetting weight calculated based on the hierarchical relationship between knowledge concepts by the learning gain to obtain the final overall knowledge gain LG_t^l:

$$w_f = \begin{cases} 1 & \text{if } p_{ij} = 0 \\ w_{tm} & \text{if } p_{ij} \neq 0 \end{cases} \tag{18}$$

$$LG_t^l = w_f \cdot \widetilde{LG}_t^c \tag{19}$$

Given that students' memory of learning materials tends to decline over time. However, when students encounter exercises similar to those they have previously encountered, they may undergo a process of knowledge reactivation, known as the review process, which aligns with the phenomena of long-term forgetting and short-term memory. We call it the forgetting-review mechanism of the learning process. We simulate this mechanism and analyze the changes in short-term knowledge mastery by computing the similarity of k adjacent knowledge states, as shown in Eq.20. And then, we apply this similarity to generate a related knowledge state h_k, as defined in Eq.21, at the current moment. This related knowledge state h_k will be further applied in the Eq.22 to updata the forgetting gate.

$$m_k = \text{sim}(\tilde{h}_{t-k}, h_{t-1}) \tag{20}$$

$$h_k = \sum_{k=1}^{T} \text{softmax}(m_k) \cdot h_{t-1} \tag{21}$$

Personalized forgetting mechanism. Individual differences can lead to variations in students' knowledge acquisition and forgetting rates when facing the same sequence of exercises. In order to calculate students' knowledge states more accurately, we incorporate students' personalized capabilities into the modeling of the forgetting process. By introducing the dynamic mechanism of the forgetting gate $\widetilde{\Gamma}_t^f$, we integrate students' personalized capabilities s_t, their previous knowledge state h_{t-1}, their knowledge mastery LG_t^l, the time interval between exercises it_t, and their short-term memory degree h_k to simulate knowledge forgetting, thereby enhancing our ability to predict knowledge states.

$$\widetilde{\Gamma}_t^f = \sigma(\mathbf{W}_4^T[h_{t-1} \oplus LG_t^l \oplus it_t \oplus s_t \oplus h_k] + b_4) \tag{22}$$

Where $\mathbf{W}_4 \in \mathbb{R}^{(5d_k \times d_k)}$ denotes the weight matrix, $b_4 \in \mathbb{R}^{d_k}$ denotes the bias term, σ denotes the non-linear *sigmoid* activation function. Similar to LPKT, we multiply the forgetting factor $\widetilde{\Gamma}_t^f$ by the previous state of knowledge h_{t-1} to eliminate the effect of forgetting. Therefore, after completing the t-th learning interaction, the student's knowledge status h_t is updated as follows:

$$h_t = \widetilde{LG}_t^c + \widetilde{\Gamma}_t^f \cdot h_{t-1} \tag{23}$$

4.2.3 *Predicting Module*

The prediction layer combines the temporal characteristics of the KT model and makes full use of students' learning history and interaction sequences to predict their performance in the next exercise. Specifically, we predict their performance in the next exercise e_{t+1} based on their knowledge state $\tilde{h}_t$ after the t-th learning interaction. We first concatenate the knowledge state embedding $\tilde{h}_t$ with the concept-perceived exercise embedding $\tilde{e}_{t+1}$, and then project them onto the output layer through a fully connected network with sigmoid activation:

$$y_{t+1} = \sigma(\mathbf{W}_5^T[\tilde{e}_{t+1} \oplus \tilde{h}_t] + b_5) \tag{24}$$

Where $\mathbf{W}_5 \in \mathbb{R}^{(2d_k \times d_k)}$ denotes the weight matrix, and $b_5 \in \mathbb{R}^{d_k}$ represents the bias term. The output y_{t+1} serves to predict the student's performance in the next exercise e_{t+1}. Additionally, we set a threshold to determine whether the student can answer e_{t+1} correctly. Specifically, if y_{t+1} surpasses the threshold, the answer is a correct response; otherwise, the answer is deemed incorrect.

4.2.4 *Objective Function*

To optimize all parameters in CPF, we employ the cross-entropy log loss between the predicted values y and the actual answers a as the objective function:

$$\mathbb{L}(\theta) = -\sum_{t=1}^{T} \left(a_t \log(y_t) + (1 - a_t) \log(1 - y_t) \right) + \lambda_\theta \|\theta\|^2 \tag{25}$$

Where θ represents all parameters in CPF model and λ_θ serves as the regularization hyper-parameter. We employ the Adam optimizer to minimize the objective function on mini-batches. More detailed information regarding the experimental settings are provided in the subsequent sections.

5 EXPERIMENTS

5.1 Experimental Setup

In this section, we provide the detailed information of the evaluation datasets, training details and other competitive methods for comparison.

5.1.1 *Dataset Description*

In our experiments, we utilize three real-world public datasets to assess the performance of CPF, *i.e.*, ASSISTments 2012, ASSISTments Challenge, EdNet-KT1. The basic statistics for these datasets are summarized in Table 1.

