

## 基于双层级对比学习的健壮知识追踪模型

陈晓岚, 毛舜, 李伟生, 林荣华, 汤庸

引用本文

陈晓岚, 毛舜, 李伟生, 林荣华, 汤庸. [基于双层级对比学习的健壮知识追踪模型](#)[J]. 计算机科学, 2026, 53(2): 31-38.

CHEN Xiaolan, MAO Shun, LI Weisheng, LIN Ronghua, TANG Yong. [Robust Knowledge Tracing Model Based on Two-level Contrastive Learning](#) [J]. Computer Science, 2026, 53(2): 31-38.

---

## 相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

### [基于用户静动态兴趣和去噪隐式负反馈的新闻推荐算法](#)

News Recommendation Algorithm Based on User Static and Dynamic Interests and Denoised Implicit Negative Feedback

计算机科学, 2026, 53(2): 152-160. <https://doi.org/10.11896/jsjcx.241200177>

### [融合对比学习的掩码图自编码器](#)

Contrastive Learning-based Masked Graph Autoencoder

计算机科学, 2026, 53(2): 145-151. <https://doi.org/10.11896/jsjcx.250100155>

### [DCL-FKT:融合双重对比学习与遗忘机制的个性化知识追踪模型](#)

DCL-FKT: Personalized Knowledge Tracing via Dual Contrastive Learning and Forgetting Mechanism

计算机科学, 2026, 53(2): 99-106. <https://doi.org/10.11896/jsjcx.250600002>

### [GTKT:融合联通主义学习和多层时序图Transformer的知识追踪模型](#)

GTKT: Knowledge Tracing Model Integrating Connectivism Learning and Multi-layer Temporal Graph Transformer

计算机科学, 2026, 53(2): 78-88. <https://doi.org/10.11896/jsjcx.250700188>

### [基于方向感知孪生网络的知识概念先序关系预测方法](#)

Direction-aware Siamese Network for Knowledge Concept Prerequisite Relation Prediction

计算机科学, 2026, 53(2): 39-47. <https://doi.org/10.11896/jsjcx.250600005>

# 基于双层级对比学习的健壮知识追踪模型

陈晓岚<sup>1,6</sup> 毛舜<sup>2</sup> 李伟生<sup>3</sup> 林荣华<sup>4</sup> 汤庸<sup>4,5</sup>

1 华南师范大学教育信息技术学院 广州 510631

2 广州航海学院人工智能学院 广州 510725

3 广东开放大学人工智能学院 广州 510091

4 华南师范大学计算机学院 广州 510631

5 广东科技学院数据智能研究院 广东 东莞 523083

6 广州市黄埔区荔园小学 广州 510700

(xlchen@m.scnu.edu.cn)

**摘要** 知识追踪是实现自适应学习的关键,它的目的是为了评估学生的知识状态并预测他们的未来表现。目前,数据的稀疏性问题使得现有的知识追踪模型在问题嵌入学习和学生知识状态模拟两个方面受到了限制。因此,一些研究引入了对比学习来缓解这一问题。然而,现有的对比学习方法依赖随机扰动图结构(用于问题嵌入)或修改学习交互序列(用于知识状态建模)生成对比视图,引入了噪声与错误自监督信号,导致问题嵌入无法良好适配学习系统下游任务。因此,提出一种创新的双层对比学习框架(Dual-level Contrastive Learning Framework, DCLF)用于同时增强知识追踪中问题的嵌入学习和学生知识状态模拟两个方面。DCLF采用了一种更有效的对比范式,这种对比范式的主要思想是不改变数据的原本信息,通过对原始数据进行关系变换或利用不同神经网络在同一数据上的输出作为对比视图。具体来说,在嵌入学习中,所提出的方法通过对数据进行关系变化,获得对比视图。在学生知识状态模拟上,所提模型使用不同的神经网络对学习交互进行编码,获得不同编码器下的知识状态。这种方法能够提取多个对比视图下丰富的自监督信号,保留数据的内在语义信息,有效避免噪声的引入。在3个常用的知识追踪数据集上进行实验验证,实验结果表明DCLF在性能上优于现有知识追踪模型。

**关键词:** 知识追踪; 自适应学习; 对比学习; 对比范式; 自监督学习

**中图分类号** TP181

## Robust Knowledge Tracing Model Based on Two-level Contrastive Learning

CHEN Xiaolan<sup>1,6</sup>, MAO Shun<sup>2</sup>, LI Weisheng<sup>3</sup>, LIN Ronghua<sup>4</sup> and TANG Yong<sup>4,5</sup>

1 School of Information Technology in Education, South China Normal University, Guangzhou 510631, China

2 School of Artificial Intelligence, Guangzhou Maritime University, Guangzhou 510725, China

3 School of Artificial Intelligence, Guangdong Open University, Guangzhou 510091, China

4 School of Computer Science, South China Normal University, Guangzhou 510631, China

5 Institute of Data Intelligence, Guangdong University of Science and Technology, Dongguan, Guangdong 523083, China

6 Liyuan Primary School, Huangpu District, Guangzhou 510700, China

**Abstract** Knowledge tracing is key to adaptive learning, aiming to assess students' knowledge states and predict their future performance. Currently, data sparsity limits existing knowledge tracing models in both question embedding learning and student knowledge state modeling. To address this, some studies have introduced contrastive learning. However, existing contrastive learning methods rely on random perturbations of graph structures(for question embedding) or modifications of learning interaction sequences(for knowledge state modeling) to generate contrastive views, introducing noise and erroneous self-supervised signals. This results in question embeddings that are poorly suited for downstream tasks in learning systems. To overcome these limitations, this study proposes an innovative Dual-level Contrastive Learning Framework(DCLF) to simultaneously enhance question embedding learning and student knowledge state modeling in knowledge tracing. DCLF adopts a more effective contrastive paradigm that avoids altering the original data information. Instead, it generates contrastive views through relational transformations of the original data or by leveraging outputs from different neural networks on the same data. Specifically, for embedding

到稿日期:2025-07-31 返修日期:2025-10-29

基金项目:国家自然科学基金(62407016)

This work was supported by the National Natural Science Foundation of China(62407016).

通信作者:林荣华(rhlin@m.scnu.edu.cn)

learning, the proposed method obtains contrastive views through relational transformations of the data. For student knowledge state modeling, it encodes learning interactions using different neural networks to obtain knowledge states under various encoders. This method extracts rich self-supervised signals from multiple contrastive views, preserving the intrinsic semantic information of the data and effectively avoiding noise introduction. Experiments conducted on three commonly used datasets demonstrate that DCLF outperforms selected existing knowledge tracing models in terms of performance.

