

杨长斌. 用于加速预训练知识追踪模型的深层特征提取器[J]. 智能计算机与应用, 2024, 14(4): 173-176. DOI: 10.20169/j.issn.2095-2163.240427

用于加速预训练知识追踪模型的深层特征提取器

杨长斌

(贵州民族大学 数据科学与信息工程学院, 贵阳 550025)

摘要: 通过对知识追踪进行预训练的嵌入与微调, 可以显著提高知识追踪任务的准确性, 但现有的微调方法通常采用全微调的方式, 即调整预训练模型的所有参数, 效率较低。本文提出了一种新的高效微调方法, 称为快速适配训练方法, 通过提取深层特征来调整模型, 以适应不同的下游任务。与全微调方法相比, 快速适配训练方法在 ASSIST 09 和 EdNet 数据集上能够实现 5.2 倍和 3.6 倍的训练时间加速, 并且同时提高了 0.49% 和 0.19% 的预测精度。

关键词: 知识追踪; 预训练模型; 微调

中图分类号: TP182

文献标志码: A

文章编号: 2095-2163(2024)04-0173-04

An accelerated deep feature extractor for pre-trained knowledge tracing models

YANG Changxiao

(School of Data Science and Information Engineering, Guizhou Minzu University, Guiyang 550025, China)

Abstract: The accuracy of the knowledge tracking task can be significantly improved by embedding and fine-tuning the pre-training for knowledge tracking, but the existing fine-tuning methods usually use full fine-tuning, i.e., adjusting all parameters of the pre-trained model, which is less efficient. In this paper, we propose a new efficient fine-tuning method, called the fast adaptation training method, to adjust the model to different downstream tasks by extracting deep features. Compared with the full fine-tuning method, the fast adaptation training method is able to achieve 5.2 times and 3.6 times training time speedups on the ASSIST 09 and EdNet datasets, and improves the prediction accuracy by both 0.49% and 0.19%.

Key words: knowledge tracing; pre-trained model; fine-tuning

0 引言

知识追踪是在线智能教育的核心, 旨在根据学生的历史互动记录捕捉学生的动态知识状态, 并预测其对新问题的回答^[1]。传统的知识追踪方法主要依赖于概率模型, 例如 Corbett 等^[2]提出的贝叶斯知识追踪(Bayesian Knowledge Tracing, BKT)模型, 通过建模学生的答题行为和学习历史, 来预测学生未来的学习表现和需求。随着深度学习技术的发展, 越来越多的研究者开始探索将深度学习引入知识追踪领域, 以应对传统方法在复杂任务上的局限性。

2015年深度知识追踪(Deep Knowledge Tracing, DKT)模型首次引入了循环神经网络(Recurrent Neural Network, RNN)及长短期记忆网络(Long Short-Term Memory, LSTM)来对学生学习知识的状态过程进行建

模; Pandey 等^[3]将自注意力机制引入到了知识追踪任务中。此后, Nakagawa 等^[4]将图神经网络(Graph Neural Networks, GNN)模型成功应用到了知识追踪任务中, 预训练模型也相继被引入到了知识追踪; 与此同时, Liu 等^[5]引入题目文本因素改善模型; Luo 等^[6]引入题目难度因素改善模型; Lalwani 等^[7]引入学习者遗忘因素改善模型。在以往的知识追踪研究中, 研究人员只关注准确度的提升, 而忽略了模型的时间, 阻碍知识追踪模型实际应用。准确度与训练时间的权衡也是知识追踪应该关注的问题, 可以通过增大模型规模、突破现有模型结构的精度局限获得较高的准确度, 而代价是大模型带来的算法性能调优时间从小时级上升到天级或更大级别的训练调优时间。

当前的预训练模型依赖于特定任务的下游数据集进行微调, 需要为每个任务训练一组专用的模型

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); 贵州省教育厅自然科学研究项目(黔教技[2022]047号)。

作者简介: 杨长斌(1999-), 男, 硕士研究生, 主要研究方向: 深度知识追踪。Email: 975926090@qq.com

收稿日期: 2023-12-01

参数,从而增加了训练时间。为了解决这个问题,本文提出了一种高效的知识追踪模型预训练模型方法,称为快速适配训练方法(Quick Adaptation Training, QAT)。快速适配训练方法(Quick Adaptation Training, QAT)。快速适配训练方法通过一种单向展开的注意力对参数的影响进行贡献度计算,并获得稀疏性嵌入,放大了结构化稀疏矩阵参数之间的距离,能够快速追踪对训练影响最大的参数,获得训练时间的提升。

