

基于细粒化矩阵的嵌入式知识追踪方法

王 瑶¹
WANG Yao

摘 要

知识追踪 (KT) 是根据学习者以前的学习记录预测其未来学习状况的技术。近年来, 基于深度学习的知识追踪 (DKT) 发展迅速, 旨在动态追踪学生学习状态进而为学生提供个性化的服务。然而, 目前的研究忽略了练习题和知识点之间的复杂联系, 为了解决这个问题, 提出了一种基于细粒化矩阵的嵌入式知识追踪方法模型 (FGM-DKT)。从数据集中提取出了学习实体以及其中隐含关系的嵌入式表达, 再将其引入到对学习者的知识建模过程中。在两个真实世界公开数据集上的实验结果表明, 提出的 FGM-DKT 可以更准确地评估学生学习状态和知识点掌握程度, 与现有知识追踪模型相比具有更高的预测能力。

关键词

深度学习; 知识追踪; 教育系统; 个性化学习; 特征提取

doi: 10.3969/j.issn.1672-9528.2024.02.004

0 引言

近年来, 受新冠疫情影响, 大多数时间学生只能在慕课网等在线教育平台学习。同时受各种硬件、软件技术以及 AI 智能迅猛发展的影响, 在线教育平台也在根据用户群体的需求不断完善自己的系统。如今在线教育平台最关键的问题是如何针对用户个体做出个性化的人物影像, 通过平台收集到的数据全面评估用户的状态, 从而提供有针对性的服务。

为了得到完整、精确的学习者信息形象, 在此方面具有举足轻重地位的知识追踪 (knowledge tracing, KT) 受到了巨大的关注。它根据学习者之前的学习记录实时追踪学习者的知识掌握状态并以此为根据预测其下一次测试的分数, 根据长期的学习记录评估学生知识状态而不是最近几次的记录^[1]。

一个有效的知识追踪模型可以帮助学生了解自己的薄弱点和长处, 从而合理地安排自己的学习时间, 提高学习效率^[2], 将时间花费在自己薄弱的知识点学习中。知识追踪模型的准确度和全面度关系着在线学习系统能否精确掌握学生学习状态, 从而指定个性化的学习内容, 在学生需要的时候能够及时给出有效的学习安排 (练习和讲座等)。因此 KT 模型的优劣是在线学习系统能否服务好学生老师的关键^[3]。

假设在一个智能教学系统中有 D 名学生和 M 个习题, 在学生 i 完成一系列学习后, 系统会从学生学习过程产生的信息中生成这名学生的学习记录 $r_i = (l_i, x_1, x_2, \dots, x_n)$, 其中包含用户 ID $l_i \in \{l_1, l_2, \dots, l_D\}$ 、学生每次学习的记录 $x_j = \{q_j, s_j, c_j\}$, q_j 代表第 j 题, s_j 为问题 q_j 涉及的技能, c_j 是一个二元组 $\in \{0, 1\}$, 本次学习做对则为 1, 否则为 0。

从知识追踪诞生到现在, 已经涌现了许多优秀的 KT 模型^[4], 主要分为传统知识追踪模型和基于深度学习的知识追踪模型^[5]。

在传统知识追踪模型中, BKT^[6] 是最早出现的一种实时反馈的用户交互模型。BKT 模型通过使用四个参数将学生学习知识点的过程建模为隐马尔可夫过程, 如图 1 所示。

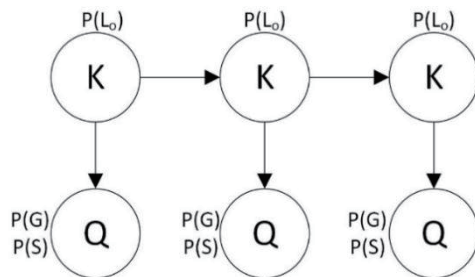


图 1 BKT 模型

用户在完成练习后得到即时反馈, 从而知识掌握水平发生变化。然而, BKT 侧重于挖掘知识点的信息, 不包括个别学生的信息和练习差异, 且 BKT 的知识掌握状态是二元状态, 学生对知识点的掌握状态会从掌握变为不掌握, 这是不现实的假设。针对这些问题, 研究者们提出了多种方法, 包括在

