

融合 GRU-TCN 的深度知识追踪方法

谭 雪¹

TAN Xue

摘 要

知识追踪 (KT) 作为智能教育系统中的核心技术, 致力于通过分析学生的历史答题记录, 动态建模其对特定知识点的理解与掌握情况。然而, 当前大多数知识追踪模型在应对长序列数据和复杂时间依赖性时仍存在显著的局限性。为此, 文章提出了一种融合时域卷积网络 (TCN) 与门控循环单元 (GRU) 的全新知识追踪模型 TGC-KT。TCN 模块通过捕捉局部时间依赖性, 显著提升了模型在处理长期依赖关系时的能力, 而 GRU 则凭借其门控机制对历史信息进行选择记忆与遗忘, 进一步优化了预测精度。此外, TGC-KT 能够有效结合局部与全局的学习信息, 动态调整模型对学生知识状态的估计。实验结果表明, 在两个公开数据集上的测试中, TGC-KT 模型在知识掌握度评估和学习效果预测方面的表现优于现有方法, 展示了其在个性化教育中的潜在应用价值。

关键词

深度学习; 知识追踪; 预测精度; 教育数据挖掘

doi: 10.3969/j.issn.1672-9528.2024.12.030

0 引言

随着个性化学习需求的日益增长, 知识追踪 (KT) 技术在教育领域的重要性愈发凸显^[1]。KT 通过分析学生的动态知识变化, 为学生提供个性化学习路径, 同时帮助教师优化教学策略。近年来, 全球范围内的新冠肺炎 (COVID-19) 疫

情对教育系统产生了深远影响, 迫使学校和教育机构迅速转向远程教学模式。这一转变不仅对教育资源的分配提出了新的要求, 也为 KT 技术的发展带来了新的机遇和挑战。

KT 旨在通过建模学生的历史序列数据来得到学生当前的知识状态。利用学生与问题之间的交互信息, 来预测学生是否能正确回答下一个相关的问题。如图 1 所示, 学生回答问题可能包含多个技能知识点, 若学生在做题过程中能将已经获得的先验知识点串联起来回答问题, 则学生大概率能对这个问题的回答。这种评价方式可以帮助教师了解学生的学习情况, 对学生答题错误的原因进行分析, 帮助教师为学生制定

1. 贵州民族大学 贵州贵阳 550000

[基金项目] 贵州省教育厅 (黔教技〔2022〕047 号), 贵州省高等学校智慧教育工程研究中心; 贵州民族大学自然科学基金 (GZMUZK〔2021〕YB22)

参考文献:

- [1] 韩永印, 王侠, 王志晓. 基于决策树的社交网络隐式用户行为数据挖掘方法 [J]. 沈阳工业大学学报, 2024, 46(3): 312-317.
- [2] 吴自博, 王波, 陈清, 等. 基于机器学习的调度操作行为挖掘与推荐技术研究及应用 [J]. 电力系统自动化, 2022, 46(8): 181-188.
- [3] 严玉婷, 薛冰, 方力谦, 等. 基于周期调整负荷成分指数的行业用电大数据价值挖掘 [J]. 电力科学与技术学报, 2022, 37(6): 181-189.
- [4] 张敬伟, 刘绍建, 杨青, 等. DMFUCP: 大规模轨迹数据通用伴随模式分布式挖掘框架 [J]. 计算机研究与发展, 2022, 59(3): 647-660.
- [5] 梁瑶, 谢春丽, 王文捷. 基于图嵌入的代码相似性度量 [J]. 计算机科学, 2022, 49(S2): 801-806.

- [6] 侯靖琳, 仇润鹤, 薛季爱, 等. 基于知识图谱嵌入和补全的电梯故障预测 [J]. 计算机工程与设计, 2022, 43(1): 224-230.