- **ASSISTments 2012 (ASSIST2012)** [1] is the largest version of the dataset, collects between September 2012 and October 2013. It includes 17,999 exercises answered by 46,674 students, with a total of 6,123,270 interactions [24]. To focus on the 265 knowledge points, we filter the dataset to include only relevant exercises.
- **ASSISTments Challenge (ASSISTChall)** [2] is a competition organized by the Assistants online Tutoring System to promote data mining and machine learning research in the field of education. The dataset contains the learning behavior and performance data of the students in the assistments system.
- **EdNet-KT1** [3] dataset contains all student system interactions collected over two years with more than 780,000 users in South Korea [27]. Each student generates an average of 441.20 interactions. It offers large-scale real-world Intelligent Assisted Instruction system (ITS) data and includes 13,169 exercises, 1,021 lectures, and labels for 293 concepts. When dealing with a problem with multiple associated knowledge concepts, only the first one is chosen.

Table 1 Statistics of all datasets.

Statistics	Datasets		
	ASSIST2012	ASSISTchall	EdNet-KT1
Students	29,018	1,709	784,309
Exercises	53,091	3,162	12,372
Concepts	265	102	141
Answer Time	26,747	1,326	9,292
Interval Time	29,748	2,839	41,830
Avg. Length	93.45	551.68	125.45

---

1) https://sites.google.com/site/assistmentsdata/home/2012-13-school-data-withaffect
2) https://sites.google.com/view/assistmentsdatamining/dataset
3) http://ednet-leaderboard.s3-website-ap-northeast-1.amazonaws.com/

5.1.2 *Training Details*

First, we sort students' learning records according to the timestamp of their answers. We then set all input sequences to a fixed length based on the average sequence length of the dataset. Specifically, in the ASSIST2012 and EdNet-KT1 datasets, we set the fixed length of the input sequence to 100, while in the ASSISTchall dataset, the fixed length of the input sequence is set to 500. For sequences longer than a fixed length, we cut them into multiple unique subsequences, each of which is a fixed length. For shorter sequences of fixed length, we pad them with zero vectors so that they reach fixed length.

We perform standard 5-fold cross-validation on all datasets. In each fold, 80% of the students are divided into the training set (80%) and the verification set (20%), the rest 20% are used as the testing set. During the training process, we randomly initialize all parameters, using uniform distribution [51]. All hyper-parameters are learned on the training set, and the model that performs best on the validation set is selected for testing set evaluation. In CPF, we added a dropout layer with a dropout rate of 0.2 to avoid overfitting. In our implementation, the parameters d_k and d_e are set to 128 and d_a to 50. For the small positive γ in the enhanced Q matrix, we set it to 0.03, and the relationship strength coefficient ρ in the P-matrix is set to 0.03, the learning rate is set to 3×10^{-3}, and the batch-size is set to 128. For a fair comparison, all models are trained on a cluster of Linux servers with NVIDIA 3090 GPUs.

5.1.3 *Other Competitive Methods*

To verify the effectiveness of CPF, we compare our model with several previous methods.

• **DKT** applies deep learning to KT for the first time. It takes the learning sequence as input to the RNN or its variant LSTM [32,62] and represents the student's state of knowledge through hidden states.

• **DKT+** is an extended and improved version of DKT that aims to solve two major problems in DKT. First, DKT cannot efficiently reconstruct unobserved input data. Second, the prediction performance of DKT is inconsistent between different time steps [38]. By enhancing the original DKT loss function with two additional regularization terms, DKT+ attempts to overcome these limitations and improve the performance and application range of the model.

• **DKVMN** is a model that introduces memory-enhancing neural networks into knowledge tracing. The model uses the learning mechanism of key-value pair memory to capture the knowledge state of students [35] and is applied to the student knowledge model and prediction model.

• **SAKT** models the student's knowledge state and learning progress by taking the student's learning interaction sequence [41] as input and using self-attention mechanisms [64].

• **AKT** is a context-aware attentive knowledge tracing model [17], utilizes dual self-attentive encoders for exercises and responses. The knowledge retriever employs attention to retrieve past knowledge relevant to the current exercise.

• **DTransformer** establishes an architecture [15] from the problem level to the knowledge level, explicitly diagnosing learners' proficiency based on the mastery of each problem. It utilizes contrastive learning to maintain the stability of diagnosing knowledge states.

• **CT-NCM** integrates dynamic forgetting [9] into students' learning process, effectively models knowledge learning and forgetting, and distinguishes positive and negative reactions.

• **LPKT** displays the modeling user's knowledge state and uses the embedding of exercises. The knowledge points contained in exercises are used to select the corresponding exercise records [6]. It reflects the consistency of students' changing knowledge state and learning process.

• **LPKT-S** considers that students generally have different progress rates [26], a student embed containing student-specific progress rates is introduced, extending LPKT to LPKT-S.

5.2 Experimental Results

In this section, we conducted several experiments to illustrate the interpretability of the model from different perspectives. Firstly, our model outperforms other competitive methods in predicting student performance. Secondly, we effectively consider students' personalized abilities, which align better with real-world application scenarios. Finally, we model the causal relationship of the forgetting process by considering the predecessor and successor relationships of knowledge concepts, and incorporate students' abilities to update knowledge states.

Table 2 Results of comparison methods on students' performance prediction.