1. 贵州民族大学 贵州贵阳 550000

[基金项目] 贵州省科技厅基础研究项目 (黔科合基础-ZK (2022) 一般 197), 基于深度学习的知识追踪技术的可解释性和个性化应用研究, 以及贵州省教育厅 (黔教技 (2022) 047 号), 贵州省高等学校智慧教育工程研究中心

不同技能之间建立层次关系^[7]、个性化初始知识参数^[8]将学生差异参数引入 BKT^[9]、捕捉问题相似度对绩效的影响^[10]。尽管研究者们尝试了各种提高 BKT 性能的方法，但它仍然受到过于简单的限制。

近年来更为主流的基于深度学习的深度知识跟踪（deep knowledge tracing, DKT）通过递归神经网络对学生的进行学习进行建模。它将递归神经网络（recurrent neural network, RNN）应用到知识追踪任务中。通过模型中的神经元以及参数来表示学习者的知识状态，从数据中学习学生、知识的潜在向量表示，从而可以动态追踪学生的知识状态变化。

本文模型在目前研究的基础上组建练习题与知识点之间的关系矩阵，并根据当前练习题与其包含的不同知识点的关联程度细粒化关系矩阵，再通过矩阵得到不同实体及其关系的隐含嵌入信息，挖掘和整合练习题和知识之间更丰富的关系，以更全面地考虑学习者学习过程中的各种相关因素。

1 模型描述

1.1 LSTM

由于学生在教育环境中与基于计算机的学习系统进行交互，每个数据集中的交互记录都包含了学生学习时的行为动作，当前学习内容的相关信息以及学生与练习题的交互结果，因此知识追踪任务需要处理长期且详细的学习信息。所以选择利用长短期记忆^[11]（long short-term memory, LSTM）基于当前输入和先前状态对学习者的信息序列进行建模。LSTM 是一种 RNN 的变体，专门设计用于处理序列数据和捕捉长期依赖关系。与传统的 RNN 相比，LSTM 通过堆叠多个 LSTM 单元来构建深层网络结构。每个 LSTM 单元都具有相同的结构和参数，但它们具有不同的隐藏状态和记忆单元。通过堆叠多个 LSTM 单元，网络可以更好地捕捉到序列数据中的复杂关系和模式。

如图 2 所示的 LSTM 可以处理知识追踪任务中存在的长期依赖问题，高效地传递和表达学习者在学习过程中产生的长序列信息，并且可以捕捉历史信息、未来信息对当前学生学习状态的影响。具体来说，将得到的隐含信息向量送入 LSTM 模型中，以跟踪学习者随时间变化的知识状态。在图 2 中， x_t 是在 t 时刻的输入信息， h_t 是在 t 时刻的输入信息，可以看到 LSTM 中的神经元会递归的调用自身并且将 $t-1$ 时刻的信息传递给 t 时刻。详细过程如下。

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i) \quad (1)$$

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f) \quad (2)$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_o) \quad (3)$$

$$g_t = \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \quad (4)$$

$$c_t = f_t \otimes c_{t-1} + i_t \otimes g_t \quad (5)$$

$$h_t = o_t \otimes \tanh(c_t) \quad (6)$$

式中： i_t 、 f_t 、 o_t 分别是输入门、记忆门和输出门， c_t 是单元存储向量， W 和 b 是网络参数。其中记忆门通过 $\tanh$ 和 sigmoid 函数来控制要保留多少之前累计的信息到最新的记忆细胞中。

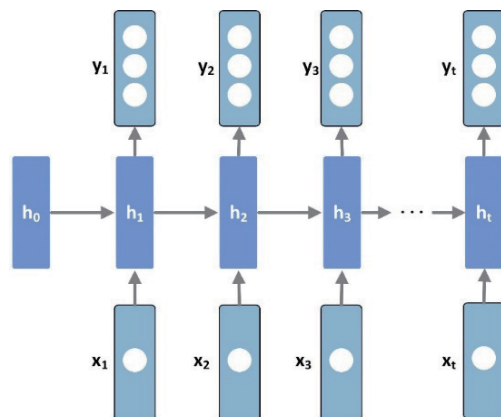


图 2 LSTM 模型图

1.2 嵌入问题难度信息

练习的难度信息在知识追踪任务中起着关键的作用，将题目的难度嵌入到模型中后，模型可以通过优化难度参数来更好地拟合学生的回答情况。这样，模型可以根据题目的难度调整对学生能力的估计，从而更好地理解学生在不同难度练习上的表现和进步情况。因此，希望最后的练习嵌入中能够包含难度信息。