【作者简介】

陈佐虎 (1989—), 男, 甘肃白银人, 本科, 工程师, 研究方向: 电网数字化。

彭振国 (1984—), 男, 甘肃正宁人, 本科, 工程师, 研究方向: 电网数字化。

何海燕 (1995—), 女, 甘肃定西人, 本科, 工程师, 研究方向: 电力大数据。

窦蕴韬 (1989—), 男, 甘肃兰州人, 本科, 工程师, 研究方向: 电网数字化。

(收稿日期: 2024-09-05)

个性化教学方案^[2]。传统的 KT 模型，在捕捉全局依赖性上有一定优势，但在处理局部细节和捕捉长期依赖性上可能不如卷积模型来得细致^[3]。例如，学生近期的学习行为可能对他们的当前知识状态有很大影响，这种短期依赖性模型需要准确捕捉的。因此。模型通过结合 TCN 和 GRU，可以构建一个兼具局部依赖捕捉能力和全局序列建模能力的知识追踪模型，从而提升 KT 模型预测的准确性以及提高模型的鲁棒性。且 TCN 卷积可以作为一种增强手段来克服 GRU 在长序列依赖、计算效率和梯度稳定性上的不足。

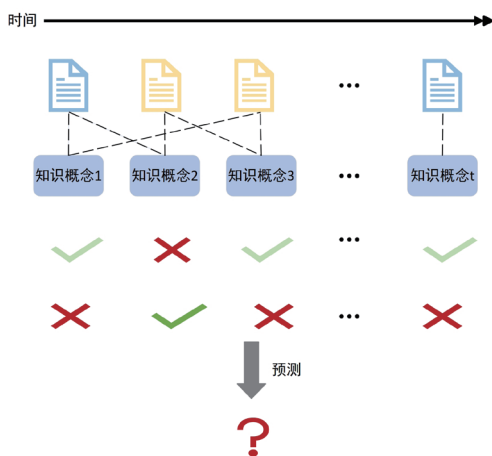


图 1 知识追踪任务示意图

1 相关工作

1.1 传统的知识追踪

传统的知识追踪方法主要基于统计模型和机器学习技术，旨在对学生的知识状态和学习进展进行建模。这些方法通常基于特征工程，采用诸如决策树和贝叶斯网络等模型^[4]。例如，基于决策树的方法能够预测学生在不同知识点上的表现，但此类方法仅能处理“一对一”的关系，无法有效捕捉多个知识点之间的复杂关联性。贝叶斯知识追踪（bayesian knowledge tracing, BKT）将学习者的潜在知识状态建模为一组二元变量，每个变量表示学习者对特定概念的掌握或未掌握状态^[5]。通过隐马尔可夫模型（hidden markov model, HMM），该方法能够动态更新每个二元变量的概率，以反映学习者对某一概念的练习回答是否正确^[6]。

1.2 深度知识追踪

然而，BKT 模型预设在学习者掌握某项技能后，便不会再有遗忘情形发生。但在真实的学习进程中，这一预设并非总能契合实际状况。因为学生的知识状态可能会随着时间的推移发生变化。

为了解决这一缺陷，受深度学习发展的启发，深度知识追踪（deep knowledge tracing, DKT）开创了基于神经网络对学生相关知识状态进行建模和预测的新方向^[7]。这标志着

知识追踪的一项重要创新。DKT 主要利用深度神经网络（如循环神经网络 [RNN] 或长短期记忆网络 [LSTM]）来建模学生的学习行为和知识水平^[8]。这些方法通过将学生的学习数据（如学习日志或答题历史）转化为时间序列，并使用 RNN 或 LSTM 捕捉序列中的时序信息和依赖关系，从而成功完成了对学生的知识状态和以及对相关知识概念的掌握度进行建模。DKT+ 是对原始 DKT 模型的改进版本，它通过引入两个额外的正则化项来增强 DKT 的损失函数。这一变体旨在解决 DKT 在重构学生答案输入方面的局限性，并减少共享相似知识概念问题的答案预测不一致性。通过这些正则化项，DKT+ 能更有效捕捉不同问题之间的联系，进而提高模型在知识追踪任务中的整体表现^[9]。在这一框架下，研究者们还提出了动态键值记忆网络模型，以提高预测结果的可解释性，并引入了知识结构信息，来应对知识追踪中的数据稀疏问题^[10]。