Method	ASSIST2012				ASSISTchall				EdNet-KT1			
	RMSE	AUC	ACC	r^2	RMSE	AUC	ACC	r^2	RMSE	AUC	ACC	r^2
DKT [1]	0.4241	0.7289	0.7360	0.1468	0.4471	0.7213	0.6907	0.1425	0.4508	0.6836	0.6889	0.1008
DKT+ [38]	0.4239	0.7295	0.7254	0.1497	0.4502	0.7101	0.6842	0.1308	0.4601	0.6429	0.6733	0.0635
DKVMN [35]	0.4261	0.7228	0.7329	0.1398	0.4503	0.7108	0.6842	0.1302	0.4538	0.6741	0.6843	0.0913
SAKT [41]	0.4258	0.7233	0.7339	0.1403	0.4626	0.6605	0.6694	0.0822	0.4524	0.6794	0.6862	0.0964
DTransformer [15]	0.4118	0.7698	0.7509	0.2004	0.4371	0.7506	0.7078	0.1791	0.4291	0.7553	0.7089	0.1837
AKT [17]	0.4121	0.7706	0.7515	0.2004	0.4364	0.7501	0.7080	0.1801	0.4297	0.7557	0.7083	0.1842
LPKT [6]	0.4089	0.7740	0.7551	0.2145	0.4179	0.7939	0.7385	0.2491	0.4290	0.7721	0.7106	0.2195
LPKT-S [26]	0.4065	0.7803	0.7584	0.2004	0.4160	0.7979	0.7420	0.2558	0.4263	0.7801	0.7158	0.2230
CT-NCM [9]	0.4013	0.7945	0.7609	0.2386	0.4096	0.8166	0.7430	0.2893	0.4293	0.7691	0.7271	0.2183
CPF	**0.3990**	**0.8026**	**0.7665**	**0.2524**	**0.4075**	**0.8206**	**0.7495**	**0.2921**	**0.4230**	**0.7980**	**0.7295**	**0.2640**

5.2.1 *Main results*

As shown in Table 2, CPF shows varying degrees of improvement on the three datasets compared with other models, indicating that the learning and forgetting behaviors emphasized by CPF are effective in knowledge tracing modeling.

DKT uses the latent vector of the LSTM model to model the overall knowledge status of students. It does not establish the connection between questions and knowledge concepts and explore the potential relationships between knowledge concepts, and cannot capture each student's mastery of knowledge concepts. Therefore, the predictive performance of DKT is lower than CPF on all three datasets. Both DKVMN and CPF can simulate students' mastery of various knowledge concepts, but DKVMN ignores the forgetting behavior during the learning process. Therefore, CPF outperforms DKVMN in predictive performance. This also shows that modeling knowledge concepts can better capture changes in knowledge status and more accurately calculate students' mastery of knowledge concepts. SAKT and AKT both consider the decay of memory over time but do not account for individual student differences and other factors that may influence the forgetting process. Compared with other models, the LPKT and LPKT-S methods comprehensively model learning and forgetting behaviors in the learning process, but LPKT does not emphasize the individual differences of students. Although LPKT-S distinguishes individual learning progress rates, it does not capture the causal relationships present in the forgetting process. CPF emphasizes the impact of students' cognitive abilities in the learning and forgetting process, deeply explores the causal forgetting caused by the relationship between knowledge concepts, it can be observed that CPF significantly outperforms the baseline model LPKT on the EdNet-KT1 dataset (*i.e.*, improves the AUC by 2.59%), indicating a more comprehensive modeling of both the learning and forgetting processes. CT-NCM, similar to CPF, is designed for modeling the forgetting process. CT-NCM adeptly incorporates the continuous and dynamic forgetting behavior into the modeling of students' learning processes, discerning the influence of both positive and negative responses on their knowledge states. However, it omits the analysis of hierarchical relationships among knowledge concepts and neglects the impact of individual differences on knowledge states. We observed that CPF outperforms the CT-NCM method on the EdNet-KT1 dataset (*i.e.*, improves the AUC by 2.89%), indicating that our modeling approach to the forgetting process is closer to real-world scenarios.

5.2.2 *Students' mastery of knowledge points*

The personalized abilities of students play a crucial role in both the learning and forgetting stages. Traditional models rely on the random initialization of student abilities. As shown in Figure 3 (a) and (b), Student2 (S2) and Student3 (S3) exhibit similar initial mastery levels of the knowledge concept c_5, demonstrating relatively consistent abilities. After completing the same sequence of exercises, there is no difference in the final mastery level of the knowledge concept c_5. However, for Student1 (S1) and Student4 (S4), due to different initial mastery levels of the knowledge concept c_4, there are differences in their abilities. Consequently, there are differences in the mastery levels after learning. This indicates that despite making similar responses to the same exercises, students with different levels of ability obtain different learning outcomes, leading to differences in the final mastery level of knowledge.
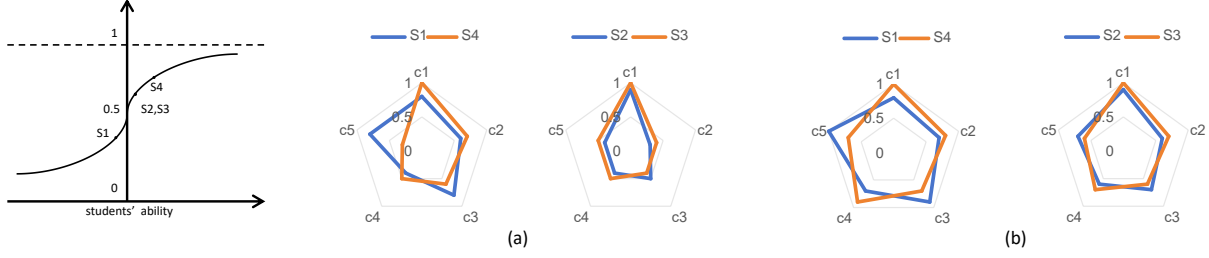
Figure 3 Illustration of the mastery level of student knowledge points. On the left side are the individual differences in students' abilities, while on the right side are the mastery levels of knowledge points after students undergo a consistent answering process. (a) and (b) represent the initial ability distribution of students and the degree of mastery of knowledge concepts after the same answering sequence respectively.

