

融合遗忘机制的多模态知识追踪模型

闫秋艳, 孙浩, 司雨晴, 袁冠

引用本文

闫秋艳, 孙浩, 司雨晴, 袁冠. [融合遗忘机制的多模态知识追踪模型](#)[J]. 计算机科学, 2024, 51(7): 133-139.

YAN Qiuyan, SUN Hao, SI Yuqing, YUAN Guan. [Multimodality and Forgetting Mechanisms Model for Knowledge Tracing](#) [J]. Computer Science, 2024, 51(7): 133-139.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[一种基于属性相似性和分布结构连通性的聚类算法](#)

Clustering Algorithm Based on Attribute Similarity and Distributed Structure Connectivity

计算机科学, 2024, 51(7): 124-132. <https://doi.org/10.11896/jsjcx.231000125>

[基于多模态视频分类任务的模态融合策略研究](#)

Modality Fusion Strategy Research Based on Multimodal Video Classification Task

计算机科学, 2024, 51(6A): 230300212-5. <https://doi.org/10.11896/jsjcx.230300212>

[基于对比学习的视觉增强多模态命名实体识别](#)

Vision-enhanced Multimodal Named Entity Recognition Based on Contrastive Learning

计算机科学, 2024, 51(6): 198-205. <https://doi.org/10.11896/jsjcx.230400052>

[异质虹膜识别研究综述](#)

Review of Heterogeneous Iris Recognition

计算机科学, 2024, 51(6): 186-197. <https://doi.org/10.11896/jsjcx.231200175>

[基于跨模态信息过滤的视觉问答网络](#)

Cross-modal Information Filtering-based Networks for Visual Question Answering

计算机科学, 2024, 51(5): 85-91. <https://doi.org/10.11896/jsjcx.230300202>

融合遗忘机制的多模态知识追踪模型

闫秋艳 孙浩 司雨晴 袁冠

中国矿业大学计算机科学与技术学院 江苏 徐州 221116

(yanqy@cumt.edu.cn)

摘要 知识追踪是构建自适应教育系统的核心和关键,常被用以捕获学生的知识状态、预测学生的未来表现。以往的知识追踪模型仅根据结构信息对问题、技能进行建模,无法利用问题、技能的多模态信息构造其相互依赖关系。同时,关于学生的记忆水平仅以时间做量化,未考虑不同模态对记忆水平的影响。因此,提出了融合遗忘机制的多模态知识追踪模型。首先,对问题、技能节点,以图文匹配作为训练任务优化单模态嵌入,并通过计算多模态融合后节点间的相似度,获得问题和技能的关联权重从而计算生成问题节点的嵌入。其次,通过长短期记忆网络获取带有遗忘因素的学生知识状态,并将其融入学生的答题记录中生成学生节点的嵌入。最后,根据学生的答题次数和不同模态的有效记忆率计算学生和问题间的关联强度,通过图注意力网络进行信息传播,预测学生对不同问题的答题情况。在两个真实课堂自采数据集上进行了对比实验和消融实验,结果表明所提方法比其他基于图的知识追踪模型具有更好的预测精度,且针对多模态和遗忘机制的设计能有效提升原始模型的预测效果。同时,通过对一个具体案例的可视化分析,进一步说明了所提方法的实际应用效果。

关键词: 知识追踪;多模态;异质图;遗忘机制

中图分类号 G40-057

Multimodality and Forgetting Mechanisms Model for Knowledge Tracing

YAN Qiuyan, SUN Hao, SI Yuqing and YUAN Guan

School of Computer Science & Technology, China University of Mining and Technology, Xuzhou, Jiangsu 221116, China

Abstract Knowledge tracing is the core and key to build an adaptive education system, and it is often used to capture students' knowledge states and predict their future performance. Previous knowledge tracing models only model questions and skills based on structural information, unable to utilize the multimodal information of questions and skills to construct their interdependence. Additionally, the memory level of students is only quantified by time, without considering the influence of different modalities. Therefore, a multimodality and forgetting mechanisms model for knowledge tracing (MFKT) is proposed. Firstly, for question and skill nodes, a image-text matching task is used to optimize the unimodal embedding, and obtain the association weight calculation of questions and skills by calculating the similarity between nodes after multimodal fusion to generate the embedding of question nodes. Secondly, the student's knowledge state is obtained through the long short-term memory network, and forgetting factors are incorporated into their response records to generate student embeddings. Finally, the correlation strength between students and questions is calculated based on the student's response frequency and the effective memory rate of different modalities. Information propagation is performed using a graph attention network to predict the student's response to different questions. Comparative experiments and ablation experiments on two real classroom self-collected datasets show that our method has better prediction accuracy compared to other graph-based knowledge tracing models, and the design of multimodality and forgetting mechanisms effectively improves the prediction performance of the original model. At the same time, through the visual analysis of a specific case, further illustrate the practical application effect of this method.

Keywords Knowledge tracing, Multimodality, Heterogeneous graph, Forgetting mechanism

知识追踪 (Knowledge Tracing, KT) 是实现自适应教育及个性化教育的有效实施途径。知识追踪可以通过学生的历史答题情况, 发掘学生的隐藏知识状态, 并根据此状态预测

学生下一步的答题情况。知识追踪一直是教育数据挖掘领域的研究热点^[1]。对学生的知识状态进行表征和推理是知识追踪的核心问题。

到稿日期:2023-10-20 返修日期:2024-03-15

基金项目:国家自然科学基金(61977061,62277046)

This work was supported by the National Natural Science Foundation of China(61977061,62277046).

