

DOI:10.13232/j.cnki.jnju.2026.02.009

一种元学习增强的早期知识追踪建模方法 MetaKT

王 晶¹, 苏健华¹, 马玉玲^{1,2*}, 于德湖¹, 崔超然³, 于志云¹

(1. 山东建筑大学计算机与人工智能学院, 济南, 250101;

2. 山东建筑大学人工智能工程技术应用研究院, 济南, 250101;

3. 山东财经大学计算机科学与技术学院, 济南, 250014)

摘要: 知识追踪根据学生的历史学习轨迹来实现对其知识掌握水平的实时评估与跟踪, 从而预测学生未来的学习表现, 是在线学习系统实现个性化学习体验的核心技术之一. 与传统方法相比, 现有的基于深度神经网络的知识追踪模型取得了显著优势, 但其通常依赖大量训练数据. 在学生答题早期, 交互数据极度稀缺, 所以训练一个复杂、有效的深度知识追踪模型十分具有挑战性. 针对此问题, 提出一种基于元学习增强的早期知识追踪框架 (Meta-Learning-Enhanced Knowledge Tracing, MetaKT). 给定目标知识追踪任务和其他若干个相关辅助任务, MetaKT 首先在辅助任务上训练模型, 然后利用目标任务的数据对预训练后的模型进行微调直至模型收敛. 在七个公开数据集上以常用的 DKT 和 DKVMN 为基准进行实验, 结果发现, 提出的 MetaKT 框架使 DKT 和 DKVMN 模型分别在 27 和 33 (共 35) 个测试场景中的 *AUC* (Area under Curve) 获得了提升.

关键词: 元学习, 深度知识追踪, 小样本学习, MAML, 个性化学生建模

中图分类号: TP391

文献标志码: A

MetaKT: A Meta-learning-enhanced early-stage knowledge tracing approach

Wang Jing¹, Su Jianhua¹, Ma Yuling^{1,2*}, Yu Dehu¹, Cui Chaoran³, Yu Zhiyun¹

(1. School of Computer and Artificial Intelligence, Shandong Jianzhu University, Ji'nan, 250101, China;

2. Institute of Applied AI Engineering and Technology, Shandong Jianzhu University, Ji'nan, 250101, China;

3. School of Computer Science and Technology, Shandong University of Finance and Economics, Ji'nan, 250014, China)

Abstract: Knowledge Tracing (KT) dynamically assesses and tracks students' knowledge mastery levels based on their historical learning trajectories, enabling the prediction of their future learning performance. As a core technology in online learning systems, KT facilitates personalized learning experiences. While existing deep neural network-based KT models (e.g., DKT, DKVMN) have demonstrated significant advantages over traditional methods, they typically require large-scale training data. Early-stage interactions, where the student response data are extremely sparse, pose substantial challenges to training complex and effective deep KT models. To address this limitation, we propose MetaKT (Meta-Learning-Enhanced Knowledge Tracing), a framework that leverages meta-learning to enhance early-stage KT performance. Given a target KT task and several related auxiliary tasks, MetaKT first pre-trains the model on auxiliary tasks, and then fine-tunes it using the target task's limited data until convergence. Experiments on seven public datasets, with DKT and DKVMN as backbones, demonstrate that MetaKT improves *AUC* for DKT and DKVMN in 27 and 33 out of 35 test scenarios, respectively.

基金项目: 国家自然科学基金(62177031), 山东省自然科学基金(ZR2021MF044), 2023年度教育部人文社会科学研究专项任务(高校辅导员研究)(2023JDSZ3174), 山东建筑大学国内访问学者经费

收稿日期: 2025-09-03

* 通信联系人, E-mail: mayuling20@sdjzu.edu.cn

Keywords: meta learning, deep knowledge tracing, few shot learning, MAML, personalized student modeling

随着互联网技术不断发展,在线教育成为一种越来越受欢迎的学习方式,为学习者提供了更加灵活和便捷的学习途径^[1]. 为了提高学生的学习效率,在线教育平台,如MOOCs^[2],通过个性化推荐系统为学生定制学习资源,以适应不同学生的学习需求^[3]. 知识追踪(Knowledge Tracing, KT)可以借助机器学习和大数据技术,根据学生的历史学习轨迹,实现对学生知识水平的实时评估与跟踪,以准确地预测学生未来的学习与答题表现,为学生个性化和自适应的学习提供技术支撑,这已成为大规模在线学习系统的核心技术之一^[4].

知识追踪的概念由心理学家 Atkinson 教授^[5]首次提出. 卡内基梅隆大学 Corbett et al^[6]较早地提出贝叶斯知识追踪方法(Bayesian Knowledge Tracing, BKT),并将其用于一个关于计算机程序设计课程的智能教学辅助系统中进行个性化的学习资源推荐. 随着深度学习在计算机视觉、自然语言处理以及语音识别等领域的广泛应用,2015年斯坦福大学 Piech et al^[7]提出深度知识追踪模型(Deep Knowledge Tracing, DKT),首次将循环神经网络(Recurrent Neural Network, RNN)用在知识追踪任务上并取得了突破性的性能提升. 此后,深度学习成为最受关注的建模方法之一,产出大量卓有成效的研究成果. Zhang et al^[8]将记忆增强神经网络引入知识追踪任务,利用一个动态存储矩阵和静态矩阵分别存储学生不断更新的知识水平状态和学生初始知识水平. Ghosh et al^[9]使用一种新颖的注意力机制来构建知识追踪模型,该注意力的权重根据试题间的相似度和上下文有关的距离呈指数级别衰减. Shen et al^[10]提出基于卷积神经网络的知识追踪模型(Convolutional Knowledge Tracing, CKT),通过多层一维卷积操作,从学生答题序列数据中提取学生个体的学习率特征. Yang et al^[11]建立试题与知识概念之间的对应关系图并利用图卷积神经网络(Graph Convolution Neural Networks, GCN)对学生知识水平进行建模. Su et al^[12]提出融合试题文本信息的

