

结合时序建模与交叉注意力的知识追踪算法研究

陈克莎¹ 霍雨佳^{1*}
CHEN Keshu HUO Yujia

摘要

知识追踪旨在通过学生的答题行为推断其潜在的知识掌握状态，但传统方法往往受限于对长序列依赖和跨题目关系的刻画不足，难以精准捕捉知识动态演化与深层关联。为此，文章提出了一种结合时序建模与交叉注意力机制的改进模型。该模型通过引入交叉注意力计算出一个注意力权重矩阵，该矩阵能够清晰地显示出当前题目与历史中每个练习的相关性，从而有效提升了对学生知识状态的动态表征能力，并在 ASSIST2009、ASSIST2015 和 EdNet 数据集上取得显著优于主流基线的表现，验证了模型的有效性与鲁棒性，为个性化学习路径推荐提供了支持。

关键词

知识追踪；交叉注意力；时序建模；智能教育

doi: 10.3969/j.issn.1672-9528.2025.12.037

0 引言

在智能教育与个性化学习的背景下，知识追踪被广泛应用于动态建模学生的学习过程，以预测其对未来题目的答题表现，从而为因材施教提供理论依据与实践价值。近年来，研究者尝试在模型中引入更多认知和行为特征来增强预测能力，例如张维等^[1]提出的模型通过整合元认知信息以及学生的猜测与失误行为，有效刻画了更复杂的学习模式；付睿等^[2]则利用图卷积网络融合多维度特征，从而提升了对题目关系的建模效果。这些工作为知识追踪在实际应用中捕捉学生多样化特征提供了新的思路。

然而，现有方法在建模序列依赖时仍面临一定挑战。Zhang 等^[3]提出的 SSAKT 模型（图1）能够有效捕捉长程依赖，但在局部交互的表达上存在不足。Yue 等^[4]通过引入学生能力属性并结合注意力机制提升模型解释性，但在复杂场景下

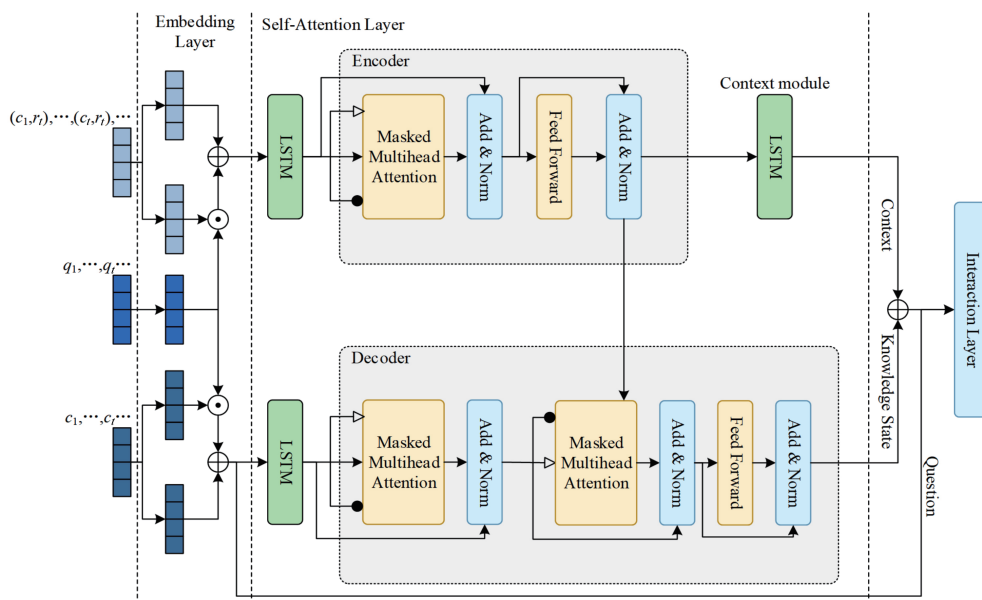


图1 SSAKT 模型架构图

可能削弱了预测精度。Pu 等^[5]提出的 EAKT 模型在认知框架与自注意力结合上取得进展，但其结构复杂度较高，不利于大规模应用。由此可见，在捕捉长程依赖、保证预测精度与维持模型可解释性之间，仍然存在需要权衡和改进的空间。