通信作者:孙浩(sunh@cumt.edu.cn)

当前大多知识追踪模型仍是以拓扑关系来建模技能、问题构成的知识空间,只有少数知识追踪模型^[2]注意到了问题的呈现形式,并通过实验研究证实了问题的内容信息对知识追踪任务是有益的。但其仅聚焦于问题文本的词嵌入,未利用问题、技能的多模态信息构建其相互关联,且未考虑时间因素和不同模态对学生记忆水平的影响。

针对上述问题,本文在多模态信息构建方面,针对问题节点,为捕获文本间关联,将文本分类(Text Classification, TC)作为训练任务,优化问题节点的文本嵌入;针对技能节点,为获取更好的单模态表示嵌入,以图像文本匹配(ITM)作为训练任务,实现对单模态表示嵌入的优化,将优化后的不同模态的嵌入进行融合,获得技能节点的最终嵌入表示。根据得到的问题、技能节点的嵌入表示计算节点间相似度,获得带有权重的问题关系和技能关系。

在学生记忆水平表达方面,受艾宾浩斯遗忘曲线的启发,本文为表达学生的遗忘问题,将学生答题的间隔时长嵌入学生的知识状态表示中,并使用长短期记忆网络^[3](LSTM)来学习学生随时间变化的知识状态。此外,由于学生对技能的记忆程度还受技能呈现的模态影响^[4-5],因此本文在信息聚合传递模块中融合问题模态对学生记忆的影响,以人脑对不同模态信息记忆的有效率定义问题-学生间的关联权重。

针对学生、问题和技能3类节点的不同结构及具有的不同模态属性,本文提出融合异构图网络、遗忘机制和多模态技术构建深度知识追踪模型(Multimodality and Forgetting mechanisms model for Knowledge Tracing, MFKT),主要贡献如下:

(1)在保留问题节点与技能节点结构信息的基础上融合节点自身的内容,设计基于卷积神经网络的图编码器和基于BERT的文本编码器分别提取节点的图像信息和文本信息,引入图像文本配对任务ITM^[6]优化节点的单模态嵌入表示,获得节点的最终融合向量,并计算节点间的相似度,重构问题和技能组成的知识空间。

(2)引入学生的答题时间间隔,将其嵌入学生的学习向量,模拟人脑的遗忘曲线,并通过学习学生的历史答题记录,获得学生的知识状态嵌入表示。结合人脑对不同模态信息记忆的有效率和学生的答题次数,为学生、问题节点赋予不同连接的权值,结合多头注意力机制通过图注意力网络GAT进行信息聚合传递,预测学生的未来答题表现。

(3)在两种不同类型的课堂自采数据集上,对所提模型的精度及模块的消融能力进行了对比实验,同时对模型在真实数据集上的应用效果进行了可视化展示。结果表明:将多模态及遗忘机制通过异构图网络进行融合,能够更加准确地建模学习者的认知规律,有效提升知识追踪的实际应用效果。

1 相关工作

1.1 知识追踪及多模态学习

知识追踪通过学生的历史答题情况,发掘学生的隐藏知识状态,根据此状态预测学生下一步的答题情况,实现进一步挖掘学生学习兴趣与认知能力的目的。知识追踪一直是教育

数据挖掘领域的研究热点。受图形神经网络等图形学习技术的代表性力量的启发,深度学习知识追踪模型开始搭载与图论相关的技术,有一类研究开始探讨知识空间中知识点的拓扑结构对知识追踪效果的影响。因此,Nakagawa等^[7]提出了图知识追踪的概念(Graph-based Knowledge Tracing, GKT)。他们引入知识图谱的相关概念,将知识结构转化为图形,从而间接地将知识追踪任务重构为图神经网络(GNN)模型中时间序列节点级分类任务,并对此猜想进行了可行性实验,验证图在知识追踪上的应用,开创了图论在知识追踪领域的研究先河。Yang等^[8]利用图卷积网络(GCN)来聚合问题嵌入和技能嵌入,提出的知识追踪模型(A Graph-based Interaction Model for Knowledge Tracing, GIKT)探讨了问题和技能间的相关性。基于会话序列的异构知识追踪模型(Session Graph-based Knowledge Tracing for student performance prediction, SGKT)^[9]使用关联图建模问题技能之间的关系,使用门控图神经网络从学生回答过程中获得学生知识状态。这代表从同构图到异构图的转化是基于图的知识追踪领域的一个主流发展方向。

2011年,Ngiam发表的《Multimodal Deep Learning》开启了多模态深度学习的大门,针对包含音频和视频的数据集进行处理,将波尔兹曼机作为模型基本单元,对视频、视频数据分别编码,通过学习重构获取两者的联合表示,实现对字母数字的识别^[10]。多模态研究重点可分为5类:多模态表征^[11]、跨模态映射^[12]、跨模态对齐^[13]、多模态融合^[14]和共同学习^[15]。而在知识追踪任务中,对不同模态数据的特征提取以及对不同模态间的元素进行对齐,是融合多模态知识追踪任务的工作重点。对不同模态的特征提取以及跨模态对齐工作有利于构造包含内容信息的复杂知识空间。