5.2.3 *Visualization of Causal Forgetting*

In CPF, we employ a novel forgetting strategy by using a matrix to represent the relationships between knowledge concepts and their prerequisites. During the modeling of the forgetting process, we determine the forgetting weight by calculating the closest prerequisite knowledge concept to the current one. Simultaneously, we also observe the phenomenon of memory enhancement, whereby students may forget certain knowledge concepts, but upon encountering similar exercises shortly thereafter, they reinforce their memory of these concepts through the review process. Fig. 4 illustrates the changes in knowledge states for CPF and LPKT. In the learning stage, the acquisition of prerequisite knowledge concepts contributes to subsequent knowledge concepts, while in the forgetting stage, the forgetting of subsequent knowledge concepts will have an impact on the mastery of prerequisite knowledge concepts. For instance, c_1 and c_2, serving as prerequisites for c_3, result in a higher knowledge state in the answer e_5 compared to the baseline. In the forgetting process, when a knowledge concept that has already forgotten another encounters a similar type of exercise, the forgetting degree for subsequent knowledge concepts may exceed simple time decay. c_5 and c_6 as prerequisites for c_7 exhibit a weaker improvement in knowledge state when answering e_{12} due to a certain level of forgetting of prerequisite knowledge concepts.
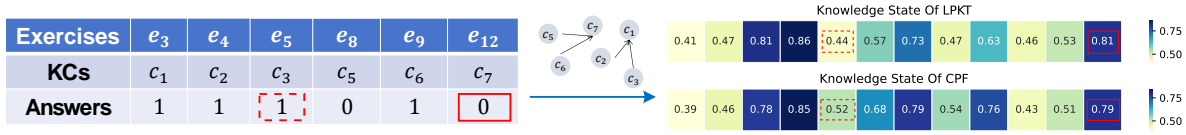

Figure 4 The evolution of students' knowledge state. The left side represents the students' answer order and results, and the right side describes the updating of the knowledge state due to the knowledge concept relationship.

5.2.4 *Ablation Study*

In this section, we conduct ablation experiments to demonstrate the efficacy of each module and the effect of parameter in CPF.

- **CPF (P)** does not consider forgetting at the conceptual level of knowledge.
- **CPF (FP)** does not consider the impact of forgetting on learning throughout the interaction.
- **CPF (L)** does not consider the learning process of students for specific knowledge concepts and overall knowledge.
- **CPF (I)** does not take into account individualized differences in students' abilities during the learning process.

Table 3 Ablation experimental results on ASSIST2012.

Model	P-matrix	Personalization	AUC	ACC	Model	Learning	Forgetting	AUC	ACC
CPF (P)	✗	✓	0.7821	0.7562	CPF (L)	✗	✓	0.7905	0.7593
CPF (I)	✓	✗	0.8008	0.7658	CPF (FP)	✗	✗	0.7725	0.7538
CPF	✓	✓	**0.8026**	**0.7665**	CPF	✓	✓	**0.8026**	**0.7665**

Apart from the settings mentioned above, the remaining components and experimental configurations of the model remain unchanged. The results in Table 3 reveal some interesting findings. Firstly, the

commonly observed phenomenon of forgetting plays a crucial role in the learning process, and ignoring the relationships between knowledge concepts when modeling the forgetting process can result in poorer predictive outcomes. Secondly, because our model explores the causal relationships of the forgetting process in more detail, ignoring the entire forgetting process would greatly diminish predictive results. This also indicates that our designed forgetting process more comprehensively considers factors influencing forgetting, better simulating real-world application scenarios. Thirdly, if the importance of knowledge concepts is not considered in modeling the learning process, there will also be a certain degree of decline in predictive results. Fourthly, neglecting student personalization would also decrease the predictive accuracy of the model.

5.3 Experimental Analysis

In this section, we analyze the distribution of relationships between knowledge concepts in the dataset, discuss the reasons for selecting directed relationships, and explore parameter sensitivity.