**Keywords** Knowledge tracing, Adaptive learning, Contrastive learning, Contrastive paradigm, Self supervised learning

## 1 引言

知识追踪 (Knowledge Tracing, KT) 在智能辅导系统 (Intelligent Tutoring Systems, ITS) 中始终扮演着重要角色, 为学习资源推荐提供了有效支持<sup>[1]</sup>。知识追踪通常利用学习交互, 包括所尝试的问题及其回答得分, 来评估学生的知识状态并预测其未来表现。其中, 获取的知识状态是支持 ITS 中自适应学习的关键信息。目前, 基于深度学习的 KT 模型已经取得了显著进展, 并已应用于多个在线教育平台, 如 Khan Academy, MOOC 以及学者网。

在知识追踪领域, 许多先进的深度学习方法被引入以推动其发展。其中, DKT<sup>[2]</sup>, DKVMN<sup>[3]</sup>, SAKT<sup>[4]</sup> 以及 GKT<sup>[5]</sup> 是最具代表性的知识追踪模型。随后, 研究者对这些模型进行了更深入的探索。然而, 这些模型采用的都是监督学习的方法, 其监督信号完全来源于学习交互历史, 因此可能面临教育数据的稀疏性问题。具体而言, 学生交互的问题仅占题库的一小部分。在训练过程中, KT 模型只能学习这一小部分题库中的信息, 难以优化高质量的问题嵌入和知识状态, 最终阻碍了知识追踪模型的预测精度的提升。

通常, 知识追踪模型在预测学生未来的表现上依赖于学生的知识状态及所预测的具体问题。知识状态和问题嵌入通常基于现有的交互数据进行训练。如前所述, 教育数据的稀疏性限制了模型在知识状态建模和问题嵌入学习过程中捕获更丰富信息的能力。因此, 对比学习 (Contrastive Learning, CL) 方法应运而生, 它作为一种自监督方法<sup>[6]</sup>, 通过学习样本间的相似性和差异性来获得更有意义的嵌入向量, 从而解决教育数据稀疏性问题<sup>[7]</sup>。

然而, 目前的一些工作仍然存在以下问题。第一, 以往模型在问题嵌入学习中使用对比学习时会引入大量噪声。其主要思路是通过定义的正负样本对, 利用生成的图嵌入视图之间的一致性进行对比。为了构建对比视图, Bi-CLKT<sup>[8]</sup> 和 S<sup>2</sup>-HHN<sup>[9]</sup> 对图结构数据 (如由问题和技能组成的问题-技能图) 进行随机扰动以生成节点对。这种扰动重构可能引入噪声信息。第二, 过去的知识追踪模型在知识状态建模中使用的对比学习方法可能引入错误信息, 误导知识状态的准确表示。其主要思路是通过修改学习交互 (如问题替换、交互顺序打乱和交互裁剪) 生成对比视图。然而, 这种方法会扰乱原始学习交互, 可能引入错误的自监督信号。

综上所述, 无论在问题嵌入学习还是知识状态建模研究中, 所有使用对比学习方法生成对比视图的模型都依赖于对原始数据的随机扰动。这可能导致关键信息丢失, 从而误导问题嵌入学习和知识状态建模。为了解决上述问题, 本文开发了一个双层对比学习框架 (Dual-level Contrastive Learning

Framework, DCLF), 分别展示了嵌入层面和知识层面的对比学习方法。不同于以往的随机扰动方法, DCLF 通过多视角捕获潜在语义信息, 从而避免改变原始数据结构。具体而言, 针对第一个问题, 问题和技能之间的关系可以构成一种异质超图<sup>[10]</sup>, 由于存在两种不同的关系, 这个超图可以分裂成两个子图, 实际上这个超图的结构信息不变, 故可以从不同视角观察问题和技能的嵌入表示。通过对比两种图视图, 自监督信号通过结构转换而非随机扰动生成, 从而提供多样且有意义的对比信息。针对第二个问题, DCLF 采用两种不同的神经网络结构对学生学习交互进行编码, 生成不同的知识状态, 从而提供多视角对比。这一方法完整保留了原始学习交互序列的语义信息, 同时引入不同编码视角下的嵌入信息。

总之, 本文模型在知识追踪领域中对对比学习进行了全面探索, 通过嵌入层面和知识层面的两个层次的对比机制, 提升了问题嵌入学习和知识状态建模过程中数据的信息量。本文的主要贡献如下:

1) 提出了一个双层级的对比学习框架 DCLF。该框架同时在问题嵌入层面和知识状态层面提升表示质量, 以减少教育数据稀疏性带来的不利影响。在此过程中, 本文引入了一种结构保持的对比学习范式替代传统的随机扰动方法, 以确保生成的对比视图与原始数据的语义一致性。

2) 在 3 个常用数据集上的广泛实验表明, DCLF 通过所提出的对比范式, 在两个层面都具有有效性。

## 2 相关工作

本章首先介绍基于深度学习的知识追踪模型, 然后讨论对比学习在知识追踪中的应用。

知识追踪旨在建模学生的知识状态, 以预测他们的未来表现<sup>[11]</sup>。这一领域的早期突破之一是 DKT<sup>[2]</sup>, 该模型利用了循环神经网络 (RNN) 对学生-问题的交互序列进行建模。此后, 其他模型相继出现, 如 DKVMN<sup>[3]</sup>, SAKT<sup>[4]</sup> 以及 GKT<sup>[5]</sup>, 研究人员探索了多种神经网络架构。Liu 等<sup>[12]</sup> 对基于深度神经网络的知识追踪方法和模型进行了综述。此外, 许多研究致力于引入辅助信息来增强这些基础模型, 如引入时间特征<sup>[13-14]</sup>、问题特征<sup>[14-15]</sup> 和教育理论<sup>[16-17]</sup>。尽管这些模型取得了一些进展, 但仍然面临教育数据稀疏性带来的挑战, 这限制了它们对学生未来表现的准确预测能力。

最近, 基于自监督方法的对比学习被用于研究知识追踪领域<sup>[18]</sup>。这些模型主要是利用对比学习优化问题嵌入或知识状态。Zhao 等<sup>[19]</sup> 针对现有知识追踪方法中存在的学生-问题交互数据稀疏等问题, 提出了一种基于双级对比学习的问题表征方法, 其改进知识追踪并有效预测学生的未来表现。Zhang 等<sup>[20]</sup> 提出了以问题为中心的多专家对比学习框架对

