



计算机科学

COMPUTER SCIENCE

GTKT:融合联通主义学习和多层时序图Transformer的知识追踪模型

李佳豪, 荆军昌, 徐茜, 刘栋

引用本文

李佳豪, 荆军昌, 徐茜, 刘栋. [GTKT:融合联通主义学习和多层时序图Transformer的知识追踪模型](#)[J]. 计算机科学, 2026, 53(2): 78-88.

LI Jiahao, JING Junchang, XU Qian, LIU Dong. [GTKT:Knowledge Tracing Model Integrating Connectivism Learning and Multi-layer Temporal Graph Transformer](#) [J]. Computer Science, 2026, 53(2): 78-88.

相似文献推荐 (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)

[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>

[基于双层级对比学习的健壮知识追踪模型](#)

Robust Knowledge Tracing Model Based on Two-level Contrastive Learning
计算机科学, 2026, 53(2): 31-38. <https://doi.org/10.11896/jsjcx.250700196>

[ETF-YOLO11n:交通图像的多尺度特征融合目标检测方法](#)

ETF-YOLO11n:Object Detection Method Based on Multi-scale Feature Fusion for Traffic Images
计算机科学, 2025, 52(12): 150-157. <https://doi.org/10.11896/jsjcx.241200021>

[基于多特征融合与集成学习的风机叶片缺陷检测方法](#)

Multi-feature Fusion and Ensemble Learning-based Wind Turbine Blade Defect Detection Method
计算机科学, 2025, 52(6A): 240900138-8. <https://doi.org/10.11896/jsjcx.240900138>

[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
计算机科学, 2025, 52(3): 197-205. <https://doi.org/10.11896/jsjcx.240700151>

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

李佳豪¹ 荆军昌¹ 徐茜¹ 刘栋^{1,2}

1 河南师范大学计算机与信息工程学院 河南 新乡 453007

2 河南省教育人工智能与个性化学习重点实验室 河南 新乡 453007

(2208183032@stu.htu.edu.cn)

摘要 知识追踪(Knowledge Tracing,KT)是根据学习者在一定学习周期内的历史答题记录,构建其知识状态,并预测其未来回答问题情况。传统知识追踪研究主要以学习者行为序列为研究对象,忽略了知识之间的拓扑结构关系。近年来,基于知识静态图的知识追踪方法取得了一定进展,但未充分考虑学习者、问题和知识点之间的动态图结构关系,忽略了学习者知识掌握过程中潜在的关联信息,导致模型泛化能力和可解释性较弱。针对以上问题,提出融合联通主义学习和多层时序图 Transformer的知识追踪模型(Graph Transformer Knowledge Tracing,GTKT)。首先,以联通主义学习理论为指导,构建学习者时序子图用于刻画学习者的历史练习序列,提出时间感知分层子图采样策略,利用邻居共现编码器挖掘节点之间的潜在关联;其次,以学习和遗忘效应理论为依据,提出了一种多频带时间编码器,用于捕捉学习者回答问题的时间特性,构建学习者-问题-知识点交互信息的多特征融合模块;再次,构建了多层时序图 Transformer 学习者知识追踪预测模块,实现学习者知识状态的动态建模与预测;最后,在6个公共数据集上的实验结果表明,GTKT在预测学习者准确率方面优于主流知识追踪模型。

关键词 知识追踪;联通主义学习;教育理论;图 Transformer;多特征融合

中图分类号 TP391.6

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

LI Jiahao¹, JING Junchang¹, XU Qian¹ and LIU Dong^{1,2}

1 School of Computer and Information Engineering, Henan Normal University, Xinxiang, Henan 453007, China

2 Key Laboratory of Educational Artificial Intelligence and Personalized Learning of Henan Province, Xinxiang, Henan 453007, China

Abstract Knowledge Tracing(KT) aims to model learners' knowledge states based on their historical exercise records and predict their future performance. Traditional KT research primarily focuses on modeling learners' behavioral sequences while overlooking the topological structure among knowledge concepts. Although recent methods using static knowledge graphs have shown progress, they fail to adequately capture the dynamic graph-structured relationships among learners, questions, and knowledge concepts, thereby ignoring potential correlations in the knowledge acquisition process and limiting model generalizability and interpretability. To address these limitations, this paper proposes a Graph Transformer Knowledge Tracing(GTKT) model that integrates connectivism learning theory with a multi-layer temporal graph Transformer. Firstly, guided by connectivism learning theory, it constructs temporal learner subgraphs to represent historical exercise sequences, proposing a time-aware hierarchical subgraph sampling strategy and a neighbor co-occurrence encoder to discover latent node relationships. Secondly, based on learning and forgetting theories, it designs a multi-band temporal encoder to capture temporal characteristics in learning behaviors and builds a multi-feature fusion module integrating learner-question-knowledge concepts interactions. Thirdly, it develops a multi-layer temporal graph Transformer module for dynamic knowledge state modeling and prediction. Experimental results on six public datasets demonstrate that GTKT outperforms mainstream knowledge tracing models in predicting learner performance.

Keywords Knowledge tracing, Connectivism learning, Educational theory, Graph Transformer, Multi-features fusion

1 引言

人工智能技术的快速发展正推动教育领域向智慧化与精准化方向变革。智慧教育通过融合大数据、机器学习等技术,

构建了覆盖教学管理、课堂实践、学习评估的全链条智能体系。知识追踪(Knowledge Tracing,KT)^[1-2]作为智慧教育中一项核心动态建模任务,旨在通过学习者特定学习周期内的答题记录,解析其知识状态的全貌,并预测未来的答题表现,

到稿日期:2025-07-30 返修日期:2025-10-25

基金项目:国家自然科学基金(62072160)

This work was supported by the National Natural Science Foundation of China(62072160).

通信作者:刘栋(liudong@htu.edu.cn)

为教师精准指导学习过程和学习者个性化定制学习路径提供 关键依据^[3]。知识追踪的概览如图 1 所示。

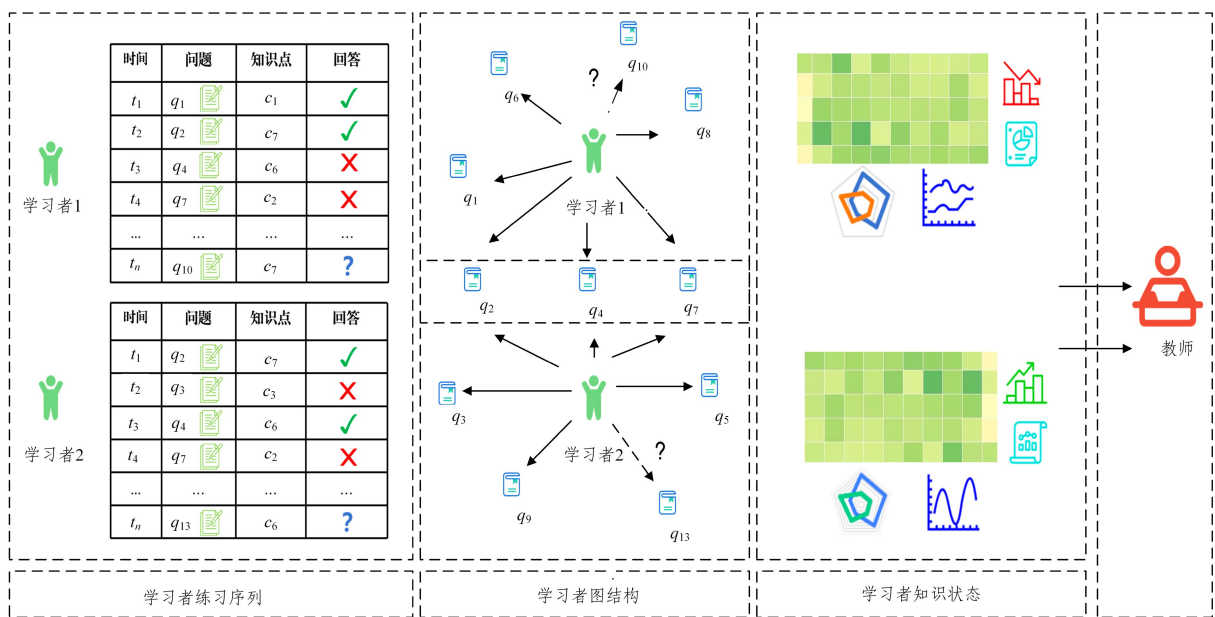


图 1 知识追踪示例图

Fig. 1 Example diagram of knowledge tracing

现有知识追踪方法主要包含两类,分别是基于学习者行为序列的知识追踪方法和基于图结构的知识追踪方法。前者将学习者的答题序列视为时序数据,结合其个体因素,直接建模知识状态的演变;后者将学习者、问题和知识点及其关系映射为图结构,依托其间的潜在关系建模知识状态。

