

基于上下文信息的知识追踪方法

何萌红¹

HE Menghong

摘要

知识追踪 (KT) 旨在通过分析学生的学习过程及其知识掌握情况来预测学生未来的学习表现。随着科技的进步, 知识追踪技术不断成熟。然而, 传统的知识追踪模型通常忽视了问题之间潜在的关联, 并且在处理长序列数据时, 模型难以有效捕捉长期依赖关系。为解决这些问题, 文章提出了一种基于上下文信息的知识追踪模型。该模型通过引入注意力机制计算问题之间的相关性, 确保当前问题获得更多相关信息, 并有效捕捉问题的长序列信息。这种方法减少了信息在传递过程中的丢失, 通过实验表明, 与现有的模型相比本模型具有更好的预测能力。

关键词

深度学习; 知识追踪; 教育系统; 上下文信息; 注意力机制

doi: 10.3969/j.issn.1672-9528.2024.12.024

0 引言

随着社会的快速发展, 人工智能在生活中的应用越来越广泛, 教育领域也因人工智能技术的快速演进而向智能化教育转型^[1], 在智能辅导系统中, 通过计算机辅助技术帮助学生进行个性化学习, 使用知识追踪来跟踪和掌握学生的知识状态。一般来说, 知识追踪通过学生的历史学习行为进行定量分析, 评估学生对知识点掌握的动态演进并以此预测学生能否回答新问题。根据此来跟踪学生在练习过程中的知识获取变化, 使得学生可以及时了解自己的知识点掌握情况, 并对知识薄弱部分进行针对性的练习。在此背景下, 知识追踪 (knowledge tracing, KT) 技术应运

而生。其核心在于通过分析学生的历史答题记录, 对其知识状态进行建模, 并预测其答对下一题的概率, 从而揭示学习过程中学生知识掌握的内在规律。这一技术极大地推动了智能辅助教育系统的发展^[2]。

现有的知识追踪模型大致可分为三类: 基于概率图模型的知识追踪、基于矩阵分解的知识追踪, 以及基于深度学习的知识追踪。Corbett 等人^[3]于 1995 年提出的贝叶斯知识追踪 (Bayesian Knowledge Tracing) 是一种典型的基于概率图模型的知识追踪方法, 它假设学生的知识状态是一个隐含的二值变量, 通过观察学生的学习行为 (例如做题记录) 来更新该知识状态的后验概率分布。而基于矩阵分解的知识追踪是通过将学生的学习记录转化为一个矩阵, 利用矩阵分解等方法来挖掘学生的知识状态和学习行为之间的潜在关联, 以实现知识追踪。然而, 这些模型仍存在局限性, 无法有效建模知识点之间的相关性, 难以捕捉学生学习过程中的复杂行为。随着技术的进步, Piech

1. 贵州民族大学 贵州贵阳 550025

[基金项目] 贵州省教育厅 (黔教技〔2022〕047 号), 贵州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)

效计算方法 [J]. 物理学报, 2024, 73(7):114-123.

[7] 张佳宁, 沈慧, 周广东. 结构监测无线传感网络数据传输优化方法 [J]. 哈尔滨工程大学学报, 2024, 45(8):1543-1551.

[8] 邵云飞, 宋友, 王宝会. 基于社交网络图节点度的神经网络个性化传播算法研究 [J]. 计算机科学, 2023, 50(4):16-21.

[9] 杨慎, 陈磊, 周绮凤. 基于图增强和图神经网络的层次社区发现方法 [J]. 厦门大学学报 (自然科学版), 2024, 63(2): 209-220.

[10] 全卫国, 曾世超, 张立峰. 多尺度卷积神经网络的电阻层

析成像算法 [J]. 计算机应用与软件, 2024, 41(5):177-182.

[11] 孙乾, 蒋楠. 基于卷积神经网络的高效量子态层析方法 [J]. 北京师范大学学报 (自然科学版), 2024, 60(3):325-330.

