

基于试题-知识点异构图和多特征融合的知识追踪模型

解培中, 李冠进, 李汀

引用本文

解培中, 李冠进, 李汀. [基于试题-知识点异构图和多特征融合的知识追踪模型](#)[J]. 计算机科学, 2025, 52(3): 197-205.

XIE Peizhong, LI Guanjin, LI Ting. [Knowledge Tracing Model Based on Exercise-Knowledge Point Heterogeneous Graph and Multi-feature Fusion](#) [J]. Computer Science, 2025, 52(3): 197-205.

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)

[基于异构图神经网络的网络切片端到端时延估计](#)

Network Slicing End-to-end Latency Prediction Based on Heterogeneous Graph Neural Network
计算机科学, 2025, 52(3): 349-358. <https://doi.org/10.11896/jsjx.240800067>

[面向异构图像压缩感知的阶数自适应多假设重构](#)

Order-adaptive Multi-hypothesis Reconstruction for Heterogeneous Image Compressive Sensing
计算机科学, 2024, 51(10): 302-310. <https://doi.org/10.11896/jsjx.230800156>

[基于关键词异构图的生成式摘要研究](#)

KHGAS: Keywords Guided Heterogeneous Graph for Abstractive Summarization
计算机科学, 2024, 51(7): 278-286. <https://doi.org/10.11896/jsjx.230500059>

[融合遗忘机制的多模态知识追踪模型](#)

Multimodality and Forgetting Mechanisms Model for Knowledge Tracing
计算机科学, 2024, 51(7): 133-139. <https://doi.org/10.11896/jsjx.231000137>

[一种基于异构图神经网络和文本语义增强的实体关系抽取方法](#)

Method for Entity Relation Extraction Based on Heterogeneous Graph Neural Networks and Text Semantic Enhancement
计算机科学, 2024, 51(6A): 230700071-5. <https://doi.org/10.11896/jsjx.230700071>

基于试题-知识点异构图和多特征融合的知识追踪模型

解培中 李冠进 李汀

南京邮电大学通信与信息工程学院 南京 210003

摘要 知识追踪要求基于学习者的历史作答情况来预测未来的答题表现,并评估知识状态的变化。探索学习者知识状态的变化有助于实现个性化服务,如课程推荐和试题推荐。然而,现有的大多数知识追踪模型在建模时考虑的特征不够全面,不能综合衡量学习者知识状态的变化。针对这一问题,提出一种新的知识追踪模型——基于试题-知识点异构图和多特征融合的知识追踪模型(EKMFKT)。具体而言,从学习者的学习过程出发,研究了两种行为特征(尝试次数和提示次数)以及两种时间特征(响应时间和间隔时间)对知识状态的影响。然后,设计了学习门和遗忘门,以模拟知识的获取和遗忘,全面更新知识状态的变化。另外,对于模型的输入,设计了基于试题-知识点异构图的图嵌入方法来预训练试题表示,使得模型的输入保持试题和知识点的关联。在两个公开数据集上的实验结果表明,EKMFKT在预测性能上优于现有模型。通过引入多个特征并确保试题与知识点的关联,EKMFKT使知识状态的变化更合理,增强了模型的可解释性。

关键词: 知识追踪;异构图;学习行为;时间特征;遗忘行为

中图分类号 G40-057

Knowledge Tracing Model Based on Exercise-Knowledge Point Heterogeneous Graph and Multi-feature Fusion

XIE Peizhong, LI Guanjin and LI Ting

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

Abstract Knowledge tracing(KT) aims to predict a learner's future performance based on their historical responses and assess changes in their knowledge state. Exploring these changes facilitates personalized services, such as course and exercise recommendations. However, most existing knowledge tracing models do not consider comprehensive features, leading to incomplete assessments of a learner's knowledge state changes. To address this issue, a new knowledge tracing model—knowledge tracing model based on exercise-knowledge point heterogeneous graph and multi-feature fusion(EKMFKT) is propose. Specifically, we study two behavioral features(attempt and hint) and two temporal features(response time and interval time) and their impact on the knowledge state. To simulate knowledge acquisition and forgetting, we design learning and forgetting gates, which comprehensively update the knowledge state. For model inputs, a graph embedding method based on the question-knowledge point heterogeneous graph is designed to pre-train question representations, preserving the associations between questions and knowledge points. Experimental results on two public datasets demonstrate that EKMFKT outperforms existing models in predictive performance. By incorporating multiple features and ensuring the connection between question and knowledge point, EKMFKT provides a more reasonable representation of knowledge state changes, enhancing the interpretability of the model.

Keywords Knowledge tracing, Heterogeneous graph, Learning behavior, Time feature, Forgetting behavior

1 引言

随着数字技术的迅速发展,智能教学系统(Intelligent Tutoring System, ITS)和大型开放式网络课程(Massive Open Online Courses, MOOCs)等在线教育平台不断涌现。这些在线教育平台因汇聚了丰富的教育资源并能降低学习成本等而备受关注。学习者在这些平台和系统上的作答行为产生了大量数据,这些教育领域的的数据推动了知识追踪

(Knowledge Tracing, KT)的发展^[1-2]。知识追踪旨在通过分析学习者的历史作答数据,来追踪在不同知识点上的知识状态。有了KT研究,学习者能够发现自己的弱点,并有针对性地准备练习,以节省精力。同时,教育系统可以定制多种个性化服务,例如指定适宜的学习路线和推荐合适的学习材料^[3-4],帮助学习者更高效地学习。

现有的知识追踪研究主要分为传统方法和深度学习方法^[5]。传统方法需要大量人工投入来构建和维护模型,因此

到稿日期:2024-07-23 返修日期:2024-10-18

基金项目:国家自然科学基金(62277032)

This work was supported by the National Natural Science Foundation of China(62277032).

通信作者:解培中(cas@njupt.edu.cn)

成本较高。近年来,深度学习技术在数据挖掘^[6]、自然语言处理^[7]和计算机视觉^[8]领域取得了显著进展,使大规模数据集的挖掘成为可能。在基于深度学习的方法中,通过引入额外的特征来建模和衡量学习者知识状态的变化一直是研究的重点。学习过程一致的知识追踪模型(LPKT)^[9]和学习行为导向的知识追踪模型(LBKT)^[10]分别引入时间特征和行为特征使得知识状态的变化更加合理,并且提高了模型的准确率。然而,以上模型在引入特征时考虑得不够全面,LPKT 主要关注时间特征对知识状态变化的影响,而 LBKT 则主要关注行为特征的影响,没有综合衡量多种特征对知识状态变化的影响。引入多个特征来建模对于真实教育场景是必要的,比如在图 1 所示的在线学习系统的场景中,学习者 1 和学习者 2 虽然都对试题 1 做出了正确的响应,但由于回答时间不同,他们对试题的熟练程度也不同。响应时间越短,代表学习者对试题越熟练,或者可能是猜测;而响应时间越长,则代表学习者对试题越不熟练。学习者 1 和学习者 2 在一定时间间隔后再回答试题 1,由于间隔时间不同,遗忘的程度也会不同,从而可能导致错误的响应。同时,尝试次数和提示次数也是影响试题响应的重要因素。例如,图 1 中学习者 1 回答试题 2 尝试了 4 次,学习者 2 回答试题 2 尝试了 1 次,这可能意味着学习者 1 反复尝试直至给出正确答案。学习者 1 独立回答了试题 3,而学习者 2 在回答试题 3 之前获取了两次提示,可能意味着学习者 1 比学习者 2 对相关知识的掌握更好。