EERNN模型,首先基于试题的文本信息,利用一个双向 LSTM (Bi-Directional Long Short-Term Memory)对试题进行编码,进一步利用RNN对学生知识水平进行跟踪. Cui et al^[13]提出一种新型双图集成学习知识追踪方法,通过超图建模与有向图建模分别构建学习交互的双图结构,从而捕获异构化的习题-概念关联与交互转移关系. 篇幅所限,本文不再一一列举,更多研究详见文献[4]. 得益于深度模型强大的表征学习能力和参数复杂性,现有的深度知识追踪方法在拥有海量历史答题数据的场景下取得了很好的性能,但在学生答题早期,答题数量极度稀缺,这一场景在知识追踪任务上普遍存在. 以公开数据集 ASSISTments2009 为例,答题数量小于10的学生人数占总人数的23.1%,64.2%的学生的答题数量不超过50个,图1给出了该数据集上学生答题数量区间分布图. 现有研究鲜有关注学生早期知识追踪的建模问题,即在答题数量较少的情况下实现对知识水平的精准建模.

为此,本文将元学习机制^[14]引入知识追踪任务,创新性地提出一种基于元学习增强的知识追踪框架(Meta-Learning-Enhanced Knowledge Tracing, MetaKT),旨在借助相关知识追踪任务中学习到的知识来提升学生答题早期的知识追踪模型的性能. 元学习^[14](Meta-Learning)可从相关任务上学习和获取有助于在给定任务上更快更准

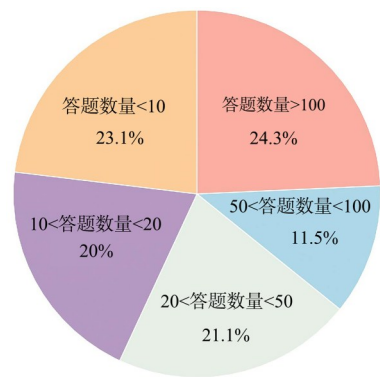


图1 ASSISTments2009 数据集中答题数量的分布图
Fig. 1 Distribution of answer counts in ASSISTments2009 dataset

地进行建模的信息,尤其在数据量较稀疏的场景中,基于元学习方法获得的模型往往具有较好的性能.受模型无关元学习算法 MAML^[15](Model Agnostic Meta Learning)的启发,在给定的多个知识追踪建模任务中,本文首先随机设定其中一个为目标任务(或测试任务),其他任务为辅助任务(或训练任务),然后整合多个辅助任务的数据集构建元训练集,基于某个特定知识追踪算法(例如 DKT),在多任务学习框架下利用元训练集对模型进行预训练.最后,将预训练得到的模型参数作为目标任务模型的初始化参数,并利用目标任务的训练数据做进一步的迭代优化,直至模型收敛到最优(或局部最优)状态.

本文的主要贡献如下.

(1)面向学生答题早期数据极度稀缺这一常见的真实应用场景,提出早期知识追踪建模预测问题,旨在答题数量较少的情况下,通过改进或设计新方法实现对学生知识水平的精准建模,以提高知识追踪模型的预见性.

(2)提出基于元学习增强的知识追踪建模方法 MetaKT,并设计了模型无关元学习知识追踪框架,通过借助其他相关知识追踪数据学习到的知识,辅助目标任务上知识追踪模型的训练,从而缓解答题数据量稀缺导致模型性能下降的问题.

(3)在七个公开数据集上的实验验证了所提 MetaKT 方法在早期知识追踪任务上的有效性.以 DKT 模型为基准,常用的指标 AUC (Area under Curve)取得了 27 升 8 降的结果;以 DKVMN 模型为基准,取得了 33 升 2 降的优异表现.相关代码和数据详见 <https://github.com/JHFSUE/MetaKT>.

1 相关工作

常用的知识追踪方法包括贝叶斯知识追踪和深度知识追踪等两类方法.

1.1 贝叶斯知识追踪 1994年 Corbett et al^[6]提出贝叶斯知识追踪模型,定义了学生的先验知识、完成一次答题后掌握相应知识概念的概率、猜测参数和粗心参数等变量,并利用隐马尔可夫模型对上述四个参数进行学习,然后采用贝叶斯公式实时地对学生的知识掌握水平进行更新.由于这

四个参数的意义明确,预测结果更容易解释. BKT 自提出以来受到了广泛关注,并成为最流行的知识追踪模型之一. Pardos and Heffernan^[16]在基础 BKT 模型中先后引入学生个体先验知识、问题难度^[17]和学生在每个问题上尝试作答的次数^[18]等信息以提升模型性能. Nedungadi and Remya^[19]利用一个随时间衰减的函数在基础 BKT 模型中引入遗忘因素. Wang et al^[20]同时引入知识状态的层次属性和时间属性对 BKT 进行改进.