而如何在复杂学习序列中有效过滤冗余信息、同时提升模型的鲁棒性，依然是知识追踪领域亟待解决的核心问题。Huang 等^[6]提出的 k-sparse attention 方法，通过在注意力计算中引入稀疏化约束，仅保留最具代表性的部分交互，从而在一定程度上降低了无关或噪声信息对预测结果的干扰。然而，这种硬性稀疏化策略也存在潜在不足，可能会舍弃掉某些对预测具有补充价值的弱相关信息，导致模型在复杂场景

1. 贵州民族大学数据科学与信息工程学院 贵州贵阳 550025
[基金项目] 贵州省教育厅、贵州省高等学校智慧教育工程研究中心和贵州思索电子有限公司支持完成（〔2022〕47号）

下表现受限。为了解决这一问题，Cui 等^[7]提出利用多关系 Transformer 结构，在建模过程中显式引入不同类型的关系边，以更精细地刻画题目间的多维交互模式。这种方法能够增强模型在区分不同题目关系时的表达能力，但其高昂的计算复杂度使得大规模应用面临挑战。此外，Lee 等^[8]提出的 MonaCoBERT 模型，通过将单调注意力与 ConvBERT 结合起来刻画学生的遗忘行为，体现出一种兼顾记忆衰减和语义建模的创新思路。然而，由于其在处理长序列时的稳定性不足，该方法在实际部署中仍存在局限。

上述研究共同揭示了一个关键矛盾：在信息冗余控制与复杂关系建模之间，如何实现计算效率、预测精度和模型稳定性的平衡，仍然是当前知识追踪模型设计需要重点突破的难点。

针对上述问题，Abdelrahman 等^[9]提出的图记忆网络展示了记忆机制在增强遗忘鲁棒性方面的潜力，而 Zhou 等^[10]提出的多头自注意力模型则强调通过多视角捕捉序列依赖以提升整体预测性能。

本文提出的模型在此基础上，设计了交互嵌入与 LSTM 结合的框架，并引入以题目嵌入为 Query、以隐状态 Key/Value 的交叉注意力机制，以突出与目标题目最相关的历史信息。在残差连接与正则化的辅助下，该模型在保证预测精度的同时缓解了冗余干扰，并在可解释性与鲁棒性上实现了平衡，从而为知识追踪任务提供了一种更具实用性的解决方案。

1 模型描述

1.1 特征处理

在本研究的模型设计中，特征处理环节首先需要将原始的学生答题记录转化为适合深度模型建模的嵌入表示。具体而言，每一道题目与其对应的作答结果（对/错）被组合为交互特征，并通过嵌入层映射到连续的低维向量空间。设题目嵌入向量为 $q_i \in \{1, 2, \dots, P\}$ ，作答结果为 $r_i \in \{0, 1\}$ ，则交互特征索引可表示为：

$$x_i = q_i + P \cdot r_i \quad (1)$$

式中： P 为题目总数，取值范围覆盖所有“题目 - 作答”组合。通过嵌入矩阵 $E \in \mathbb{R}^{2Q \times d}$ ，交互特征最终被表示为 $e_i = E[x_i]$ ，为后续的时序建模提供输入。

此外，为进一步捕捉题目的语义信息与时序依赖，模型在交互嵌入之外，还为题目本身构建了单独的题目嵌入向量 q_i 。这种处理方式有助于在注意力计算时区分“题目语义”与“交互状态”。设 $H \in \mathbb{R}^{L \times d}$ 为 LSTM 得到的隐状态序列， $P \in \mathbb{R}^{L \times d}$ 为题目嵌入序列，则交叉注意力机制中的上下文表示为：

$$\text{Attention}(P, H, H) = \text{softmax}\left(\frac{PH^T}{\sqrt{d}}\right)H \quad (2)$$

该过程通过题目嵌入作为查询向量（Query），在隐状态序列中选择与目标题目最相关的历史信息，实现特征的动态筛选与增强表达。在模型架构中所用的 LSTM 模型架构如图 2 所示。

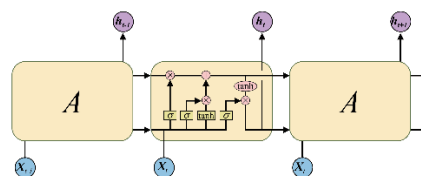


图 2 所示 LSTM 模型架构

该结构清晰地展示了 LSTM 在时序建模中的核心机制：通过输入门、遗忘门和输出门的动态调节，模型能够在长期依赖与短期信息之间实现平衡，从而有效缓解传统 RNN 在长序列建模时出现的梯度消失或爆炸问题。结合交叉注意力机制，LSTM 不仅保留了对历史交互的全局记忆，还能突出与当前题目最相关的上下文表示，这种设计进一步增强了模型在知识追踪任务中的稳定性与预测能力。