问题进行显式建模,以解决交互数据稀疏导致的问题嵌入过拟合等问题。

在优化问题嵌入的知识追踪方法中,Bi-CLKT<sup>[8]</sup>和S<sup>2</sup>-HHN<sup>[9]</sup>使用了相同的对比方法,即随机扰动图数据以构造对比视图,从而增强问题嵌入。不同之处在于,S<sup>2</sup>-HHN<sup>[9]</sup>使用的图数据更加复杂,它采用由模板、问题和技能组成的超图来执行对比学习。在优化知识状态方面,CLAKT<sup>[21]</sup>和CCKT<sup>[22]</sup>通过数据增强技术(如随机替换问题或去除问题)来增强学生的知识状态。SPCLKT<sup>[23]</sup>采用了类似的方法,但其与CLAKT的不同之处在于,它基于问题难度替换问题,而非随机替换。

总之,过去的方法在解决教育数据稀疏性方面的效果往往有限。对比学习的出现为这一问题引入了新的解决方案。然而,目前的方法所依赖的随机扰动的数据增强方式可能会破坏原始教育数据中的重要信息,可能会误导模型的训练。本文提出的DCLF采用了一种不同的策略,能够保留原始教育数据的语义信息,提取更强健的自监督信号,从而帮助生成高质量的问题嵌入,并提升知识状态建模的准确性。

### 3 准备工作

#### 3.1 问题定义

在介绍具体的方法之前,首先对知识追踪进行形式化的定义。具体来说,每位学生的学习交互可表示为 $Z^s = (z_1^s, z_2^s, \dots, z_n^s)$ ,其中 $z_t^s$ 表示学生 $s$ 在时间步 $t$ 的交互。每个交互 $z_t^s$ 是一个元组 $(q_{i,t}^s, p_i^s)$ ,其中 $q_{i,t}^s$ 表示学生在时间步 $t$ 与问题 $q_i$ 的交互。如果学生正确回答了问题 $q_{i,t}^s$ ,则 $p_i^s$ 为1,否则为0。基于学生的历史学习交互,知识追踪的目的是评估其知识状态。为了验证知识状态评估的有效性,一般做法是将知识状态作为依据,预测学生在未来的表现。

#### 3.2 对比视图构建

为了获得有效的问题嵌入,在本文方法中构建了一个

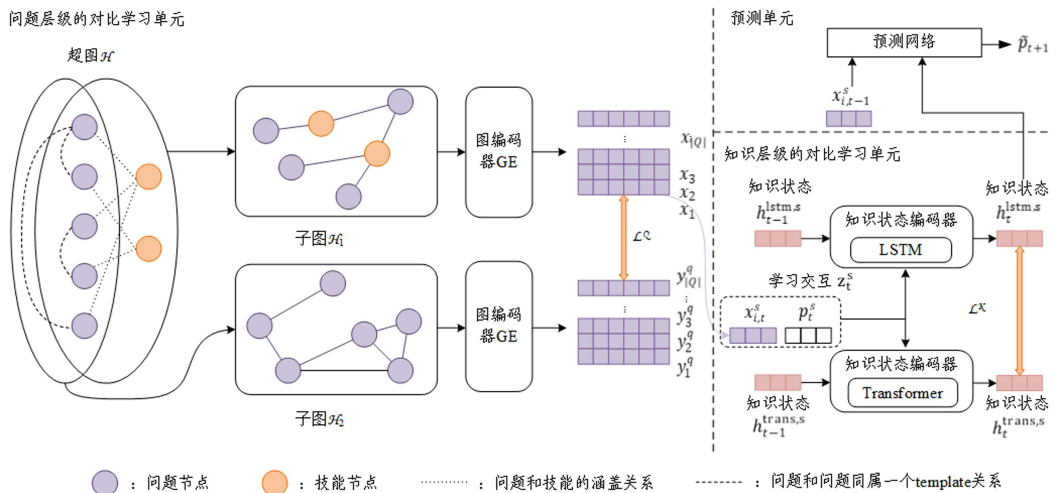


图1 DCLF框架

Fig. 1 Framework of DCLF

#### 4.1 问题嵌入层级的对比学习单元

##### 4.1.1 图编码器

为了充分挖掘图结构中蕴含的丰富信息,本文引入了一个图编码器,用于编码基于图结构数据的特征。由于近年来

超图,用于表示问题和技能之间的关系。设 $q \in Q$ 表示问题集合, $k \in K$ 表示技能集合。可以用数学公式的形式定义超图 $H$ ,具体来说,一个超图 $H = \{Q, K, E\}$ 包含以下3个部分:

顶点集: $Q = \{q_1, q_2, \dots, q_n\}$ 表示问题集合, $K = \{k_1, k_2, \dots, k_n\}$ 表示技能集合。边集: $E = E_q \cup E_{qk}$ ,其中 $E_q = \{e_q \mid e_q \subseteq E_q\}$ 表示问题之间的关系边。若 $q_i$ 和 $q_j$ 属于同一模板(Template),则 $\{q_i, q_j\} \in E_q$ 。这里还给出表示问题与技能之间的关系边: $E_{qk} = \{e_{qk} \mid e_{qk} = q, k, q \in Q, k \in K\}$ 。故在此图中存在两个附加约束:1) $e_{qk} \in E_{qk}$ , $q$ 与 $k$ 之间的关系在数据集中已知;2) $e_q \in E_q, q_i, q_j \in e_q \Leftrightarrow q_i, q_j$ 属于同一模板(在数据集中也是已知的)。

通过上述定义,超图 $H$ 将问题和技能组织在一起,并通过问题之间的模板关系 $E_q$ 和问题与技能之间的关联关系 $E_{qk}$ 表示两种类型的边。在后面的内容中, $H_1$ 和 $H_2$ 是分别由超图 $H$ 生成的子图,其中 $H_1$ 是具有问题与技能之间的关联关系 $E_{qk}$ , $H_2$ 是具有问题之间的模板关系 $E_q$ 。

### 4 具体方法

本文提出的DCLF的结构如图1所示,由以下4个关键部分组成。1)问题嵌入层级的对比学习单元:该层利用两种图视图,捕捉问题的共性和个性信号,从而生成有效的问题嵌入。2)知识状态层级的对比学习单元:该层通过两个知识编码器从学生的历史学习交互中获得两个知识状态,并引入交叉对比机制。3)预测单元:利用获得的知识状态预测学生在下一步的表现。4)联合优化:从上述3个单元中计算出问题嵌入层级的损失、知识状态层级的损失和预测损失,并将这些损失整合到模型的整体优化过程中。在正式介绍方法之前,首先初始化问题嵌入矩阵 $E^q \in \mathbf{R}^{Q \times d}$ 和技能嵌入矩阵 $E^k \in \mathbf{R}^{K \times d}$ ,其中 $d$ 表示嵌入维度。

图卷积网络(Graph Convolutional Networks,GCN)<sup>[24]</sup>取得显著进展,因此采用原始的GCN模型作为图编码器(Graph Encoder,GE):

$$Y = GE(E, A) = GCN(E, A) = D^{-\frac{1}{2}} A D^{-\frac{1}{2}} E W_1 \quad (1)$$