练习的难度是指学生在解答该题目时所需的认知和技能水平。基于历史学生表现的数据（学生的正确率、作答时间等），通过分析学生在过去回答题目的结果，计算出每个题目的难度指标，并将题目的难度作为模型的参数之一加入练习嵌入中。

对于一个问题 q ，设置了问题属性特征 s' 和问题技能特征 j' 。问题属性特征包含题目的正确率和学生平均作答时间信息。假设与 q 相关的技能 s_j 共有 T 个，则问题特征被表示为：

$$s' = \frac{1}{|T|} \sum_{s_j \in T} s_j \quad (7)$$

$$Z = \text{ReLU}(s' + j' + b) \quad (8)$$

接着利用问题技能特征和问题属性特征生成问题 q 的线性信息 Z ，其中 b 为偏置参数。将难度信息转化为特征表示形式后，将其与练习嵌入融合到一起并输入到知识追踪模型中。

1.3 基于细粒化矩阵的嵌入式知识追踪模型

本文通过 KC（knowledge component）矩阵提取练习题和技能之间的隐含信息，并以嵌入的方式提取矩阵的潜在向量表示，将嵌入的信息作为额外的特征输入到知识追踪模型中，从而模型可以通过学习题目特征和技能特征之间的关联，

更准确地预测学生的能力水平。

KC 矩阵是一个二维矩阵，其中行表示学生，列表示不同的知识组件或知识点。每个元素表示学生对于相应知识组件的掌握情况，通常用 0 或 1 表示。例如，如果学生对某个知识组件掌握了，则对应的元素值为 1；如果学生尚未掌握该知识组件，则对应的元素值为 0。嵌入的是将高维数据映射到一个低维向量空间的技术，同时保持数据的特定结构或语义关系。通过这种映射，数据的相似性在向量空间中可以通过距离或夹角等度量进行度量。这种连续向量的表示可以包含更丰富的信息。本文提出的模型框架如图 3 所示。

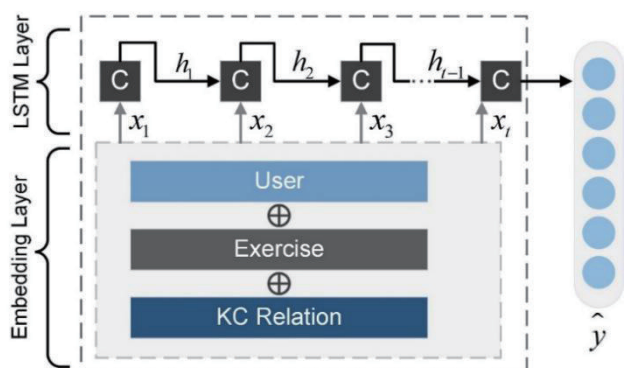


图 3 基于细粒化矩阵的嵌入式知识追踪模型

首先，根据给定的学生练习记录 $r_i = (I_i, x_1, x_2, \dots, x_n)$ 探索其中潜在的知识与练习的关联对学生知识掌握状态的影响，该影响权重用 γ_i 表示。具体而言，建立了练习与练习的关系表 $E \in \mathbb{R}^{n \times n}$ 、技能和技能的关系表 $R \in \mathbb{R}^{v \times v}$ 以及练习和技能的关系表 $T \in \mathbb{R}^{n \times v}$ 用于储存这些权重。练习与练习关系表中，对于时间 t ，如果练习 i 与练习 j 都与同一个知识概念 k 相关，则认为练习 i 与练习 j 也是相关的，即 $\gamma_t^{ij} = 1$ 否则 $\gamma_t^{ij} = 0$ ， $0 \leq i, 0 \leq j$ ，技能表和练习 - 技能表也是同样的设置。此外，在实际情况中知识和习题之间其实存在着关系紧密和只是稍微沾边的现象，现有模型这种只有“相关”“不相关”两种状态的关系缺乏更丰富细致的表达。因此，在本文的模型中，根据关系的紧密程度设置了新权重 ϵ_i 以量化关系表中的相关性权重。例如，练习 i 对技能 u 的相关权重被计算为练习 i 所包含的所有技能的平均值。