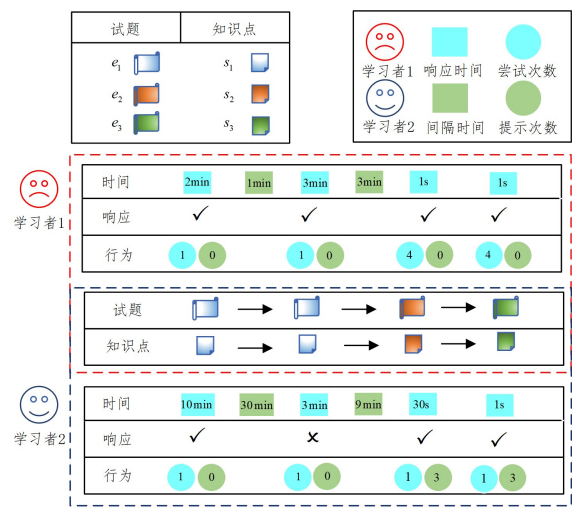


图 1 在线学习系统中,两名学习者在回答习题时的学习行为和
学习时间示例

Fig. 1 Examples of learning behaviors and learning times of two learners when answering exercises on the online learning system

为契合真实教育场景,综合衡量学习者知识状态的变化,本文提出了一种基于试题-知识点异构图和多特征融合的知识追踪模型(Knowledge Tracing Model Based on Exercise-Knowledge Point Heterogeneous Graph and Multi-Feature Fusion,EKMFKT),研究了两种行为特征(尝试次数和提示次数)以及两种时间特征(响应时间和间隔时间)对学习者的知识状态的影响。首先,使用行为模块提取关键学习行为特征。将提取的学习行为特征与时间特征输入学习模块,以研究它们对学习者的知识获取的影响。然后,结合遗忘因素与知识

获取,全面更新知识状态的变化。最后,结合下一时刻的试题输出预测结果。

本文贡献总结如下:

- 1)提出了一种基于试题-知识点异构图和多特征融合的知识追踪模型(EKMFKT)。该模型将学习者作答序列中的时间特征、行为特征、试题知识点特征以及响应特征作为输入,从学习者的学习过程出发建模,来综合评估它们对学习者的知识状态的影响。
- 2)提出了一种基于试题-知识点异构图的预训练方法来表示试题,使得相同知识点的试题表示相似,增强了模型输入的可解释性。
- 3)针对学习过程中常见的遗忘和知识获取现象,EKMFKT 设计了遗忘门和学习门来模拟学习者的遗忘和知识获取。

2 相关工作

EKMFKT 模型主要研究了两种行为特征以及两种时间特征对学习者的知识状态的影响,另外设计了一种基于试题-知识点异构图的图嵌入方法来预训练试题表示。本章主要从知识追踪模型和异构网络图嵌入两个维度加以介绍。

2.1 知识追踪

知识追踪的模型分为传统方法和基于深度学习的方法。传统方法的代表包括贝叶斯知识追踪(Bayesian Knowledge Tracing,BKT)、性能因素分析(Performance Factors Analysis,PFA)和项目反应理论(Item Response Theory,IRT)。BKT^[11]通过将学习者的学习过程建模为一个隐马尔可夫过程,利用贝叶斯推断来动态地估计学习者的知识状态。PFA^[12]提供了一种更加全面的学习建模方法,不仅考虑学习者对特定知识点的掌握程度,还考虑到其他影响学习表现的因素,例如学习者的注意力、动机、学习策略。IRT^[13]模型考虑了两个关键因素:学习者的学习能力和试题的特征。试题的特征包括难度、区分度和猜测情况,这些因素共同影响了学习者的答题情况。传统方法需要大量的人工投入来进行模型的构建和维护,因此成本较高。近年来,随着深度学习技术的不断发展,越来越多的研究开始利用深度学习方法来实现知识追踪,并取得了不错的效果。DKT^[14](Deep Knowledge Tracing)是深度知识追踪模型领域的开创性工作,它将深度学习的方法应用到知识追踪领域,取得了比传统方法更为显著的性能提升。但 DKT 忽略了许多因素的影响,比如额外的输入信息(试题文本特征、时间特征、行为特征等)、遗忘因素以及模型的可解释性等。PEBG^[15](Pre-training Embeddings via Bipartite Graph)从试题-知识点关联矩阵出发,设计 3 个损失函数来联合优化知识点间关系、试题间关系和试题难度,得到试题的预训练编码表示特征,并基于此进行建模。EKT^[16](Exercise-aware Knowledge Tracing)采用自然语言处理领域中的 Word2vec 方法来处理试题的文本信息进行建模。SPKT^[17](Student-Problem Association Based Heterogeneous Graph Knowledge Tracing Model)使用图注意力网络提取学习者和问题间的关系,并基于此进行建模。AKT^[18](Attention Knowledge Tracing)将 Rasch 模型与试题编码

相结合,生成了包含试题难度特征的试题表示。LPKT(Learning Process-consistent Knowledge Tracing)模型从学习者的学习过程出发建模,研究时间特征对学习者的知识状态变化的影响。LBKT(Learning Behavior-oriented Knowledge Tracing)则考虑了学习过程中行为特征对学习者的知识状态的影响。DKT+[19]模型提出了一种融合试题响应时间、试题间隔时间以及题目作答次数的方法,以此来模拟学习者的遗忘。GFLDKT[20](Gating-controlled Forgetting and Learning Mechanisms for Deep Knowledge Tracing)模型引入了两个门控网络,即遗忘门和学习门,用于捕捉不同知识水平的学习者在学习和遗忘方面的特征。通过引入不同的特征,上述模型在一定程度上保证了知识状态变化的合理性以及模型的精度。但是学习者知识状态的变化受到多种因素的影响,现有的大多数知识追踪模型在建模时考虑的特征不够全面,导致不能综合衡量学习者知识状态的变化。本文提出的 EKM-FKT 模型综合考虑了两种时间特征和两个行为特征对学习者的知识状态变化的影响。