1.2 异构网络

国内外研究学者已在同构网络上进行了针对发掘复杂网络中的隐藏信息的多角度深入钻研。然而,同构网络中,对实体类别和实体间连接的表示定义往往是单一的,无法表示复杂的现实世界。在建模过程中往往会丢失很多关键信息。因此,异构网络依靠自身可表示不同类型节点和不同类型的边的特点,逐渐在多个领域崭露头角,如社交网络^[16]、生物分子网络^[17]、物理网络^[18]、引文网络^[19]、知识图谱^[20-21]等。异构网络的诞生催生了针对网络数据进行数据挖掘的研究,并且高质量的异构网络对机器学习任务也有辅助增强作用,可以有效提升机器学习任务的性能表现如节点分类^[17]、数据可视化^[22]、推荐系统^[23]等。现有的异构网络表示学习大致可分为基于网络结构的嵌入方法^[24]和基于异构图神经网络的嵌入方法^[25]。其中基于网络结构的嵌入方法利用网络中拓扑结构的相似性生成节点嵌入向量,根据获取相似性方式的不同,分为基于随机游走的方法和基于一阶/二阶相似度的方法;与同构网络不同,异构图神经网络中不同类型的节点包含的信息在意义、维度上有所差异。因此,异构图神经网络方法通常采用层次聚合的方式,即节点级别的聚合以及语义级别的聚合。

2 模型框架

2.1 整体介绍

融合遗忘机制的多模态知识追踪模型(MFKT)整体架构如图1所示,主要分为4个模块:问题节点嵌入模块、学生节点嵌入模块、学生-问题边模块、信息传递聚合和预测模块。

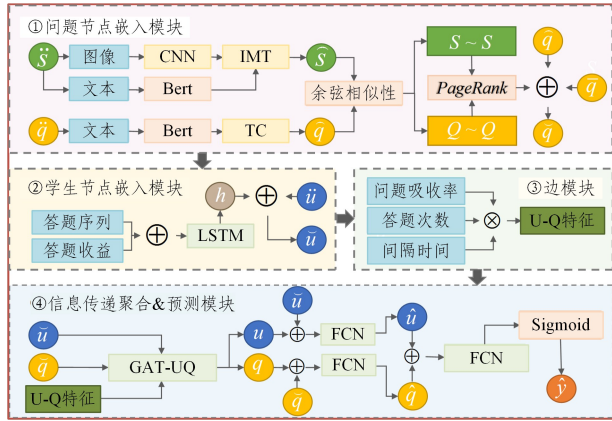


图1 MFKT模型框架图

Fig.1 Framework of MFKT

与现有知识追踪模型相比,MFKT可以捕获技能的多模态信息,获得包含更丰富语义信息和学生记忆水平的知识空间。除此之外,本文将融合技能多模态的问题嵌入并结合遗忘机制通过GAT对学生的知识状态进行细化处理,可以更好地跟进学生对不同技能、问题的掌握程度,为下游任务如知识状态评估、习题及课程资源推荐等奠定基础。

2.2 具体设计

2.2.1 问题节点嵌入表示

首先,选用Bert模型,对问题节点的文本长度采取“长切短补”的方式进行统一,结合TC的下游任务,使用交叉熵损失函数对问题节点的文本表示进行优化,根据Bert提取问题文本向量 \hat{q} ,并根据余弦相似度计算问题节点间权重,获得带有权重的问题关联图($Q \sim Q$)。

其次,选用CNN提取技能节点的图像特征,使用Bert提取技能节点的文本特征,引入文献[26]的图像文本对齐方法。对于每一个图像和文本,分别计算图像-文本和文本-图像的相似性,如式(1)和式(2)所示:

$$\gamma_c^{I \rightarrow T}(I) = \frac{\exp(s(I, T_c) / \zeta)}{\sum_{c=1}^C \exp(s(I, T_c) / \zeta)} \quad (1)$$

$$\gamma_c^{T \rightarrow I}(T) = \frac{\exp(s(T, I_c) / \zeta)}{\sum_{c=1}^C \exp(s(T, I_c) / \zeta)} \quad (2)$$

其中, ζ 是可训练的参数, $s(\cdot)$ 表示图像文本在低维空间的相似度函数, I 表示图像, T 表示文本, C 表示样本数量。

再次,对于得到的图像文本特征,本小节通过图像文本对比损失与图像文本配对任务的结合优化技能节点的单模态表示。通过拼接操作得到技能节点的多模态表示向量 \hat{S} ,并根据余弦相似度计算问题节点间权重,获得带有权重的技能关联图($S \sim S$)。

最后,本文基于PageRank算法^[27],定义问题节点的两个

属性:技能丰富度和问题区分度。计算式分别如式(3)和式(4)所示:

$$ss_j = \sum_{k \in Z_j} \frac{PR(s_k)}{\sum_{i=1}^{L(s_k)} \theta_{ki}} \quad (3)$$