尽管深度知识追踪方法在学生知识状态建模方面取得显著进展，但仍面临一些挑战。本文针对知识追踪数据集具有图结构的特点，提出了一种新的结构，将 GRU 和 TCN 相融合，旨在提高知识追踪预测的准确性。

2 模型描述

2.1 GRU 模型

在知识追踪任务中，学生与基于计算机的学习系统之间的交互记录通常是长期且详细的，这些记录包含了学生的行为动作、当前学习内容的信息以及与练习题的交互结果。为了有效处理这些复杂的序列数据，并准确预测学生对各个知识点的掌握程度，本文采用了 GRU 模型。GRU 模型通过其独特的门控机制，能够有效捕捉和记忆学生在学习过程中的长期依赖关系，同时缓解了传统 RNN 中常见的梯度消失和梯度爆炸问题。

此外，GRU 相比 LSTM 具有更简洁的结构，这不仅提高了模型的计算效率和内存使用率，还使得模型在处理长序列数据时表现更加稳定和高效。GRU 模型的灵活性和适应性使其能够实时处理学生的学习数据，并根据最新的学习行为动态调整对学生知识状态的预测。

GRU 在知识追踪任务中展现了出色的性能，尤其是在处理长期依赖问题方面。它能够有效传递和表达学习者在整个学习过程中生成的长序列信息，同时捕捉历史学习行为对当前知识状态的影响。具体而言，隐含信息向量被输入到 GRU 模型中，用于跟踪学习者知识状态随时间的动态变。GRU 模型通过递归调用自身，将 $t-1$ 时刻的隐藏状态传递给 t 时刻，从而能够在序列中持续捕捉学习者知识状态的变化。详细过程为：

$$R_t = \sigma(X_t W_{xr} + H_{t-1} W_{hr} + b_r) \quad (1)$$

$$Z_t = \sigma(X_t W_{xz} + H_{t-1} W_{hz} + b_z) \quad (2)$$

$$\tilde{H}_t = \tanh(X_t W_{xh} + (R_t \cdot H_{t-1}) W_{hh} + b_h) \quad (3)$$

$$H_t = Z_t \cdot H_{t-1} + (1 - Z_t) \cdot \tilde{H}_t \quad (4)$$

式中： Z_t 是更新门向量； R_t 是重置门向量； $\tilde{H}_t$ 是候选隐藏状态输出，最终的隐藏状态 H_t 通过候选隐藏状态和前一时刻隐藏状态之间的加权和得到。 W 和 b 是相关网络参数。

2.2 知识状态提取模块

因果卷积处理输入的问技能时间序列时，通过一维卷积核在时间域内进行时间序列的卷积。每一层的输出由上一层的未知输入和下一位置的输入共同获得，并严格遵守时间约束，只使用时间 t （即时间 $0 \sim t-1$ ）的元素之前的元素来预测当前问题的是否能正确回答的情况的概率分布。

定义单个的知识状态的输入为 $x_t=(q_t, a_t)$ ，其中， q_t 代表问题， a_t 代表学生的回答。那么当前的时刻 t 知识状态的输入序列 $Q=[q_1, q_2, \dots, q_t]$ 表示的是与时间序列对应的输入输出对，目标是通过深度学习模型预测学生未来的表现。

本文将 TCN 应用于知识追踪模型，假设卷积核 $F=[f_1, f_2, \dots, f_k]$ 是被用于提取输入特征，那么卷积公式及其描述为：

$$(F * X)_{(xt)} = \sum_{k=1}^K f_k x_{t-K+k} \quad (5)$$

式中： x_{t-K+k} 表示输入序列中的第 $t-K+k$ 个值； K 表示卷积核的长度。卷积核用于设置网络的采样频率，并适当地填充边缘位置以确保输出和输入的维度一致。

TCN 提取的深度局部知识状态信息首先通过 GRU 模型进行处理，GRU 的门控机制使其能够捕捉输入序列中的长短期依赖关系，并动态地调整信息流动。处理后的输出序列特征随后经过全连接层，以便将 GRU 的输出转化为固定维度的特征表示。这一步骤不仅有助于模型学习复杂的非线性关系，还能够优化梯度更新过程，使梯度在反向传播时保持稳定，从而提升模型的预测性能并避免梯度消失或爆炸的问题。