1.2 交叉注意力机制

在完成交互特征与题目特征的处理之后，模型通过交叉注意力机制来实现题目语义与学生历史答题行为的对齐与融合。与传统自注意力不同，交叉注意力的核心思想是利用题目嵌入作为查询向量，以 LSTM 输出的隐状态序列作为键和值，从而突出与目标题目最相关的历史信息。注意力得分的计算公式为：

$$\alpha_{i,j} = \frac{\exp\left(\frac{q_i \cdot h_j}{\sqrt{d}}\right)}{\sum_{j=1}^L \exp\left(\frac{q_i \cdot h_j}{\sqrt{d}}\right)} \quad (3)$$

式中： $q_i \in \mathbb{R}^d$ 表示时刻 i 的题目嵌入； $h_i \in \mathbb{R}^d$ 表示 LSTM 在时刻 i 的隐状态； $\alpha_{i,j}$ 则表示题目 q_i 对历史答题交互 h_j 的注意力权重。注意力架构如图 3 所示。

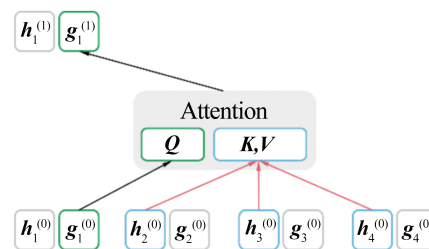


图 3 注意力架构图

在得到注意力权重之后，模型会将其作为加权系数，作用于学生过往的交互表示。随后，模型将所有历史交互在权重的引导下进行加权求和，从而得到一个综合的上下文表示。该上下文表示不仅整合了学生长期与短期的学习轨迹信息，还在特征层面上与当前题目的语义保持一致性，最终形成与

目标题目高度匹配的动态表示, 为后续的预测提供更具针对性的输入, 用公式表示为:

$$\mathbf{c}_t = \sum_{i=1}^L \alpha_{t,i} \mathbf{h}_i \quad (4)$$

式中: $\mathbf{c}_t \in \mathbf{R}^d$ 是在时刻 t 得到的上下文表示, 不仅保留了时序依赖信息, 还通过注意力机制实现了与目标题目的对齐。这种表示有效地缓解了冗余历史信息对预测的干扰。

最后, 模型将题目嵌入 $\mathbf{q}_t$ 与上下文向量 $\mathbf{c}_t$ 融合, 得到综合的表示, 用于预测学生在该题目的答题结果。融合方式采用残差连接与非线性变换, 以增强表达能力和稳定性:

$$\mathbf{z}_t = \tanh(\mathbf{W}[\mathbf{q}_t + \mathbf{c}_t] + \mathbf{b}) \quad (5)$$

式中: $\mathbf{W}$ 与 $\mathbf{b}$ 为可训练参数; $\mathbf{z}_t$ 为最终的特征表示输入至预测层。

通过这种设计, 交叉注意力机制在保持模型可解释性的同时, 有效提升了对学生知识状态的建模精度。

1.3 训练目标与损失函数

模型在每个时刻 t 输出对所有题目的正确率预测向量 $\hat{\mathbf{p}}_t = [\hat{p}_{t,1}, \dots, \hat{p}_{t,Q}] \in (0,1)^Q$ 。

由于训练与评估关注的是“下一时刻的题目”是否答对, 因此需要从 $\hat{\mathbf{p}}_t$ 中选出对应题目的预测概率。令 $\mathbf{e}(\mathbf{q}_{t+1}) \in \{0,1\}^Q$ 为 $\mathbf{q}_{t+1}$ 的 one-hot 向量, 则对下一题的预测概率可以写成点积或对所有题目求和的形式:

$$\hat{y}_t = \sum_{j=1}^Q \hat{p}_{t,j} \mathbf{1}\{\mathbf{q}_{t+1} = j\} = \hat{\mathbf{p}}_t^T \mathbf{e}(\mathbf{q}_{t+1}) \quad (6)$$