式(3)表示问题节点的技能丰富度,其中 θ_{ki} 表示第 k 个技能节点和第 i 个技能节点的关联权重,即两个技能节点间的相似度。其中 Z_j 为问题 j 所包含的技能集合, $L(s_k)$ 表示技能节点 s_k 在技能-问题图的出度,即为技能节点 s_k 直接关联的问题节点数量,该公式主要用于计算和问题 j 相关的技能在技能关联图 $S \sim S$ 上的PR值的和。

$$qq_j = \sum_{k \in Q_j} \frac{\omega_{jk} \cdot PR(q_k)}{\sum_{i=1}^{L(q_k)} \omega_{ki}} \quad (4)$$

式(4)表示问题节点的问题区分度,其中 ω_{ki} 表示第 k 个问题节点和第 i 个问题节点间的相似度。该公式主要用于计算问题 j 在问题关联图 $Q \sim Q$ 中的加权PR值的和。

最后,综合考虑问题节点的丰富度和关联度,将问题的重要性定义为 b_j ,计算式见式(5)。其含义是,一个问题包含的技能丰富度越高、与其他问题的关联度越低,其越重要。

$$b_j = ss_j / qq_j \quad (5)$$

由于题目的重要性是客观存在的,并不影响题目和技能间的关系,因此对问题节点的嵌入定义如式(6)所示。

$$\check{q}_j = \hat{q}_j \oplus \bar{q}_j \oplus b_j \quad (6)$$

其中, \hat{q}_j 为根据Bert提取问题文本向量, \bar{q}_j 为问题的技能映射表示, \oplus 表示连接操作。

2.2.2 学生节点嵌入表示

在模型MFKT中,为学习答题顺序对学生的影响,将回答的问题、回答问题的正误,以及回答问题所获得的收益作为学生练习的特征,并结合LSTM网络学习学生的答题时序。

首先,加入学生答题积分,度量学生从问题中获取的收益。该积分由在线答题平台结合学生过去的成绩生成,即将学生回答一个问题的即时收益用正误结果与答题积分做乘积表示,计算过程如式(7)所示:

$$\eta_t = r_t \cdot g_t \quad (7)$$

其中, η_t 表示学生在 t 时刻的答题收益; $r_t \in \{0,1\}$ 代表学生在 t 时刻的答题结果,回答正确则为1,回答错误则为0; g_t 代表学生在 t 时刻所获得的答题积分。

其次,使用 e 来表示学习过程中根据学生回答问题的时刻划分的练习元组, e_t 即为学生在 t 时刻的练习元组,其计算式如式(8)所示:

$$e_t = \text{LeakyReLU}(W_1 [q_t \oplus \eta_t] + b_1) \quad (8)$$

其中, \oplus 表示连接操作, q_t 为学生在 t 时刻回答的问题嵌入, W_1 为权重矩阵, b_1 为偏置量。

为发掘练习间的潜在关联,保留学生答题顺序对知识状态的影响,将练习元组 e_t 作为输入数据投入LSTM中来学习学生的知识状态,得到隐藏状态 h_t 作为学生在 t 时刻的能力表示。因为其体现了答题序列和答题收益随时间的变化程度,所以在本文中针对学生节点的最终嵌入表示如式(9)所示:

$$\hat{u}_t = (\check{u}_t \cdot Q_s) \oplus h_t \quad (9)$$

其中, \ddot{u}_t 表示学生在 t 时刻的历史答题序列, Q_s 为问题与技能的邻接矩阵, \cdot 表示点乘。

2.2.3 学生-问题边表示

根据艾宾浩斯遗忘曲线的定义, 学生在某一时刻对一道题目的印象与距离上次答题的时间长度呈指数级负相关。此外, 麦拉宾法则指出, 学生对不同模态信息的有效记忆率有所不同, 人类会对图片信息更为敏感。因此对于学生回答一次问题对问题的记忆程度 β , 定义如式(10)所示:

$$\beta = \frac{100\delta}{(\log(l))^\ell + \delta} \cdot \rho_\varphi \quad (10)$$

其中, l 为当前时刻 t 与学生答题时刻的间隔时长, 常数 ℓ 和 δ 分别为 1.25 和 1.84。 ρ_φ 为问题吸收率, 表示人脑对该问题包含信息的接收程度, φ 表示问题的隐含模态, 根据梅拉比安沟通模型^[28] 结合当代的教育研究, 设置 $\rho \in \{0.07, 0.55, 0.62\}$, 即当 φ 为纯文本信息时, 问题吸收率为 0.07, 当 φ 为纯图像信息时, 问题吸收率为 0.55, 当 φ 为图文结合信息时, 问题吸收率为 0.62。

由于在答题过程中, 同一道问题可能重复出现, 因此学生对一个问题的总体记忆程度, 由每次答题的记忆程度累加表示。本文使用 sigmoid 函数规范化一个学生对一个问题的总体记忆程度, 并将其表示为学生和问题间的权重。

$$\epsilon = \sigma\left(\sum_{i=0}^n \beta_i\right) \quad (11)$$

其中, ϵ 表示单个学生和单个问题的关联权重, n 表示该学生回答该问题的答题次数, β_i 为学生第 i 次回答问题时对问题的记忆程度。

2.2.4 信息传递与聚合

由于学生-问题异构图中的问题节点和学生节点分别处于不同的特征空间中, 因此对每种类型为 Φ_i (问题、学生节点) 的节点, 首先引入节点类型转换矩阵 W_{Φ_i} , 将节点投影到同一向量空间中进行计算, 如式(12)所示:

$$\mathbf{h}_i = W_{\Phi_i} \cdot \mathbf{F}_i \quad (12)$$

其中, \mathbf{F}_i 和 \mathbf{h}_i 分别表示节点 i 的初始属性特征和投影后的属性特征 (\mathbf{F}_i 可表示学生节点嵌入 \check{u}_i 或问题节点嵌入 \check{q}_j)。