2.3 预测和优化模块

通过 TCN 提取出的特征表示为 x_t ，它将作为 GRU 的输入，GRU 将序列中的每一个时刻的输入特征 x_t 和前一时刻的隐藏状态 H_{t-1} 结合，产生新的隐藏状态 H_t ，输出知识掌握的预测，公式为：

$$p_t = \sigma(w_p H_t + b_p) \quad (6)$$

式中： w_p 和 b_p 是用于预测的权重矩阵和偏置项； σ 是激活函数。

使用模型对学生进行交互建模后，为了优化知识追踪模型，采用了交叉熵损失函数作为优化目标，为了减少预测答案 p_t 和实际答案之间的差异，文中使用 Adam 优化器在小批

量上最小化，目标函数定义为：

$$L = -\sum_{t=1}^N (a_t \log p_t + (1 - a_t) \log(1 - p_t)) \quad (7)$$

式中： a_t 表示学生对第 t 个问题的真实回答； p_t 是指模型预测学生能否准确回答第 t 个问题的概率。

3 实验与结果分析

为了评估模型在知识追踪任务上的可行性和性能表现，将模型与主流的 DLKT 模型在两个真实的公开数据集上进行了比较，以评估模型的优劣。

3.1 数据集描述

为了更好地评估模型，文中在 KT 中两个广泛使用的数据集上进行了实验，即 Assist 09 和 Assist 12。所有数据集均为公开且真实的数据集，包含学习者和真实的计算机辅助教学系统之间的时间交互记录。

(1) Assist09 数据集是在线教育平台在 2009—2010 年收集的真实的 learners 交互记录数据，从中选取“skillbuilder”数据集上进行实验，此数据集中，大多数练习只含有一种技能，有一小部分与两种或三种技能有关。

(2) Assist12 数据集是在 2012—2013 年从 Assist09 数据集同一平台收集。在该数据集中，各个问题仅与单一技能存在关联，但一个技能仍然对应着多个问题。

对于这两个数据集，除了过滤掉损坏的样本外，并未做太多修改。详细的统计数据如表 1 所示。

表 1 数据集介绍

数据集	Statistics		
	# 学生数	# 技能	# 交互数
Assist09	3852	123	282 619
Assist12	27 485	265	2 709 436

3.2 实施细节

本研究使用 TensorFlow 实现所有比较方法，模型采用 50 个 epoch 进行评估。技能、问题和答案的嵌入大小被固定为 100，并且在训练过程中随机初始化和更新。采用保持概率为 0.8 的 dropout 以防止过拟合。所有可训练参数均通过 Adam 算法进行优化，学习率设为 0.001，最小批量大小为 32。为了评估模型对学习者的下一个反应的预测准确性，我们采用了广泛应用的评价指标：预测准确性（ACC）和受试者操作特征（ROC）曲线下面积（AUC），较高的 AUC 和 ACC 值表明模型的性能更好。

3.3 实验结果

为了评估所提方法的有效性，本文使用以下模型作为基线。

BKT：通过贝叶斯推理进行预测，将技能的知识状态建模为二元变量（掌握或未掌握）。

DKT: 是一个深度学习模型。它使用递归神经网络对序列数据的强大学习能力来模拟学生整个学习过程。它的输入是学生概念表征的独热嵌入。

DKVMN: 受记忆增强神经网络的启发，使用静态矩阵和动态矩阵来存储所有概念并分别更新学生的学习状态，而不是使用单一的隐藏状态。

SAKT: 是第一个在 KT 模型中加入注意力机制的人，它提出了一个自我关注的模型来捕捉学生的学习记录之间的长期依赖关系。将每个问题划分为小部分，并通过自注意机制进行加权，从而可以捕捉到更准确的知识状态