5.3.1 *Construction of Knowledge Concept Relation*

Inspired by Relation Map Driven Cognitive Diagnosis (RCD), we introduce the following matrix to build dependencies between concepts [53,61]. The first is the Answer Matrix, denoted by A. From the exercise record, we calculate the matrix A, where A_{ij} represents the number of times the concept j is answered correctly immediately after the concept i is answered correctly, and $A_{i,j} = \frac{n_{i,j}}{\sum_k n_{i,k}}$ (if $i \neq j$), else, it is 0. In addition, we have a Transition Matrix, denoted by T, which is a binary matrix [23,60]. Where $T_{i,j} = 1$ indicates that there is an edge from concept i to concept j. To obtain the transfer matrix T, we first calculate the normalized matrix of matrix A, denoted as $\widetilde{T}$. Specifically, $\widetilde{T}_{ij} = \frac{A_{ij} - min(A)}{max(A) - min(A)}$ represents the probability that some educational relationship exists between concept i and concept j. Then, we determine the relations by $T_{i,j} = 1$ if $\widetilde{T}_{ij} > threshold$. we set $threshold$ as third power of the average value of matrix T. If $T_{i,j} = 1$ but $T_{j,i} \neq 1$, then the concept j is a successor to the concept i. The transition matrix T calculated in this way can represent the hierarchical relationship between concepts.
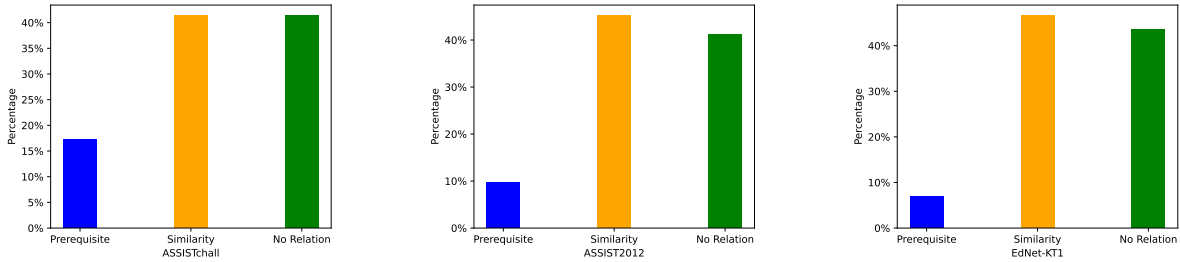

Figure 5   Comparison of exercise-concept correlation study

5.3.2 *Knowledge Concept Relationship Analysis*

Fig.5 depicts three types of relationships between knowledge concepts in three datasets: (1) the c_i concept is similar to the c_j concept; (2) The concept of c_i is the prerequisite knowledge concept of c_j; (3) There is no clear relationship between the c_i concept and the c_j concept. The distribution trends of the three datasets are similar, with the lowest proportion of prerequisite relationships, and the higher proportions of similar relationships and no clear relationships. We rely on these relationships to measure the degree of forgetting of relevant knowledge concepts. Due to the large proportion of similar relationships, choosing to use similar relationships for weighting may have an impact on many exercises in the datasets, which will cause the weighting effect to be insignificant. Therefore, we decided to use prerequisite relationships to weight the forgetting process. From the results in Table 2, it can also be observed that the experimental results of ASSISTchall show relatively minor improvements, which reinforces the validity of our choice of prerequisite relationships as the weighting method.

5.3.3 *Parameter Sensitivity*

Additionally, we conducted experiments to assess the impact of γ on the LPKT and CPF enhanced Q matrix. We tried five different γ values: 0, 0.01, 0.03, 0.05 and 0.1. The experimental results are
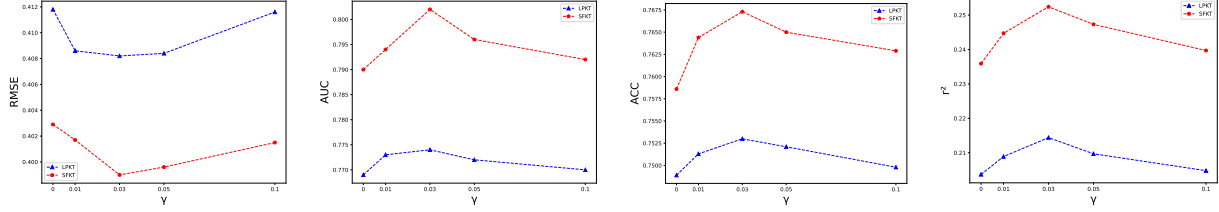
Figure 6 The influence of the small positive value $\boldsymbol{\gamma}$ in the enhanced Q-matrix on the performance of LPKT and CPF on ASSIST2012.

shown in Fig. 6. It can be observed that setting γ to a small positive value enhances the performance of LPKT and CPF, with the maximum gain achieved when γ is set to 0.03. When γ approaches zero, it becomes challenging for γ to bridge the potential correlation between different knowledge concepts. On the other hand, if γ increases, more errors may occur in the enhanced Q matrix, which would impair the performance of LPKT and CPF. From the graph, it is evident that CPF and LPKT exhibit distinct slopes. In the CPF model, we first use the Q matrix to match the knowledge concepts present in the current exercise and then utilize the predecessor-successor knowledge concept matrix to identify the prerequisite knowledge concepts of the current one. This step emphasizes that our model relies more on the relationships between knowledge concepts when computing the forgetting process, further validating the effectiveness of the concept-driven forgetting mechanism proposed in our work.

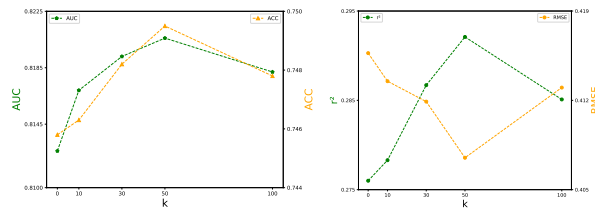| K Metrics	0	10	30	50	100
AUC	0.8126	0.8169	0.8193	**0.8206**	0.8182
ACC	0.7458	0.7463	0.7482	**0.7495**	0.7478
RMSE	0.4157	0.4135	0.4119	**0.4075**	0.4130
r^2	0.2760	0.2783	0.2867	**0.2921**	0.2851

Figure 7 Experimental results with different K values on ASSISTchall.