1.2 深度知识追踪 2015年斯坦福大学 Piech et al^[7]首次将 RNN 用在知识追踪任务上,提出深度知识追踪模型 DKT. 得益于深度神经网络强大的表征学习能力和函数拟合能力,DKT 模型在公开数据集 ASSISTments 上的 AUC 指标高出传统 BKT 模型近 25%. 此后,深度神经网络得到了研究者广泛的关注,涌现了一批成效卓著的深度知识追踪建模方法. 动态键值记忆网络(Dynamic Key-Value Memory Networks, DKVMN)^[8]通过两个内存矩阵来建模知识点的底层概念结构并跟踪每个概念的掌握状态. 注意力知识追踪(Attentive Knowledge Tracing, AKT)^[9]采用单调注意力机制,根据交互序列与当前时间的距离,动态调整交互过程中试题的注意力权重. 基于图的知识追踪(Graph-Based Knowledge Tracing, GKT)^[21]将知识追踪任务转化为图神经网络中针对时间序列节点的分类问题,构建了一个知识点关系图并明确地学习了它们的关系. Su et al^[12]和 Liu et al^[22]通过引入试题文本、遗忘因素^[23]等辅助信息来建模学生的知识状态. Liu et al^[24]引入试题与概念的对应关系来构建知识追踪模型. Nagatani et al^[25]通过引入遗忘有关的信息对 DKT 模型进行扩展.

现有多数知识追踪方法聚焦于海量的学生答题数据进行建模,并取得了显著成效. 然而,鲜有研究关注在答题数量较少的情况下来实现对学生知识水平的精准建模的问题.

2 所提方法 MetaKT

2.1 问题描述 在知识追踪任务中,设 $Q = \{q_1, q_2, \dots, q_M\}$, 其中, $q \in Q$ 表示问题, M 表示问

题总数. 学生从 Q 中选择一些试题进行回答, 学生回答的记录为:

$$X = \{(q_1, r_1), (q_2, r_2), \dots, (q_t, r_t)\}$$

其中, $r \in \{0, 1\}$. 如果学生回答正确则 $r = 1$, 否则 $r = 0$. 模型的输出是 y_t , 表示学生在 t 时刻正确回答试题的概率.

从训练数据集 D_{train} 中随机采样学生样本数据, 构建元训练数据集 $T = \{T_1, T_2, \dots, T_n\}$, 其中, n 表示元训练数据集数量, $T_i = \{S_i, Q_i\}$, S_i 表示支持集, Q_i 表示查询集, 并且 $S_i \cap Q_i \neq \emptyset$. 在元训练阶段, 使用支持集 S_i 对模型进行训练, 使用查询集 Q_i 优化模型的初始参数. 在元测试阶段, 将测试数据集 D_{test} 划分为支持集 d_{train} 和查询集 d_{test} , 并且 $d_{\text{train}} \cap d_{\text{test}} = \emptyset$, 利用在元训练中学习到的先验知识, 通过支持集 d_{train} 微调, 最终在查询

集 d_{test} 上测试模型性能.

2.2 模型整体框架 图 2 给出了所提方法 MetaKT 的基本框架, 右下角的知识状态是基于元学习优化后的学生知识掌握概率分布, 作为元测试阶段的最终预测结果, 三角形的形态(如大小、颜色层次)反映知识掌握的不同维度与水平. 如图 2 所示, 为了增强模型预测小样本数据的能力, MetaKT 通过模型无关元学习算法 MAML^[15] 对基学习模型参数进行优化. 元学习参数优化分为元训练和元测试两个阶段, 在元训练阶段, 首先从知识追踪数据集中随机抽取一定数量的学生交互数据构建元训练数据集, 并基于初始基学习器进行双层优化. 在元测试阶段, 利用在元训练中学习到的先验知识进行微调, 微调后, 评估模型在小样本知识追踪数据集上的性能.

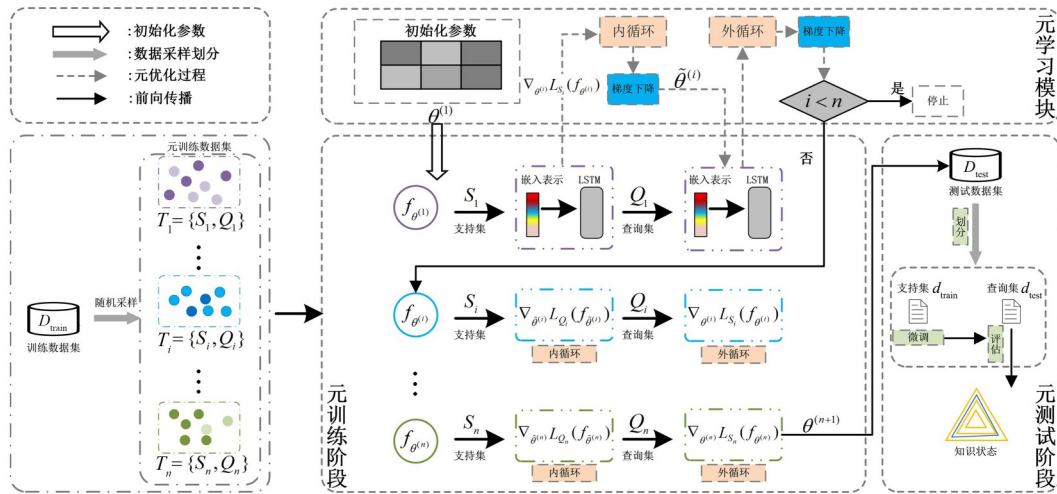


图 2 MetaKT 的框架

Fig. 2 The framework of MetaKT

2.3 元优化学习 首先, 构建元训练数据集, 并在元训练阶段模型进行参数优化; 然后, 在元测试阶段, 利用在元训练中学习到的先验知识, 对小样本数据进行测试, 以验证模型在新任务上的预测效果. 以下算法给出了所提 MetaKT 方法的详细描述.