GKT: 将 KT 问题重新表述为时间序列节点级分类问题，并使用标准图学习技术（如消息传递 GNN）解决 KT 问题。

BKT、DKT、DKVMN、SAKT、GKT 和 TGC-KT 模型在数据集 Assisit12、Assisit09 上的性能对比如表 2，与五个基线模型相较而言，本研究提出的模型展现出了最为优异的预测性能。这表明，通过融合 TCN 的 GRU 知识追踪模型很好地解决了传统模型的长序列依赖问题。结合 GRU 和 TCN 的优势，文中模型成功利用 TCN 在长序列建模任务中的局部提取信息的优势，能更好地抓取长历史时间的答题记录，从而减轻 KT 模型中的记忆丢失问题。

表 2 不同模型的性能比较

模型	Assisit12		Assisit09	
	AUC	ACC	AUC	ACC
BKT	0.620 4	0.650 2	0.657 1	0.663 5
DKT	0.690 1	0.712 8	0.672 3	0.662 3
DKVMN	0.701	0.686	0.755 0	0.648 9
SAKT	0.683 1	0.715 6	0.697 2	0.673 1
GKT	0.697 2	0.719 4	0.734	0.656
TGC-KT	0.859 9	0.791 2	0.816 3	0.757 0

通过实验结果观察到，TGC-KT 模型能够更精确地评估学生对知识的掌握情况，进一步验证了模型的有效性。这种结合为知识追踪领域带来了新的视角和方法，提高了预测性能和模型的泛化能力。

4 总结

本文提出了一种名为 TGC-KT 的模型，旨在提升知识追踪模型的可解释性和预测准确性。该模型不仅能够有效捕捉学生的知识状态，还能跟踪其学习进展，从而加深对个体学习过程的理解。通过结合 TCN，TGC-KT 模型能够通过局部卷积核提取细粒度的局部特征，同时保持输入时间序列的完整性。这一特性解决了传统模型在处理长期依赖和局部细节时的不足。同时，结合 GRU，TGC-KT 能够动态调整学习状态，准确捕捉学生在长期学习过程中的知识状态变化。这一设计为个性化学习路径推荐提供了新的视角，能够为不同学生提供个性化的学习建议。未来，希望将该模型应用于各种

智能教育平台，实现对学生知识掌握变化的实时追踪，并为学生推荐个性化的学习策略。

参考文献:

[1] WANG L S, SY A, LIU L, et al. Deep knowledge tracing on programming exercises[C]//Proceedings of the fourth (2017) ACM conference on learning@ scale. New York:ACM,2017: 201-204.

[2] CHENG S, LIU Q, CHEN E H, et al. Adaptkt: a domain adaptable method for knowledge tracing[C]//Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining. New York:ACM, 2022: 123-131.

[3] LIU T Y, ZHANG M, ZHU C Y, et al. Transformer-based convolutional forgetting knowledge tracking[J]. Scientific reports, 2023, 13(1): 19112.

[4] SHEN S H, LIU Q, CHEN E H, et al. Convolutional knowledge tracing: modeling individualization in student learning process[C]//Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM, 2020:1857-1860.

[5] ALBERT T C, JOHN R A. Knowledge tracing: modeling the acquisition of procedural knowledge[J]. User modeling and user-adapted interaction, 1994, 4: 253-278.

[6] XIONG X L, ZHAO S Y, ERIC G, et al. Going deeper with deep knowledge tracing[EB/OL].[2024-07-12]. https://www.educationaldatamining.org/EDM2016/proceedings/paper_133.pdf.

[7] CHRIS P, JONATHAN S, HUANG J, et al. Deep knowledge tracing[DB/OL].(2015-06-19)[2024-08-21].<https://doi.org/10.48550/arXiv.1506.05908>.

[8] SEPP H, JÜRGEN S. Long short-term memory[J]. Neural computation, 1997,9(8): 1735-1780.

[9] 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.

[10] ZHANG J N , SHI X J, KING I, et al. Dynamic key-value memory networks for knowledge tracing[DB/OL].(2017-02-17) [2024-08-11].<https://doi.org/10.48550/arXiv.1611.08108>.

【作者简介】

谭雪(1998—)，女，贵州毕节人，统计学硕士，研究方向：智慧教育。

(收稿日期：2024-09-09)