【作者简介】

谢鸿博 (2004—), 男, 河北秦皇岛人, 本科, 研究方向: 计算机科学与技术。

(收稿日期: 2024-08-09)

等人^[4]首次将深度神经网络引入知识追踪领域。如图1所示,与传统知识追踪模型相比,基于深度学习的知识追踪模型能够动态追踪学生知识熟练度随时间的变化过程,更好地对学生学习过程中长序列信息的建模,提供更准确的知识状态反映,并具备更强的预测能力。

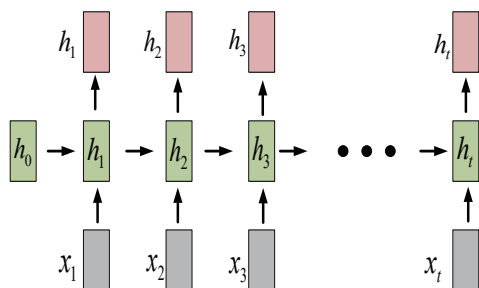


图1 DKT模型图

随着研究的深入,越来越多的知识追踪模型被提出。例如,由于学生的记忆会随时间衰减,DKT-forget^[5]模型考虑了这种遗忘行为;KSGKT^[6]模型则着重考虑了技能之间的先决条件。这些模型表明,当前的知识追踪方法多集中于对问题所涉及知识点的建模,并不断演进以更好地捕捉学生的学习动态。在现有的知识追踪模型中,往往忽视了问题本身在学习过程中所蕴含的潜在信息。然而,学生在回答当前问题时,通常需要依赖对历史问题的理解和掌握程度,才能准确判断和作答。如果未能有效建立当前问题与历史问题之间的关联,可能导致学生无法充分理解当前问题,从而影响预测的准确性。此外,尽管RNN^[7]模型及其变体在处理长序列信息时具有一定的优势,但在捕捉学生学习历史的长期依赖性方面仍然存在不足。RNN模型在面对长序列时,容易出现信息丢失或衰减的现象,导致模型在处理较长学习序列时性能下降,且受到记忆遗忘的影响也会使当前问题对历史知识点的掌握出现一些偏差。

为了解决这些问题,本模型在现有研究的基础上进行了优化和改进。具体而言,模型充分考虑了问题之间的潜在关系,并引入了注意力机制^[8],对当前问题与历史问题进行建模,深入挖掘二者间的潜在关联。通过注意力机制,模型能够对每个历史问题的重要性进行加权,突出与当前问题密切相关的历史问题,从而提供更丰富、更精准的知识表示。此外,模型还纳入了遗忘因子的影响,有效捕捉学生的历史学习信息,减少信息传递过程中的丢失,显著提升了模型的预测能力。通过这些改进,模型不仅能够准确捕捉问题之间的上下文信息,还解决了长序列信息的长期依赖性问题,从而大幅提升了对未来问题回答情况的预测精度。