其中,  $\mathbf{E}$  和  $\mathbf{A}$  分别表示图视图的嵌入矩阵和邻接矩阵,  $\mathbf{D}$  表示对角度矩阵。

#### 4.1.2 问题嵌入层的对比机制

为了联合子图  $H_1$  和  $H_2$  增强问题嵌入, 应用对比学习, 从  $H_1$  和  $H_2$  中提取自监督信号, 从而获得增强问题嵌入的判别性信息。首先, 通过 GCN 从这些图中获取嵌入结果, 公式如下:

$$\mathbf{Y}^{q^k} = GE((\mathbf{E}^q, \mathbf{E}^k), H_2) \quad (2)$$

$$\mathbf{Y}^q = GE(\mathbf{E}^q, H_2) \quad (3)$$

其中,  $\mathbf{Y}^q = (\mathbf{y}_1^q, \mathbf{y}_2^q, \dots, \mathbf{y}_{|Q|}^q)$  表示问题特征矩阵,  $\mathbf{Y}^{q^k}$  表示包含问题嵌入和技能嵌入的矩阵。可以根据索引从  $\mathbf{Y}^{q^k}$  中提取问题嵌入, 得到  $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{|Q|})$ 。可以看出, 问题节点被融入到了两个不同的图中。这种设计使我们能够从这两个图中为每个问题提取判别性信息作为自监督信号, 这些信息能够有效补充问题的嵌入表示。在此过程中, 来自两个图视图中相同问题的特征被视为正样本对, 而不同问题的特征则被视为负样本对:

$$L^Q = \frac{1}{|Q|} \sum_{i=1}^{|Q|} \log \frac{\gamma(\mathbf{x}_i, \mathbf{y}_i^q)}{\sum_j \gamma(\mathbf{x}_i, \mathbf{y}_j^q) + \sum_{j \neq i} \gamma(\mathbf{x}_i, \mathbf{x}_j)} \quad (4)$$

其中,  $\gamma(\cdot, \cdot) = \exp(\text{sim}(\cdot, \cdot) \xi)$ ,  $\text{sim}(\cdot, \cdot)$  表示计算余弦相似度,  $\xi$  是温度参数。

#### 4.2 知识嵌入层级的对比学习单元

与以往研究注重通过单个学习交互的随机数据增强不同, DCLF 利用多个知识编码器的多样化视图来丰富知识状态。DCLF 中, 学生的知识状态通过两个知识编码器获得, 作为额外的自监督信号。同时, 在对比机制中, 不同的学生被视为负样本对。

##### 4.2.1 知识状态编码器

知识状态编码器旨在获取用于预测的知识状态。由于 RNN 在知识追踪领域中被广泛应用, 并取得了显著成果, 故在 DCLF 中采用其变体长短时记忆网络 (Long Short-Term Memory, LSTM)<sup>[25]</sup> 来建模学生的知识状态。

给定学生的一段交互序列, 首先通过索引选择获取其嵌入表示  $(\mathbf{z}_1^s, \mathbf{z}_2^s, \dots, \mathbf{z}_i^s)$ , 其中  $\mathbf{z}_i^s = \delta(\mathbf{x}_{i,t}^s, \mathbf{p}_i^s)$  表示学习交互  $\mathbf{z}_i^s$  的嵌入表示,  $\delta(\cdot, \cdot)$  表示拼接操作,  $\mathbf{p}_i^s \in \{0\}^d$ 。如果学生  $s$  正确回答了问题  $q_{i,t}^s$ , 则  $\delta(\mathbf{x}_{i,t}^s, \mathbf{p}_i^s) = [\mathbf{x}_{i,t}^s, \mathbf{p}_i^s]$ ; 否则,  $\delta(\mathbf{x}_{i,t}^s, \mathbf{p}_i^s) = [\mathbf{p}_i^s, \mathbf{x}_{i,t}^s]$ 。故知识状态可以通过以下方式获得:

$$\mathbf{h}_i^{\text{LSTM},s} = \text{LSTM}(\mathbf{h}_{i-1}^{\text{LSTM},s}, \mathbf{z}_i^s) \quad (5)$$

其中,  $\mathbf{h}_i^{\text{LSTM},s} \in \mathbf{R}^{2d}$  表示学生  $s$  在时间步  $t$  的知识状态。

为了获得知识状态层级的对比知识状态, 以及利用自注意力机制的 KT 模型在知识追踪领域取得了显著进展, DCLF 结合了基于 Transformer 的知识编码器:

$$\mathbf{h}_i^{\text{trans},s} = \text{Transformer}(\mathbf{h}_{i-1}^{\text{trans},s}, \mathbf{z}_i^s) \quad (6)$$

该知识编码器生成的知识状态  $\mathbf{h}_i^{\text{trans},s} \in \mathbf{R}^{2d}$ 。

##### 4.2.2 知识状态层级的对比机制

基于上述两个知识编码器, 学生在时间步  $t$  的知识状态由两个向量表示, 每个向量来自不同的知识编码器。这两个编码器视图中提取的判别性信息作为自监督信号, 帮助提高知识状态建模的准确性。为此, 我们首先将这些状态映射到

一个向量空间, 并使用前馈网络计算对比损失:

$$\mathbf{H}_i^{\text{LSTM},s} = \mathbf{W}_2 \cdot \mathbf{h}_i^{\text{LSTM},s} + \mathbf{b}_2 \quad (7)$$

$$\mathbf{H}_i^{\text{trans},s} = \mathbf{W}_2 \cdot \mathbf{h}_i^{\text{trans},s} + \mathbf{b}_2$$

然后, 同一学生的来自两个知识编码器的知识状态被视为一对正样本, 不同学生的知识状态则被视为一对负样本。计算知识级别的对比损失如下:

$$L^K = -\frac{1}{S} \sum_{s=1}^S \frac{1}{T} \sum_{t=1}^T \log \frac{\gamma(\mathbf{H}_i^{\text{LSTM},s}, \mathbf{H}_i^{\text{trans},s})}{\sum_l \gamma(\mathbf{H}_i^{\text{LSTM},s}, \mathbf{H}_i^{\text{trans},(l)}) + \sum_{l \neq i} \gamma(\mathbf{H}_i^{\text{LSTM},s}, \mathbf{H}_i^{\text{LSTM},(l)})} \quad (8)$$