算法 元优化算法	
输入:	内循环学习率 η , 外循环学习率 γ , 测试数据集 D_{test} , 元训练数据集 D_{train}
输出:	正确回答的概率 y_t

1. 初始化 $\theta^{(1)}$

2. 元训练

3. 当 $i < Epoch$ 时

4. 从 D_{train} 中随机取样 $T = \{S, Q\}$

5. 支持集的损失梯度为 $\nabla_{\theta^{(i)}} L_S(f_{\theta^{(i)}})$

6. 用梯度下降法使用 η 更新参数:

$$\tilde{\theta}^{(i)} \leftarrow \theta^{(i)} - \eta * \nabla_{\theta^{(i)}} L_S(f_{\theta^{(i)}})$$

7. 计算查询集的损失为 $L_Q(f_{\tilde{\theta}^{(i)}})$

8. 使用 γ 更新参数: $\theta^{(i+1)} \leftarrow \theta^{(i)} - \gamma * \nabla_{\tilde{\theta}^{(i)}} L_Q(f_{\tilde{\theta}^{(i)}})$

9. 结束

10. 元测试

11. 将 D_{test} 划分为 d_{train} 和 d_{test}
12. 利用元训练优化后的模型对 d_{train} 进行微调
13. 基于微调后的模型评估 d_{test}

具体地,构建元训练数据集的过程如下.

训练数据集 D_{train} 包含多个知识追踪数据集,用于模型在元学习过程中的训练阶段;测试数据集 D_{test} 为目标数据集,表示用于微调和评估的模型. $p(\text{train})$ 和 $p(\text{test})$ 分别对应数据集 D_{train} 和 D_{test} 的分布,并且训练集和测试集的分布不相交,即 $p(\text{train}) \cap p(\text{test}) = \emptyset$.

从 D_{train} 中随机采样学生样本数据来构建元训练数据集 T . T_i 由支持集 S_i 和查询集 Q_i 组成,即 $T_i = \{S_i, Q_i\}$, 其中, i 表示第 i 个支持集和查询集,且 $S_i \cap Q_i \neq \emptyset$. 最终构建的元训练数据集为 $T = \{T_1, T_2, \dots, T_n\}$.

在元训练阶段,受模型无关元学习算法 MAML^[15] 在处理小样本方面的成功启发,首先对元训练数据集进行训练,然后,将元训练优化后的模型参数作为元测试阶段模型的初始化参数.

具体地,首先初始化模型参数为 $\theta^{(1)}$, 经过 n 个元训练集 T 对模型参数进行优化. 在第 i 次元优化步骤中,在支持集 S_i 上进行适应,计算损失 $L_{S_i}(f_{\theta^{(i)}})$, 如式(1)所示:

$$\tilde{\theta}^{(i)} \leftarrow \theta^{(i)} - \eta * \nabla_{\theta^{(i)}} L_{S_i}(f_{\theta^{(i)}}) \quad (1)$$

其中, S_i 表示元训练数据集 T_i 中的支持集; 函数 f 代表基学习器的映射关系,其数学定义为 $f_{\theta}: X \rightarrow y$, y 表示预测学生正确回答知识点的概率; $\tilde{\theta}^{(i)}$ 表示经过支持集梯度下降更新后的模型参数, $L_{S_i}(f_{\theta^{(i)}})$ 表示使用模型参数 $\theta^{(i)}$ 在支持集 S_i 上计算的交叉熵损失; η 是内循环学习率. 接着,通过查询集来评估模型并再次计算损失,如式(2)所示:

$$\theta^{(i+1)} \leftarrow \tilde{\theta}^{(i)} - \gamma * \nabla_{\tilde{\theta}^{(i)}} L_{Q_i}(f_{\tilde{\theta}^{(i)}}) \quad (2)$$

其中, Q_i 表示元训练数据集 T_i 中的查询集, $\theta^{(i+1)}$ 表示经过查询集梯度下降更新后的模型参数, $L_{Q_i}(f_{\tilde{\theta}^{(i)}})$ 表示使用支持集更新后的模型参数在查询集 Q_i 上计算的交叉熵损失, γ 是外循环学习率. 经过 n 次元训练之后,可以获得具有先验知识 $\theta^{(n+1)}$ 的模型.

最后进入元测试阶段. 由于元训练优化后的

模型参数不完全适应测试数据集,因此,需要对测试数据集 D_{test} 进行微调. 将 D_{test} 划分为支持集 d_{train} 和查询集 d_{test} , 使用元训练优化后的参数模型在支持集 d_{train} 上进行微调. 微调时,根据支持集损失计算梯度并更新参数,使模型收敛到更优状态,如式(3)所示:

$$\theta \leftarrow \theta^{(n+1)} - \alpha * \nabla_{\theta^{(n+1)}} L_{d_{\text{train}}}(f_{\theta^{(n+1)}}) \quad (3)$$

其中, α 表示元测试阶段的学习率, $L_{d_{\text{train}}}(f_{\theta^{(n+1)}})$ 表示使用元训练优化后的参数 $\theta^{(n+1)}$ 在支持集 d_{train} 上计算的交叉熵损失, θ 表示经过支持集 d_{train} 微调后的模型参数. 最后,使用查询集 d_{test} 评估微调后模型的性能.