2.2 异构网络图嵌入

异构网络的图嵌入可以定义为:学习一个 d 维的矩阵表示 $\mathbf{A} \in \mathbb{R}^{|V| \times d}$, $d \ll |V|$,使得其中的向量能够捕获异构网络的结构和语义信息。异构网络图嵌入在推荐系统、社交网络分析和知识图谱等领域取得了显著成果[21]。Mikolov 等[22]提出了 Word2Vec 模型,采用连续词袋模型(Continuous Bag of Words, CBOW)和跳字模型(Skip-gram)学习单词的向量表示。Perozzi 等[23]在 Word2Vec 模型的基础上提出了 DeepWalk,用于学习图节点的向量表示。Tang 等[24]提出了一种用于学习异构网络节点向量表示的方法,通过最大化邻接节点之间的相似性和最小化非邻接节点之间的相似性来实现。Grover 等[25]提出了 node2vec 模型,它允许用户通过参数来调整生成的图节点序列的特征。Dong 等[26]提出的 meta-path2vec 模型利用元路径来采样异构网络中节点之间的关系。基于以上研究,我们设计了一种基于试题-知识点异构图的图嵌入方法来预训练试题表示,使得模型的输入保持试题和知识点的关联。

3 问题定义

本章将从知识追踪和异构网络图嵌入两个维度对问题的定义加以介绍。

3.1 知识追踪

假设在线教育系统中存在 $|L|$ 个学习者、 $|E|$ 道试题和 $|S|$ 个知识点,学习者在特定学科上的历史作答记录描述为 $X = \{x_1, it_1, x_2, \dots, it_{t-1}, x_t\}$,其中 $x_t \in (e_t, s_t, r_t, at_t, b_t)$ 代表学习者的基本作答要素, $e_t \in |E|$ 代表学习者 t 时刻回答的试题, $s_t \in |S|$ 代表学习者 t 时刻回答试题所对应的知识点, r_t 代表对试题的响应(1 代表正确,0 代表错误), at_t 代表回答试题所花费的时间, $b_t \in (a_t, h_t)$ 代表 t 时刻回答试题时的行为特征即尝试次数和提示次数, it_t 代表 t 时刻回答的试题与 $t-1$ 时刻回答的试题的时间间隔。知识追踪通常是基于历史作答情况预测下一时刻试题正确作答的概率 $p(r_{t+1}=1|e_{t+1}, X)$ 。

3.2 试题-知识点异构图嵌入

假设在线教育系统中 $E = \{e_i\}_{i=1}^{|E|}$ 表示所有的试题 $|E|$, $S = \{s_j\}_{j=1}^{|S|}$ 表示所有的知识点 $|S|$,试题可能与一个或多个知识点相关联,而知识点可能与多道试题相关。试题-知识点关系可以自然地表示为一个异构图 $G = (E, S, C)$ 。在这个图中,试题节点 E 和知识点 S 被视为不同类型的节点,它们之间的关系由边 C 来表示,即 $C = \{r_{ij}\} \in \{0, 1\}^{|E| \times |S|}$,其中 1 代表相关,0 代表不相关。试题-知识点异构图嵌入可以定义为学习一个 d 维矩阵表示 $\mathbf{A} \in \mathbb{R}^{|V| \times d}$,其中 $d \ll |V|$,使得其中的向量捕捉到试题-知识点异构网络的结构和语义特征。

4 方法描述

为了更好地介绍所提模型,本章首先介绍知识追踪中的基本表示,然后详细说明模型的细节,最后阐述试题-知识点异构图嵌入的预训练过程。

4.1 知识追踪嵌入表示

4.1.1 知识状态嵌入表示

对于知识状态的嵌入表示,与文献[16]一样,使用矩阵 $\mathbf{H}_t \in \mathbb{R}^{|S| \times d_h}$ 表示 t 时刻学习者的知识状态,其中 $|S|$ 表示知识点的数量, d_h 代表其维度。因此,学习者掌握知识点的熟练程度可以由 $\mathbf{H}$ 矩阵的每一行来表示。

4.1.2 时间嵌入表示

时间嵌入指将试题的响应时间和间隔时间转换为向量表示。这两个时间可以影响知识的获取和遗忘。我们以分钟为单位离散化响应时间和间隔时间。使用矩阵 $\mathbf{AT} \in \mathbb{R}^{n_{at} \times d_{at}}$ 和 $\mathbf{IT} \in \mathbb{R}^{n_{it} \times d_{it}}$ 表示离散化后的矩阵表示,其中 n_{at} 和 n_{it} 表示离散化数量, d_{at} 和 d_{it} 表示离散化后的向量的维度。因此,交互序列 x_t 中的 at_t 和 it_t 可以向量化为 $\mathbf{at}_t \in \mathbb{R}^{d_{at}}$ 和 $\mathbf{it}_t \in \mathbb{R}^{d_{it}}$ 。

4.1.3 交互嵌入表示

将试题-知识点-响应时间-响应作为基本的交互元素,试题的嵌入矩阵为 $\mathbf{E} \in \mathbb{R}^{|E| \times d_e}$,知识点的嵌入矩阵为 $\mathbf{S} \in \mathbb{R}^{|S| \times d_s}$,其中 $|E|$ 和 $|S|$ 表示试题和知识点的数量, d_e 和 d_s 表示其维度。假设试题和知识点矩阵已经通过一个预训练的嵌入模块进行了预训练(该模块将在第 4.3 节中介绍), t 时刻的问题和知识点的向量表示可以根据其对应的 id 来查表表示: $\mathbf{e}_t \in \mathbb{R}^{d_e}$, $\mathbf{s}_t \in \mathbb{R}^{d_s}$ 。对于响应,使用 $\mathbf{R} \in \mathbb{R}^{2 \times d_r}$ 矩阵来将两种响应结果(正确、错误)表示为一个 d_r 维的向量, t 时刻的响应 r_t 可以向量化为 $\mathbf{r}_t \in \mathbb{R}^{d_r}$ 。使用全连接层来融合以上嵌入,得到交互嵌入表示:

$$\mathbf{i}_t = \text{Leaky ReLU}(\mathbf{W}_1[\mathbf{e}_t \oplus \mathbf{s}_t \oplus \mathbf{at}_t \oplus \mathbf{r}_t] + \mathbf{b}_1) \quad (1)$$

其中, $\mathbf{W}_1 \in \mathbb{R}^{d_h \times (d_e + d_s + d_{at} + d_r)}$, $\mathbf{b}_1 \in \mathbb{R}^{d_h}$ 是可训练的参数; $\oplus$ 表示拼接操作; Leaky ReLU 作为激活函数,用来避免输入小于零时的神经元死亡的问题。

4.1.4 Q-矩阵

Q-矩阵 $\mathbf{Q} = \{q_{i,j}\} \in \mathbb{R}^{|E| \times |S|}$ 表示试题和知识点的对应关系。其中, $q_{i,j} = 1$ 代表试题 e_i 和知识点 s_j 相关联,反之则不相关。 $\mathbf{q}_e \in \mathbb{R}^{|S|}$ 代表试题 e_t 所对应的知识点向量。根据教育评估领域理论[27],试题通常考查学习者对特定知识点的理解程度,知识点通常以问题的形式出现,学习者对特定