其中,  $S$  表示每个批次中的学生数量,  $l$  表示批次中除学生  $S$  外的其他学生,  $T$  表示学生  $s$  尝试过的问题数量。

#### 4.3 预测单元

为了预测学生未来的表现, DCLF 将需要预测的问题与当前的知识状态相结合, 以获得预测结果:

$$\mathbf{f}_{t+1} = \text{ReLU}(\mathbf{W}_3 [\mathbf{h}_t^{\text{LSTM},s}, \mathbf{x}_{t,t+1}^s] + \mathbf{b}_3) \quad (9)$$

$$\tilde{p}_{t+1} = \sigma(\mathbf{W}_4 \cdot \mathbf{f}_{t+1} + \mathbf{b}_4)$$

其中,  $\mathbf{W}_3 \in \mathbf{R}^{3d \times d}$ ,  $\mathbf{W}_4 \in \mathbf{R}^{d \times 1}$ ,  $\mathbf{b}_3 \in \mathbf{R}^{d \times 1}$  和  $\mathbf{b}_4 \in \mathbf{R}^1$  是可训练参数;  $\tilde{p}_{t+1}$  表示预测的表现。最后, 交叉熵被用作目标函数来优化预测结果:

$$L^P = -(p_t \log \tilde{p}_t + (1 - p_t) \log(1 - \tilde{p}_t)) \quad (10)$$

#### 4.4 联合优化

基于上述讨论, 可以获得来自 3 个单元的各自的损失值, 即联合优化的总体目标函数定义为问题层级的对比损失  $L^Q$ 、知识状态层级的对比损失  $L^K$  和预测损失  $L^P$  的线性组合, 具体如下:

$$L = L^P + \lambda_1 L^Q + \lambda_2 L^K \quad (11)$$

其中,  $\lambda_1$  和  $\lambda_2$  是超参数, 用于控制每个损失项在总损失中的权重。

## 5 实验及结果分析

为了评估 DCLF 的有效性, 本文在 3 个广泛使用的知识追踪数据集上进行了实验。首先概述基准模型和实验设置, 然后讨论实验结果, 以展示模型的有效性。

### 5.1 数据集介绍和实验设置

本文实验在 ASSISTments2009, EdNet 和 Algebra05 这 3 个经典的数据集上进行, 以验证模型的性能<sup>[26]</sup>。这些数据集均为知识追踪领域常用的数据, 具体的统计信息如表 1 所列。值得注意的是, 在 EdNet 数据集中, 遵循之前工作的方法, 随机抽取了 5000 名学生的学习交互信息进行实验。

表 1 数据集统计

Table 1 Dataset statistics

	ASSIST09	EdNet	Algebra05
学习交互	110 200	394 900	809 700
问题数量	16 900	11 400	1 100
技能数量	111	188	138
学生数量	3 700	5 000	574

对于实验设置, 本文采用了 Area Under Curve (AUC) 和 Root Mean Square Error (RMSE) 作为评估指标。同时, 采用

标准的  $k$  折交叉验证( $k=5$ )展示平均实验结果。在每折交叉验证中,学生数据按照 60%,20%,20% 的比例划分为训练集、验证集和测试集。设置嵌入维度  $d=128$ ,批量大小  $L=32$ ,学习率为 0.001,最大序列长度为 200。此外,超参数 $\lambda_1$ 和 $\lambda_2$ 的候选值设定为 $\{0.01,0.02,0.05,0.1,0.5,1.0,2.0\}$ 。使用 Adam 优化器来最小化损失函数  $L$ 。所有模型均在一台具有 NVIDIA RTX 4090 显卡的 Linux 服务器上,利用 PyTorch 实现。

## 5.2 对比模型

为了验证 DCLF 的有效性,选择以下知识追踪模型作为对比模型。

1)DKT<sup>[2]</sup>:基于深度学习的开创性工作,本文用 LSTM 网络构建 DKT 模型。

2)DKVMN<sup>[3]</sup>:一个经典的 KT 模型,使用改进的内存增强神经网络来建模知识状态。

3)DKT+<sup>[27]</sup>:在 DKT 基础上引入了两个正则化项,以改善预测的一致性。

4)SAKT<sup>[4]</sup>:一个典型的基于自注意力机制的知识追踪模型。

5)DKT+Forgetting<sup>[12]</sup>:在 DKT 中引入了学生遗忘行为的建模,包括序列时间间隔、重复时间间隔和过去尝试次数。

6)AKT<sup>[13]</sup>:另一个以自注意力机制为核心的知识追踪模型,实现了问题层面的预测。

7)GIKT<sup>[14]</sup>:首次引入图卷积网络来获取问题的嵌入向量。

8)Bi-CLKT<sup>[8]</sup>:首次使用了基于随机扰动的对比学习来增强问题嵌入。

9)CLAKT<sup>[18]</sup>:其构建新的学习交互作为对比学习中的对比视图,以改善知识状态嵌入的准确性。

10) $S^2$ -HHN<sup>[9]</sup>:构建了一个问题和技能组成的异质超图并结合基于随机扰动的对比学习来增强问题嵌入。

11)CCKT<sup>[22]</sup>:通过设置能力与知识共注意力层以模拟二者相互影响,同时引入对比学习丰富数据表征并提升知识状态追踪可靠性。

12)HKT<sup>[28]</sup>:通过构建技能层次树来建模跨层关系,构建技能关联图来建模同层关系,并结合跨图层次对比学习提升表示能力。同时,设计权重分配网络动态融合不同层次的知识影响。

## 5.3 性能比较

实验对比了 3 个数据集上所有基线模型在 AUC 和 RMSE 指标上的表现,结果如表 2 和表 3 所列。从表 2 和表 3 中可以得出以下关键结论。

1)技能层级的知识追踪模型(如 DKT,DKVMN,SAKT 等)的性能远不如基于问题层级的 KT 模型(如 GIKT,Bi-CLKT, $S^2$ -HHN 等)。这也意味着针对问题嵌入的研究已成为推动知识追踪领域进步的关键方向。虽然基于问题层级的模型采用了不同的方法获取问题嵌入,但其核心思想都是围绕通过引入问题的相关信息来获取和优化问题嵌入。因此,问题嵌入的研究至关重要<sup>[29]</sup>。

表 2 AUC 指标下的实验结果

Table 2 Experiment results under AUC metric

Models	ASSIST09	EdNet	Algebra05
DKT	0.7521	0.6278	0.7638
DKVMN	0.7363	0.6153	0.7873
DKT+	0.7545	0.6178	0.7643
SAKT	0.6961	0.6025	0.7868
DKT+Forgetting	0.7576	0.6491	0.7681
AKT	0.7533	0.6477	0.7760
GIKT	0.7690	0.6816	0.7854
Bi-CLKT	0.7758	0.6885	0.7932
CLAKT	0.7624	0.6643	0.7893
$S^2$ -HHN	0.7738	0.6893	0.7916
CCKT	0.7814	0.6955	0.8047
HKT	0.7829	0.6914	0.8015
DCLF	<b>0.7879</b>	<b>0.7042</b>	<b>0.8089</b>