$$\epsilon_t^{iu} = \frac{1}{\sum_{u=1}^v \gamma_t^{iu}} \quad (9)$$

其次将三个关系表通过嵌入算法生成了关系嵌入矩阵 $K_i \in \mathbb{R}^{128 \times M}$ 。

然后将从知识嵌入矩阵中取出的知识嵌入向量 x_i 、学生嵌入 p_i 、和关系嵌入 r_i 合并为学习者的信息序列 v_i 。将三种特征编码为一个特征，作为新的学习者信息序列，它结合了实体信息和习题与知识之间隐含的信息，以理解在学习过程中不同的学习者对知识的学习能力，跟踪学生的知识状态。

学习者的信息序列获取为：

$$v_i = \text{Concat}(x_i, p_i, r_i) \quad (10)$$

学习者信息序列经过 LSTM 层后输出特征还会进入一个残差连接层以及一个全连接层，以更有效地进行梯度更新。

最后，通过交叉熵损失函数对模型进行了优化，具体地，通过 Adam 优化最小化在每个交互作用下真实答案 c_i 和预测结果 $\tilde{c}_i$ 之间的以下目标函数。

$$\mathcal{L} = -\sum_i (c_i \log \tilde{c}_i + (1 - c_i) \log (1 - \tilde{c}_i)) \quad (11)$$

与以往模型相比，本模型不仅考虑了学习者和知识点的信息，而且考虑了练习题和知识点之间的关系。将特征拼接以形成复合特征，然后通过残差层和全连接层获得目标向量以指示学生的知识掌握程度。最后，输出学生正确回答问题的概率 $\tilde{P}_i$ 。

2 实验与结果分析

为了评估本文的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 模型在两个公开数据集上进行了比较，以评估它们的优劣。

2.1 数据集

实验数据集包括 ASSISITments09 和 ASSISITments12。ASSISIT09 数据集是由 ASSISTment 在线教育系统在 2009—2010 年收集的真实的 learners 之间的交互记录。在该数据集中，大部分每道练习题只由一种技能构成，另外小部分可能与两种或者三种技能相关。ASSISIT12 数据集通过 2012 年全年收集更新了 Assist09。这是一个每道问题只包含一项技能的数据集。在这些数据集中，每次学生与练习题交互为一行数据（即，如果学生 S 回答了具有两个技能的问题 P，则这两个技能将被折叠为格式 skill1_skill2 并以单行表示）。为增加数据的真实性和可靠度，筛选出了交互条目大于 4 的学习者，本文删除了技能为空的数据。

2.2 评价指标

为了对模型性能优劣进行定量，选择了现在研究中广泛使用的预测准确度（ACC）和 receiver operating characteristic（ROC）曲线下面积（AUC）来评估预测学习者下一步反应的准确度。两个指标的值越大则表示模型的性能越好。

2.3 预处理

在数据划分步骤，以 0.8 为分界点分成了测试集和验证集，即提取 80% 的学习者数据用于模型训练，模型在训练集上训练后，再用剩下 20% 的数据作为测试集，用其误差近似模型在现实场景中的泛化误差。因此，训练集和测试集都是来自同一分布的数据。

将每个学生的交互序列设定为 200 的长度，即如果学习

者的交互次数大于 200 则只保留前 200 次记录, 不足的则用空符号填充以将长度拓展为 200。问题和技能表征的嵌入维度被设定为 128。

本文模型在 200 个 epochs 上进行了实验, 学习率设置为 0.001, LSTM 的隐藏层大小和批量都被取为 128。除了标量参数以外, 网络中的技能和问题嵌入矩阵和其他参数矩阵都使用正态分布随机初始化。

BKT、DKT、SAKT、DKVMN 和提出的 FGM-DKT 模型在两个数据集上的性能对比被总结在图 4 和图 5 中。在所有方法中, 提出的 FGM-DKT 都表现出了最好的性能。这说明通过在输入端加入更多有意义的信息(比如对学生学习过程有潜在影响的技能和练习之间的关系)可以有效地提升知识追踪模型的性能。在两个真实数据集上的实验结果表明, FGM-DKT 的预测准确度高于大部分现有的 KT 模型, 进一步验证了该模型在模拟学习者学习过程问题上的优越性。