问题的回答情况会反映他们对相关知识点的掌握程度。因此,知识状态矩阵 $\mathbf{H}_{t-1}$ 是通过试题所对应的知识点向量来进行更新的。

$$\tilde{\mathbf{H}}_{t-1} = \mathbf{q}_{e_t} \cdot \mathbf{H}_{t-1} \quad (2)$$

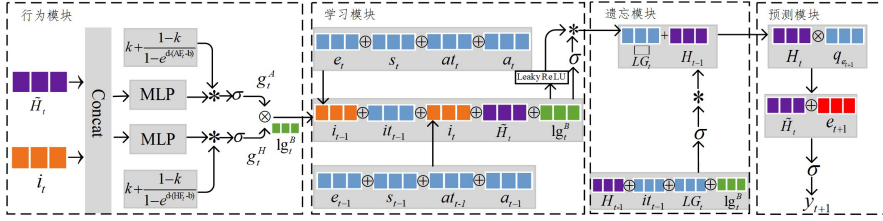


图 2 EKMFKT 的框架与细节

Fig. 2 Framework and details of EKMFKT

4.2.1 行为模块

分析学习过程中两个主要行为: 试题尝试次数和获取提示次数。通常情况下, 多数学习者会尝试一次并且不获取任何提示来回答问题。然而, 一些学习者会持续不断地尝试或试图寻求提示^[28]。这种行为可能表明了学习者对有些概念不熟悉, 或者急于追求正确答案。因此, 这些学习者的知识增长可能与只尝试一次且不获取提示的学习者不同。与文献[10]一样, 假设尝试次数和提示次数近似服从泊松分布, 其概率分布为 $p_{ji}^A \sim \pi(\lambda_j^A)$, $p_{ji}^H \sim \pi(\lambda_j^H)$, 其中 p_{ji}^A 和 p_{ji}^H 分别表示学习者在试题 e_j 上的尝试和提示次数, λ_j^A 和 λ_j^H 分别代表可以从数据集中学习到的参数。

尝试和提示因子为:

$$\begin{aligned} AF_{ji} &= 1 - P(\pi(\lambda_j^A) \geq p_{ji}^A) \\ HF_{ji} &= 1 - P(\pi(\lambda_j^H) \geq p_{ji}^H) \end{aligned} \quad (3)$$

有了尝试和提示因子后, 可以计算这两个行为对知识获取的影响。

$$\begin{aligned} c_t^A &= k + \frac{1-k}{1+e^{d \cdot (AF_t - b)}} \\ c_t^H &= k + \frac{1-k}{1+e^{d \cdot (HF_t - b)}} \end{aligned} \quad (4)$$

$$\mathbf{g}_t^A = \sigma(c_t^A \cdot (\mathbf{W}_2^A [\tilde{\mathbf{H}}_{t-1} \oplus \mathbf{i}_t] + \mathbf{b}_2^A))$$

$$\mathbf{g}_t^H = \sigma(c_t^H \cdot (\mathbf{W}_2^H [\tilde{\mathbf{H}}_{t-1} \oplus \mathbf{i}_t] + \mathbf{b}_2^H))$$

其中, $\mathbf{W}_2^A, \mathbf{W}_2^H \in \mathbb{R}^{d_h \times 2d_h}$, $\mathbf{b}_2^A, \mathbf{b}_2^H \in \mathbb{R}^{d_h}$, 知识获取向量 $\mathbf{g}_t^A, \mathbf{g}_t^H$ 受尝试因子和提示因子 AF_{ji}, HF_{ji} 的影响。

学习行为对知识获取的影响不是孤立的, 而是相互作用的。例如, 一个学习者回答试题的响应时间很短, 可能是对该知识点掌握得很熟练或者是猜测所致, 但是通过结合学习行为, 比如尝试多次或者提示多次, 则更可能是猜测所致, 所以考虑学习行为间的作用是必要的。因此对这两个向量进行外积运算^[29]。

$$\mathbf{B}_t = \begin{bmatrix} \mathbf{g}_t^A \\ 1 \end{bmatrix} \otimes \begin{bmatrix} \mathbf{g}_t^H \\ 1 \end{bmatrix} \quad (5)$$

其中, $\otimes$ 代表外积运算; $\mathbf{B}_t$ 既保留了每个行为的特征, 又保留了两个行为相互作用的特征。最终使用全连接层获得综合行为特征向量。

$$\mathbf{lg}_t^B = \text{Leaky ReLU}(\mathbf{W}_3 \cdot \mathbf{B}_t + \mathbf{b}_3) \quad (6)$$

其中, $\mathbf{W}_3 \in \mathbb{R}^{d_h \times (d_h + 1) \times (d_h + 1)}$, $\mathbf{b}_3 \in \mathbb{R}^{d_h}$ 。将得到的行为特征

4.2 EKMFKT 模型

本节从学习者的学习过程出发进行建模。模型如图 2 所示。该模型在每个学习步骤中由 4 个模块组成: 1) 行为模块; 2) 学习模块; 3) 遗忘模块; 4) 预测模块。

向量 $\mathbf{lg}_t^B$ 传递给学习模块, 来综合衡量结合其他特征对知识获取的影响。

4.2.2 学习模块

基于循环神经网络的设计理念^[30-31], 我们设计了一个时间域通道, 来模拟学习者在长期学习过程中的知识获取。将前一时刻交互嵌入的表示、当前时刻交互嵌入的表示、间隔时间、知识状态以及行为特征结合起来, 来模拟学习者知识获取的过程。

$$\mathbf{lg}_t = \text{Leaky ReLU}(\mathbf{w}_4 [\mathbf{i}_{t-1} \oplus \mathbf{i}_t \oplus \tilde{\mathbf{H}}_{t-1} \oplus \mathbf{lg}_t^B] + \mathbf{b}_4) \quad (7)$$

其中, $\mathbf{w}_4 \in \mathbb{R}^{d_h \times 5d_h}$, $\mathbf{b}_4 \in \mathbb{R}^{d_h}$ 。

考虑到并非所有获取的知识都能转化为知识增长, 设计了一个学习门来调节知识获取的能力。

$$\mathbf{I}_t^L = \sigma(\mathbf{w}_5 [\mathbf{i}_{t-1} \oplus \mathbf{i}_t \oplus \tilde{\mathbf{H}}_{t-1} \oplus \mathbf{lg}_t^B] + \mathbf{b}_5) \quad (8)$$

其中, $\mathbf{w}_5 \in \mathbb{R}^{d_h \times 5d_h}$, $\mathbf{b}_5 \in \mathbb{R}^{d_h}$ 。

真实的学习增益为:

$$\mathbf{LG}_t = \mathbf{I}_t^L \cdot \mathbf{lg}_t \quad (9)$$

4.2.3 遗忘模块

根据遗忘曲线的理论, 学习者掌握知识点的熟练程度随着时间的推移也会下降, 他们的知识状态的更新受到遗忘因素的影响^[32]。因此, EKMFKT 设计了一个遗忘门来建模遗忘现象。我们认为学习者的遗忘与其学习行为和时间间隔密切相关。将学习者的学习行为特征、前一时刻的知识状态、间隔时间、知识增长和学习行为结合到遗忘门中, 模拟他们的遗忘过程。

$$\mathbf{I}_t^F = \sigma(\mathbf{w}_6 [\mathbf{H}_{t-1} \oplus \mathbf{i}_t \oplus \mathbf{LG}_t \oplus \mathbf{lg}_t^B] + \mathbf{b}_6) \quad (10)$$