3 实验结果与分析

分别将两个经典的深度知识追踪模型 DKT^[7] 和 DKVMN^[8] 作为基准模型放入所提方法 MetaKT 中开展实验. 需要注意的是, MetaKT 是一个面向早期知识追踪问题的普适性框架, 现有的其他深度知识追踪模型, 例如 AKT^[9] 和 GKT^[21], 也可以作为基准模型放入该框架进行训练. 受算力和时间所限, 将 MetaKT 与其他更为先进的深度知识追踪模型进行融合的研究留待以后开展. 为了分析所提方法的性能表现, 设计了如下两个问题.

(1) MetaKT 框架是否可以带来基准模型的性能提升?

(2) 答题数据量的变化对模型性能产生怎样的影响?

首先介绍实验所需数据集和实验设置, 然后, 围绕上述两个问题进行实验结果的分析.

3.1 数据集 为了评估 MetaKT 的有效性, 本文采用七个公开的数据集, 这些数据均来源于真实教学平台的学生答题数据, 包括 EdNet, ASSISTments2009 (ASSIST09), ASSISTments2012 (ASSIST12), ASSISTments2015 (ASSIST15), ASSISTments2017 (ASSIST17), JunyiAcademy (Junyi) 和 algebra05. 表 1 给出了七个数据集的详细情况.

3.2 实验 操作系统为 Windows 10 专业版 64-bit (Build 19045), 硬件配置为 Intel Xeon W-

表 1 实验数据集的基本统计

Table 1 Basic statistics of the experimental datasets

数据集	学生数量	知识概念数量	试题数量	每个学生平均答题数量
EdNet	10000	1792	1173020	117
ASSIST09	4151	110	325637	78
ASSIST12	28834	198	2629870	91
ASSIST15	19840	100	683801	34
ASSIST17	1709	102	941107	550
Junyi	7101	1326	1577903	222
algebra05	574	138	809694	1410

2255 CPU (20 个核心, 基准频率为 3.7 GHz), 64 GB 内存, 显卡为 NVIDIA GeForce RTX 3090. 开发环境为 PyCharm 2021.3. 所有模型均使用 PyTorch 实现, 优化器为 Adam. 默认设置学习率为 0.001. 批量大小为 64, 深度学习追踪模型的 dropout 为 0.1. 使用 5 折交叉验证来评价模型的性能. 设置元测试阶段学习率 $\alpha = 0.001$, 内循环学习率 η 的取值范围为 $\{0.01, 0.001, 0.0001\}$, 外循环学习率 γ 的取值范围为 $\{0.01, 0.001\}$.

对于动态键值记忆网络(DKVMN)模型, 键矩阵(d_k)和值内存矩阵(d_v)维度的取值范围为 $\{10, 50, 100, 200\}$. 同时, 该模型的内存维度 N 的取值范围为 $\{10, 20, 50\}$.

3.3 模型的评价指标 为了能够准确地预测学生未来的学习与答题表现, 将其建模为一个二元分类问题, 选取 AUC 和 ACC 作为评价指标. AUC (Area Under ROC Curve) 是 ROC (Receiver Operating Characteristic Curve) 曲线下面积, 通

通过对 ROC 曲线下各部分的面积求和而得到. ROC 曲线以假正例率(False Positive Rate, FPR) 为横轴、真正例率(True Positive Rate, TPR) 为纵轴. AUC 衡量了学习器在不同分类阈值下的整体性能, 其取值范围为 $[0, 1]$, 越接近 1 表示模型的区分能力越强. ACC (Accuracy, 准确率) 反映了分类器整体预测的正确性, 其取值范围同样为 $[0, 1]$, 其值越大说明分类器的整体预测精度越高.

3.4 与基准模型比较 通过比较所提方法与基准方法的性能, 对 MetaKT 框架是否可以提升基准模型的性能进行解答. 为了便于分析, 将所提方法与 DKT (DKVMN) 结合后的模型命名为 MetaDKT (MetaDKVMN), 逐一进行分析说明.

(1) 以 DKT 为基准模型. 表 2 给出了以 DKT 模型为基准, 在七个公开数据集上的比较实验结果, 表中箭头的上和下分别表示 MetaDKT 和 DKT 模型相比, 性能升高或降低. 由表可见, MetaDKT 的整体表现优于基准 DKT 模型. 在最大答题数量为 10 时, 和基准模型相比, 均有不同幅度的提升. 在最大答题数量为 5, 20, 30, 50 时, 和基准模型相比, 均取得了五胜二负的结果.

(2) 以 DKVMN 为基准模型. 表 3 给出了以 DKVMN 模型为基准, 在七个公开数据集上的比较实验结果, 表中箭头的上和下分别表示 MetaDKVMN 和 DKVMN 模型相比, 性能升高或降低. 由表可见, MetaDKVMN 的整体表现优于基准 DKVMN 模型. 在最大答题数量为 5, 20, 50 时, 和基准模型相比, 算法性能均有不同幅度的提

表 2 以 DKT 模型为基准在七个数据集上的 AUC Table 2 AUC of the DKT model as the backbone across seven datasets