Fig. 7 illustrates the impact of different values of k on the prediction results when considering the preceding k neighboring knowledge states. Due to the presence of numerous knowledge concepts in practical applications, considering the similarity between all knowledge concepts in the sequence may lead to an increase in computational complexity. Therefore, we experimented with five different values of k: 0, 10, 30, 50, and 100. When k is small, it is equivalent to disregarding the memory enhancement process described above. As k increases, the complexity of the relationships between knowledge concepts grows, potentially involving a large number of knowledge concepts in the enhancement process, which could destabilize the model.

6 Conclusion

In this paper, we propose a novel concept-driven personalized forgetting knowledge tracking model (CPF) aimed at addressing personalized learning and forgetting processes in real-world applications. CPF integrates students' personalized capabilities into both the learning and forgetting processes and models the causal relationships of forgetting processes through the hierarchical relationships between knowledge concepts. Extensive experiments on three public datasets demonstrate the superiority of CPF over existing methods. Future research directions include further strengthening the CPF model, considering the multi-concept effects, more comprehensively characterizing the intrinsic relationships between knowledge concepts, and delving deeper into the impact of knowledge concepts on the model.

References

1 Piech, Chris and Bassen, Jonathan and Huang, Jonathan and Ganguli, Surya and Sahami, Mehran and Guibas, Leonidas and Sohl-Dickstein, Jascha. *Deep Knowledge Tracing.* Proceedings of the 28th International Conference on Neural Information Processing Systems , 2015, MIT Press, Cambridge, MA, USA.

2 Lord, F.M. (1980). *Applications of Item Response Theory To Practical Testing Problems (1st ed.).* Routledge.

3 Hiromi Nakagawa, Yusuke Iwasawa, Yutaka Matsuo. *Graph-based Knowledge Tracing: Modeling Student Proficiency Using Graph Neural Network.* In *2019 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*, 2019.

4   Thomas M. Moerland, Joost Broekens, Aske Plaat, Catholijn M. Jonker. *Model-based Reinforcement Learning: A Survey. 2023.*

5   Bai, Yujie; Chen, Shujuan; Xing, Zhenchang, et al. *ArgusDroid: detecting Android malware variants by mining permission-API knowledge graph. Sci. China Inf. Sci.*, 2023, 66, 192101.

6   Shuanghong Shen, Qi Liu, Enhong Chen, Zhenya Huang, Wei Huang, Yu Yin, Yu Su, and Shijin Wang. *Learning Process-Consistent Knowledge Tracing.* Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, 2021.

7   Yan Zhao, Huifang Ma, Jing Wang, Xiangchun He, Liang Chang. *Question-response representation with dual-level contrastive learning for improving knowledge tracing.* In: *Information Sciences*, Volume 658, 2024, 120032.

8   Yuling Ma, Chaoran Cui, Xiushan Nie, Gongping Yang, Kashif Shaheed, Yilong Yin. *Pre-course Student Performance Prediction with Multi-instance Multi-label Learning.* In: *Sci. China Inf. Sci.*, Volume 62, Issue 2, 2019

9   Ma, Haiping, Wang, Jingyuan, Zhu, Hengshu, Xia, Xin, Zhang, Haifeng, Zhang, Xingyi, Zhang, Lei. *Reconciling Cognitive Modeling with Knowledge Forgetting: A Continuous Time-aware Neural Network Approach.* In International Joint Conference on Artificial Intelligence, 2022.

10 Liu, H., Zhang, T., Wu, P., et al. *A Review of Knowledge Tracking.* In *Journal of East China Normal University*, 2019.

11 Shengzhe Dai, Zhiheng Li, Li Li, Dongpu Cao, Xingyuan Dai, Yilun Lin. *Investigating the Dynamic Memory Effect of Human Drivers via ON-LSTM.* In: *Sci. China Inf. Sci.*, 2020.

12 Jing-Jing Bai. *Using Self-Efficacy Theory to Stimulate Middle School Students' Learning Motivation. Education, Psychology*, 2021.

13 Abdelrahman, Ghodai and Wang, Qing and Nunes, Bernardo. *Knowledge Tracing: A Survey.* ACM Comput. Surv., 2023.

14 Philip I. Pavlik, Hao Cen, and K. Koedinger. *Performance Factors Analysis - A New Alternative to Knowledge Tracing.* In: *International Conference on Artificial Intelligence in Education*, 2009.

15 Yin, Y., Dai, L., Huang, Z., Shen, S., Wang, F., Liu, Q., Chen, E., Li, X. *Tracing Knowledge Instead of Patterns: Stable Knowledge Tracing with Diagnostic Transformer.* In *Proceedings of the ACM Web Conference 2023*, ACM, 2023.

16 Qiqi Su, Christos Kloukinas, Artur d'Avila Garcez. *Modular Neural Networks for Time Series Forecasting: Interpretability and Feature Selection using Attention. ArXiv*, 2023.

17 Aritra Ghosh, Neil T. Heffernan, Andrew S. Lan. *Context-Aware Attentive Knowledge Tracing. Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2020.

18 Hongbo Gao, Hang Su, Yingfeng Cai, Renfei Wu, Zhengyuan Hao, Yongneng Xu, Wei Wu, Jianqing Wang, Zhijun Li, Zhen Kan. *Trajectory Prediction of Cyclist Based on Dynamic Bayesian Network and Long Short-Term Memory Model at Unsignalized Intersections.* In: *Sci. China Inf. Sci.*, 2021.