其中, $\mathbf{w}_6 \in \mathbb{R}^{d_h \times 4d_h}$, $\mathbf{b}_6 \in \mathbb{R}^{d_h}$ 。

通过结合学习门和遗忘门, 完成知识状态的更新。

$$\mathbf{H}_t = \widetilde{\mathbf{LG}}_t + \mathbf{I}_t^F \cdot \mathbf{H}_{t-1} \quad (11)$$

4.2.4 预测模块

知识状态更新之后, 使用知识状态 $\mathbf{H}_t$ 来预测下一个时刻的试题作答情况。考虑真实的学习场景, 当 $t+1$ 时刻给出一个问题 e_{t+1} 时, 学习者会使用问题所对应的知识点来解决问题, 因此我们拼接下一时刻的试题, 然后通过带有 Sigmoid 激活函数的全连接层来得到预测结果。

$$\tilde{\mathbf{H}}_t = \mathbf{q}_{e_{t+1}} \cdot \mathbf{H}_t \quad (12)$$

$$y_{t+1} = \sigma(\mathbf{w}_7 [\tilde{\mathbf{H}}_t \oplus \mathbf{e}_{t+1}] + \mathbf{b}_7)$$

其中, $\mathbf{w}_7 \in \mathbb{R}^{d_h \times 2d_h}$; $\mathbf{b}_7 \in \mathbb{R}^{d_h}$; y_{t+1} 表示对试题 e_{t+1} 的预测情况, 值在 0~1 之间。

我们使用交叉熵对数损失来训练模型。该损失衡量了预测值 y_{t+1} 与正确答案 a_t 之间的差异。

$$\mathcal{L} = - \sum_{n=1}^{|L|} \sum_{t=1}^T (r_t \log(y_t) + (1-r_t) \log(1-y_t)) \quad (13)$$

4.3 试题-知识点异构图嵌入

Q-矩阵用于描述试题与知识点之间的关联性(第 4.1.4 节), 但是直接利用 Q-矩阵表示试题和知识点的是不可取的, 存在以下两个明显的缺点^[20]: 1) 它是手动标记的, 可能存在错误和主观偏见; 2) 它是十分稀疏的, 直接用 Q-矩阵表示试题和知识点对模型的训练存在困难。因此, 设计了基于试题知识点异构图的方法来对试题和知识点进行预训练嵌入表示。具体来说, 利用 Q-矩阵这种关系, 可以构建试题-知识点的异构图。随后, 通过基于元路径的采样技术从异构图中获取采样数据^[33], 然后利用异构网络的图嵌入技术对采样点进行训练, 生成试题和知识点的密集值向量表示。该向量仍然保持试题与知识点之间的关联, 增强了模型输入的可解释性。

预训练模块的流程图如图 3 所示。首先, 随机初始化一个嵌入矩阵来存放训练得到的结果 $\mathbf{A} \in \mathbb{R}^{|V| \times d}$ 。然后, 根据 Q-矩阵生成试题知识点异构图^[33], 指定元路径为 $s-e-s$, 并根据元路径对异构图进行采样, 生成预训练数据。例如, 对知识 s_1 采样, 生成的训练数据为: $s_1-e_3-s_3-e_2-s_1$ 。通过最大化条件概率来得到训练结果。

$$p(c_t | v; \theta) = \frac{e^{x_{c_t} \cdot x_v}}{\sum_{u \in V} e^{x_u \cdot x_v}} \quad (14)$$

其中, $\mathbf{x}_v$ 是矩阵 $\mathbf{A}$ 的第 v 行, 代表第 v 个点的嵌入表示; $\mathbf{x}_{c_t}$ 代表点 v 的邻域的向量表示。比如当训练到点 e_3 时, 需要最大化邻域节点 s_1 和 s_3 , 由此得到的向量表示既能克服稀疏问题, 又能同时保持试题和知识点对应的关系。

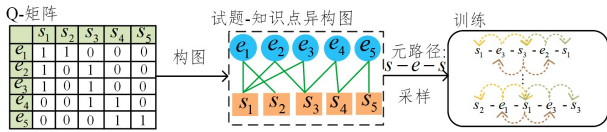


图 3 预训练流程

Fig. 3 Pre-training workflow

通过最大化条件概率可以逐步得到优化的嵌入矩阵 $\mathbf{A} \in \mathbb{R}^{|V| \times d}$, 该矩阵包含了试题和知识点的嵌入表示。将该矩阵拆分为 $\mathbf{E} \in \mathbb{R}^{|E| \times d_e}$ 和 $\mathbf{S} \in \mathbb{R}^{|S| \times d_s}$, 分别用于存放试题和知识点的嵌入表示。其中, $|\mathbf{E}|$ 和 $|\mathbf{S}|$ 表示试题和知识点的数量, d_e 和 d_s 表示其维度。这两个向量可以用于知识追踪中的试题、知识点查找表。例如, 在第 t 个时间步, 学习者做的练习

以及练习对应的知识点为 e_t 和 s_t , 它们是一个 one-hot 向量, 取出向量中为 1 的值对应索引, 去查找表中获得对应的嵌入表示。换句话说, 如果 e_t 的第 m 项是 1 (其他项是 0), 那么学习者在第 t 个时间步中做的是整套练习中的第 m 个练习。练习 e_t 的嵌入被记为 $\mathbf{e}_t \in \mathbb{R}^{d_e}$, 可以通过式(15)获得。

$$\begin{aligned} \mathbf{e}_t &= \mathbf{E}^T \times \mathbf{e}_t \\ \mathbf{s}_t &= \mathbf{S}^T \times \mathbf{s}_t \end{aligned} \quad (15)$$

5 实验

5.1 数据集

考虑到 EKMFKT 模型需要多种特征作为输入, 在两个公共数据集上对其进行了评估: 1) ASSISTChall¹⁾; 2) ASSIST2012²⁾。这两个数据集包含了行为特征和时间特征。表 1 列出了这两个数据集的基本统计信息。其他现有数据集由于缺少关键特征, 不能满足实验要求。例如, ASSIST2009³⁾ 和 Junyi⁴⁾ 数据集缺少关于题目间隔时间的特征, EdNet-KT1⁵⁾ 数据集缺少学习行为特征, 因此它们不适合本次研究。

1) ASSISTChall 是为了 2017 年的一个 ASSISTments 数据挖掘竞赛而收集的, 具有最高的数据完整性和正常性。在这个数据集中, 学习者的学习序列相对较长, 比其他 ASSISTments 数据集拥有更多的记录数。数据预处理时, 筛选掉没有知识点和试题的记录。

2) ASSIST2012 是由 2004 年创建的在线辅导系统 ASSISTments 收集而来, 这些数据包括了 2012—2013 学年学习者的练习记录。实验中使用的版本是“Skill-builder”, 同样筛选掉没有知识点和试题的记录。