基于学习行为序列的知识追踪方法早期主要使用教育心理学理论和概率模型刻画知识状态。例如,贝叶斯知识追踪(Bayesian Knowledge Tracing, BKT)^[4]利用隐马尔可夫模型模拟知识点掌握状态的隐变量变化。项目反应理论(Item Response Theory, IRT)^[5]引入问题难度与学习者能力参数。表现因子分析(Performance Factor Analysis, PFA)^[6]进一步纳入遗忘等因素。深度知识追踪^[7](Deep Knowledge Tracing, DKT)首次将循环神经网络(RNN)^[8]引入知识追踪,实现端到端的建模。AKT^[9], DTransformer^[10]等模型则利用 Transformer 架构^[11]的自注意力机制捕获问题间的长程依赖,克服了 DKT 等模型将问题视为同等难度进行处理的局限,显著提升了序列建模能力。

尽管基于学习行为序列的知识追踪方法取得了一定进展,但它们忽略了问题与知识点之间天然存在的先决条件,模型难以捕捉知识体系的拓扑特性,导致模型无法利用教育先验知识增强推理,并限制了模型可解释性。

近年来,基于图结构的知识追踪方法利用问题-知识之间存在的拓扑结构和图神经网络(GNN)可以聚合丰富局部特征的特性,凸显其在建模学习者知识状态的优越性。例如, GKT^[12]将知识结构转换为图的数据结构,首次将 GNN 引入知识追踪任务。另外, MFKT^[1], EKMFKT^[2], GAKT-IRT^[3], DGEKT^[13], GIKT^[14], SGKT^[15]和 DyGKT^[16]不仅建模问题-知识点之间的拓扑关系,而且与多种教育理论融合,构建更清晰的教与学的拓扑结构。

尽管基于图结构的知识追踪方法能够捕捉知识的拓朴

结构,但传统知识静态图建模难以充分刻画知识状态的动态演变过程,以及跨时间步的关联知识点连锁遗忘效应,难以适应真实教育场景中的知识关联动态性与异质性,限制了模型的泛化能力。

针对上述挑战,本文提出了融合联通主义学习和多层时序图 Transformer 的知识追踪模型 GTKT。强调联通主义学习理论中“知识网络化”“动态连接性”等观点,将学习者历史练习序列构建为时序子图,提出基于时间感知的分层子图采样策略,设计一种邻居共现编码器挖掘节点之间的潜在关联。同时,依据学习和遗忘效应理论设计一种基于多频带的时间编码器,用于捕捉练习的时间特性,构建学习者-问题-知识点交互信息的多特征融合模块。最终,构建了基于多层时序图 Transformer 学习者知识追踪预测模块,实现学习者知识状态的动态建模与预测。本文的主要贡献如下:

1)以联通主义学习理论为指导,提出了动态时序子图构建方法,包含时间感知分层子图采样策略及邻居共现编码器,有效挖掘节点关联;

2)以学习和遗忘理论为依据,设计了基于多频带的时间编码器,用于捕捉学习时间特性,构建异构多特征融合模块,整合知识点依赖、学习增益、遗忘效应等教育先验;

3)提出了基于多层时序图 Transformer 的知识追踪预测模型,实现精准动态知识状态建模;

4)在 6 个公开数据集上的实验结果表明, GTKT 在预测学习者表现方面,其 AUC 和 ACC 优于主流知识追踪模型 2.64%~6.84%和 1.42%~5.71%。

2 相关工作

知识追踪的核心目标是通过学习者的历史答题记录建模其动态知识状态并预测未来表现。现有研究主要沿着两大技术路线演进:基于学习行为序列的知识追踪和基

于图结构的知识追踪。本章将系统梳理这两类方法的研究进展。

2.1 基于学习行为序列的知识追踪

基于学习行为序列的知识追踪,其核心是将学习者的答题序列视为时间序列数据,直接建模知识状态的时序演化^[17]。知识追踪是从认知科学^[18]、迁移学习^[19]、增强学习^[20]、学习曲线^[21-23]和教育心理学^[24]等教学理论角度建模的,它将学习者蕴含在大脑中的知识体系量化为知识状态,并从过去的学习、作答、经验、遗忘^[22]、犯错^[4]等各种学习行为中持续更新知识状态。知识追踪依据研究者的技术路线可细分为基于概率模型的知识追踪和基于深度学习序列模型的知识追踪。

2.1.1 基于概率模型的知识追踪

早期知识追踪研究深受教育心理学和认知科学^[18]等启发,采用概率模型计算知识状态的变化。例如,Baker等^[4]提出BKT,利用隐马尔可夫模型,通过引入先验学习率、猜测、失误参数模拟知识状态的动态变化;Embretson等^[5]提出IRT,通过问题难度与学习者能力的线性组合预测正确率,但未考虑知识状态的时序依赖性;因子分析模型^[6]在IRT基础上通过多因子和潜在变量分解量化学习过程,以捕捉学习过程中的遗忘效应。这类方法虽具可解释性,但难以有效捕捉复杂知识状态的动态累积效应和序列长程依赖。

2.1.2 基于深度学习序列模型的知识追踪

近年来,基于深度学习的知识追踪模型发展迅速。例如,Piech等^[7]提出的DKT将学习者历史答题表现转化为时间序列输入至RNN^[8]/LSTM^[25]中,通过端到端学习从答题序列到知识状态的映射关系,显著提升预测性能。但DKT简单地假设学习者知识状态只有单个隐藏知识点。为了克服这一局限,DKT+^[26]通过正则化项增强损失函数,约束相似知识点问题的预测一致性。

Transformer的自注意力机制克服了RNN/LSTM在长序列依赖建模上的局限。例如,SAKT^[27]首次将注意力机制引入知识追踪任务,有效捕捉长序列依赖;AKT^[9]考虑了遗忘效应,进一步融合单调注意力机制和Rasch模型^[28],实现追踪过程的精细化表示;Liu等^[29]同样受到Rasch模型的启发,提出了simpleKT,通过建模题目差异捕捉包含相同知识点的问题之间的差异,采用点积注意力机制捕捉学习交互中的时间感知。

此外,研究者也从多种角度关注问题和知识点。例如,Shen等^[30]提出的LPKT通过监测学习者的学习增益和学习遗忘,观测学习者在连续学习进程中的变化,从而捕捉学习者知识状态的演变;Liu等^[31]提出的EKT引入文本内容,采用问题嵌入模块生成来提高模型的性能;Chen等^[32]提出的QIKT通过知识获取模块和问题解决模块获得了以问题为中心的认知表征,同时采用基于项目反应理论的预测层生成可解释预测结果;Cai等^[33]提出的MDKT-GCCF基于学习者知识水平和群体邻近信息,采用协同过滤算法挖掘群体学习偏好。但上述方法均未能显式建模拓扑结构。

现有方法虽在学习行为序列建模上取得进展,但多数基于学习序列的知识追踪假设单一隐藏状态或隐式处理知识

点,未能显式建模问题-知识点间的先决条件与拓扑结构,限制了模型利用教育先验知识进行推理的能力和可解释性。另外,尽管部分工作尝试融入遗忘效应,但整体上对认知科学、迁移学习、学习曲线等丰富教育规律的深度整合仍显不足。

2.2 基于图结构的知识追踪

图作为具有高度表达能力的数据结构,广泛用于表示各个领域的复杂数据,它们从数据中捕捉结构和语义信息,促进了社交媒体、知识图谱、推荐系统^[34]、问答系统^[35]、情感分析^[36]、文本生成^[37]和信息检索等任务的发展。

基于图结构的知识追踪将学习过程映射为“学习者-问题-知识点”的图结构,利用GNN聚合节点特征并更新知识状态。例如,Nakagawa等^[12]提出的GKT首次将GNN引入知识追踪,将知识点建模为图节点,通过图注意力网络(GAT)传播知识状态,通过预定义的静态邻接矩阵传播知识状态;Yang等^[14]提出的GIKT构建问题-知识点二部图,利用图卷积网络(GCN)更新知识状态;Liu等^[38]提出的PEBG通过预训练问题-知识点图学习问题嵌入;Wu等^[15]提出的SGKT采用了异构图建模,通过门控图神经网络更新知识状态;Cheng等^[16]提出的DyGKT首次将动态图引入知识追踪任务,通过GRU更新边权重以模拟知识状态演变。

部分研究者通过深入探究问题-知识点异构图,并尝试融合更多教育信息。例如,Yan等^[1]提出的MFKT聚焦多模态数据,通过预训练问题和知识点节点嵌入构建异构图,从中获取技能丰富度和问题区分度,差异化地反映不同模态的遗忘效应,并通过LSTM获取学习者知识状态;Xie等^[2]提出的EKMFKT预训练问题-知识点异构图,构建学习门和遗忘门来衡量学习者学习增益和遗忘,设计了一个时间域通道获取学习者知识状态;Dong等^[3]提出的GAKT-IRT获得异构图嵌入后,通过图注意力机制与IRT相结合,采用LSTM进行预测。