1 模型描述

本文将从历史交互模块和预测模块这两个模块来解释模型,如图2所示。

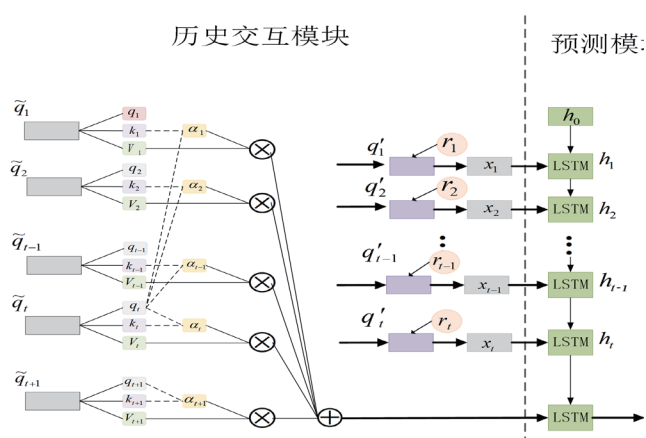


图2 整体框架

1.1 历史问题交互模块

学生的知识状态在学习过程中是动态变化的,随着时间的推移,学生对一些已学问题的记忆可能逐渐减弱,甚至被“遗忘”,但这些历史问题仍可能对学生当前的学习表现产生重要影响。这种遗忘效应导致学生在回答当前问题时,历史问题的作用未能得到充分的利用。LSTM(长短期记忆网络)^[9]虽然擅长处理序列数据,尤其是通过记忆单元来保留一些较为近期的状态信息,但它在处理长时间跨度的依赖关系时存在明显的局限性。当序列长度增加时,LSTM可能难以有效保留先前的学生状态,导致对与当前问题相关的早期历史问题的“遗忘”。

为了解决这个问题,本文引入了相关历史问题交互模块,通过注意力机制来学习当前问题与历史问题之间的相关性。这一机制能够捕捉不同历史问题对当前问题的影响,从而有效考虑历史知识对当前学习状态的影响。具体来说,通过注意力机制计算当前问题与历史问题之间的相关性权重。这些权重表征了历史问题对当前问题的影响程度。基于这些权重,可以加权汇聚历史问题的影响,帮助当前状态更准确地反映出学生在当前问题上的能力。这样,当前问题的表示不仅包含了当前的特征信息,还融入了历史知识的影响,从而弥补了LSTM在捕获长期依赖关系方面的不足。

具体来说,将当前时间步的题目嵌入 q_t 映射为查询向量(query),历史问题 $q_i(i=1,2,\dots,t-1)$ 映射为键向量(key)和值向量(value),且通过研究发现,人们的记忆会随着时间的推移呈指数衰减,鉴于人脑的遗忘机制,由此,计算得到当前时间步 t 的问题与历史时间步 i 的历史问题的注意力权重:

$$\hat{o}_{ii} = \text{softmax}(w_{ii}) = \frac{\exp(w_{ii})}{\sum_{i=1}^{t-1} \exp(w_{ii})} \quad (1)$$

$$w_{ii} = \frac{\exp(-\theta |t_i - t_t|)(W^Q q_t)^T \times W^K q_i}{\sqrt{D_k}} \quad (2)$$

式中： $|t_{i+1}-t_i|$ 是步骤 i 和 $t+1$ 之间的时间间隔； $\theta > 0$ 是随时间的可学习衰减率； $W^Q \in \mathbb{R}^{d \times D_k}$ 、 $W^K \in \mathbb{R}^{d \times D_k}$ 是可学习的权重矩阵，也是自注意力机制中 query 和 key 的映射矩阵； $\sqrt{D_k}$ 是缩放因子，用于防止点积结果过大，从而避免梯度消失或爆炸。

通过以上的计算从而得到当前问题的最终表示：

$$q'_i = \sum_{i=1}^{t-1} \partial_{t,i} W^K q_i \quad (3)$$

对于每个历史时间步 i ，将问题和答案的嵌入向量进行拼接，并通过非线性变换将其映射到 d 维度，得到练习表示：

$$x_i = \text{ReLU}(W_1([q'_i, r_i]) + b_1) \quad (4)$$

式中：用 $[\]$ 表示向量的拼接； W_1 和 b_1 分别是全连接层和 ReLU 激活层中的参数。

1.2 预测模块

学生的学习能力随着时间的推移不断积累，这表明当前的学习状态不仅与当下的知识掌握情况有关，更深受历史知识的影响。学生在学习过程中的每一次题目作答都可以看作是知识积累的一部分，逐渐形成一个反映其学习进程的动态时间序列。为了更好地理解和建模这种复杂的动态关系，传统的机器学习方法可能无法有效捕捉数据中的时间依赖性。然而，LSTM（长短期记忆网络）因其专门设计的记忆单元，能够在长时间序列中保持关键信息，具有强大的捕捉时间依赖关系的能力，特别适合处理这种序列数据。