表 1 数据集信息统计

Table 1 Statistics of datasets

统计	数据集	
	ASSISTChall	ASSIST2012
记录数	864 713	2 630 080
学习者数	1 709	28 539
试题数	2 210	50 988
知识点数	101	245
回答时间数	1 265	650
间隔时间数	33 079	301 052
序列平均长度	551.682	93.453
平均尝试数	2.718	1.354
平均提示数	1.267	0.393

5.2 训练细节

实验中, 按照学习者人数将数据集按 8:1:1 的比例划分, 分别得到训练集、验证集和测试集, 按照学习者回答问题的开始时间排序, 将输入序列的长度设置为 100, 对超出这一长度的序列进行截断处理, 而对于长度较短的序列, 采用填充 0 值的方法进行处理。我们使用了 Xavier 初始化的方法来初始化

¹⁾ <https://sites.google.com/view/assistmentsdatamining/dataset>

²⁾ <https://sites.google.com/site/assistmentsdata/home/2012-13-school-data-withaffect>

³⁾ <https://sites.google.com/site/assistmentsdata/home/assistent-2009-2010-data/skillbuilder-data-2009-2010>

⁴⁾ <https://pslscdatashop.web.cmu.edu/DatasetInfo?datasetId=1198>

⁵⁾ <http://ednet-leaderboard.s3-website-ap-northeast1.amazonaws.com/>

模型参数^[34]。在设置行为模块的超参数时,参考了 LBKT 中的设置。实验中,知识状态嵌入矩阵、问题嵌入、知识点嵌入以及时间特征的嵌入维度均设置为 128,响应嵌入的维度设置为 50,学习率设置为 0.001,为了提高训练的稳定性,统一将 Q 矩阵的值增加 0.01。选择 Adam 优化器,并在每个 epoch 进行了学习率衰减。所有的实验都在 NVIDIA GeForce RTX 3090 GPU 的 Linux 服务器上进行。

5.3 对比实验及结果分析

为了验证所提 EKMFKT 模型的性能,将其与以下基准方法进行比较。

1)DKT^[14]。这是第一个将深度学习方法应用于知识追踪的模型。实现中使用 LSTM 来预测学习者的表现。

2)DKT+^[35]。这是 DKT 模型的变体,旨在解决 DKT 中的两个主要问题。第一个问题是 DKT 无法重构观察到的输入,第二个问题是 DKT 在不同时间步上预测的性能不一致。

3)AKT^[18]。该模型将注意力机制与 IRT 模型相结合,利用自注意力机制学习问题和回答的上下文感知表示,并使用知识检索器模块获取与当前问题相关的知识。

4)LPKT^[9]。该模型旨在通过直接建模学习者的学习过程来解决学习者的知识状态与学习过程的一致性問題,并且结合了试题响应时间和试题间隔时间来建模遗忘因素。

5)LBKT^[10]。该模型考虑了学习过程中学习行为对学习 者知识状态的影响,将遗忘因素与学习者的知识获取相结合,全面更新他们知识状态的变化。

实验采用 ROC 曲线下的面积(AUC)、准确率(ACC)和均方根误差(RMSE)指标评估了不同知识追踪模型在预测学习者答题情况方面的表现。实验结果如表 2 所列,研究结果表明 EKMFKT 模型在选用的数据集上的指标优于其他 KT 方法。这表明,引入时间特征和学习行为特征能够提高模型对学习 者知识状态的估计。具体而言,与传统的抽象建模方法(如 DKT 和 DKT+)不同,模型从学习者的学习过程出发的建模方法,能够更有效地捕捉学习者的知识获取、遗忘过程和行为因素,增强了模型的可解释性和性能。此外,该模型的性能也优于基于注意力机制的模型(如 AKT),这表明充分挖掘特定的时间特征对于提高模型性能至关重要。最后,与仅考虑时间因素或学习行为的模型相比,该研究提出的模型综合考虑了这两个因素,并获得了性能的提 升。这表明挖掘学习者学习过程中的多维特征是必要性。

表 2 对比实验结果

Table 2 Comparison experiment results

模型	ASSISTChall			ASSIST2012		
	AUC	ACC	RMSE	AUC	ACC	RMSE
DKT	0.7213	0.6907	0.4471	0.7289	0.7360	0.4241
DKT+	0.7101	0.6842	0.4502	0.7295	0.7254	0.4239
AKT	0.7536	0.7117	0.4342	0.7515	0.7412	0.4204
LPKT	0.7771	0.7283	0.4237	0.7512	0.7401	0.4198
LBKT	0.7927	0.7423	0.4171	0.7539	0.7415	0.4189
EKMFKT	0.8099	0.7552	0.4071	0.7586	0.7439	0.4167

研究表明,EKMFKT 模型在 ASSISTChall 数据集上的

表现明显优于其他模型,其中 AUC 平均提高 5.89%,ACC 平均提高 4.38%,RMSE 平均降低 2.74%,这说明 EKMFKT 模型能够更有效地捕获长序列中的信息。然而,在 ASSIST2012 数据集上,EKMFKT 模型略优于其他模型,其中 AUC 平均提高 1.56%,ACC 平均提高 0.71%,RMSE 平均降低 0.47%,这表明 EKMFKT 模型具有更好的适应更大数据集的能力。然而,可以注意到性能提升的幅度有限,我们认为这可能是由于 ASSIST2012 数据集中交互序列较短,并且学习者在解决问题时几乎没有尝试多次,也几乎没有寻求提示,从而导致模型在该数据集中提取行为特征不明显。为了更好地验证模型的性能,对各个模块进行消融实验。

5.4 消融实验

为了研究本文提出的模型中不同模块的作用,在两个数据集上进行了消融实验。将原始模型与几种变体进行比较,结果如表 3 所列。所有对比的模型都使用相同的超参数,以消除超参数不同带来的影响。

表 3 消融实验结果

Table 3 Ablation experiment results

模型	ASSISTChall			ASSIST2012		
	AUC	ACC	RMSE	AUC	ACC	RMSE
EKMFKT(w/o at)	0.8058	0.7534	0.4091	0.7574	0.7435	0.4172
EKMFKT(w/o it)	0.8032	0.7494	0.4119	0.7579	0.7439	0.4168
EKMFKT (w/o behavior)	0.7868	0.7354	0.4188	0.7576	0.7427	0.4173
EKMFKT (w/o forget gate)	0.7887	0.7371	0.4188	0.7568	0.7434	0.4171
EKMFKT (w/o learning gate)	0.8072	0.7527	0.4093	0.7585	0.7436	0.4166
EKMFKT	0.8099	0.7552	0.4071	0.7586	0.7439	0.4167

首先,对于 ASSISTChall 数据集,我们发现模拟遗忘行为和学习行为可以在一定程度上提高模型的性能。如果不考虑这两种特征,将导致预测结果较大幅度地下降。对于 ASSIST12 数据集,由于学习者在解决问题时的平均尝试次接近于 1,平均提示次数接近于 0,可能使得模型提取的行为特征和遗忘特征不太明显,从而导致性能提升不显著。其次,通过学习模块来建模学习者的知识获取也能提高模型性能,尽管提升微不足道,但提高了模型的可解释性。最后,答题时间和间隔时间是整个学习过程中的重要特征,如果被省略,将对准确建模学习过程产生不利影响。但是在两个数据集中,这种提升微不足道。我们认为这是因为数据集中的时间间隔大多是相同的值 0,因此没有丰富的响应时间,导致特征区分度不足。消融结果表明,不同的特征提供了不同程度的性能提升。值得注意的是,行为特征和遗忘特征能够使模型性能获得较大幅度的提升。