尽管以上方法采用了图结构构建学习者的知识结构,然而多数方法依赖静态图结构,难以充分捕捉知识状态随时间的动态演变,忽略了“学习者-问题-知识点”三者作为有机整体的关联,忽视了真实教育场景中的动态性,部分方法仍然未脱离基于线性时间序列的学习行为序列模型的本质,限制了模型的泛化能力。此外,尽管动态图方法有所尝试,但如何高效、精准地建模长时序、细粒度的知识状态动态交互,仍是亟待解决的问题。

3 问题描述

在一个在线学习系统中,每个学习者的学习活动主要由练习记录组成,即一系列的问题和相应的回答。对于一个学习者 l ,在时间步 t 内回答问题 $q_t^l \in \mathcal{Q}$,问题包含知识点 $c_t^l \in \mathcal{C}$,得到问题回答的响应 $r_t^l \in \{0, 1\}$,表示学习者是否正确地回答了这个问题。当 $r_t^l = 1$ 时,表示学习者正确回答了该问题;当 $r_t^l = 0$ 时,表示学习者对该问题的回答错误。将学习者回答问题看作学习者和问题之间的交互,通过学习者-问题-知识点之间的交互序列构建图,设交互图为:

$$G = \{(l, q_t^l, c_t^l, r_t^l), \dots, (l, q_T^l, c_T^l, r_T^l)\}$$

其中, L 是学习者的数量, T 是学习序列的长度, \mathcal{Q} 是所有问

题的集合, C 是所有知识点的集合。图包含节点特征矩阵 $\mathbf{X}_v \in \mathbb{R}^{N \times d_v}$ 、边特征矩阵 $\mathbf{X}_e \in \mathbb{R}^{M \times d_e}$ 、以及时间戳向量 $t \in \mathbb{R}^B$ 。 M 为交互次数, N 为问题与知识点节点总数, d_v 为问题难度、知识点层级等维数, d_e 为记录答题正误、响应时间等维数, B 为批次大小。知识追踪的任务目标是:

- 1) 追踪学习者的知识状态;
- 2) 预测学习者对下一个问题 q_{t+1} 的响应 r_{t+1} 。

表 1 列出了本文使用的部分公式符号,方便读者阅读。

表 1 数学符号和描述

Table 1 Mathematical notations and descriptions

符号	含义
$q_i^t r_i^t$	学习者 i 在 t 时间回答的问题、对应的知识点、回答
$w_{q,c}^t$	t 时间问题与知识点边权重
v_{src}, u_{dst}	源节点、目标节点
σ	非线性激活函数 <i>sigmoid</i>
T_{fused}	多频带时间衰减
H_{out}	学习者知识状态
\hat{y}	学习者正确回答问题的概率

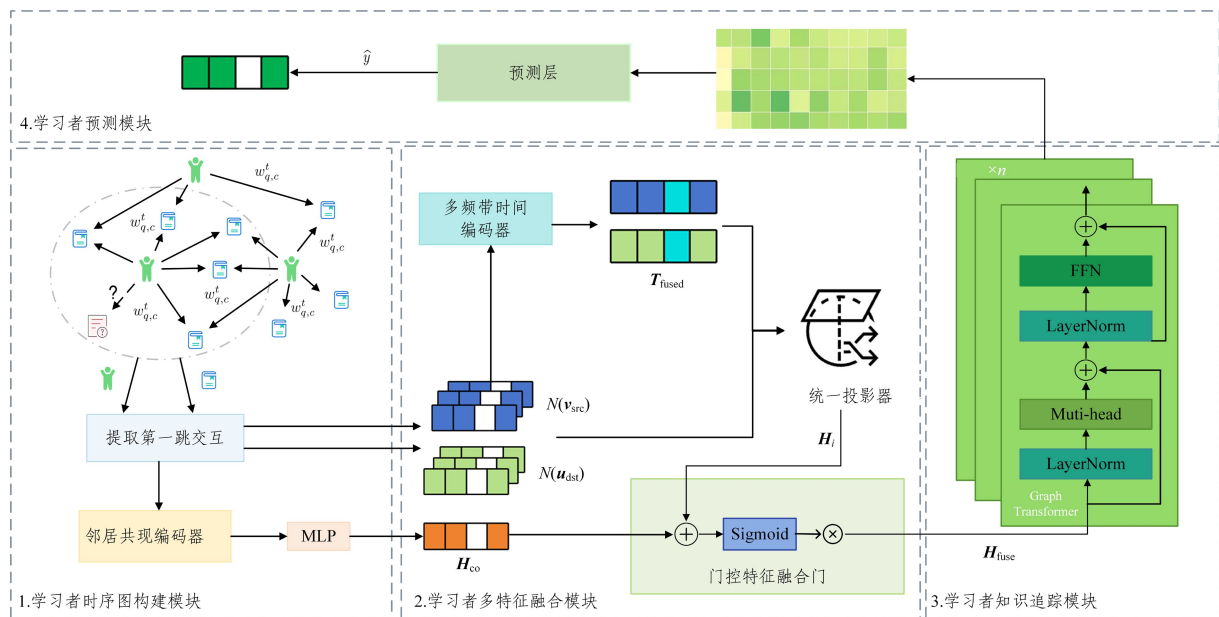


图 2 GTKT:融合联通主义学习和多层时序图 Transformer 的知识追踪模型

Fig. 2 GTKT: knowledge tracing combining connectivism learning and multi-layer temporal graph transformer

模型通过学习者时序图构建模块将学习者历史练习序列转化为学习者时序子图,通过时间感知子图采样后,采用邻居共现编码器挖掘节点之间的潜在关联;学习者特征融合模块通过多频带时间编码捕捉学习者练习的时间特性,并通过多特征融合门融合学习者-问题-知识点的丰富信息;学习者知识追踪模块采用多层次图 Transformer 层对学习者的知识状态进行编码,实现知识状态的动态演变;学习者预测模块最终实现学习者对下一个问题回答情况的预测。

4.1 学习者时序图构建模块

学习者时序图构建模块分为 3 部分来处理学习者时序图:学习者时序子图构建、时间感知子图采样、邻居共现编码器。

4.1.1 学习者时序子图构建

联通主义学习理论强调,在个体头脑与外部网络间通过

4 GTKT:融合联通主义学习和多层时序图 Transformer 的知识追踪模型

“学习就是连接”。在动态图的构建过程中,联通主义学习理论提供了可行的理论支撑。联通主义学习理论^[39-40]认为,学习是节点间连接的形成与优化过程,是一种群智协同、动态生成、螺旋进化的知识新形态。知识状态的本质是学习者对问题-知识点网络拓扑的认知映射。联通主义学习理论的核心观点包括知识网络化、动态连接性、多特征交互。结合联通主义学习理论研究基于图的知识追踪模型,能够更加强知识体系的网络化本质与动态连接特性,引导知识追踪任务更加符合教育的核心理念和学习者的实际需求。

鉴于 Transformer^[11]强大的表示能力和图表示学习优良的特征提取能力,本文提出的 GTKT(Graph Transformer-based Knowledge Tracing)是一种以联通主义学习理论为指导,融合动态时序图结构与多层图 Transformer 等模块的知识追踪框架,其核心架构如图 2 所示。

感知、交互等作用方式建立连接,以实现个体知识的增长与能力的发展^[39]。

受联通主义学习理论的启发,GTKT 首先将学习者历史练习序列处理为图的形式,采用时序联合图(Temporal Union Graph, TUG)建模实现“知识网络化”和“动态连接化”,节点分为学习者、问题与知识点,学习者的第一跳邻居节点为学习者回答的所有问题,问题的第一跳邻居包含问题包含的知识点。边包含响应时间、答题记录、依赖关系等信息,TUG 可动态更新与学习者关联的信息。

在每个时间步 t , TUG 的结构随学习者的交互进行动态演化。具体表现为,若问题 q_i 被学习者所回答,则 TUG 更新与其关联知识点 c_j 的边权重 $w_{q,c}^t$:

$$w_{q,c}^t = \sigma(\alpha \cdot r(q_i) + \beta \cdot \Delta t) \quad (1)$$

其中, σ 表示非线性激活函数 *sigmoid*; α, β 为可学习参数, Δt

为当前时间与学习者上一次练习相同知识点问题的时间间隔。若知识点 c_j 是 c_i 的先决条件, 则添加边 $e_{c_i \rightarrow c_j}$, 边权重通过共现频率动态调整。

4.1.2 时间感知子图采样

针对学习者长序列的学习历史交互, GTKT 采用基于时间感知的子图采样策略, 通过时间感知邻居采样的方法, 采用滑动窗口策略构建时序局部子图。对于每个节点 v_{src} , 根据时间窗口 $v_{v,u}$ 采样最近 I 次交互, 从历史记录中提取第一跳邻居 u_{dst} 并建立序列从中学习。具体做法是:

$$N = \{(v_{src}, e_{v,u}, u_{dst}, t_{v,u}) \mid u \in Top-I\} \quad (2)$$

其中, $e_{v,u}$ 为源节点学习者 v_{src} 到目标节点学习者回答的问题 u_{dst} 的边。设置 I 可在训练中降低内存消耗, 适配大规模学习者学习日志。为了确保批处理内的一致性和高效处理, 使用 padding 技术来标准化处理长度小于 I 的学习序列。