LSTM 不仅可以记住学生先前答题时的表现，还能够适应知识掌握水平随着时间的变化趋势，从而为预测学生未来答题的准确性提供更加精准的依据。通过保留长期和短期记忆，LSTM 能够有效地从学生与一系列题目之间的交互中提取出潜在的学习模式，捕捉到历史答题记录中的重要信息，并利用这些信息来预测学生未来的表现。基于这些特性，选择使用 LSTM 来学习和预测学生的学习状态：

$$i_t = \sigma(W_i[x_t, h_{t-1}, c_{t-1}] + b_i) \quad (5)$$

$$f_t = \sigma(W_f[x_t, h_{t-1}, c_{t-1}] + b_f) \quad (6)$$

$$o_t = \sigma(W_o[x_t, h_{t-1}, c_{t-1}] + b_o) \quad (7)$$

$$c_t = f_t c_{t-1} + i_t \tanh(W_c[x_t, h_{t-1}] + b_c) \quad (8)$$

$$h_t = o_t \tanh(c_t) \quad (9)$$

式中： i_t 、 f_t 、 o_t 分别是输入门、遗忘门和输出门； c_t 是细胞记忆向量； W_* 和 b_* 是权重矩阵和偏置项。

第 $t+1$ 步的学生表现可以通过问题表示 q_{t+1} 和当前知识状态 h_{t+1} 进行预测，公式为：

$$p_{t+1} = \sigma(W_2[q_{t+1}, h_{t+1}] + b_2) \quad (10)$$

式中： W_2 和 b_2 分别是全连接层和 sigmoid 激活层中的参数。

为了优化模型，使用梯度下降来更新模型中的参数，具

体来说，最小化预测正确概率 p_{t+1} 与学生答案的真实标签 r_{t+1} 之间的交叉熵损失 L 来优化：

$$L = -(\sum_t r_{t+1} \log p_{t+1} + (1 - r_{t+1}) \log(1 - p_{t+1})) \quad (11)$$

2 实验与结果分析

在 3 个数据集进行一系列实验来评估本模型在知识追踪上的可行性。

2.1 数据集

实验数据集有 Assist09、Assist12 和 EdNet。详细的统计数据如表 1 所示，Assist09 数据集是于 2009—2010 年期间从 Assistment 在线教育平台收集。该数据集包含 3852 名学生、123 种技能、17 737 道问题和 282 619 次练习。

表 1 数据集信息

Dataset	Assist09	Assist12	EdNet
#students	3852	27 485	5000
#questions	17 737	53 065	12 161
#skills	123	265	189
#exercises	282 619	2 709 436	676 974

Assist12 数据集是数据集于 2012—2013 年期间从 Assist09 相同的平台收集。是 Assist09 数据集的更新版，在该数据集中，每个问题只与一种技能相关，但一种技能仍然对应多个问题。经过与 Assist09 相同的数据处理后，该数据集包含 27 485 名学生、265 种技能、53 065 道问题和 2 709 436 次练习。

EdNet 数据集是最近公开的数据集，由于整个数据集规模过大，随机选择了 5000 名学生，其中包含 189 种技能、12 161 道问题和 676 974 次练习。