1 相关工作

在引入深度学习模型之前,贝叶斯知识追踪^[2]是传统机器学习方法中具有代表性的一种知识追踪模型。Corbett等^[2]于1995年提出的贝叶斯知识追踪建模过程是利用隐马尔科夫模型(Hidden Markov Model, HMM)对学生知识点下题目的作答情况进行建模。BKT是对知识点建模,缺乏更丰富细致与个性化的表达,Käser等^[8]提出的动态贝叶斯知识跟踪(Dynamic Bayesian Knowledge Tracking, DBKT)模型能够在—个模型中联合表示多个不同的知识点,从而增强了知识追踪的表示能力;此外,Hawkins等^[9]基于贝叶斯知识追踪模型加入习题相关性进行优化;Qiu等^[10]加入遗忘参数模拟学生学习中的遗忘过程;Agarwal^[11]将知识掌握状态分级,模拟学生学习状态。深度学习模型在知识追踪领域最早的应用是Piech等^[12]于2015年提出的深度知识追踪模型,确定了深度知识追踪模型的基本模型架构。随着深度学习的兴起,出现了许多解决知识追踪问题的深度学习模型,如Choi等^[13]基于编码器-解码器(Encoder-Decoder, ED)提出的知识追踪模型;Shin等^[14]通过添加时间特征来模拟学生学习过程的知识追踪模型;Song等^[15]提出基于双图对比学习来优化知识点之间关系的知识追踪模型。

大规模预训练模型在深度学习各个领域取得成功,深度知识追踪也走上了预训练时代。Liu等^[16]通过丰富问题表示并利用多因素对学习者的学习进度进行建模;Wang等^[17-18]着重于分解交互中的基础关系信息与交互信息的表示能力,以提高模型精度;Jiang等^[19]基于问题-技能序列关系建模知识点之间的相关性;Sun等^[20]基于学生-问题交互信息捕捉问题之间的相似性。这些模型使得深度知识追踪的性能大幅上升,但缺乏参数高效迁移的方法,造成无法投入实际部署,本文将根据知识追踪特点设计适用于教育数据的高效参数微调方法。

2 方法

本文通过计算编码器当前输出(t)之前的($\overleftarrow{h}_{t-\delta}, \dots, \overleftarrow{h}_{t-1}$)步的注意力,并得到长度为 δ 的输出向量,即注意力窗口长度。具体的计算过程如式(1)所示:

$$a_i^t = \mathbf{v}^T \tanh(\mathbf{W}_\beta [h_t, c_i] + \mathbf{W}_\gamma o_i) \quad (1)$$

其中,矩阵 $\mathbf{W}_\beta, \mathbf{W}_\gamma$ 和向量 \mathbf{v} 都是可训练参数, h_t 表示前一个时间步 t 的隐藏状态输出; c_i 表示当前时间步的细胞状态; o_i 是编码器中第 i^{th} 步的输出。

当前时间步 t 的注意力函数输出,记为 \tilde{h}_t ,通过对前 δ 个隐藏状态及其对应的注意力分数进行加权求和得到,公式(2):

$$\tilde{h}_t = \sum_{i=t-\delta}^{t-1} s_i^t o_i \quad (2)$$

其中, s_i 表示注意力分数,通过对相关系数进行Softmax操作得到。

将前一个时间步的 \tilde{h} 作为输入的一部分传递给后续步骤, \tilde{h} 的信息传递会持续地递归进行,直到训练结束。当前时间步的 \tilde{h}_t 限制了下一个 \tilde{h}_{t+1} 的计算。

通过对预训练参数进行循环计算注意力分数,选中重要的预训练参数,并且这些关于参数重要性的信息将传递给未来的步骤。具体来说, o_t 是要传递给编码器的下一个步骤的输出,公式(3):

$$o_t = \mathbf{W}_\sigma [h_t, \tilde{h}_t] \quad (3)$$

其中, \mathbf{W}_σ 是一个权重矩阵, σ 表示一个激活函数,完整的输入包括将 \tilde{h}_t 与输入 h_t 进行拼接。

可以选择不同的 δ 值来生成不同稀疏程度的稀疏矩阵,较大的 δ 值能够一次选择更多的参数来计算注意力分数。然而, δ 的选择并不影响参数在编码器中经历的迭代次数,因为迭代次数与仅使用编码器结构时的迭代次数保持一致。本文的快速适配训练方法结构示意图如图1所示,其中绿色线表示当前时间步长 t 处的单元信息,蓝线是当前时间步长和所有先前 δ 时间步长($\overleftarrow{h}_{t-\delta}, \dots, \overleftarrow{h}_{t-1}$),紫色线表示当前时间步 t 的输出和单元信息的连接,红线是注意力加权后的前一个时间步的输出,其中虚线表示较小的权重,实线表示较大的权重,对当前时间步产生较大的影响,黄色的块表示可学习的参数矩阵, o_t 是要传递到编码器下一步的输出。