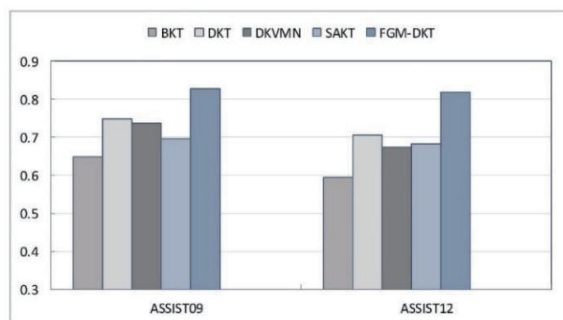


图 4 AUC 性能比较图

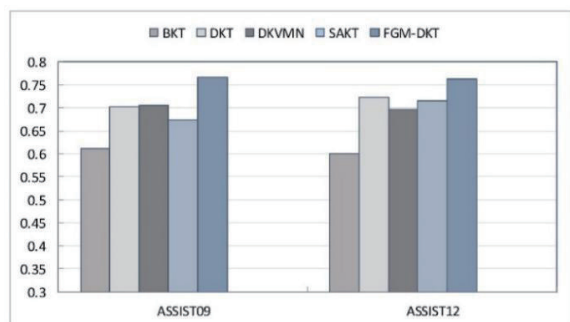


图 5 ACC 性能比较图

3 总结

本文从建模练习题与其中知识点之间关系紧密程度的角度出发, 提出了一种基于细粒化矩阵的嵌入式知识追踪方法 (FGM-DKT)。该模型对以往知识追踪模型做补充, 使得在线学习系统可以更准确地判断学习者的知识掌握状态和学习状态, 从而更精准地为每位学生提供个性化的教学安排, 也便于老师掌握学生的学习情况。今后的工作重点是要进一步探究学习过程中存在的各种因素, 整合学习相关因素, 丰富信息表征, 构建可解释性更强、更有效的 KT 模型。

参考文献:

- [1] PIECH C, BASSEN J, HUANG J, et al. Deep knowledge tracing[EB/OL]. (2015-06-19)[2023-08-08].<https://arxiv.org/abs/1506.05908>.
- [2] CHENG S, LIU Q, CHEN E, 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] WANG W, MA H, ZHAO Y, et al. Tracking knowledge proficiency of students with calibrated Q-matrix[J]. Expert systems with applications, 2022, 192(4): 116454.1-116454.11.
- [4] HE L, TANG J, LI X, et al. Multi-type factors representation learning for deep learning-based knowledge tracing[J]. World Wide Web, 2022, 25(3): 1343-1372.
- [5] ABDELRAHMAN G, WANG Q, NUNES B. Knowledge tracing: a survey[J]. ACM computing surveys, 2023, 55(11): 1-37.
- [6] CORBETT A T, ANDERSON J R. Knowledge tracing: modeling the acquisition of procedural knowledge[J]. User modeling and user-adapted interaction, 1994, 4: 253-278.
- [7] KÄSER T, KLINGLER S, SCHWING A G, et al. Beyond knowledge tracing: modeling skill topologies with bayesian networks[C]//Intelligent Tutoring Systems: 12th International Conference. Berlin: Springer, 2014: 188-198.
- [8] PARDOS Z A, HEFFERNAN N T. Modeling individualization in a bayesian networks implementation of knowledge tracing[C]//User Modeling, Adaptation, and Personalization: 18th International Conference. Berlin: Springer, 2010: 255-266.
- [9] YUDELSON M V, KOEDINGER K R, GORDON G J. Individualized bayesian knowledge tracing models[C]//Artificial Intelligence in Education: 16th International Conference, AIED 2013. Berlin: Springer, 2013: 171-180.
- [10] HAWKINS W J, HEFFERNAN N T. Using similarity to the previous problem to improve bayesian knowledge tracing[EB/OL]. (2014-06-12)[2023-08-15].https://ceur-ws.org/Vol-1183/bkt20y_paper04.pdf.
- [11] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural computation, 1997, 9(8): 1735-1780.

【作者简介】

王瑶 (1999—), 女, 四川宜宾人, 硕士, 研究方向: 统计学、智慧教育。

(收稿日期: 2023-11-10)