4.1.3 邻居共现编码器

联通主义学习理论的动态连接性认为, 学习行为会动态调整节点间的连接强度。GTKT 通过构建邻居共现编码器, 将源节点和目标节点序列中每个邻居出现的频率编码, 强化节点之间的关联。统计源节点 v_{src} 和目标节点 u_{dst} 的邻居共现频率, 映射为高阶交互模式:

$$H_{co} = MLP\left(\frac{|N(v_{src}) \cap N(u_{dst})|}{|N(v_{src}) + N(u_{dst})|}\right) \quad (3)$$

其中, $N(\cdot)$ 表示邻居集合, $MLP(\cdot)$ 为两层全连接网络, 采用 ReLU 激活函数。

联通主义本体论^[39]提出, 教育是不断建立网络节点之间的连接并保证信息畅通流动的联通过程。通过时序图的构建方法, 能够保证节点间建立连接, 学习者通过不断进行学习和问题回答, 从而影响自身知识状态变化。

4.2 学习者多特征融合模块

学习者多特征融合模块将学习者丰富的学习信息共同融合, 参与到学习者知识追踪的过程中, 使其更加符合真实教育场景。学习者多特征融合模块主要分为 3 个部分: 多频带时间编码器 (Muti-band Time Encoder, MTE)、统一投影器 (Unity-project Layer, UL)、门控特征融合门 (Feature Fusion Gate, FFG)。

4.2.1 多频带时间编码器

学习者某时刻回答问题后可能会复习正确答案或纠正错误, 也可能准备回答下一个问题, 在下次回答包含同样知识点的问题之前, 学习者会存在遗忘的情况, 即学习者的遗忘效应。

作为真实教育场景中最重要的因素之一, 诸多研究者结合艾宾浩斯遗忘曲线, 采用传统时间编码方式简单地将时间差 Δt 输入网络。然而这种做法存在一定缺陷, 记忆衰减可能遵循类似艾宾浩斯遗忘曲线的规律, 直接输入时间差很难让模型自己学会这种周期性模式, 且几分钟内的多次答题这样细微的时间变化和相隔几天的宏观时间变化应该有不同表示, 模型还需要耗费大量参数从时间差中学习复杂的时间函数。

GTKT 设计了多频带时间编码器, 结合学习者短期练习效应和长期遗忘规律等学习与遗忘特性, 通过采用多频带傅

里叶变换捕捉学习者复杂的时间衰减效应引入了遗忘过程:

$$T_i = \text{Concat}[\sin(\omega_k \Delta t), \cos(\omega_k \Delta t)]_{k=1}^{I_0} \quad (4)$$

$$T_{\text{fused}} = T_i + \gamma \cdot \sigma(W_d T_i) \quad (5)$$

其中, $\omega_k = e^{-\frac{k}{I_0} \log \tau}$ 为频率参数, T_i 控制衰减速率, k 为频带数, γ 为缩放因子, W_d 为投影矩阵。傅里叶级数表明, 任何复杂的周期性函数都可以分解为一系列简单正弦和余弦函数的加权和。多频带傅里叶变换编码提供一组能够表示复杂时间模式的基础函数, 通过不同的频带捕捉不同尺度的时间变化。对于任意时间差 Δt , $(\sin \Delta t, \cos \Delta t)$ 定义了一个唯一的点, 避免了不同 Δt 映射到相同编码的情况。时间 Δt 的微小变化, 会导致 (\sin, \cos) 点在单位圆上连续移动, 使得编码输出也是平滑变化的, 符合记忆随时间平滑衰减的遗忘特性。模型通过后续的线性层学习每个频带组件的权重, 从而拟合出最优的且与数据匹配的时间衰减函数。

4.2.2 统一投影器

联通主义学习理论认为, 知识并非孤立存在, 而是通过依赖、共现、语义关联等形成复杂网络, 并在个体头脑与外部网络间通过感知、交互等作用建立连接, 以实现个体知识的增长与能力的发展。

结合上述丰富的教育信息, GTKT 通过学习节点历史交互序列来训练, 并将每个节点的序列分割成多个片段, 用于捕捉长期的时间依赖性, 通过统一投影器将节点、边、时间等异质特征映射至统一空间:

$$H_i = W_i \cdot X_{\text{patch}}^i + b_i, i \in \{v, e, t, co\} \quad (6)$$

其中, $W_i \in R^{d_{\text{channel}} \times (P \cdot d_i)}$, $b_i \in R^{d_{\text{channel}}}$ 为可学习参数; d_{channel} 为投影至统一通道的通道维度; P 为分块数量; X_{patch}^i 为分块处理后的特征。

4.2.3 门控特征融合门

从统一投影器中得到学习者中间隐藏状态, 提取学习者时序图构建模块中邻居共现编码器计算出的共现隐藏状态, 将二者同时输入至门控机制动态聚合多特征:

$$G = \sigma(W_g \cdot \text{Concat}(H_i, H_{co})) \quad (7)$$

$$H_{\text{fuse}} = G \otimes H_i + (1 - G) \otimes (W_{co} \cdot H_{co}) \quad (8)$$

其中, σ 表示非线性激活函数 sigmoid, \otimes 为逐元素相乘。

4.3 学习者知识追踪模块

学习者知识追踪模块是 GTKT 的核心模块, 通过图 Transformer 层和预测层完成知识追踪的核心任务。

受到 Yu 等^[41]提出的动态图学习架构的启发, 依托该架构设计了基于图 Transformer 的动态图学习知识追踪层, 模型将多个图 Transformer 层进行合并, 通过结构-时序解耦, 分离时间维度和图结构注意力, 避免了传统 Transformer 在长序列中的过平滑化问题, 从而优化模型的性能。

图 Transformer 采用了多头注意力机制, 每层包含 h 个注意力头, 计算跨块全局依赖, 得到学习者的知识状态:

$$\text{head}_h = \text{Softmax}\left(\frac{Q_h K_h^T}{\sqrt{d_{\text{channel}}/H_{\text{fuse}}}}\right) V_h \quad (9)$$

$$H_{\text{GT}} = \text{GraphTransformer}(W_o \cdot (\text{head}_1, \dots, \text{head}_h)) \quad (10)$$

$$H_{\text{out}} = \text{LayerNorm}(H_{\text{GT}} + \text{dropout}(H_{\text{GT}})) \quad (11)$$

$$\mathbf{H}_{out} = LayerNorm(\mathbf{H}_{out} + dropout(GeLU(W_2 \cdot GeLU(W_1 \cdot \mathbf{H}_{out})))) \quad (12)$$

其中, $\mathbf{Q}_h, \mathbf{K}_h, \mathbf{V}_h$ 为第 h 头的查询、键、值矩阵; \mathbf{W}_0 为输出投影矩阵; $\mathbf{W}_1, \mathbf{W}_2$ 为可训练矩阵。

图 Transformer 的注意力机制包含丰富的时间编码后, GTKT 可以捕捉短期练习效应和长期遗忘规律等学习与遗忘特性, 符合学习者学习遗忘的规律。

4.4 学习者预测模块

知识追踪的核心任务是根据学习者的历史回答状况生成每一个时刻的知识状态, 而最终目的是预测下一个问题学习者的回答情况。学习者预测模块通过状态更新器提取学习者的知识状态:

$$\mathbf{h}_t = Update(\mathbf{W}_h \cdot \mathbf{H}_{out}) \quad (13)$$

其中, \mathbf{W}_h 为更新器参数, \mathbf{H}_{out} 为图 Transformer 输出。根据模型生成的学习者知识状态, GTKT 将源节点和目标节点嵌入拼接后映射至预测空间, 预测学习者未来回答问题的正确概率:

$$\hat{y} = \sigma(\mathbf{W}_{out} \cdot \mathbf{h}_t) + b_k \quad (14)$$

其中, $\mathbf{W}_{out} \in R^{2d_{channel} \times 1}$ 表示可训练输出矩阵, σ 表示非线性激活函数 sigmoid, b_k 表示偏置。

4.5 模型训练

通过训练 GTKT 预测得到的回答概率, 采用二值交叉熵损失函数与学习者的真实回答记录比较, 从而进行损失估计。

$$Loss_{GTKT} = - \sum_{i=1}^T (r_i \log y_i + (1 - r_i) \log(1 - y_i)) \quad (15)$$

采用的 AdamW^[42] 优化方法是 Adam^[43] 优化器的改进版本, 其主要改进在于, 提出了解耦权重衰减的权重衰减处理方式, 能够更好地控制权重的幅度, 提高模型的训练效率和泛化能力, 减少过拟合的风险, 解决了原始 Adam 优化器在权重衰减方面的一些问题, 并且对超参数的选择更鲁棒。

5 实验分析

本章旨在全面评估 GTKT 模型。首先介绍实验所用的数据集和基本设置, 随后, 通过一系列实验回答以下关键研究问题。

RQ1: GTKT 在预测任务上是否优于最先进的知识追踪模型?