针对投影后的属性特征 \mathbf{h}_i , 计算学生节点 i 与其邻居问题节点 j 之间的注意力分数 e_{ij} 。 N_i 表示学生 i 回答的问题集合。同时对注意力分数进行规范化, 如式(13)、式(14)所示:

$$e_{ij} = \text{LeakyReLU}(\mathbf{h}_i \cdot \mathbf{h}_j^\top) \quad (13)$$

$$\alpha_{ij} = \text{softmax}(e_{ij}) = \frac{\exp(e_{ij})}{\sum_{t \in N_i} \exp(e_{it})} \quad (14)$$

将学生节点 i 的投影特征和注意力权重系数相乘, 计算出节点 i 的异质邻居聚合特征。为使模型学习到的嵌入更加稳定, 按照上述公式独立计算 K 次, 其中, k 表示第 k 个注意力头, 并将每次计算的结果拼接起来作为节点 i 最终的异质邻居聚合特征 \mathbf{H}_i , 如式(15)所示:

$$\mathbf{H}_i = \parallel_{k=1}^K \sigma\left(\sum_{j \in N_i} \alpha_{ij} \mathbf{h}_j^\top\right) \quad (15)$$

其中, $\sigma(\cdot)$ 表示 ELU 激活函数。

由于上述过程仅聚合了学生节点的问题邻居特征, 忽略

了学生节点自身的属性特征, 因此将问题邻居聚合特征 \mathbf{H}_i 与学生节点属性特征 \mathbf{F}_i 拼接, 并通过全连接层进行特征融合, 如式(16)所示:

$$\mathbf{Z}_i = \sigma(g(\mathbf{H}_i \oplus \mathbf{F}_i)) \quad (16)$$

其中, $g(\cdot)$ 表示全连接层, \oplus 表示拼接操作。 \mathbf{Z}_i 表示最终获得的问题向量嵌入 \hat{q}_j 和学生向量嵌入 \hat{u}_i 。

为预测学生 i 在下一时刻是否能正确回答问题 j , 将问题节点和学生节点的嵌入拼接, 计算出预测概率 \hat{y}_{ij} , 如式(17)所示:

$$\hat{y}_{ij} = \text{sigmoid}(f(\hat{u}_i \oplus \hat{q}_j)) \quad (17)$$

其中, $f(\cdot)$ 表示全连接层, $\text{sigmoid}(\cdot)$ 表示非线性激活函数。

本文采用交叉熵损失计算模型的预测值与真实标签间的损失。

$$\mathcal{L} = - \sum_{i,j \in y^+ \cup y^-} (y_{ij} \log \hat{y}_{ij} + (1 - y_{ij}) \log(1 - \hat{y}_{ij})) \quad (18)$$

其中, y_{ij} 表示问题 j 与学生 i 间的真实关联标签; y^+ 表示正样本中的节点集, 即回答正确的学生问题关联; y^- 表示负样本中的节点集, 即回答错误的学生问题关联。最后, 采用反向传播算法对整个模型进行训练。

3 实验

为了验证 MFKT 模型及其优化的 3 个模块的有效性, 在真实课堂环境下在线答题平台收集的数据上进行了对比实验和消融实验, 选择 AUC, ACC, recall, F1 值作为模型精度的评价指标。

3.1 实验说明

3.1.1 实验平台及参数设置

实验环境为: Windows11 操作系统, CPU 型号为 Intel Core i7-12700H, 核心数为 14 核, 内存为 24GB。显示芯片型号为 RTX3060, 显存类型为 GDDR6。本文在 pytorch 框架下实现 MFKT 模型及其变体模型 MFKT-Q, MFKT-U, MFKT-W。

3.1.2 数据集介绍

实验所需的多模态数据集由真实课堂下的在线答题平台收集。其中数据集 DPA_2023 为“数据库原理”课程 2023 学年第一学期学生的答题记录, 包括 91 名学生、68 个问题、21 个技能、3195 条学生答题记录。数据集 PP_2023 为“Python 程序设计”课程 2023 学年第一学期学生的答题记录, 包括 117 名学生、54 个问题、15 个技能、2775 条学生答题记录。两门课程分别为计算机专业基础课和全校公共基础课, 采用不同性质的课程验证本文所提 MFKT 模型的有效性和通用性。

3.2 对比实验结果

鉴于本文模型 MFKT 与 3 个对比实验均为分类预测任务, 选取了 3 个基于图结构并针对问题或学生节点进行探讨的知识追踪模型 (GIKT^[8], SGKT^[9], GKT^[7]) 进行对比。所有实验的 epoch 设为 300, 为避免偶然性, 所有结果均由五折