对于每个数据集，将 80% 的序列作为训练集，20% 的序列作为测试集。为了评估这些数据集上的结果，采用了曲线下面积（AUC）作为评估指标。

2.2 基线

为了准确评估模型的性能，选择了以下模型作为基线进行比较。

(1) BKT 基于贝叶斯理论进行预测，将学生的知识状态建模为一组二进制变量。

(2) DKT 首个将知识追踪与深度神经网络结合的模型，使用 RNN 为基础模型建模学生的知识状态。

(3) DKVMN 通过扩展记忆增强网络，使用静态和动态矩阵分别存储不同概念的矩阵来建立学生的知识状态。

(4) DKT-Q DKT 模型的一种变体，将 DKT 的技能嵌入改为问题的独热编码作为输入进行预测。

(5) SAKT 采用自注意力机制捕捉输入序列中不同位置之间的关系的模型。

2.3 评价指标

为了准确评论模型的优劣，使用曲线下面积 (AUC) 来评估和预测模型的性能。

2.4 实验细节

在本文中，使用 TensorFlow 实现了所有对比方法，将技能、问题和答案的嵌入大小固定为 100，为了提高模型的计算能力，在嵌入模块，将学生交互长度设置为 200，嵌入维度 d 设置为 100. 最大聚合层 L 设置为 3，在 LSTM 模块，将 LSTM 中记忆单元大小分别设计为 200 和 100，其输出维度 d 设置为 100，表示学生知识状态的维度大小。在整个模型中，为了防止过拟合，使用概率为 0.3 的 dropout，并用 Adam 算法来对所有可训练的参数进行优化，学习率为 0.001，批次处理大小为 32。

2.5 整体性能

使用 BKT、DKT、DKT-Q、DKVMN^[10]、SAKT^[11] 模型在 3 个数据集上进行性能上的对比。表 2 记录了各个模型在 3 个数据集的 AUC 结果。

表 2 AUC 性能表

模型	Assist09	Assist12	EdNet
BKT	0.637 5	0.610 4	0.607 3
DKT	0.756 1	0.728 6	0.682 2
DKT-Q	0.732 8	0.762 1	0.728 5
DKVMN	0.755 9	0.724 7	0.692 1
SAKT	0.757 7	0.726 2	0.689 8
ours	0.776 3	0.763 0	0.738 4

由表 2 中可以看出，BKT、DKT、DKT-Q、SAKT、DKVMN 和提出的模型在 3 个数据集上的性能对比结果被总结在表 2 中。在所有方法中，模型都表现出了最佳的性能。这不仅验证了方法的有效性，还表明通过对问题与历史问题之间的相关性计算，能够有效缓解数据依赖性问题，从而提升知识追踪模型的预测能力。实验结果显示，模型在 3 个真实数据集上的预测准确度均高于大部分现有的 KT 模型，进一步证明了该模型在模拟学习者学习过程问题上的优越性。

3 结语

在本文中，通过引入问题之间的上下文信息，旨在减少数据的稀疏性，并有效捕捉长序列中问题之间的长期依赖关系。这一方法有助于更准确地把握学生的学习状态，从而显著提升模型的预测能力。通过这种方式，不仅增强了模型对学生学习过程的理解，也提高了对未来表现的预测准确性，

为知识追踪问题提供了更有效的解决方案。今后的工作重点是进一步探索问题与知识点之间的联系，丰富信息表征，提供可解释更强、更有效的知识追踪模型。

参考文献：

[1] 郑庆华,董博,钱步月,等. 智慧教育研究现状与发展趋势[J]. 计算机研究与发展, 2019,56(1):209-224.

[2] 吴水秀,罗贤增,熊键,等. 知识追踪研究综述[J]. 计算机科学与探索, 2023,17(7):1506-1525.

[3] CORBETT A T, ANDERSON J R. Knowledge tracing: modeling the acquisition of procedural knowledge[J].User modeling and user-adapted interaction, 1994,12(4):253-278.

[4] PIECH C, SPENCER J, HUANG J, et al. Deep knowledge tracing[C]//Proceedings of the 28th International Conference on Neural Information Processing Systems.New York: ACM, 2015: 505-513.

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

[6] GAN W B, SUN Y, SUN Y. Knowledge structure enhanced graph representation learning model for attentive knowledge tracing[J].International journal of intelligent systems, 2022, 37(3): 2012-2045.

[7] WILLIAMS R J, ZIPSER D. A learning algorithm for continually running fully recurrent neural networks[J]. Neural computation, 1989,1(2):270-280.

[8] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]// Proceedings of the 31st International Conference on Neural Information Processing Systems. NewYork: ACM, 2017: 6000-6010.

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

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

[11] PANDEY S, KARYPIS G. A self-attentive model for knowledge tracing[DB/OL].(2019-07-16)[2024-05-12]. <https://doi.org/10.48550/arXiv.1907.06837>.

【作者简介】

何萌红（1998—），女，贵州铜仁人，硕士，研究方向：智慧教育。

（收稿日期：2024-09-13）