19 Guo Xie, Anqi Shangguan, Rong Fei, Wenjiang Ji, Weigang Ma, Xinhong Hei. *Motion Trajectory Prediction Based on a CNN-LSTM Sequential Model.* In: *Sci. China Inf. Sci.*

20 Deqing Huang, Yuanzhe Fu, Na Qin, Shibin Gao. *Fault Diagnosis of High-Speed Train Bogie Based on LSTM Neural Network.* In: *Sci. China Inf. Sci.*, Volume 64, Issue 1, 2021.

21 Ji, Zongcheng; Chen, Kai; He, Yonggang; et al. *Heterogeneous memory enhanced graph reasoning network for cross-modal retrieval. Sci. China Inf. Sci.*, 2022.

22 Habimana, Olivier; Li, Yunchao; Li, Rui; et al. *Sentiment analysis using deep learning approaches: an overview. Sci. China Inf. Sci.*, 2020.

23 Weibo Gao, Qi Liu, Zhenya Huang, Yu Yin, Haoyang Bi, Mu-Chun Wang, Jianhui Ma, Shijin Wang, and Yu Su. *RCD: Relation Map Driven Cognitive Diagnosis for Intelligent Education Systems.* Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2021.

24 Mingyu Feng, Neil Heffernan, Kenneth Koedinger. *Addressing the Assessment Challenge with an Online System That Tutors as It Assesses.* User Modeling and User-Adapted Interaction, August 2009. Kluwer Academic Publishers.

25 Wang, Shitong; Xu, Gang; Du, Ruofei. *Restricted Bayesian classification networks. Sci. China Inf. Sci.*, 2013, 56, 1–15.

26 Shuanghong Shen, Enhong Chen, Qi Liu, Zhenya Huang, Wei Huang, Yu Yin, Yu Su, Shijin Wang. *Monitoring Student Progress for Learning Process-Consistent Knowledge Tracing. IEEE Trans Knowl Data Eng*, 2023.

27 Youngduck Choi, Youngnam Lee, Dongmin Shin, Junghyun Cho, Seoyon Park, Seewoo Lee, Jineon Baek, Chan Bae, Byungsoo Kim, Jaewe Heo. *Ednet: A Large-Scale Hierarchical Dataset in Education.* International Conference on Artificial Intelligence in Education, 2020. Publisher: Springer.

28 Zhang Zheng, Wu Le, Liu Qi, Liu Jiayu, Huang Zhenya, Yin Yu, Zhuang Yan, Gao Weibo, Chen EnHong. *Understanding and Improving Fairness in Cognitive Diagnosis.* In: *Sci. China Inf. Sci.*, 2023.

29 Liu, Xiaozhong; Pan, Sinno Jialin; Zhang, Qi; et al. *Reformulating natural language queries using sequence-to-sequence models. Sci. China Inf. Sci.*, 2019.

30 Albert T Corbett, John R Anderson. *Knowledge tracing: Modeling the acquisition of procedural knowledge.* User Modeling and User-Adapted Interaction, 1994-1995.

31 Xia, Min; Weng, Long; Wang, Zheng. *Sequence memory based on an oscillatory neural network. Sci. China Inf. Sci.*, 2014.

32 Sepp Hochreiter, Jürgen Schmidhuber. *Long Short-Term Memory.* Neural Computation, November 1997.

33 Wei Wang, GuoSun Zeng. *Bayesian cognitive trust model based self-clustering algorithm for MANETs.* In: *Sci. China Inf. Sci.*, 2010.

34 Nakagawa, Hiromi, Iwasawa, Yusuke, and Matsuo, Yutaka. *Graph-Based Knowledge Tracing: Modeling Student Proficiency Using Graph Neural Network.* In IEEE/WIC/ACM International Conference on Web Intelligence (WI), New York, NY, USA, 2019. Association for Computing Machinery.

35 Jiani Zhang, Xingjian Shi, Irwin King, and Dit-Yan Yeung. *Dynamic Key-Value Memory Networks for Knowledge Tracing.* Proceedings of the 26th International Conference on World Wide Web, 2017.

36 Anderson, John R., Boyle, C. Franklin, Corbett, Albert T., and Lewis, Matthew W. *Cognitive Modeling and Intelligent Tutoring.* Artificial Intelligence, 42(1), 1990.

37 Fu, Xiaolan; Cai, Longyan; Liu, Yiqun; et al. *A computational cognition model of perception, memory, and judgment. Sci. China Inf. Sci.*, 2014, 57, 1–15.

38 Chun-Kit Yeung and D. Y. Yeung. *Addressing two problems in deep knowledge tracing via prediction-consistent regularization.* Proceedings of the Fifth Annual ACM Conference on Learning at Scale, 2018.

39 Villano, Michael. *Probabilistic Student Models: Bayesian Belief Networks and Knowledge Space Theory.* In Intelligent Tutoring Systems, Second International Conference, ITS, Montréal, Canada, Proceedings (Lecture Notes in Computer Science), Vol. 608. Springer, 1992.

40 Wang, Dongwei and Han, Zhi and Wang, Yanmei and Chen, Xiai and Liu, Baichen and Tang, Yandong. *Review helps learn better: Temporal Supervised Knowledge Distillation.* arXiv 2023.