交叉验证所得,实验结果的具体数值如表 1 所列。

表 1 中最好的指标值加粗表示。结果显示,在两个数据集的 4 项指标上,MFKT 的表现均优于对比实验中的基线方法。相比基线方法,在数据集 DPA_2023 上,本文模型的 AUC 平均提高 6.15%,ACC 平均提高 16.83%,recall 平均提高 11.38%,F1 平均提高 11.63%;在数据集 PP_2023 上,AUC 平均提升 3.76%,ACC 平均提升 12.31%,recall 平均提升 4.61%,F1 平均提升 9.64%。其中,在 ACC,recall 和 F1 指标上,相比基线方法提升最为显著。由实验结果可知,加入多模态数据和遗忘机制会对知识状态的预测精度产生积极影响。

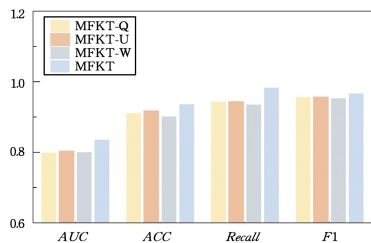
表 1 MFKT 模型对比实验

Table 1 MFKT model comparison experiment

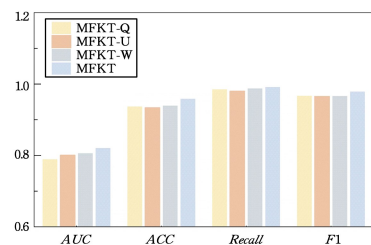
Dataset	Model	AUC	ACC	recall	F1
DPA_2023	GIKT	0.7772	0.7864	0.8978	0.8668
	SGKT	0.7941	0.7670	0.8512	0.8474
	GKT	0.7503	0.7500	0.8590	0.8369
	MFKT	0.8354	0.9361	0.9831	0.9667
PP_2023	GIKT	0.7586	0.8153	0.9433	0.8357
	SGKT	0.7680	0.8282	0.9267	0.8924
	GKT	0.8229	0.8631	0.9667	0.9190
	MFKT	0.8208	0.9586	0.9917	0.9788

3.3 消融实验结果

为验证 MFKT 模型中优化的 3 个模块的有效性,本节设计了 3 个消融实验,分别独立验证每个模块的优化性能,精度结果如表 2 所列。图 2 展示了消融实验在两个数据集下的效果图,其中 MFKT 为本文完整模型。MFKT-Q 表示问题节点嵌入模块中,未增加技能节点的多模态特征,仅保留文本特征;MFKT-U 表示学生节点嵌入模块中,未使用 LSTM 网络学习学生的能力表示,仅通过学生做题序列生成学生节点的嵌入表示;MFKT-W 则代表在保留问题、学生节点嵌入模块的基础上,以无权异构图来表示问题和学生之间的关系,消除梅拉比安沟通模型和艾宾浩斯遗忘曲线对模型的性能影响。结果表明,知识追踪的 3 个改进对模型性能均有提升作用。



(a)数据集 DPA_2023 下的实验结果



(b)数据集 PP_2023 下的实验结果

图 2 消融实验效果

Fig. 2 Ablation experiment effects

表 2 MFKT 模型消融实验

Table 2 MFKT model ablation experiment

Dataset	Model	AUC	ACC	recall	F1
DPA_2023	MFKT-Q	0.7985	0.9104	0.9435	0.9569
	MFKT-U	0.8029	0.9172	0.9432	0.9564
	MFKT-W	0.8001	0.9013	0.9351	0.9526
	MFKT	0.8354	0.9361	0.9831	0.9667
PP_2023	MFKT-Q	0.7895	0.9370	0.9850	0.9673
	MFKT-U	0.8004	0.9334	0.9800	0.9653
	MFKT-W	0.8062	0.9397	0.9875	0.9665
	MFKT	0.8208	0.9586	0.9917	0.9788

综合来看,MFKT-Q 与 MFKT 间的差距最大,可知问题节点模块对于本文模型 MFKT 而言最为必要。MFKT-U 在 DPA_2023 上与 MFKT 的差值整体略低于 MFKT-M,在 PP_2023 数据集上的差值略高于 MFKT-M。这证明在 DPA_2023 上,MFKT-W 对模型的提升效果高于 MFKT-U,这是由于 DPA_2023 数据集集中所涉及的技能以文本表示居多,且技能先后顺序不强,学生对其掌握水平更易受到时间和问题涉及模态的影响。而在 PP_2023 数据集上,MFKT-U 对模型的提升效果更胜一筹,这是由于 PP_2023 数据集所涉及的技能存在顺序学习的特点,即学生在掌握高阶技能前,必须学习与之相关的基础技能,学生对技能的掌握情况更易受学生答题顺序支配。

3.4 学生知识状态分析

本节从数据集 DPA_2023 中选取某一学生的一段答题序列,追踪其在答题过程中对技能掌握程度的变化,说明遗忘机制和多模态因素的引入对实际知识状态的影响效果。

图 3 是知识状态变化热力图,图中横坐标表示学生每次答题的结果,纵坐标表示技能编号,热力图中每一个单元格的数字代表学生的具体掌握程度。单元格颜色越深,说明学生对知识的掌握程度越好。其中,技能 A 表示数据库定义,技能 B 表示数据库管理系统定义,技能 C 表示数据库系统定义,技能 D 表示实体联系模型,技能 E 表示逻辑数据模型,技能 F 表示三级模式与数据独立性。其中,技能 A,B,C 均为定义类技能,其主要表现形式为文本;技能 D,E,F 均为模型类技能,其表现形式多为文本与结构化图像相结合的表示。学生的答题记录由 4 个连续时间段拼接而成,t-1-t-4,t-5-t-8,t-9-t-13,t-14-t-20 分别为一段连续时间内的答题记录。当学生答对一道题目时,对应技能的知识状态数值升高,证明学生对该技能的掌握程度有所提升;当学生答错一道题目时,对应技能的知识状态降低,修正模型对学生知识状态的预测结果。