在损失函数上, 本文采用二元交叉熵 (binary cross-entropy) 作为监督信号, 并只在有效的标签位置上计算。对整个批次 / 序列将损失按有效位置平均, 同时加入 L_2 正则化以抑制过拟合, 目标函数为:

$$L = -\frac{1}{\sum_i m_i} \sum_i m_i (y_i \log \hat{y}_i + (1 - y_i) \log(1 - \hat{y}_i)) + \lambda \|\theta\|_2^2 \quad (7)$$

式中: y_i 为真实标签; θ 为模型参数; λ 为权重衰减系数。

训练时使用小批量随机梯度下降 (如 Adam) 最小化 L , 并在验证集上基于 AUC 指标进行模型选择与早停, 从而保证在不平衡数据与填充序列下的稳健性与泛化能力。

2 实验结果与分析

2.1 数据集介绍

本研究采用了三类典型的教育数据集, 分别为 ASSIST09、ASSIST15 以及 EdNet。这些数据集均来源于真实的在线学习平台, 广泛用于知识追踪任务的建模与评测。ASSIST09 数据集收集自 ASSISTments 平台, 包含 2009 年期间学生的答题日志, 记录了学生不同知识点上的交互信息, ASSIST15 数据集则是该平台在 2015 年的扩展版本, 数据规

模显著提升, 涵盖了更多学生、题目和知识点。相比之下, EdNet 数据集由大规模智能学习平台 Santa 收集, 包含超过 1.3 亿条交互记录, 是目前公开教育数据集中规模最大、覆盖面最广的一类, 本文选取了 5 000 条数据来验证实验的结果。通过在 3 个数据集上的实验, 可以综合考察模型在不同规模场景下的表现, 从而更全面地验证其有效性。

2.2 评价指标

在二分类任务中, AUC (area under the ROC curve) 是衡量模型整体性能的重要指标之一。表示模型在所有可能的分类阈值下, 区分正负样本能力的平均水平。具体而言, AUC 值等价于随机抽取一个正样本和一个负样本时, 模型将正样本判定为正的的概率。其计算公式为:

$$\text{AUC} = \int_0^1 \text{TPR}(\text{FPR}^{-1}(x)) dx \quad (8)$$

式中: TPR 为真正率; FPR 为假正率; ROC 曲线即以 FPR 为横坐标、TPR 为纵坐标绘制而成。

2.3 实验细节

在实验过程中, 首先对数据集进行了预处理与划分, 同时固定测试集以确保评估的可重复性。模型采用 PyTorch 框架实现, 训练过程中启用了 GPU 加速, 并结合多线程数据加载以提升训练效率。

在超参数设定方面, 模型训练了 30 个 epoch, 优化器选择 Adam, 初始学习率设为 $1e-3$, 并使用 Cosine Annealing 调度策略进行动态调整。批大小设置为 128, 在训练过程中同时监控损失函数的变化情况。最终的实验结果通过在测试集上的表现进行综合评估, 其整体的实验细节统计表格如表 1 所示。

表 1 实验细节统计表格

参数类别	设置内容
框架与工具	PyTorch, Python
数据划分	训练集 80%, 测试集 20%
批大小	128
训练轮数	30
优化器	Adam
初始学习率	$1e-3$

2.4 实验结果

为全面评估所提方法有效性, 本文选取 BKT (贝叶斯理论建模知识状态转移)、DKT (循环神经网络捕捉时序依赖)、DKT+ (DKT 基础上增正则化与特征增强以缓解过拟合)、DKVMN (动态键值记忆结构显式存储更新知识点表征)、SAKT (自注意力机制建模题目间依赖) 等典型知识追踪模

型为对比基线，从不同角度反映了该领域发展脉络。其实验结果如表 2 所示。

表 2 对比实验结果

模型	ASSIST2009	ASSIST2015	Ednet
BKT	0.635 7	0.600 2	0.601 9
DKT	0.748 1	0.689 0	0.692 3
DKT+	0.752 6	0.693 2	0.703 6
DKVMN	0.753 8	0.712 3	0.708 2
SAKT	0.757 2	0.721 2	0.701 5
Ours	0.777 1	0.765 5	0.742 9

从表 2 的对比结果可以看出，本文提出的模型在 3 个典型数据集上均展现出了超越现有方法的优异性能。相较于 BKT 模型，DKT、DKT+、DKVMN 通过神经网络与记忆结构提升预测精度，凸显深度特征提取与动态记忆机制的价值；SAKT 则突破循环网络局限，可直接建模题目依赖，在捕捉远程关联上更优。这些对比结果表明，随着建模方法的不断演进，知识追踪任务在时序依赖捕捉、关系建模与特征表示方面逐步得到了提升。