表 3 RMSE 指标下的实验结果

Table 3 Experiment results under RMSE metric

Models	ASSIST09	EdNet	Algebra05
DKT	0.4304	0.4813	0.3795
DKVMN	0.4377	0.4864	0.3718
DKT+	0.4287	0.4802	0.3786
SAKT	0.4511	0.4893	0.3791
DKT+Forgetting	0.4356	0.4791	0.3773
AKT	0.4292	0.4752	0.3465
GIKT	0.4217	0.4693	0.3422
Bi-CLKT	0.4236	0.4678	0.3415
CLAKT	0.4271	0.4719	0.3407
$S^2$ -HHN	0.4223	0.4691	0.3381
CCKT	0.4156	0.4667	0.3358
HKT	0.4138	0.4681	0.3367
DCLF	<b>0.4106</b>	<b>0.4618</b>	<b>0.3327</b>

2)DKT+Forgetting 和 CLAKT 在性能比较中的表现也不错。尽管这些模型也是基于问题层级的知识追踪模型,但它们主要依赖已有方法获取练习嵌入,更注重的是知识状态的建模。这表明,虽然知识状态建模方面的进展有助于知识追踪领域的发展,但其影响不如通过优化问题嵌入所取得的提升显著。尽管如此,探索知识状态建模仍然是一个有效且有价值的研究方向<sup>[30]</sup>。

3)在 3 个数据集上,DCLF 明显优于所有基线模型,尤其是在 EdNet 数据集上的表现尤为突出。这种优势不仅归功于本文对问题嵌入学习和知识状态建模这两个关键过程的深入探索,还得益于引入了一种新颖的对比范式<sup>[31]</sup>。与传统对比学习方法依赖随机扰动生成对比视图不同,这些方法通常会引入噪声信息,而本文方法在避免引入噪声信息的同时能有效获取可靠的自监督信号。因此,DCLF 实现了高质量的问题嵌入和知识状态<sup>[32]</sup>。

## 5.4 消融实验

本节进行了消融实验,以观察 DCLF 各个组成部分的有效性。在实验中,对完整模型及其 6 个变体进行比较。

1)DCLF-EK:移除两个对比学习层,直接使用随机初始化的技能嵌入进行预测。

2)DCLF-E:仅移除嵌入层级的对比学习层。

3)DCLF-K:移除知识层级的对比学习层。

4)DCLF-CL:同时移除嵌入层级和知识层级的对比机制。

5) DCLF-ECL: 移除嵌入层级的对比机制。

6) DCLF-KCL: 移除知识层级的对比机制。

消融实验结果如图 2 所示, 可观察得出以下结果。

1) 在引入对比机制后, 模型的性能显著提升。这表明, 模型捕获的自监督信号有助于提升问题嵌入和知识状态的质量。

2) 两个对比学习层均对提高预测准确性有所贡献, 特别

是嵌入层级的对比学习层的影响更为显著。这表明, 引入不同的图视图能够提供更丰富的信息, 从而作为有价值的自监督信号。

3) 随着学生数量的增加, 知识层级的对比学习层在 EdNet 数据集上表现出明显优势。这表明, 利用知识编码器提供的多视角信息, 有助于显著提升知识状态建模的效果。