从图 3 可知,由于技能 A,B,C 三者相关性高,因此学生答对某一个技能上的问题时,其他两个技能的状态都会随之提升。例如在 t-19,学生回答了和技能 C 相关的问题,技能 A 和技能 B 的知识状态也随之提升,说明 MFKT 模型可以模拟出知识间关联对学生知识状态的影响。

同时,t-5,t-9,t-14 均为新一组练习的起始位置,与上一时刻练习是不连续的,且中间相隔时间长,学生受到遗忘因素的影响,知识状态都应该有所降低。MFKT 模型添加了遗忘机制,可以捕捉到遗忘行为对知识状态产生的影响,较好地表达了这种下降趋势,证明了模型的改进与设计是有效的。

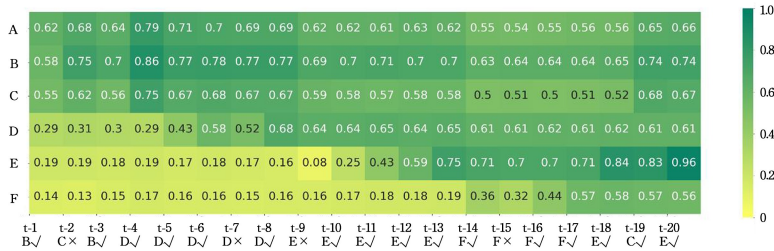


图3 MFKT模型知识状态热力图

Fig. 3 Knowledge state thermodynamic diagram of MFKT model

图4为图3的点线图表示,相比热力图,点线图可以更好地观测知识状态的变化趋势。技能A,B,C均为定义类技能,其数据的主要表现形式为文本,且相关性高,三者呈现近乎一致的变化趋势。技能D,E,F均为模型类技能,其数据的表现形式多为文本与结构化图像相结合的表现。

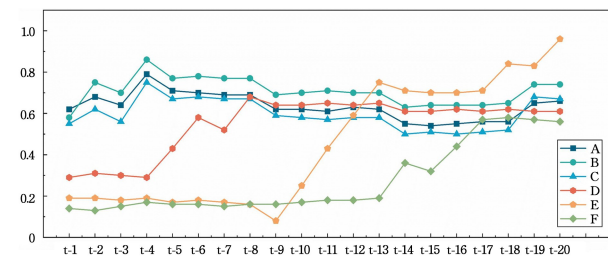


图4 学生知识状态点线图

Fig. 4 Dot line diagram of students' knowledge state

从图4的趋势中可看出,t-5,t-9,t-14为每组练习的起始位置。由于t-1-t-4处的学生仅学习了技能A,B,C,因此本节将观测t-8-t-9和t-13-t-14处的知识状态变化。在t-8-t-9处,学生已经学习了技能A,B,C,D,故此处将对技能D与技能A,B,C的知识状态变化趋势。可观测到,4个技能均受遗忘曲线作用呈下降趋势,但D的下降趋势相比技能A,B,C呈现出更为平缓的状态。在t-13-t-14处,学生完成了对技能E的学习,可观测到,在此处D和E的下降趋势近乎一致,均呈现出比A,B,C更为平缓的趋势。这符合梅拉比安沟通模型和艾宾浩斯遗忘曲线的研究结果:人脑对事物的印象55%来自于视觉记忆,学生在某一时刻对一道题目的印象与距离上次答题的时间长度呈指数级负相关。这证明了本文模型对学生知识状态的建模贴合现实场景,是合理有效的。

结束语 本文提出了一种融合遗忘机制的多模态知识追踪模型MFKT。在GAT网络的基础上,分别从问题节点嵌入、学生节点嵌入和学生-问题边表示3个方面进行改进。首先,对问题、技能节点,以图文匹配作为训练任务优化单模态嵌入,并通过计算多模态融合后节点间的相似度,获得问题和技能的关联权重,结合问题和技能的映射关系以及问题的重要性生成问题节点的嵌入。其次,通过长短期记忆网络获取带有遗忘因素的学生知识状态,并将其融入学生的答题记录中生成学生节点的嵌入。最后,根据学生的答题次数和不同模态的有效记忆率计算学生和问题间的关联强度,通过图注意力网络进行信息传播,预测学生对不同问题的答题情况。实验部分,在两个数据集上采用了4个分类任务的指标对所提模型的效果进行验证,与对比模型相比,MFKT均获得了