RQ2: GTKT 的核心模块设计是否有效提升了模型性能和可解释性?

RQ3: GTKT 所使用的核心架构是否优于其他模型?

RQ4: 模型中的超参数对 GTKT 模型性能产生何种影响?

RQ5: GTKT 建模的学习者知识状态是否能够反映真实

教育场景中的各类教学理论?

5.1 实验数据

为全面评估 GTKT 在不同学习场景下的泛化能力, 在 6 个广泛使用的公开真实数据集上进行了实验。这些数据集涵盖了多样化的学习者和知识领域, 其关键统计信息如表 2 所列。具体数据集描述如下。

1) ASSISTment2009¹⁾: 从 ASSISTment 教育平台收集到的在 2009—2010 学年的学习者答题记录, 采用了 skill-builder 版本。

2) ASSISTment2012²⁾: 从 ASSISTment 教育平台收集到的在 2012—2013 学年的学习者答题记录。

3) ASSISTment2017³⁾: 该数据集来自于 2017 ASSISTments Longitudinal Data Mining 竞赛, 其特点是学习者序列普遍较长, 且允许学习者在同一问题上进行多次尝试直至答对。

4) Junyi⁴⁾: 源自 2015 年基于 Khan Academy 发布的开源代码构建的 Junyi Academy 电子学习平台, 包含详细的问题日志, 采用了 junyi_ProblemLog_original.csv 版本。

5) Algebra2005⁵⁾: 在 KDDcup 2010 教育数据挖掘挑战上提出, 包含了 2005—2006 年学习者对代数问题的回答。

6) Algebra2006⁶⁾: 在 KDDcup 2010 教育数据挖掘挑战上提出, 包含了 2006—2007 年学习者对代数问题的回答。

表 2 数据集统计信息

Table 2 Dataset statistical information

数据集	学习者数量	问题数量	知识点数量	练习总个数
ASSISTment2009	4 217	26 688	123	346 860
ASSISTment2012	27 485	53 065	265	2 711 813
ASSISTment2017	1 700	3 200	102	9 428 000
Junyi	721	191 874	48	25 853 987
Algebra2005	574	1 084	138	809 694
Algebra2006	1 840	549 165	1 701	2 289 726

数据集预处理方面, 对于每个数据集, 按 8:1:1 的比例将数据集划分为训练集、验证集和测试集, 选取学习者作为源节点, 其回答的问题作为目标节点, 选取回答题目初始作答时间作为时间戳, 选取回答问题正误作为标签, 建立边索引确定交互总次数。删除了上述字段为空的记录, 同时删除了数据集中知识点字段为空的记录。

由于 Algebra2005 和 Algebra2006 中并未体现学习者回答问题 ID 的字段, 而是用字段“problem_name”和“step_name”共同表示, 因此通过构建复合唯一标识符来表示学习者回答的问题。

5.2 基线模型

为全面评估 GTKT 的性能, 本文选取以下基线模型进行对比。在选取基线模型时, 本文充分考虑设计 GTKT 的特

¹⁾ <https://sites.google.com/site/assistmentsdata/home/assistment-2009-2010-data/skillbuilder-data-2009-2010>

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

³⁾ <https://sites.google.com/view/assistmentsdatamining/dataset>

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

⁵⁾ <https://pslclatashop.web.cmu.edu/KDDCup/Download?datasetId=1>

⁶⁾ <https://pslclatashop.web.cmu.edu/KDDCup/Download?datasetId=2>

点,依据现有知识追踪研究主要的两大技术路线,即基于学习行为序列的知识追踪和基于图结构的知识追踪,并从研究者广泛选取的经典模型和近年来性能表现良好的模型中,共选取 11 个代表性基线模型进行对比。

1) 基于学习行为序列的知识追踪

DKT^[7]:首个基于深度学习的知识追踪模型,使用单层 LSTM 模型来预测学习者表现。

AKT^[9]:融合情境感知注意机制和 Rasch 模型,建模遗忘效应。

LPKT^[30]:关注学习者的学习过程和知识状态更新。

QIKT^[32]:通过知识获取和问题解决双模块学习问题认知表示。

SimpleKT^[29]:结合 Rasch 模型对与时间感知点积注意力。

2) 基于图结构的知识追踪

GKT^[12]:首个基于 GNN 的知识追踪模型,利用 GAT 传播知识状态。

SGKT^[15]:采用异构图和门控图神经网络建模学习交互。

DGEKT^[13]:通过知识点关联超图和有向转移图捕获序列关系,采用 GRU 更新学习者知识状态。

EKMFKT^[2]:构建问题-知识点异构图,构建学习门和遗忘门衡量学习者学习增益和遗忘,设计了一个时间域通道获取学习者知识状态。

GAKT-IRT^[3]:将图注意力机制与 IRT 相结合,采用 LSTM 预测学习者知识状态。

DyGKT^[16]:引入动态图建模答题序列,采用 GRU 更新学习者的知识状态。

5.3 实验设置

模型基于 PyTorch^[44] 和 DyGLib^[41] 框架实现了 GTKT 及基线模型。实验将邻居节点数设置为 200,使用训练集来训练模型,并使用验证集来调整超参数,在所有实验中使用了 AdamW^[42] 优化器,学习速率设置为 0.003,训练、验证和测试时的批量大小固定为 2000,嵌入维度设置为 64,dropout 设为 0.1。此外,还对模型的权值进行了衰减为 10^{-5} 的 L2 正则化处理。在训练中设置了早停机制,如果模型在验证集上的表现没有提升超过 20 个 epoch,则提前停止训练,以避免过拟合。

由于基线模型中,不少知识追踪模型中学习者输入序列长度不同,因此统一将所有序列设置固定长度为 200。由于在设计模型时添加了滑动窗口以进行邻居采样,在后文针对不同的滑动窗口长度进行了评估(详见对 RQ4 的回答)。

对于评估知识追踪模型的有效性,本文使用 AUC 和 ACC 作为评估预测性能的指标。AUC 被定义为 ROC 曲线与下坐标轴围成的面积,0.5 的 AUC 值表示随机猜测获得的预测性能,高 AUC 值表明模型具有较高的预测性能。ACC 为准确率,即正确预测结果占全部结果的百分比,高 ACC 值表明模型具有较高的预测性能。

5.4 实验结果与分析

为了评估 GTKT 的有效性,根据实验结果回答 5 个问题。

RQ1:GTKT 在预测任务上是否优于最先进的知识追踪模型?

为了评估知识追踪模型的有效性,在学习者未来表现预测任务中,将 GTKT 模型与所有的基线模型进行了比较,进行了 5 次实验并计算了平均结果^[16,45-46],最优结果以粗体显示,次优结果以下划线标记。实验结果如表 3 所列。

表 3 GTKT 与其他知识追踪方法性能比较

Table 3 Performance comparison of GTKT with other knowledge tracing methods

model	ASSISTment2009		ASSISTment2012		ASSISTment2017		Junyi		Algebra2005		Algebra2006	
	AUC	ACC	AUC	ACC	AUC	ACC	AUC	ACC	AUC	ACC	AUC	ACC
DKT	0.5713	0.6596	0.6342	0.7798	0.7028	0.6058	0.5874	0.8352	0.6439	0.8459	0.6731	0.8581
AKT	0.5915	0.6733	0.6430	<u>0.7899</u>	0.7348	<u>0.6579</u>	0.6688	0.8321	0.6561	0.8506	0.6711	0.8344
LPKT	0.6078	0.6253	0.6084	0.7228	0.6355	0.5715	0.6058	0.8409	0.6344	0.7958	0.6109	0.8698
QIKT	0.5965	0.6611	0.6288	0.7856	0.6572	0.5945	0.5924	0.8399	0.6547	0.8530	0.6615	0.8565
SimpleKT	<u>0.6256</u>	<u>0.7192</u>	0.6694	0.7875	0.7465	0.6247	0.6487	0.8092	0.6569	<u>0.8540</u>	<u>0.6812</u>	0.8309
GKT	0.6266	0.7090	<u>0.6861</u>	0.7733	0.7283	0.6298	0.6294	0.8194	<u>0.6723</u>	0.8212	0.6655	0.8398
SGKT	0.6048	0.6476	0.6472	0.7606	0.7364	0.6374	0.6114	0.8334	0.6398	0.8021	0.6532	0.8091
DGEKT	0.6193	0.6618	0.6384	0.7770	0.6858	0.6060	0.6378	0.8226	0.6372	0.8452	0.6615	0.8322
EKMFKT	0.6198	0.6358	0.6738	0.7133	0.7243	0.6286	0.6169	0.8341	0.6318	0.7956	0.6137	0.8706
GAKT-IRT	0.6133	0.6514	0.6465	0.7302	0.6974	0.6123	0.6268	0.8357	0.6268	0.8357	0.6094	0.8669
DyGKT	0.6252	0.6327	0.6494	0.7685	<u>0.7672</u>	0.6471	0.6261	<u>0.8419</u>	0.6642	0.8078	0.6162	<u>0.8718</u>
GTKT	0.7164	0.7456	0.7385	0.7980	0.7747	0.7044	0.7361	0.8522	0.7312	0.8623	0.7288	0.8796