数据集	最大答题量 5		最大答题量 10		最大答题量 20		最大答题量 30		最大答题量 50	
	DKT	MetaDKT	DKT	MetaDKT	DKT	MetaDKT	DKT	MetaDKT	DKT	MetaDKT
EdNet	63.80%	64.00% ↑	60.24%	60.36% ↑	66.71%	66.90% ↑	67.18%	67.19% ↑	67.57%	67.60% ↑
Junyi	64.54%	65.23% ↑	72.77%	73.12% ↑	67.12%	67.19% ↑	67.77%	67.95% ↑	68.61%	68.66% ↑
ASSIST09	79.17%	79.21% ↑	74.42%	74.95% ↑	79.42%	79.61% ↑	79.71%	79.70% ↓	79.74%	79.70% ↓
ASSIST17	62.53%	63.28% ↑	62.69%	63.51% ↑	62.97%	63.13% ↑	63.87%	63.92% ↑	64.82%	64.90% ↑
algebra05	70.24%	69.79% ↓	77.25%	77.54% ↑	71.18%	71.14% ↓	71.69%	72.33% ↑	73.72%	74.16% ↑
ASSIST12	62.50%	62.90% ↑	69.22%	69.36% ↑	65.35%	65.40% ↑	66.07%	66.22% ↑	66.81%	66.78% ↓
ASSIST15	70.03%	69.98% ↓	75.30%	75.63% ↑	71.58%	71.57% ↓	71.84%	71.76% ↓	72.02%	72.05% ↑

表 3 以 DKVMN 模型为基准在七个数据集上的 AUC

Table 3 AUC of the DKVMN model as the backbone across seven datasets

数据集	最大答题量 5		最大答题量 10		最大答题量 20		最大答题量 30		最大答题量 50	
	DKVMN	Meta-DKVMN	DKVMN	Meta-DKVMN	DKVMN	Meta-DKVMN	DKVMN	Meta-DKVMN	DKVMN	Meta-DKVMN
EdNet	65.80%	65.87% ↑	65.72%	65.75% ↑	68.05%	68.13% ↑	68.83%	68.96% ↑	69.18%	69.24% ↑
Junyi	64.75%	64.97% ↑	66.84%	66.78% ↓	67.53%	67.57% ↑	68.21%	68.14% ↓	68.88%	68.92% ↑
ASSIST09	75.42%	75.75% ↑	75.77%	76.10% ↑	76.32%	76.98% ↑	76.65%	76.92% ↑	77.15%	77.32% ↑
ASSIST17	62.82%	63.48% ↑	62.94%	63.05% ↑	63.43%	63.56% ↑	64.49%	64.66% ↑	65.27%	65.35% ↑
algebra05	68.07%	69.93% ↑	69.04%	70.81% ↑	72.68%	73.02% ↑	73.46%	73.62% ↑	74.82%	74.87% ↑
ASSIST12	62.93%	62.98% ↑	64.50%	65.53% ↑	65.93%	66.04% ↑	66.82%	66.84% ↑	67.64%	67.69% ↑
ASSIST15	70.16%	70.18% ↑	71.07%	71.13% ↑	71.50%	71.60% ↑	71.72%	71.80% ↑	71.93%	72.02% ↑

升. 在最大答题数量为 10 和 30 时, 和基准模型相比, 均取得了六胜一负的结果. 以最大答题数量是 5 为例, 在 algebra05 数据集上, MetaDKVMN 的 AUC 比基准 DKVMN 模型高 1.86%; 在 ASSIST2017 数据集上, MetaDKVMN 的 AUC 比基准 DKVMN 模型高 0.66%. 表明 MetaDKVMN 通过引入元学习机制, 在一定程度上提升了 DKVMN 在小样本场景下的性能, 也证明 MetaKT 的元学习优化策略在多数场景下能够提升基准模型的性能.

(3) 显著性分析. 为了进一步验证 MetaKT 的性能提升是否具有统计学显著性, 在对比实验中引入非参数检验方法, 即符号检验 (Sign Test). 该方法不依赖结果分布的正态性假设, 能够稳健地检验 MetaKT 和基准模型相比, 改进是否显著. 具体地, 符号检验通过统计 MetaKT 在多个实验场景中“优于”或“劣于”基准模型的次数, 并基于二项分布计算显著性水平.

实验结果如下.

(1) 实验 1 (MetaDKT vs. DKT), 35 个场景中 MetaKT 提升 27 次, 下降 8 次. 符号检验得到的 $p = 0.00094$ ($P(X \geq 27) \approx 0.00094$), 显著小于 0.001, 可以认为 MetaKT 对 DKT 的提升具有显著性.

(2) 实验 2 (MetaDKVMN vs. DKVMN), 35 个场景中 MetaKT 提升 33 次, 下降 2 次. 符号检验得到的 $p = 1.84 \times 10^{-8}$ ($P(X \geq 33) \approx 1.84 \times 10^{-8}$), 远小于 0.0001, 可以认为 MetaKT 对 DKVMN 的

提升具有极显著性.

综上, 即使在部分场景下, 数值提升的幅度有限, MetaKT 对基准模型的改进仍然具有统计学显著性, 进一步增强了实验结论的可信度.

3.5 答题数量对模型性能的影响 为了验证答题数据量对模型性能的影响, 从源数据集中截取 5, 10, 20, 30, 50 等不同长度的序列开展实验. 图 3 和图 4 显示七个数据集上模型随着数据量的增加, 其 AUC 的变化. 由图可见, 数据量过少时, 所有方法的性能都不理想, 随着训练数据增加, 所有模型的 AUC 整体呈上升趋势, 表明更多的训练数据有助于提升模型性能. 根据所提方法在不同数据集上的表现, 可以看到以下趋势.

(1) 随着答题交互数量的增加, 所有模型的 AUC 均有所提升, 其中所提方法和基准 DKVMN 模型的融合在多数情况下取得了较好的性能. 这说明通过元训练阶段利用其他知识追踪数据集进行的参数优化是有效的, 能够快速有效地处理新的知识追踪任务.