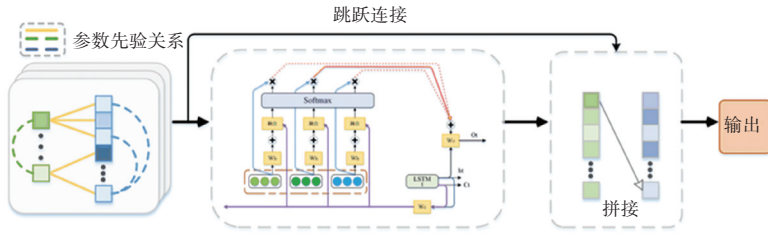


图 1 快速适配训练方法结构示意图

Fig. 1 Schematic diagram of the quick adaptation training method structure

3 实验

3.1 实验设置

本文将在 ASSISTments 2009、EdNet 两个基准数据集上进行实验。

ASSIST09 来自 ASSISTMENTS 在线辅导系统, 去掉重复记录之后, 包含 4 151 个学生在 110 个问题上的 325 673 个交互信息。

EdNet 是由 Santa(一个人工智能导学系统) 收集的大规模分层的学生活动数据集, 包含 784 309 名学生 131 317 236 个交互信息, 是迄今为止发布的最大的公共交互教育系统数据集。

3.2 实验细节

为了评估全微调方法和本文提出的方法在每个数据集上的性能, 采用 ROC 曲线下面积(AUC) 作为评估指标。在超参数的设置中, 问题嵌入的维度为 $d = 128$ 。使用相同的优化器 Adam, 在批量大小为 $b = 128$ 和学习率为 $l = 0.001$ 的条件下, 将本文的方法与全微调方法进行了比较实验。实验通过逐步增加局部上下文注意力长度 ($\delta = 4, 8, 16$ 和 64), 并相应地训练结构化稀疏嵌入来评估本文所提出方法的有效性。将每个数据集分为 80% 的训练和验证和 20% 的测试集, 使用基于二分图的预训练嵌入 (Improving knowledge tracing via pre-training question embeddings, PEBG) 模型作为研究参数有效的迁移学习的基线, 比较了训练和调整速度。

3.3 实验结果

模型预测精度对比见表 1, PEBG 模型优于先前的知识追踪模型。在 ASSIST09 数据集上, 精度达到了 82.78%, 相比于经典模型 DKT 和 DKVMN 平均提高了 9.03%; 在 EdNet 数据集上, PEBG 模型的精度平均提高了 8.19%。

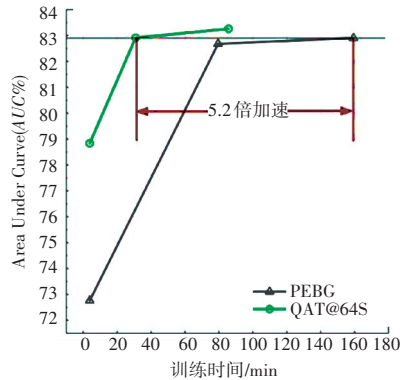
本文提出的 QAT 方法进一步改善了 PEBG 模型的性能。当将 QAT 方法应用于 PEBG 模型时, 在 ASSIST09 和 EdNet 数据集上分别达到了 83.27% 和 77.39% 的精度, 相对于原模型最佳精度提高了

0.49% 和 0.19%。

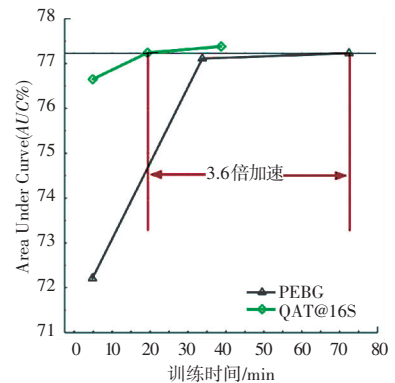
表 1 模型预测精度对比

模型	ASSIST09	EdNet
BKT	0.647 6	0.562 1
DKT	0.735 6	0.690 9
DKVMN	0.739 4	0.689 3
PEBG	0.827 8	0.772 0
pre-training+QAT	0.832 7	0.773 9

实验设置的 $\delta = 4, 8, 16$ 和 64 四组实验, 在 ASSIST 09 数据集与 EdNet 数据集上精度与训练时间的对比折线图如图 2 所示, 本文提出的方法加快了 PEBG 模型的训练速度并提高了准确率。



(a) ASSIST 09 数据集



(b) EdNet 数据集

图 2 精度与训练时间的对比折线图

Fig. 2 Line chart comparing accuracy and training time

快速适配训练方法(QAT)使PEBG模型获得更好的性能,在达到与PEBG相同精度的同时,在ASSIST 09数据集上获得大于5倍的训练时间提升,在EdNet数据集上获得大于3倍的训练时间提升;在ASSIST 09数据集上获得0.36%的AUC提升,速度快1.86倍,在EdNet数据集上获得0.2%的AUC提升,速度快1.17倍。