结果表明:

1)GTKT 的性能优于对比的所有基线模型。在进行对比实验时,所有基线模型均在与 GTKT 相同的数据集划分、数据预处理及评估指标下进行训练与测试,以确保所有对比实验的公平性与一致性。

2)与基于学习行为序列的方法相比,GTKT 的 AUC 和 ACC 在 6 个数据集上分别提升了 2.82%~14.80% 和 1.82%~11.91%。这说明 GTKT 能够捕捉更复杂的“学习者-问题-知识点”拓扑结构,可以很好地处理基于图结构的知识追踪任务。

3)与基于图的知识追踪模型相比,GTKT 的 AUC 和 ACC 在 6 个数据集上分别提升了 3.83%~8.98% 和 0.42%~

6.86%,证明了 GTKT 中设计的学习者知识追踪模块中多层次时序图 Transformer 相较于其他核心架构更优秀。同时,与基于静态图结构的模型相比,采用动态图结构的 GTKT 和 DyGKT 的性能表现更优,这说明动态图结构可以更好地捕获学习者随时间不断增长的知识状态,符合联通主义学习理论中指出的“动态连接性”。

4)尽管在部分较大规模数据集上,GTKT 在 AUC 指标上的绝对提升幅度为 1%~2%,但提升在真实教育场景中仍具有实际意义和应用价值。对于知识追踪任务而言,其核心主要是反映学习者随时间变化的知识状态,为学习者过滤掉其已经掌握知识点的冗余题目,并优先推荐其最需要练习的薄弱环节。

5)模型的计算效率与资源消耗是评估其实用性与可部署性的关键指标,实验在 ASSISTment2017 数据集上固定其他超参数,仅调整图 Transformer 的层数,统计其参数量及单个训练轮次的平均耗时,结果如表 4 所列。

表 4 不同层数图 Transformer 的参数量和单个 epoch 训练时间

层数	参数量	每个 epoch 训练时间/s
Layer=1	0.962×10^6	202.9
Layer=2	1.719×10^6	347.2
Layer=3	2.475×10^6	465.6

可以看出,从 1 层到 3 层,参数量增长约 1.57 倍,训练时间增长约 1.3 倍,AUC 和 ACC 均有不同程度增长。这表明模型虽然引入了相对复杂的多层图 Transformer 结构,但其资源消耗在可控范围内,且带来的性能提升证明这种计算开

表 5 GTKT 消融实验性能比较(使用配对 t 检验,显著性水平为 0.01)

Table 5 Performance comparison of GTKT ablation experiment(using paired t-test at 0.01 significance level)

model	ASSISTment2009		ASSISTment2012		ASSISTment2017		Junyi		Algebra2005		Algebra2006	
	AUC	ACC	AUC	ACC	AUC	ACC	AUC	ACC	AUC	ACC	AUC	ACC
w/o MTE	0.6534*	0.7482	0.6485*	0.7334*	0.7045*	0.7032	0.7031*	0.8227*	0.7077*	0.8338*	0.6559*	0.8310*
w/o UL	0.6432*	0.6639*	0.6483*	0.7121*	0.6342*	0.7037	0.6439*	0.7824*	0.6419*	0.8101*	0.6661*	0.8321*
w/o FFG	0.6209*	0.6537*	0.6112*	0.6983*	0.6639*	0.7045	0.6121*	0.8201*	0.6213*	0.7913*	0.6439*	0.8166*
w/o M&U	0.6323*	0.6598*	0.6543*	0.7349*	0.6434*	0.7012	0.6188*	0.7983*	0.6491*	0.8122*	0.6451*	0.8008*
w/o M&F	0.6398*	0.6748*	0.6584*	0.7337*	0.6129*	0.6811*	0.6309*	0.7831*	0.6441*	0.8091*	0.7104	0.7941*
w/o U&F	0.6012*	0.6232*	0.6345*	0.7328*	0.5626*	0.6721*	0.6232*	0.7992*	0.6296*	0.7997*	0.6369*	0.8339*
GTKT	0.7164	0.7456	0.7385	0.7980	0.7747	0.7044	0.7361	0.8522	0.7312	0.8623	0.7288	0.8796

分析表 5 中的实验结果可知:

1)w/o MTE 在 AUC 上有一定程度下降,平均 AUC 降幅约 7.15%,这是因为 MTE 主要捕捉时间衰减效应,通过生成短期或长期的时间间隔特征,提高模型性能,去除 MTE 模块会导致模型时间建模能力下降。

2)w/o UL 在 AUC 上有一定程度的下降,平均 AUC 降幅约 6.83%,这表明 UL 层的特征空间对齐对追踪学习者学习状态具有较好的表现,统一有助于异构特征融合,显著提升鲁棒性,而去除 UL 模块会导致特征对齐能力减弱。

3)w/o FFG 对于模型有显著性能下降,平均 AUC 降幅约 8.72%,这表明 FFG 有效地聚合了子图中学习者-问题-知识点节点之间包含的丰富信息,对于学习者所尝试回答的问题及其相应知识点的掌握情况建模具有良好的效果,而去除 FFG 会导致静态特征融合无法适应不同学习者-问题-知识点的动态交互。

4)w/o MTE&UL 性能低于单独移除 MTE 或 UL,表明时间编码和特征对齐具有协同作用。

5)w/o MTE&FFG 性能降幅较大,这表明 FFG 可降低

销是合理且必要的。

RQ2:GTKT 各个模块的设计是否为模型的性能和可解释性带来提升?

为了回答该问题,主要对 GTKT 中涉及的 3 个模块分别进行了消融实验,验证多频带时间编码器(MTE)、统一投影层(UL)、门控特征融合门(FFG)在 GTKT 中的有效性。为了进一步验证上述消融实验的准确性,同时进行了单个模块和两个任意模块的消融,并在上述 6 个数据集上评估了不同变体的性能,结果如表 5 所列,w/o 代表去除了某个模块,最优结果以粗体显示,次优结果以下划线标记。

为验证性能提升的统计显著性,实验进行了配对 t 检验。设定显著性水平为 0.01,针对 GTKT 与每个数据集上的各个消融模型进行了比较,结果如表 5 所列,* 表示模型的性能达到了统计显著性水平($p < 0.01$)。

UL 缺失的影响。

6)w/o UL&FFG 性能最差,因为 UL 和 FFG 共同构建了特征融合的基础框架,二者缺失导致模型退化为最基础的拼接,进而导致模型的泛化能力降低。

7)尽管不同数据集存在差异,但结果均证明了多频带时间编码器(MTE)、统一投影层(UL)和门控特征融合门(FFG)在模型中发挥的重要性,GTKT 在所有数据集中表现稳定且可靠。

RQ3:GTKT 所使用的核心架构是否优于其他模型?

接下来将探讨不同核心架构在知识追踪上的性能,比较图 Transformer 与知识追踪中常用的 4 种核心架构 LSTM,Transformer,AKT-Attention,GRU,分别表示为 GTKT-L,GTKT-T,GTKT-A,GTKT-G。对于上述非图结构的核心架构,参照研究工作的最优性能设置,并将预处理后的数据集中所代表的学习者、问题、标签等数据输入至模型中进行实验,以便适应不同架构的模型。实验将 GTKT 中的核心架构进行了更换,4 种架构选取 RQ1 中性能表现最优的进行更换,结果如表 6 所列。

表 6 GTKT 更换核心架构性能比较

Table 6 Performance comparison of GTKT replaced core architecture

model	ASSISTment2009		ASSISTment2012		ASSISTment2017		Junyi		Algebra2005		Algebra2006	
	AUC	ACC	AUC	ACC	AUC	ACC	AUC	ACC	AUC	ACC	AUC	ACC
GTKT-L	0.5384	0.6397	0.6328	0.6263	0.6348	0.7168	0.6093	0.7583	0.5958	0.8063	0.5676	0.8143
GTKT-T	0.6454	0.7052	0.7096	0.6594	0.7392	0.6564	0.6839	0.8078	0.6719	0.8237	0.6898	0.8391
GTKT-A	0.6658	0.7389	0.7230	0.7459	0.7548	0.6519	0.7588	0.7921	0.7050	0.8161	0.7011	0.8304
GTKT-G	0.5426	0.6394	0.6261	0.7183	0.6283	0.6204	0.6294	0.7809	0.6623	0.7813	0.6455	0.8029
GTKT	0.7164	0.7456	0.7385	0.7980	0.7747	0.7044	0.7361	0.8522	0.7312	0.8623	0.7288	0.8796

可以得出以下结论:

1)从整体性能比较来看,性能排序从好到差依次为:图Transformer > AKT-Attention > Transformer > GRU \approx LSTM,其中以图Transformer为核心架构的模型在大多数情况下表现最佳。这表明所提出的图Transformer架构适合基于图的知识追踪。

2)AKT-Attention和Transformer优于LSTM和GRU建模,这与以往的研究结果^[7]一致,且通过RQ1中DGEKT和DyGKT基线模型的对比(因为二者核心架构也采用了GRU)也可得出该结论。这可能是由于基于图的知识追踪更适合基于注意力机制的架构,以便其分配注意力权重。尽管两种基线方法也同样表现出良好的学习者知识状态预测性能,但基于RNN的核心架构更符合连续性学习者序列和知识状态更新的特点,不能很好地处理基于图的知识追踪模型。

3)AKT-Attention始终优于Transformer,突显了考虑学习者遗忘行为的重要性。

RQ4:模型中的超参数对GTKT模型性能产生何种影响?