(2) 在 ASSISTments2009 数据集上, 所提方法与基准 DKT 的融合性能较好. 在 ASSISTments2015 和 algebra05 数据集上, 所提方法与基准 DKT 模型相比, 性能偶有下降, 可能是因为 ASSISTments2015 包含更多的学生, 而且知识点数量最少, 平均每个知识点的回答率更高. algebra05 拥有的学生数量最少, 平均每个学生完成的试题最多.

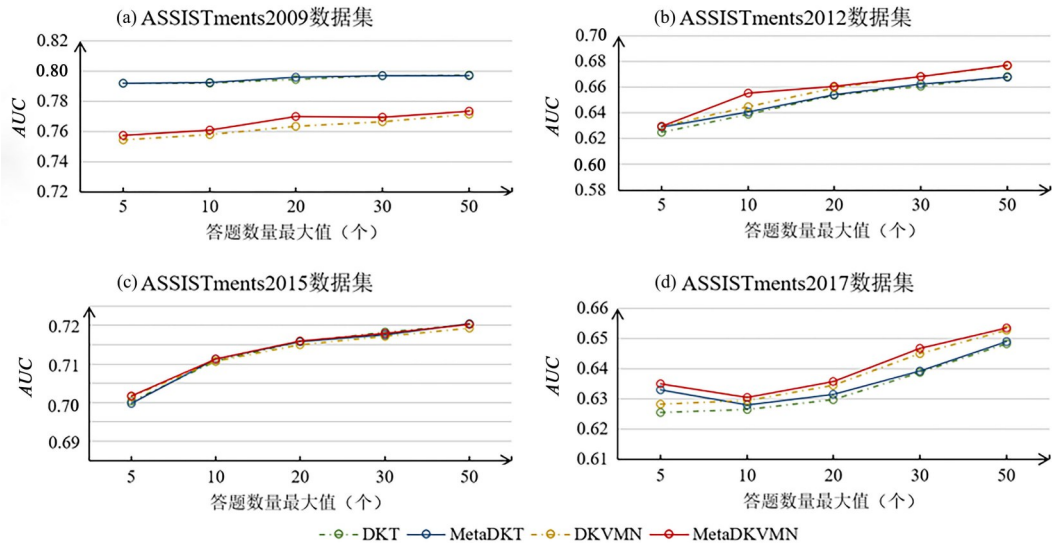


图 3 在各数据集上模型性能随答题数量的变化情况

Fig. 3 Performance of models with increasing response counts on different datasets

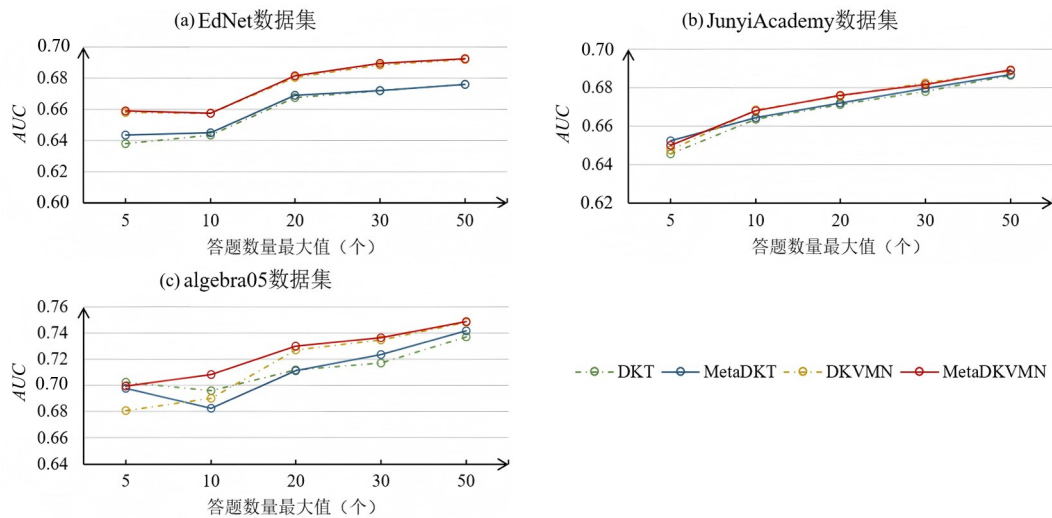


图 4 在EdNet等三个数据集上模型性能随答题数量的变化情况

Fig. 4 Performance of models with increasing response counts on three datasets, e. g., EdNet

4 结论

为了解决小样本环境下学生知识状态评估的难题,本文提出一种基于元学习增强的知识追踪模型 MetaKT,通过引入模型无关元学习算法来提高模型在小样本条件下的预测性能.在七个真实教育数据集上进行了广泛实验,经过非参数检验,结果表明,相对于基准方法,MetaKT算法带来的性能提升具有统计学上的显著性,这说明通过引入元学习机制,所提方法能够有效借助其他

相关任务上学习到的知识来提升模型性能.未来将优化 MetaKT 模型的架构,并将 MetaKT 与其他更先进的深度知识追踪模型进行融合,实现在学生答题早期对其知识掌握水平的精准建模,提升模型的预见性.