快速适配训练方法的加速比具有很大的实用价值。知识追踪预训练模型需要进行小时级的训练,因此3-5倍的加速可以使小时级的训练消耗降低到分钟级。这种加速有助于算法模型在实际应用中的部署。

4 结束语

本文提出一种快速适配训练的方法来对知识追踪预训练模型进行高效微调。通过提取深层特征来调整面对不同下游任务时的权重分布;与全微调相比,使用本文所提出的快速适配训练方法在ASSIST 09和EdNet上获得5.2倍与3.6倍加速比的训练时间提升,同时获得了0.49%和0.19%的平均预测精度提升。

参考文献

- [1] 刘铁园,陈威,常亮,等. 基于深度学习的知识追踪研究进展[J]. 计算机研究与发展,2022,59(1):81-104.
- [2] 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.
- [3] PANDEY S, KARYPIS G. A self-attentive model for knowledge tracing[J]. arXiv preprint arXiv:1907.06837, 2019.
- [4] NAKAGAWA H, IWASAWA Y, MATSUO Y. Graph-based knowledge tracing: modeling student proficiency using graph neural network [C]//Proceedings of IEEE/WIC/ACM International Conference on Web Intelligence. 2019: 156-163.
- [5] LIU Q, HUANG Z, YIN Y, et al. EKT: Exercise-aware knowledge tracing for student performance prediction[J]. IEEE Transactions on Knowledge and Data Engineering, 2021, 33(1): 100-115.
- [6] LUO R, LIU F, LIANG W, et al. Dagkt: Difficulty and attempts boosted graph-based knowledge tracing [C]//Proceedings of International Conference on Neural Information Processing. Cham: Springer, 2022: 255-266.
- [7] LALWANI A, AGRAWAL S. What does time tell? Tracing the forgetting curve using deep knowledge tracing [C]//Proceedings

- of Artificial Intelligence in Education: 20th International Conference. Chicago, USA: Springer, 2019: 158-162.
- [8] KÄSER T, KLINGLER S, SCHWING A G, et al. Dynamic bayesian networks for student modeling[J]. IEEE Transactions on Learning Technologies, 2017, 10(4): 450-462.
- [9] HAWKINS W J, HEFFERNAN N T. Using similarity to the previous problem to improve bayesian knowledge tracing [C]//Proceedings of Educational Data Mining and Learning Analytics. 2014:136-140.
- [10] QIU Y, QI Y, LU H, et al. Does time matter? Modeling the effect of time with bayesian knowledge tracing [C]//Proceedings of Educational Data Mining and Learning Analytics. 2011: 139-148.
- [11] AGARWAL D, BAKERR S, MURALEEDHARAN A. Dynamic knowledge tracing through data driven recency weights [C]//Proceedings of the 13th International Conference on Educational Data Mining (EDM 2020). 2020:725-729.
- [12] PIECH C, BASSEN J, HUANG J, et al. Deep knowledge tracing [J]. Computer Science, 2015, 3(3): 19-23.
- [13] CHOI Y, LEE Y, CHO J, et al. Towards an appropriate query, key, and value computation for knowledge tracing [C]//Proceedings of the Seventh ACM Conference on Learning@ Scale. 2020: 341-344.
- [14] SHIN D, SHIM Y, YU H, et al. Saint+: Integrating temporal features for ednet correctness prediction [C]//Proceedings of LAK21: 11th International Learning Analytics and Knowledge Conference. 2021: 490-496.
- [15] SONG X, LI J, TANG Y, et al. JKT: A joint graph convolutional network based deep knowledge tracing[J]. Information Sciences, 2021, 580: 510-523.
- [16] LIU Y, YANG Y, CHEN X, et al. Improveing knowledge tracing via pre-training question embeddings [C]//Proceedings of the Twenty-ninth International Conference on International Joint Conference on Artificial Intelligence. 2021:1577-1583.
- [17] WANG W, MA H, ZHAO Y, et al. SEEP: Semantic-enhanced question embeddings pre-training for improving knowledge tracing [J]. Information Sciences, 2022, 614: 153-169.
- [18] WANG W, MA H, ZHAO Y, et al. PERM: Pre-training question embeddings via relation map for improving knowledge tracing [C]//Proceedings of International Conference on Database Systems for Advanced Applications. Cham: Springer, 2022: 281-288.
- [19] JIANG C, GANB W, SUA G, et al. Improving knowledge tracing through embedding based on metapath [C]//Proceedings of the 29th International Conference on Computers in Education. 2021:11-20.
- [20] SUN J, ZHOU J, ZHANG K, et al. Collaborative embedding for knowledge tracing [C]//Proceedings of Knowledge Science, Engineering and Management: 14th International Conference. Cham: Springer, 2021: 333-342.