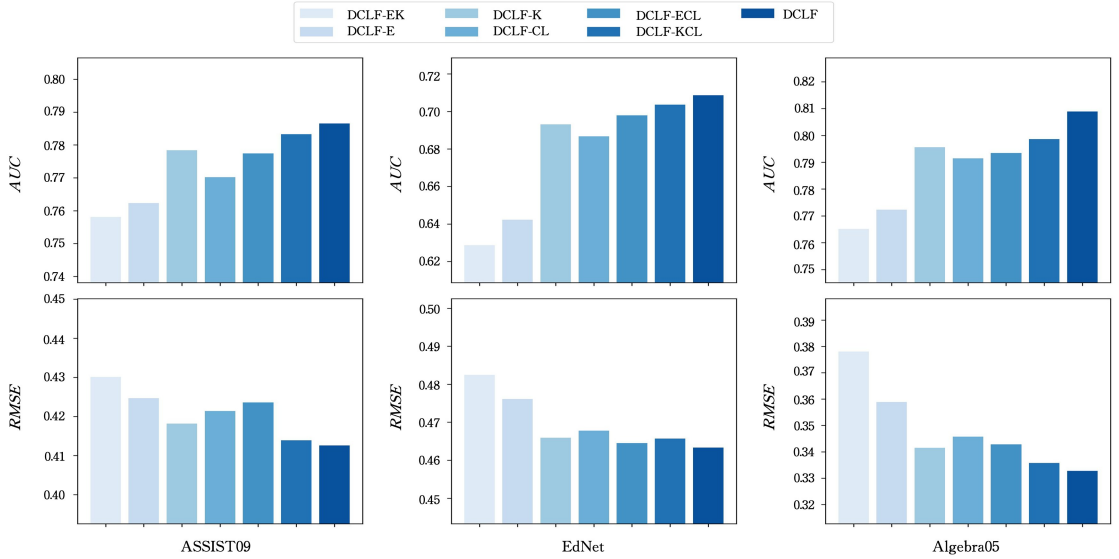


图 2 3 个数据集下的消融实验结果

Fig. 2 Results of ablation study on three datasets

## 5.5 超参数 $d$ 的分析

图 3 展示了 DCLF, Bi-CLKT 和 GIKT 这 3 个模型在

3 个数据集上的不同维度  $d \in \{64, 128, 256, 512\}$  下的 AUC 和 RMSE 表现。

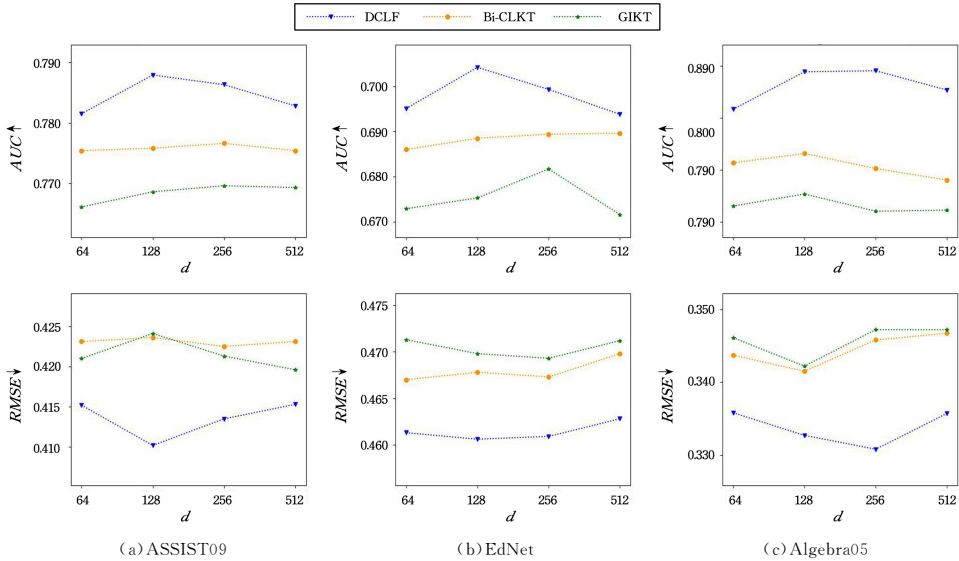


图 3 嵌入维度  $d$  的分析

Fig. 3 Analysis of embedding dimension  $d$

从 AUC 指标来看, DCLF 模型在 3 个数据集上的表现始终优于 Bi-CLKT 和 GIKT, 且在维度为 128 时达到最优。Bi-CLKT 的 AUC 表现较为稳定, 而 GIKT 的 AUC 随维度增加波动较大, 尤其在 EdNet 数据集上表现相对较差。

在 RMSE 指标方面, DCLF 同样表现出较低的误差, 尤其在 ASSIST09 和 Algebra05 数据集上, 随着维度增加, RMSE

呈下降趋势, 分别在维度为 128 和 256 时达到最小值。相比之下, Bi-CLKT 和 GIKT 的 RMSE 变化不大, 整体误差水平略高于 DCLF。

综上所述, DCLF 模型在不同维度下均展现出较好的鲁棒性和优越的性能, 尤其在中等维度时取得了最佳表现, 说明其对维度变化具有较强的适应能力。

## 5.6 效率分析

如图4所示,本文还进行了效率分析,选择了5个相关性较高的模型进行对比,展示了6个模型在运行时间(min)和内存占用(GB)上的性能比较。

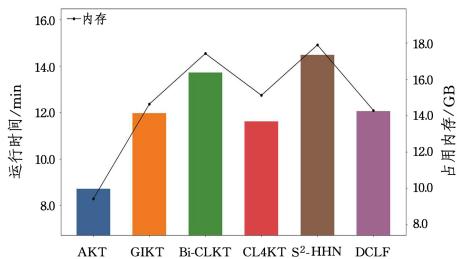


图4 ASSIST09数据集上的效率分析

Fig.4 Efficiency analysis on ASSIST09 dataset

从柱状图可以看出,模型S<sup>2</sup>-HHN的运行时间最长,接近15min同时其内存占用也最高,接近18GB。这主要是因为S<sup>2</sup>-HHN采用了大规模的图数据进行实验,这一做法通常会消耗较大的内存和时间。AKT模型的运行时间最短,不到

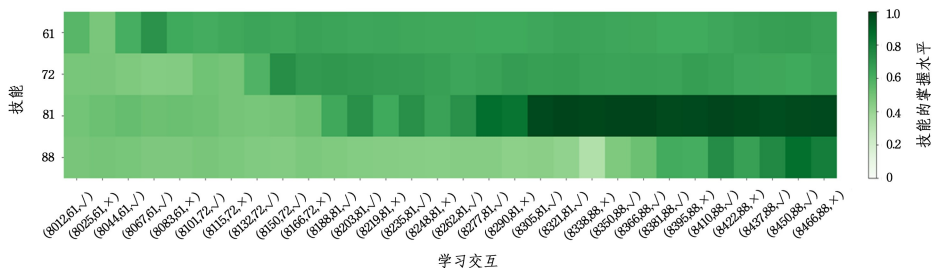


图5 知识状态可视化案例

Fig.5 Case study of knowledge state visualization

从图6中可以看出,模型不仅能够区分不同技能的掌握差异,还能捕捉学习者在具体交互下的知识状态变化。例如,随着学习者不断地正确回答与技能81相关的习题,技能81的掌握水平随着学习交互的增加逐渐加深,呈现出稳定的提升趋势;而从整体上看,技能72和技能88则表现出一定的波动,反映了学习者在这些技能上掌握过程的复杂性。特别是在答错交互后,模型能够及时体现掌握度的下降,这说明其预测结果具有较强的合理性和可解释性。

相比于仅给出预测正确率的模型,本文方法能够在技能层面揭示“学习者在哪些技能上已经掌握良好,在哪些技能上仍存在不足”。这种可解释性不仅增强了模型结果的可信度,也为教育实践提供了有价值的参考:教师可以据此识别薄弱技能并进行有针对性的教学干预,学习者也能够更清晰地了解自身的学习状态。

因此,该可视化结果表明,所提出的模型不仅具备较强的预测能力,而且在可解释性上具有明显优势,能够更好地支持个性化教学与学习决策。

**结束语** 本研究提出了一种创新的双层对比学习框架(DCLF),以缓解数据稀疏性对知识追踪任务带来的挑战。DCLF通过构建语义保留的对比视图,从多个层面增强了问题嵌入学习与学生知识状态建模的能力。与传统基于随机扰动的对比方法不同,DCLF利用关系变换和网络输出构造更稳定有效的自监督信号,显著降低了噪声干扰。实验结果验

证了该方法在多个基准数据集上优于现有先进模型的性能表现,展示出其在提升知识追踪准确性和泛化能力方面的潜力。

## 5.7 可解释性分析

为了更好地说明DCLF的可解释性,本文采用过去工作提到的方法<sup>[3]</sup>,从ASSIST09数据集中随机选取一名学生的交互过程,将学生的知识状态进行可视化,如图5所示。该图展示了不同技能在学习交互过程中的掌握水平变化情况。横轴表示学习交互(问题,技能,回答正误),纵轴表示技能编号,颜色深浅对应模型预测的技能掌握水平。通过这种可视化方式,可以直观地观察模型在个体学习过程中的动态追踪效果。

证了该方法在多个基准数据集上优于现有先进模型的性能表现,展示出其在提升知识追踪准确性和泛化能力方面的潜力。

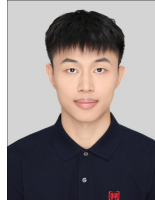
## 参考文献

- [1] ABDELRAHMAN G, WANG Q, NUNES B. Knowledge Tracing: A Survey [J]. ACM Computing Surveys, 2023, 55(11): 1-37.
- [2] PIECH C, BASSEN J, HUANG J, et al. Deep Knowledge Tracing [C]// Proceedings of Advances in Neural Information Processing Systems. Curran Associates Inc., 2015.
- [3] ZHANG J, SHI X, KING I, et al. Dynamic Key-Value Memory Networks for Knowledge Tracing [C]// Proceedings of the 26th International Conference on World Wide Web. New York: ACM, 2017: 765-774.
- [4] PANDEY S, KARYPIS G. A Self-Attentive Model for Knowledge Tracing [J]. arXiv:1907.06837, 2019.
- [5] NAKAGAWA H, IWAWA Y, MATSUO Y. Graph-Based Knowledge Tracing: Modeling Student Proficiency Using Graph Neural Network [C]// Proceedings of IEEE/WIC/ACM International Conference on Web Intelligence. New York: ACM, 2019: 156-163.
- [6] KHOSLA P, TETERWAK P, WANG C, et al. Supervised Contrastive Learning [C]// Proceedings of Advances in Neural Information Processing Systems. Curran Associates Inc., 2020:

- 18661-18673.
- [7] PENG W, LI S, ZHANG W. Investigation and Analysis on E-Learning Behavior of Spare-Time Students [C]//Proceedings of 2011 International Conference on Internet Computing and Information Services. New York: IEEE, 2011: 381-384.
- [8] SONG X, LI J, LEI Q, et al. Bi-CLKT: Bi-Graph Contrastive Learning-Based Knowledge Tracing [J]. Knowledge-Based Systems, 2022, 241: 108274.
- [9] WU T, LING Q. Self-Supervised Heterogeneous Hypergraph Network for Knowledge Tracing [J]. Information Sciences, 2023, 624: 200-216.
- [10] SUN X, YIN H, LIU B, et al. Heterogeneous Hypergraph Embedding for Graph Classification [C]//Proceedings of the 14th ACM International Conference on Web Search and Data Mining. New York: ACM, 2021: 725-733.
- [11] SHEN S, LIU Q, HUANG Z, et al. A Survey of Knowledge Tracing: Models, Variants, and Applications [J]. IEEE Transactions on Learning Technologies, 2024, 17: 1858-1879.
- [12] LIU Z, GUO T, LIANG Q, et al. Deep Learning Based Knowledge Tracing: A Review, A Tool and Empirical Studies [J]. IEEE Transactions on Knowledge and Data Engineering, 2025, 37(8): 4512-4536.
- [13] NAGATANI K, ZHANG Q, SATO M, et al. Augmenting Knowledge Tracing by Considering Forgetting Behavior [C]//Proceedings of The World Wide Web Conference. New York: ACM, 2019: 3101-3107.
- [14] GHOSH A, HEFFERNAN N, LAN A S. Context-Aware Attentive Knowledge Tracing [C]//Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York: ACM, 2020: 2330-2339.
- [15] YANG Y, SHEN J, QU Y, et al. GIKT: A Graph-Based Interaction Model for Knowledge Tracing [C]//Proceedings of Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Cham: Springer, 2020: 299-315.
- [16] TSUTSUMI E, KINOSHITA R, UENO M. Deep-IRT with Independent Student and Item Networks [C]//Proceedings of the International Educational Data Mining Society Conference. ERIC, 2021: 510-517.
- [17] ZHOU N, DONG Y Q, YAN L K, et al. Investigation on Exercise Recommendation Integrating Student Knowledge State and Chaos Firefly Algorithm [J]. Journal of Frontiers of Computer Science & Technology, 2025, 19(6): 1620-1631.
- [18] ZHAO Y J, MENG F J, XU X J. A Review of Knowledge Tracking for Online Education Learners [J]. Journal of Computer Applications, 2024, 44(6): 1683-1698.
- [19] ZHAO Y, MA H, WANG J, et al. Question-Response Representation with Dual-Level Contrastive Learning for Improving Knowledge Tracing [J]. Information Sciences, 2024, 658: 120032.
- [20] ZHANG H, LIU Z, SHANG C, et al. A Question-Centric Multi-Experts Contrastive Learning Framework for Improving the Accuracy and Interpretability of Deep Sequential Knowledge Tracing Models [J]. ACM Transactions on Knowledge Discovery from Data, 2025, 19(2): 1-25.
- [21] LEE W, CHUN J, LEE Y, et al. Contrastive Learning for Knowledge Tracing [C]//Proceedings of the ACM Web Conference 2022. New York: ACM, 2022: 2330-2338.
- [22] ZHENG N, SHAN Z. Co-Attention and Contrastive Learning Driven Knowledge Tracing [C]//Proceedings of Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Cham: Springer, 2024: 177-194.
- [23] DAI H, YUN Y, ZHANG Y, et al. Self-Paced Contrastive Learning for Knowledge Tracing [J]. Neurocomputing, 2024, 609: 128366.
- [24] YAO L, MAO C, LUO Y. Graph Convolutional Networks for Text Classification [C]//Proceedings of the AAAI Conference on Artificial Intelligence. AAAI, 2019: 7370-7377.
- [25] HOCHREITER S, SCHMIDHUBER J. Long Short-Term Memory [J]. Neural Computation, 1997, 9(8): 1735-1780.
- [26] ZHAO Y, TUO J, SHAN K, et al. Overview of Knowledge Tracking Models in the Field of Intelligent Education [J]. Journal of Computer System & Applications, 2025, 34(6): 1-11.
- [27] YEUNG C K, YEUNG D Y. Addressing Two Problems in Deep Knowledge Tracing via Prediction-Consistent Regularization [C]//Proceedings of the Fifth Annual ACM Conference on Learning at Scale. New York: ACM, 2018: 1-10.
- [28] LI Q, HUANG Z, SUN J, et al. HKT: Hierarchical Structure-based Knowledge Tracing [J]. Information Processing & Management, 2025, 62(5): 104206.
- [29] XIE P Z, LI G J, LI T. Knowledge Tracing Model Based on Exercise-Knowledge Point Heterogeneous Graph and Multi-feature Fusion [J]. Computer Science, 2025, 52(3): 197-205.
- [30] ZHANG W, LUO P H, GONG Z W, et al. Multi Relationship and Time Enhanced Knowledge Tracking Model [J]. Application Research of Computers, 2025, 42(3): 728-734.
- [31] WU Z, HUANG L, HUANG Q, et al. SGKT: Session Graph-Based Knowledge Tracing for Student Performance Prediction [J]. Expert Systems with Applications, 2022, 206: 117681.
- [32] TONG H, WANG Z, ZHOU Y, et al. Introducing Problem Schema with Hierarchical Exercise Graph for Knowledge Tracing [C]//Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM, 2022: 405-415.



**CHEN Xiaolan**, born in 1997, postgraduate. Her main research interests include knowledge tracing, educational data analysis and applications, and AI-empowered primary education.



**LIN Ronghua**, born in 1994, Ph.D, associate research fellow, is a member of CCF(No. C5693M). His main research interests include educational data mining, recommender system and social network analysis.