此外,我们还进行了不同序列长度情况下预测性能的实验,以评估 EKMFKT 在不同序列长度上能否比当前基于学习过程建模的模型表现得更好。一般来说,更长的学习序列代表更完整的学习过程。然而,数据集中通常存在不完整的学习过程,其中较短的学习序列较为常见。因此,研究不同长度的序列是有必要的。我们在 ASSISTchall 数据集中比较了

EKMFKT 与 LBKT 和 LPKT 在不同学习序列长度下对学习 者表现的预测结果。图 4 显示了比较结果。具体而言,我 们设定了 5 种不同的长度:200,100,50,20 和 10。从图 4

可以看出,在不同序列的长度下,EKMFKT 的预测性能优于 LBKT 和 LPKT,说明该模型能够更好地建模学习者的学习 过程,在实际学习环境中具有更大的应用潜力。

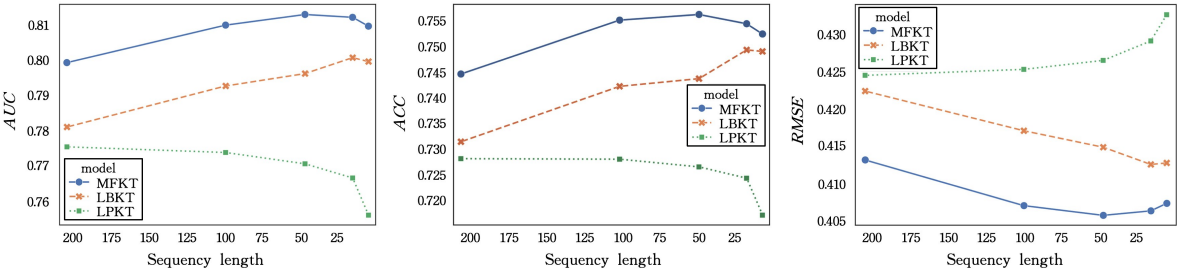


图 4 在 ASSISTchall 数据集中,序列长度对 EKMFKT, LBKT 和 LPKT 性能影响的比较

Fig. 4 Comparison of impact of sequence length on performance of EKMFKT, LBKT and LPKT in ASSISTchall dataset

5.5 实验结果可视化分析

5.5.1 知识状态可视化

知识追踪的主要目的是通过建模学习者的学习过程来预测答题情况,因此我们对模型捕捉学习者知识状态变化的过程进行展示。本节从 ASSISTChall 数据集中截取某一学习者的答题序列,来追踪其知识状态的变化情况,从而说明引入时间特征和学习行为特征对知识状态的影响效果。

图 5 为知识状态的热力图,其中横坐标展示了学习者答题的信息,纵坐标代表具体知识点,单元格反映了对特定知识点的掌握程度,颜色越深表明掌握程度越高。试题 $e_1 - e_6$ 代表了对知识点 15 的回答情况, $e_7 - e_{16}$ 代表对知识点 1 的回答情况, $e_{17} - e_{20}$ 代表对知识点 60 的回答情况。从图中可以看出,在 $e_1 - e_6$ 的回答中,虽然前 4 题的回答是错误的,但其

知识状态却逐渐提升,这可能是由于解答问题所花费的时间较长并且学习者在答题过程中获得了提示信息,这表明学习者在积极思考问题,因此知识状态得以提升。尽管学习 者对试题 e_7 的答案是正确的,但其知识状态却下降了,这可能是由于答题间隔时间较长并且学习者在答题时获得了提示,导致了遗忘。对于试题 e_{13} ,学习者尽管回答正确,但是由于答题时间较长,知识状态也下降了。学习者在试题 e_{16} 的答案虽然是正确的,但知识状态并未显著提升,从答题时间较短并且学习者在答题过程中获得了提示可以看出,这可能是猜测所致。因此,我们可以得出,学习者知识状态的变化受多个因素共同影响,引入多个特征可以更好地反映学习者知识状态的变化,从而增强模型知识状态的变化的可解释性。

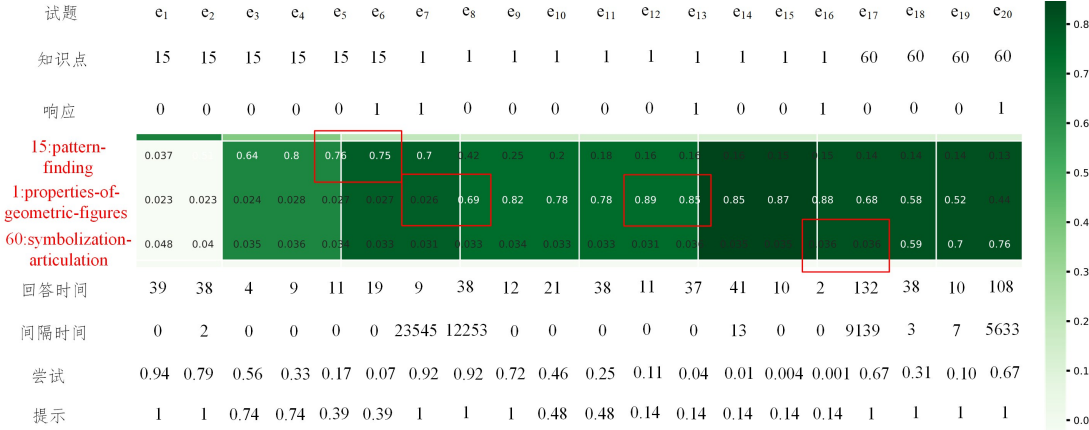


图 5 知识状态变化热力图

Fig. 5 Heatmap of knowledge state change

5.5.2 试题聚类

经过预训练后,相同知识点的试题应该聚类到同一类别中。为了更清晰地展示试题的聚类效果,选取了 ASSIST17 数据集中的 10 个知识点进行可视化分析。利用 T-SNE 方法对这些试题的嵌入进行降维,并将其投影到二维空间中进行展示,相同知识点的试题使用相同的颜色表示。如图 6 所示,该预训练模型能够将不同知识点的问题映射到不同的区域,从而使不同知识点的区域具有一定的区分度,增强了模型输入的可解释性。

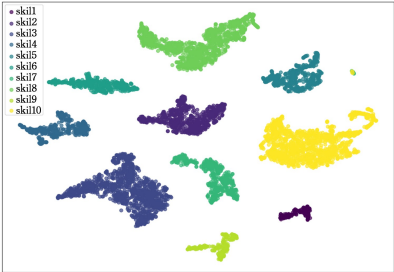


图 6 ASSIST12 数据集的试题聚类可视化

Fig. 6 Visualization of exercise clustering on ASSIST12 dataset

结束语 本文提出了一种基于试题-知识点异构图和多特征融合的知识追踪模型,综合研究了两种时间特征和两种行为特征对学习者的知识状态的影响。在公开的数据集上的实验结果表明,模型取得了较好的预测性能。在未来的研究中,可以考虑从以下两个角度入手,以改进模型。

1)预训练试题表示部分和知识追踪模型部分的训练是分开,训练比较繁琐,后续将考虑从联合损失函数角度出发,将两个训练过程联合训练。

2)学习者的学习是一个复杂的过程,知识状态的变化受多个因素共同影响,我们仅考虑了两种时间因素和两种行为因素,后续将考虑从试题的难度以及试题文本信息等角度出发,研究更多的因素对知识状态的影响。

参 考 文 献

[1] WANG Y,ZHU M X,YANG S H,et al. Review and Performance Comparison of Deep Knowledge Tracing Models[J]. Ru-an Jian Xue Bao/Journal of Software,2023,34(3):1365-1395.