较优的表现。此外,针对多模态和遗忘机制的设计均对模型有提升效果。同时,对一个具体案例进行可视化分析,进一步说明了本文方法的实际应用效果。

参考文献

- [1] LI F M, YE Y W, LI X F, et al. Application of knowledge tracking model in Education: A Review of Relevant Studies from 2008 to 2017 [J]. Distance Education in China, 2019(7): 86-91.
- [2] SU Y, LIU Q, LIU Q, et al. Exercise-Enhanced Sequential Modeling for Student Performance Prediction [C] // National Conference on Artificial Intelligence, 2018.
- [3] HOCHREITER S, SCHMIDHUBER J. Long Short-Term Memory [J]. Neural Computation, 1997, 9(8): 1735-1780.
- [4] CHEN Z. An experimental study on the influence of information presentation style and students' cognitive style on science learning in multimedia environment [D]. Chongqing: Southwestern Normal University, 2004.
- [5] CHEN X L. Visual Computing—An Extension of Human Perception [J]. Measurement & Control Technology, 2000, 19(5): 7-14.
- [6] WANG L, LI Y, HUANG J, et al. Learning Two-Branch Neural Networks for Image-Text Matching Tasks [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 41(2): 394-407.
- [7] NAKAGAWA H, IWASAWA Y, MATSUO Y. Graph-based knowledge tracing: modeling student proficiency using graph neural network [C] // IEEE/WIC/ACM International Conference on Web Intelligence, 2019: 156-163.
- [8] YANG Y, SHEN J, QU Y, et al. GIKT: a graph-based interaction model for knowledge tracing [C] // Machine Learning and Knowledge Discovery in Databases: European Conference. Springer International Publishing, 2021: 299-315.
- [9] WU Z, HUANG L, HUANG Q, et al. SGKT: Session graph-based knowledge tracing for student performance prediction [J]. Expert Systems with Applications, 2022, 206: 117681.
- [10] NGIAM J, KHOSLA A, KIM M, et al. Multimodal deep learning [C] // Proceedings of the 28th International Conference on Machine Learning (ICML-11), 2011: 689-696.
- [11] KARPATHY A, JOULIN A, FEI-FEI L F. Deep fragment embeddings for bidirectional image sentence mapping [C] // Proceedings of the 27th International Conference on Neural Information Processing Systems, 2014: 1889-1897.
- [12] MALHOTRA P, RAMAKRISHNAN A, ANAND G, et al.

- LSTM-based encoder-decoder for multi-sensor anomaly detection[J]. arXiv:1607.00148,2016.
- [13] MITCHELL M,DODGE J,GOYAL A, et al. Midge: Generating image descriptions from computer vision detections[C]// Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics/2012:747-756.
- [14] TENNEY D,LIU L,VAN DEN HENGEL A. Graph-structured representations for visual question answering[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017.
- [15] BALTRUŠAITIS T,AHUJA C,MORENCY L P. Multimodal machine learning: A survey and taxonomy[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 41(2):423-443.
- [16] TYLEND A T,ANGELOVA R,BEDATHUR S. Towards time-aware link prediction in evolving social networks[C]// Proceedings of the 3rd Workshop on Social Network Mining and Analysis. 2009.
- [17] THEOCHARIDIS A,VAN DONGEN S,ENRIGHT A J, et al. Network visualization and analysis of gene expression data using BioLayout Express 3D[J]. Nature Protocols,2009,4(10):1535-1550.
- [18] BATTAGLIA P,PASCANU R,LAI M, et al. Interaction networks for learning about objects, relations and physics[C]// Proceedings of the 30th International Conference on Neural Information Processing Systems. 2016:4509-4517.
- [19] ATWOOD J,TOWSLEY D. Diffusion-convolutional neural networks[C]// Proceedings of the 30th International Conference on Neural Information Processing Systems. 2016:2001-2009.
- [20] SUN Z,DENG Z H,NIE J Y, et al. Rotate: Knowledge graph embedding by relational rotation in complex space[J]. arXiv: 1902.10197,2019.
- [21] LI J H,WANG C,LI M. Knowledge Tracking Algorithm Driven by Graph Ripple Feature[J]. Journal of Chinese Computer Systems. 2023,44(7):1419-1427.
- [22] WU S Q,DONG Y H,WANG X, et al. Learning attribute network algorithm based on high-order similarity[J]. Telecommunications Science,2020,36(12):20-32.
- [23] ZHOU C,LIU Y,LIU X, et al. Scalable graph embedding for asymmetric proximity[C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2017.
- [24] DONG Y,CHAWLA N V,SWAMI A. metapath2vec: Scalable representation learning for heterogeneous networks[C]// Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2017:135-144.
- [25] WANG X,JI H,SHI C, et al. Heterogeneous graph attention network[C]// The World Wide Web Conference. 2019:2022-2032.
- [26] LI J,SELVARAJU R,GOTMARE A, et al. Align before fuse: Vision and language representation learning with momentum distillation[J]. Advances in Neural Information Processing Systems,2021,34:9694-9705.
- [27] PAGE L,BRIN S,MOTWANI R, et al. The PageRank citation ranking: Bringing order to the web[R]. Stanford infolab,1999.
- [28] FENG H Y. Reserch of Teachers-Students' s Space Distacne in the University[D]. Zhengzhou: Henan University,2015.



YAN Qiuyan, born in 1978, Ph.D, professor, is a member of CCF (No. 13244S). Her main research interests include multimodal image behavior recognition, education big data analysis and time series data mining.



SUN Hao, born in 1999, undergraduate, is a student member of CCF (No. P9283G). His main research interests include knowledge tracing and so on.

(责任编辑:何杨)