而本文提出的方法有效弥补了上述方法各自存在的不足。一方面，LSTM 结构在建模学生历史学习轨迹时为知识状态的动态演化提供了坚实的基础；另一方面，交叉注意力机制则以目标题目为核心，在历史隐状态中有选择性地提取与之最相关的信息，从而突出了“目标导向”的知识对齐过程。这种双重建模方式不仅提升了预测的准确性，也增强了模型在面对不同规模与不同场景数据集时的泛化性与鲁棒性，更加符合智能教育中个性化学习路径推荐的实际需求。因此，本文工作不仅在实验结果上验证了方法的优越性，同时也为知识追踪模型在理论与应用层面的发展提供了新的思路与方向。

3 结语

综上所述，本文针对知识追踪任务提出了一种结合时序建模与交叉注意力机制的改进模型。该方法不仅能够有效捕捉学生答题行为的动态依赖，还能突出与目标题目最相关的历史信息，从而提升预测的准确性与解释性。通过与多种经典模型的对比实验，验证了所提模型在不同场景下的泛化能力与鲁棒性。整体而言，本研究为智能教育中个性化学习路径推荐与学习状态诊断提供了有价值的参考。

参考文献:

[1] 张维, 张登, 龚中伟, 等. 元认知信息和猜测与失误行为增强的知识追踪模型 [J/OL]. 小型微型计算机系统, 1-12[2025-09-13].<https://doi.org/10.20009/j.cnki.21-1106/TP.2025-0272>.

[2] 付睿, 李晓兰, 徐焱, 等. 融合多维度特征的图卷积网络知识追踪模型 [J/OL]. 计算机应用研究, 1-10[2025-09-13].
<https://doi.org/10.19734/j.issn.1001-3695.2025.06.0156>.

[3] ZHANG X L, ZHANG J T, LIN N Z, et al. Sequential self-attentive model for knowledge tracing[C]//Artificial Neural Networks and Machine Learning – ICANN 2021. Berlin:Springer,2021:318-330.

[4] YUE Y Q, SUN X Q, JI W D, et al. Augmenting interpretable knowledge tracing by ability attribute and attention mechanism[EB/OL].(2023-02-04)[2025-11-22].<https://doi.org/10.48550/arXiv.2302.02146>.

[5] PU Y J, WU W J, PENG T H, et al. Embedding cognitive framework with self-attention for interpretable knowledge tracing[J]. Scientific reports, 2022, 12(1): 17536.

[6] HUANG S Y, LIU Z T, ZHAO X Y, et al. Towards robust knowledge tracing models via k-sparse attention[C]//Proceedings of the 46th international ACM SIGIR conference on research and development in information retrieval. New York: ACM, 2023: 2441-2445.

[7] CUI J J, CHEN Z Y, ZHOU A M, et al. Fine-grained interaction modeling with multi-relational transformer for knowledge tracing[J]. ACM transactions on information systems, 2023, 41(4): 1-26.

[8] LEE U, PARK Y, KIM Y, et al. MonaCoBERT: monotonic attention based ConvBERT for knowledge tracing[EB/OL].(2022-09-07)[2025-11-22].<https://doi.org/10.48550/arXiv.2208.12615>.

[9] ABDELRAHMAN G, WANG Q. Deep graph memory networks for forgetting-robust knowledge tracing[J]. IEEE transactions on knowledge and data engineering, 2022, 35(8): 7844-7855.

[10] ZHOU Y, WU F Z, HUANG G M, et al. A deep knowledge tracing model based on multi-head self-attention mechanism[C]//International Conference on Algorithms, High Performance Computing, and Artificial Intelligence (AHPCAI 2023). POD Publisher:SPIE,2023:652-657.

【作者简介】

陈克莎（2001—），女，贵州安顺人，硕士研究生，研究方向：深度知识追踪、海量数据统计与分析、深度学习。

霍雨佳（1982—），通信作者（email:2645327286@qq.com），男，贵州遵义人，博士，副教授、硕士生导师，研究方向：深度知识追踪、智慧教育、自然语言处理等。

（收稿日期：2025-10-16 修回日期：2025-12-12）