参考文献

[1] Miao F C, Wayne H, Huang R H, et al. AI and education: A guidance for policymakers. Paris: United Nations Educational, Scientific and Cultural

- Organization. <https://unesdoc.unesco.org/ark:/48223/pf0000379707>, 2021.
- [2] Anderson A, Huttenlocher D, Kleinberg J, et al. Engaging with massive online courses//Proceedings of the 23rd International Conference on World wide web. New York, NY, USA: Association for Computing Machinery, 2014:687–698.
- [3] Desmarais M C, Baker R S J D. A review of recent advances in learner and skill modeling in intelligent learning environments. *User Modeling and User - Adapted Interaction*, 2012, 22(1):9–38.
- [4] Shen S H, Liu Q, Huang Z Y, et al. A survey of knowledge tracing: Models, variants, and applications. *IEEE Transactions on Learning Technologies*, 2024, 17:1898–1919.
- [5] Atkinson R C. Ingredients for a theory of instruction. *American Psychologist*, 1972, 27(10):921–931.
- [6] Corbett A T, Anderson J R, et al. Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modeling and User - adapted Interaction*, 1994, 4:253–278.
- [7] Piech C, Bassen J, Huang J, et al. Deep knowledge tracing. *Advances in Neural Information Processing Systems*, 2015, 28:505–513.
- [8] Zhang J N, Shi X J, King I, et al. Dynamic key-value memory networks for knowledge tracing//Proceedings of the 26th International Conference on World Wide Web. Geneva, Switzerland: International World Wide Web Conferences Steering Committee, 2017:765–774.
- [9] Ghosh A, Heffernan N, Lan A S. Context-aware attentive knowledge tracing//Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York, NY, USA: Association for Computing Machinery, 2020:2330–2339.
- [10] Shen S H, Liu Q, Chen E H, et al. Convolutional knowledge tracing: Modeling individualization in student learning process//Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval. New York, NY, USA: Association for Computing Machinery, 2020:1857–1860.
- [11] Yang Y, Shen J, Qu Y R, et al. GIKT: A graph-based interaction model for knowledge tracing//Machine Learning and Knowledge Discovery in Databases: European Conference. Heidelberg, Germany: Springer, 2020:299–315.
- [12] Su Y, Liu Q W, Liu Q, et al. Exercise-enhanced sequential modeling for student performance prediction//Proceedings of the 32nd AAAI Conference on Artificial Intelligence, 13th Innovative Applications of Artificial Intelligence Conference and 8th AAAI Symposium on Educational Advances in Artificial Intelligence. Menlo Park, NJ, USA: AAAI Press, 2018:2435–2443.
- [13] Cui C R, Yao Y M, Zhang C Y, et al. DGEKT: A dual graph ensemble learning method for knowledge tracing. *ACM Transactions on Information Systems*, 2004, 42(3):1–24.
- [14] Lemke C, Budka M, Gabrys B. Metalearning: A survey of trends and technologies. *Artificial Intelligence Review*, 2015, 44(1):117–130.
- [15] Finn C, Abbeel P, Levine S. Model-agnostic meta-learning for fast adaptation of deep networks//Proceedings of the 34th International Conference on Machine Learning. Volume 70. Sydney, Australia: JMLR.org, 2017:1126–1135.
- [16] Pardos Z A, Heffernan N T. Modeling individualization in a Bayesian networks implementation of knowledge tracing//Proceedings of the 18th International Conference on User Modeling, Adaptation, and Personalization. Heidelberg, Germany: Springer, 2010:255–266.
- [17] Pardos Z A, Heffernan N T. KT-IDEM: Introducing item difficulty to the knowledge tracing model//Proceedings of the 19th International Conference on User Modeling, Adaption, and Personalization. Heidelberg, Germany: Springer, 2011:243–254.
- [18] Pardos Z A, Bergner Y, Seaton D T, et al. Adapting Bayesian knowledge tracing to a massive open online course in edX//Proceedings of the 6th International Conference on Educational Data Mining. Memphis, TE, USA: International Educational Data Mining Society, 2013:137–144.
- [19] Nedungadi P, Remya M S. Incorporating forgetting in the personalized, clustered, bayesian knowledge tracing model//2015 International Conference on

- Cognitive Computing and Information Processing. Noida, India: IEEE, 2015: 1–5.
- [20] Wang Z, Zhu J L, Li X, et al. Structured knowledge tracing models for student assessment on Coursera// Proceedings of the 3rd ACM Conference on Learning @Scale. New York, NY, USA: Association for Computing Machinery, 2016: 209–212.
- [21] Nakagawa H, Iwasawa Y, Matsuo Y. Graph-based knowledge tracing: Modeling student proficiency using graph neural network//2019 IEEE/WIC/ACM International Conference on Web Intelligence. Thessaloniki, Greece: IEEE, 2019: 156–163.
- [22] Liu Q, Huang Z Y, Yin Y, et al. Ekt: Exercise-aware knowledge tracing for student performance prediction. IEEE Transactions on Knowledge and Data Engineering, 2019, 33(1): 100–115.
- [23] Huang Z Y, Liu Q, Chen Y Y, et al. Learning or forgetting? A dynamic approach for tracking the knowledge proficiency of students. ACM Transactions on Information Systems, 2020, 38(2): 1–33.
- [24] Liu Y F, Yang Y, Chen X Y, et al. Improving knowledge tracing via pre-training question embeddings. <https://arxiv.org/abs/2012.05031>, 2020–12–09.
- [25] Nagatani K, Zhang Q, Sato M, et al. Augmenting knowledge tracing by considering forgetting behavior//The World Wide Web Conference. New York, NY, USA: Association for Computing Machinery, 2019: 3101–3107.

(责任编辑 杨可盛)