[2] ABDELRAHMAN G,WANG Q,NUNES B. Knowledge tracing: A survey[J]. ACM Computing Surveys 2023,55(11):1-37.

[3] WU Z,LI M,TANG Y,et al. Exercise recommendation based on knowledge concept prediction [J]. Knowledge-Based Systems, 2020,210:106481.

[4] REN Y,LIANG K,SHANG Y,et al. MulOER-SAN:2-layer multi-objective framework for exercise recommendation with self-attention networks [J]. Knowledge-Based Systems, 2023, 260:110117.

[5] SONG X,LI J,CAI T,et al. A survey on deep learning-based knowledge tracing [J]. Knowledge-Based Systems, 2022, 258: 110036.

[6] ZHANG Y,YANG Q. A survey on multi-task learning [J]. IEEE Transactions on Knowledge and Data Engineering,2021, 34(12):5586-5609.

[7] DEVLIN J,CHANG M W,LEE K,et al. BERT:Pre-training of deep bidirectional transformers for language understanding [C]//Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1(Long and Short Papers). Minneapolis,MN: Association for Computational Linguistics,2019:4171-4186.

[8] HE K,ZHANG X,REN S,et al. Deep residual learning for image recognition [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR). Las Vegas, NV:IEEE,2016:770-778.

[9] SHEN S,LIU Q,CHEN E,et al. Learning process-consistent knowledge tracing [C]//Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining. 2021: 1452-1460.

[10] XU B,HUANG Z,LIU J,et al. Learning behavior-oriented knowledge tracing[C]//Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. 2023: 2789-2800.

[11] CORBETT A T,ANDERSON J R. Knowledge tracing: Modeling the acquisition of procedural knowledge [J]. User Modeling and User-adapted Interaction,1994,4:253-278.

[12] PAVLIK JR P I,CEN H,KOEDINGER K R. Performance Factors Analysis--A New Alternative to Knowledge Tracing[C]// Conference on Artificial Intelligence in Education: Building Learning Systems That Care:from Knowledge Representation to Affective Modelling. 2009:531-538.

[13] EMBRETSON S E,REISE S P. Item response theory[M]. Psychology Press,2013.

[14] PIECH C,BASSEN J,HUANG J,et al. Deep knowledge tracing [J]. Advances in Neural Information Processing Systems,2015, 25:505-513.

[15] LIU Y,YANG Y,CHEN X,et al. Improving knowledge tracing via pre-training question embeddings [C]// Proceedings of the 30th International Joint Conference on Artificial Intelligence(IJ-CAI). 2021:1577-1583.

[16] LIU Q,HUANG Z,YIN Y,et al. Ekt:Exercise-aware knowledge tracing for student performance prediction [J]. IEEE Transactions on Knowledge and Data Engineering,2019,33(1): 100-115.

[17] YAN Q Y,SI Y Q,YUAN G. Student-Problem Association Based Heterogeneous Graph Knowledge Tracing Model[J]. Acta Electronica Sinica,2023,51(12):3549-3556.

[18] 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. 2020:2330-2339.

[19] NAGATANI K,ZHANG Q,SATO M,et al. Augmenting knowledge tracing by considering forgetting behavior [C]//The World Wide Web Conference. 2019:3101-3107.

[20] ZHAO W,XIA J,JIANG X,et al. A novel framework for deep knowledge tracing via gating-controlled forgetting and learning mechanisms [J]. Information Processing & Management,2023, 60(1):103-114.

[21] WANG X,JI H,SHI C,et al. A survey on heterogeneous network representation learning [J]. IEEE Transactions on Knowledge and Data Engineering,2020,32(12):2306-2324.

[22] MIKOLOV T,SUTSKEVER I,CHEN K,et al. Distributed representations of words and phrases and their compositionality [J]. Advances in Neural Information Processing Systems,2013, 26:3111-3119.

[23] PEROZZI B,AL-RFOU R,SKIENA S. Deepwalk:Online learning of social representations [C] // Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2014:701-710.

[24] TANG J,QU M,WANG M,et al. Line:Large-scale information network embedding [C]//Proceedings of the 24th International Conference on World Wide Web. 2015:1067-1077.

[25] GROVER A,LESKOVEC J. node2vec:Scalable feature learning for networks [C]//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mi-

ning. 2016;855-864.

[26] DONG Y, CHAWLA N V, SWAMI A. metapath2vec: Scalable representation learning for heterogeneous networks [C]// Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2017;135-144.

[27] FALMAGNE J C, COSYN E, DOIGNON J P, et al. The assessment of knowledge, in theory and in practice [C]// IEMC'03 Proceedings: Managing Technologically Driven Organizations: The Human Side of Innovation and Change (IEEE Cat. No. 03CH37502). IEEE, 2003;609-615.

[28] BAKER R S, ROLL I, CORBETT A T, et al. Do performance goals lead students to game the system? [C]// International Conference on Artificial Intelligence in Education. 2005;57-64.

[29] ZADEH A, CHEN M, PORIA S, et al. Tensor fusion network for multimodal sentiment analysis [C]// Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. 2017;1103-1114.

[30] HOCHREITER S, SCHMIDHUBER J. Long short-term memory [J]. Neural Computation, 1997, 9(8):1735-1780.

[31] CHO K, VAN MERRIENBOER B, GULCEHRE C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation [C]// Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2014;1724-1734.

[32] LOFTUS G R. Evaluating forgetting curves [J]. Journal of Experimental Psychology: Learning, Memory, and Cognition, 1985, 11(2);397-406.

[33] WAN G M, ZHENG D, YE Z, et al. Deep Graph Library: A Graph-Centric, Highly-Performant Package for Graph Neural Networks [J]. arXiv:1909.01315, 2019.

[34] GLOROT X, BENGIO Y. Understanding the difficulty of training deep feedforward neural networks [C]// Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics. JMLR Workshop and Conference Proceedings, 2010;249-256.

[35] 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. 2018;1-10.

XIE Peizhong, born in 1968, Ph.D, associate professor. Her main research interests include signal processing and educational data mining.

(责任编辑:柯颖)