为了更好地解释GTKT模型的合理性,这里介绍了敏感性分析,包含不同子图邻居节点数 I 和MTE中不同频带数 k 对GTKT的影响。

子图的邻居节点数是基于图的知识追踪模型中的一项重要因素,因为它与学习者历史交互序列的长度有密切的关系,且包含来自邻居的有价值信息与噪声。GTKT设计了滑动窗口进行子图采样,将学习者序列控制在一定范围内,这一部分探究子图邻居节点数(即滑动窗口长度)对模型性能的影响。

本文测试了不同子图邻居节点数 I 在6个模型中的敏感性,结果如图3所示。实验所采用的数据集为ASSISTment2017,其特点是学习者序列普遍较长,且允许学习者在同一问题上进行多次尝试直至答对。

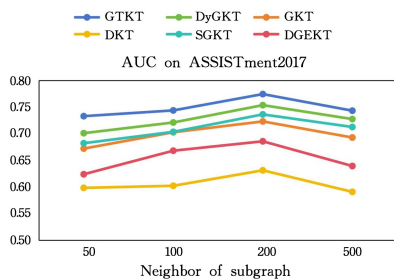


图3 子图邻居节点对GTKT的影响

Fig. 3 Effect of subgraph neighbor nodes on GTKT

通过图3可得出以下结论:

1)随着邻居序列变长,GTKT的性能随邻居节点数增长而上升,到达200时最好,但200后性能随之下降;

2)当 $I < 200$ 时,模型主要聚焦强关联,降低了噪声影响,但忽略了学习者知识状态的长距离依赖,容易导致过拟合;

3)当 $I > 200$ 时,模型能够覆盖全局依赖,反映出能够关注学习者更长时间的学习状况,但计算成本高,且容易引入大量弱相关的噪声;

4)当 $I = 200$ 时,模型能平衡广度与深度,更好地关注

学习者学习过程中的知识状态变化。

MTE将时间衰减规律以数学模型的形式注入模型中,降低了模型的学习难度,提供了可观的时序推理能力。这同样解释了为何在消融实验中,移除MTE会导致模型性能下降。而MTE中的频带数是区分学习者短期练习效应和长期遗忘规律的一项重要因素,GTKT通过设置不同频带数,改变模型在低、中、高频3个状态下对学习者的知识状态的预测状况。

为了深入探究时间编码器的作用,设计实验分析了模型在不同时间间隔下的预测性能,如图4所示,测试了在6个不同数据集下,不同频带数对模型的敏感性。

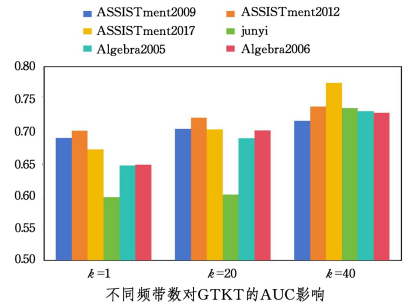


图4 不同频带数对GTKT的AUC影响

Fig. 4 Effect of different frequency bands on AUC of GTKT

通过图4可以得出以下结论:

1)6个数据集上的AUC随频带数减小均有不同程度减小。 $k=1$ 时,模型性能具有不错的表现,但其性能随着时间间隔的增大而下降,表明其无法有效建模长期依赖。 $k=20$ 的性能表现次优,这可能是因为 $k=1$ 和 $k=20$ 时MTE处理的时间范围仅涉及短回答,忽略了长时间回答间隔对学习者的遗忘程度的影响。而 $k=40$ 时,GTKT模型在所有时间尺度上都保持了最优的性能,这证明40个频带提供了足够的时间尺度来同时捕捉短期练习效应和长期遗忘规律。

2)增加频带数对模型性能有持续的正向影响。在6个不同数据集上, $k=40$ 相比 $k=1$ 和 $k=20$ 分别带来了5.03%~1.24%的AUC提升。这表明更细粒度的多尺度时间编码能更有效地捕捉学习过程中的复杂时间依赖关系。

RQ5:GTKT所构建的学习者知识状态是否符合真实教育场景中的各类教学理论?

为了更好地理解GTKT所构建的学习者知识状态是如何演变的,本节进行了个案分析。采用学习者知识状态热力图的可视化方法展示,通过单个学习者知识状态的演变,体现GTKT模型的学习者个体适应性,其能够更好地适应真实教育场景中各类教学理论。

在实验中,选取了两位学习者14和824的真实学习序列回答进行采样,并分别在图5(a)和图5(b)中展示了使用GTKT对每个知识概念掌握水平演变的估计结果。

图5中,每一行分别显示了在一定学习序列内单个概念的知识状态演变估计值,图的正上方和左侧的圆圈显示了每一行或列所关联的知识概念,不带白色点的圆圈表示学习者对这个问题的回答是正确的,带有白色点的圆圈表示学习者对这个问题的回答是错误的。从知识状态热力图中可以看出:

1)随着时间的推移,学习者会逐渐忘记之前学到的内容,

知识状态随时间衰减和累积努力考虑了遗忘曲线、学习曲线等教育理论;

2)从学习者 824 在 $t=14\sim t=27$ 时的知识状态可以看出,学习者对于回答错误的问题,知识掌握程度并不是简单地下降,是根据问题-知识点的相关性与依赖程度而有条件地下降甚至是上升,模型能够生成稳定且符合教育理论的知识状态。

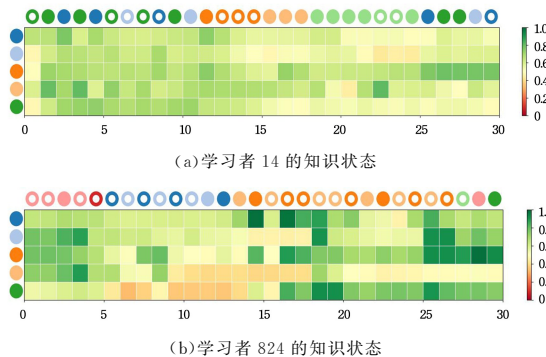


图5 不同学习者的知识状态

Fig.5 Knowledge state of different learners

结束语 本文提出了融合联通主义学习和多层时序图 Transformer 的知识追踪模型,综合研究了如何建模知识体系的拓扑结构和在真实教育场景中图模型捕捉知识状态的动态演变与异质性关联。在公开数据集上的实验表明,模型在预测学习者准确率方面优于主流知识追踪模型。在未来的研究中,可以考虑从以下角度改进模型。

1)针对构建学习者时序子图仅考虑动态角度,尚未考虑学习者情感、学习风格等学习者个体因素,以及学习者在答题时可能潜在的多模态类型题目,后续将考虑融入更多因素,如学习者个体因素、多模态数据,以及将静态的个体特征、多模态数据特征与动态的学习行为相结合,使模型更具学习者个体适应性和多模态数据适应性。

2)尽管本文以联通主义学习理论为启发,强调了“知识网络化”和“动态连接性”,但当前模型主要聚焦于“学习者-问题-知识点”三者交互结构的动态演化,对理论中“信息畅通流动”“连接优化”等更深层次的内涵体现仍显不足。

3)探索“学习者-问题-知识点”构图多样性。当前图结构仅局限于有向有权边,然而学习者和问题及知识点可能存在无向有权边、有向无权边、无向无权边等情况,且目前仅捕捉图的单跳邻居节点,未能充分挖掘多跳邻居所蕴含的丰富拓扑信息。未来工作可探索更具解释性的动态图构建机制。

4)本文虽通过个案热力图展示了知识状态的演变,但可解释性分析仍不够系统,缺乏对不同群体的对比分析,未来需采用更系统化的可解释性评估框架进行总体评估。

参考文献

[1] YAN Q Y, SUN H, SI Y Q, et al. Multimodality and Forgetting Mechanisms Model for Knowledge Tracing [J]. Computer Science, 2024, 51(7): 133-139.