41 Shalini Pandey and George Karypis. *A Self Attentive model for Knowledge Tracing.* ArXiv, 2019.

42 Albert T. Corbett, John R. Anderson. *Knowledge tracing: Modeling the acquisition of procedural knowledge.* User Modeling and User-Adapted Interaction, 2005.

43 Steven P. Reise. *Item Response Theory.* In: The Encyclopedia of Clinical Psychology. John Wiley & Sons, Ltd, 2014.

44 Xie, Zhiwei; Zeng, Zhiwei; Zhou, Guangyou, et al. *Topic enhanced deep structured semantic models for knowledge base exercise answering. Sci. China Inf. Sci.*, 2017, 60, 110103.

45 Shaul Markovitch and Paul D. Scott. *The Role of Forgetting in Learning.* In: Proceedings of the Fifth International Conference on Machine Learning. Morgan Kaufmann Publishers Inc., 1988.

46 Zhenya Huang, Qi Liu, Yuying Chen, Le Wu, Keli Xiao, Enhong Chen, Haiping Ma, Guoping Hu. *Learning or Forgetting? A Dynamic Approach for tracing the Knowledge Proficiency of Students.* In: ACM Trans. Inf. Syst.2020.

47 J. Liu, G. Xu, Z. Ying. *Data-Driven Learning of Q-Matrix.* In: Applied Psychological Measurement, 2012.

48 Geng, Xiangyu; Ouyang, Dantong; Zhang, Yu. *Model-based diagnosis of incomplete discrete-event system with rough set theory. Sci. China Inf. Sci.*, 2017.

49 G. R. Loftus. *Evaluating forgetting curves.* In: Journal of Experimental Psychology: Learning, Memory, and Cognition, 1985.

50 Oiwa, Hiroshi; Matsushima, Shinpei; Nakagawa, Hiroshi. *Feature-aware regularization for sparse online learning. Sci. China Inf. Sci.*, 2014, 57, 1–21.

51 Xavier Glorot and Yoshua Bengio. *Understanding the difficulty of training deep feedforward neural networks.* In: International Conference on Artificial Intelligence and Statistics, 2010.

52 Shuanghong Shen, Qi Liu, Enhong Chen, Han Wu, Zhenya Huang, Weihao Zhao, Yu Su, Haiping Ma, and Shijin Wang. *Convolutional Knowledge Tracing: Modeling Individualization in Student Learning Process.* Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, 2020.

53 Hiromi Nakagawa, Yusuke Iwasawa, Yutaka Matsuo. *Graph-Based Knowledge Tracing: Modeling Student Proficiency Using Graph Neural Network.* In: IEEE/WIC/ACM International Conference on Web Intelligence, 2019.

54 Shiwei Tong, Qi Liu, Wei Huang, Zhenya Huang, Enhong Chen, Chuanren Liu, Haiping Ma, Shijin Wang. *Structure-Based Knowledge Tracing: An Influence Propagation View.* In: 2020 IEEE International Conference on Data Mining (ICDM).

55 Pan, Long; Shao, Huijuan; Xi, Yu; et al. *Bipartite Consensus Problem on Matrix-valued Weighted Directed Networks. Sci. China Inf. Sci.*, 2021.

56 Hermann Ebbinghaus. *Memory: A contribution to experimental psychology. Annals of Neurosciences*, vol. 20, no. 4, 2013.

57 Koki Nagatani, Qian Zhang, Masahiro Sato, Yan-Ying Chen, Francine Chen, Tomoko Ohkuma. *Augmenting knowledge tracing by considering forgetting behavior.* In: WWW, 2019.

58 Zhang, Xiaochuan; Qiu, Xipeng; Pang, Jianmin; Liu, Fudong; Li, Xingwei. *Dual-axial self-attention network for text classification. Sci. China Inf. Sci.*, 2021, 64.

59 Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin. *Attention is All You Need.* In: Advances in Neural Information Processing Systems 30, 2017.

60 S. Wang, Z. Zeng, X. Yang, and X. Zhang. *Self-Supervised Graph Learning for Long-Tailed Cognitive Diagnosis.* Proceedings of the AAAI Conference on Artificial Intelligence, 2023.

61 S. Wang, Z. Zeng, X. Yang, K. Xu and X. Zhang. *Boosting Neural Cognitive Diagnosis with Student's Affective State Modeling.* Proceedings of the AAAI Conference on Artificial Intelligence, 2024.

62 Yang, Yang; Qin, Na; Huang, Deqing; Fu, Yuanzhe. *Fault Diagnosis of High-Speed Railway Bogies Based on LSTM.* In: *2018 5th International Conference on Information, Cybernetics, and Computational Social Systems (ICCSS)*

63 Abdelrahman Ghodai, Qing Wang. *Deep Graph Memory Networks for Forgetting-Robust Knowledge Tracing. arXiv preprint arXiv:2108.08105*, 2021.

64 Vie, Jill-Jênn, and Kashima, Hisashi. *Knowledge Tracing Machines: Factorization Machines for Knowledge Tracing.* In The Thirty-Third Conference on Artificial Intelligence, The Thirty-First Innovative Applications of Artificial Intelligence Conference, The Ninth Symposium on Educational Advances in Artificial Intelligence, Honolulu, Hawaii, USA, 2019. AAAI.