[2] 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.

[3] DONG Y F, HUANG G, XUE W R, et al. Graph Attention Deep Knowledge Tracing Model Integrated with IRT [J]. Computer Science, 2023, 50(3): 173-180.

[4] BAKER R S J D, CORBETT A T, ALEVEN V. More Accurate Student Modeling Through Contextual Estimation of Slip and Guess Probabilities in Bayesian Knowledge Tracing [C] // Proceedings of the 9th International Conference on Intelligent Tutoring Systems (ITS). 2008: 406-415.

[5] EMBRETSON S E, REISE S P. Item Response Theory [M]. New York: Psychology Press, 2013.

[6] PAVLIK P I, CEN H, KOEDINGER K R. Performance Factors Analysis - A New Alternative to Knowledge Tracing [C] // Proceedings of the 2009 Conference on Artificial Intelligence in Education (AIED). 2009: 531-538.

[7] PIECH C, BASSEN J, HUANG J, et al. Deep Knowledge Tracing [C] // Proceedings of the 28th Conference on Neural Information Processing Systems (NeurIPS). 2015: 505-513.

[8] LIPTON Z C, BERKOWITZ J, ELKAN C. A Critical Review of Recurrent Neural Networks for Sequence Learning [J]. arXiv: 1506.00019, 2015.

[9] 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 (KDD). ACM, 2020: 2330-2339.

[10] YIN Y, DAI L, HUANG Z Y, et al. Tracing Knowledge Instead of Patterns: Stable Knowledge Tracing with Diagnostic Transformer [C] // Proceedings of the 26th ACM Web Conference 2023 (WWW). ACM, 2023: 855-864.

[11] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is All You Need [C] // Proceedings of the 31st Conference on Neural Information Processing Systems (NeurIPS). ACM, 2017: 6000-6010.

[12] NAKAGAWA H, IWASAWA Y, MAEDA Y. Graph-based Knowledge Tracing: Modeling Student Proficiency Using Graph Neural Network [C] // Proceedings of the 2019 IEEE/WIC/ACM International Conference on Web Intelligence (WI). 2019: 156-163.

[13] CUI C R, YAO Y M, ZHANG C Y, et al. DGEKT: A Dual Graph Ensemble Learning Method for Knowledge Tracing [J]. ACM Transactions on Information Systems, 2024, 42(3): 1-24.

[14] YANG Y, SHEN J J, QU Y R, et al. GIKT: A Graph-based Interaction Model for Knowledge Tracing [C] // Proceedings of the 2020 European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML-PKDD). 2020: 299-315.

[15] 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.

[16] CHENG K, PENG L Z, WANG P Y, et al. DyGKT: Dynamic Graph Learning for Knowledge Tracing [C] // Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD). ACM, 2024: 409-420.

[17] CHEN Z Y, SHAN Z L. Research Advances in Knowledge Tracing [J]. Computer Science, 2022, 49(10): 83-95.

- [18] ANDERSON J R, BOYLE C F, CORBETT A T, et al. Cognitive Modeling and Intelligent Tutoring[M]// Artificial Intelligence. 1990;7-49.
- [19] PERKINS D N, SALOMON G. Transfer of Learning[J]. International Encyclopedia of Education, 1992, 2: 6452-6457.
- [20] ROEDIGER H L, KARPICKE J D. Test-enhanced Learning: Taking Memory Tests Improves Long-term Retention[M]// Psychological Science. 2006;249-255.
- [21] MAYO R C M, RUSSELL S, NUGENT T, et al. Revealing the learning in learning curves[C]// Proceedings of the 16th International Conference on Artificial Intelligence in Education (AIED). 2013;473-482.
- [22] EBBINGHAUS H. Memory: A Contribution to Experimental Psychology[J]. Annals of Neurosciences, 2013, 20(4): 155-156.
- [23] YELLE L E. The Learning Curve: Historical Review and Comprehensive Survey[J]. Decision Sciences, 1979, 10(2): 302-328.
- [24] BRUNER J S. Toward a Theory of Instruction[M]. Cambridge: Harvard University Press, 1966.
- [25] HOCHREITER S, SCHMIDHUBER J. Long Short-term Memory[J]. Neural Computation, 1997, 9(8): 1735-1780.
- [26] YEUNG C K, YEUNG D Y. Addressing Two Problems in Deep Knowledge Tracing via Prediction-Consistent Regularization [C]// Proceedings of the 5th Annual ACM Conference on Learning at Scale (L@S). ACM, 2018: 1-10.
- [27] PANDEY S, KARYPIS G. A Self Attentive Model for Knowledge Tracing[C]// Proceedings of the 12th International Conference on Educational Data Mining (EDM). 2019;384-389.
- [28] RASCH G. Probabilistic Models for Some Intelligence and Attainment Tests [M]. Chicago: University of Chicago Press, 1993.
- [29] LIU Z T, LIU Q Q, CHEN J H, et al. simpleKT: A Simple but Tough-to-Beat Baseline for Knowledge Tracing [C]// Proceedings of the 11th International Conference on Learning Representations (ICLR). 2023.
- [30] SHEN S H, LIU Q, CHEN E H, et al. Learning Process-Consistent Knowledge Tracing[C]// Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining (KDD). ACM, 2021;1452-1460.
- [31] LIU Q, HUANG Z Y, YIN Y, et al. EKT: Exercise-aware Knowledge Tracing for Student Performance Prediction [J]. IEEE Transactions on Knowledge and Data Engineering, 2021, 33(1): 100-115.
- [32] CHEN J H, LIU Z T, HUANG S Y, et al. Improving Interpretability of Deep Sequential Knowledge Tracing Models with Question-Centric Cognitive Representations[C]// Proceedings of the 37th AAAI Conference on Artificial Intelligence. AAAI, 2023;14196-14204.
- [33] CAI L Q, LIU Y C, REN B, et al. A personalized knowledge recommendation method integrating multi-knowledge point and group characteristics [J]. Journal of Chongqing University of Posts & Telecommunications (Natural Science Edition), 2024, 36(5): 1023-1031.
- [34] WU S, SUN F, ZHANG W, et al. Graph Neural Networks in Recommender Systems: A Survey [J]. ACM Computing Surveys, 2023, 55(5): 1-37.
- [35] ALADE Y A, FENG C, MAO X, et al. Graph Neural Networks for Visual Question Answering: A Systematic Review [J]. Multimedia Tools and Applications, 2023, 83: 55471-55508.
- [36] DAS N, SADHUKHAN B, CHATTERJEE R, et al. Integrating Sentiment Analysis with Graph Neural Networks for Enhanced Stock Prediction: A Comprehensive Survey [J]. Decision Analytics Journal, 2024, 10: 100417.
- [37] LIN Y, RUAN T, LIU J, et al. A Survey on Neural Data-to-Text Generation [J]. IEEE Transactions on Knowledge and Data Engineering, 2024, 36(4): 1431-1449.
- [38] LIU Y F, YANG Y, CHEN X Y, et al. Improving Knowledge Tracing via Pre-training Question Embeddings [C]// Proceedings of the 29th International Joint Conference on Artificial Intelligence. 2021;1577-1583.
- [39] CHEN L, XU Y Q. The Philosophical Outlook of Connectivism and Its Enlightenment on Educational Reform [J]. Education Research, 2023, 3(14): 16-25.
- [40] DOWNES S. Recent Work in Connectivism [J]. European Journal of Open, Distance and E-Learning, 2020, 22(2): 113-133.
- [41] YU L, SUN L L, DU B W, et al. Towards Better Dynamic Graph Learning: New Architecture and Unified Library [C]// Proceedings of the 36th Conference on Neural Information Processing Systems (NeurIPS). 2023;67686-67700.
- [42] LOSHCHILOV I, HUTTER F. Decoupled Weight Decay Regularization. [C]// Proceedings of the 7th International Conference on Learning Representations (ICLR). 2019.
- [43] KINGMA D P, BA J. Adam: A Method for Stochastic Optimization [J]. arXiv:1412. 6980, 2014.
- [44] ADAM P, SAM G, FRANCISCO M, et al. Pytorch: An imperative style, high-performance deep learning library [C]// Proceedings of the 33rd Conference on Neural Information Processing Systems (NeurIPS). 2019;8026-8037.
- [45] GHODAI A, WANG Q. Deep graph memory networks for forgetting-robust knowledge tracing [J]. IEEE Transactions on Knowledge and Data Engineering, 2022, 35(8): 7844-7855.
- [46] SHEN X X, YU F H, LIU Y Q, et al. Revisiting Knowledge Tracing: A Simple and Powerful Model [C]// Proceedings of the 32nd ACM International Conference on Multimedia (MM). 2024;263-272.



LI Jiahao, born in 1999, postgraduate. His main research interests include big data in education and knowledge tracing.



LIU Dong, born in 1976. Ph.D, professor, Ph.D supervisor, is a senior member of CCF (No. 21678S). His main research interests include big data mining in education, social network analysis, etc.