

# 基于学生-问题关联的异构图知识追踪模型

闫秋艳, 司雨晴, 袁 冠, 王志晓

(中国矿业大学计算机科学与技术学院, 江苏徐州 221116)

**摘 要:** 知识追踪旨在评估学习者的学习状态, 并根据先前的答题情况预测他们未来的答题表现. 然而现有的知识追踪模型大多仅关注问题或技能间的关联, 忽略了学生与问题间的结构关系. 为此我们提出了基于学生-问题关联的异构图知识追踪模型(Student-Problem association based heterogeneous graph Knowledge Tracing model, SPKT). 该模型在知识追踪中融合了学生的学习能力和问题的重要性, 并使用图注意力网络学习学生问题间的关联, 获得学生、问题的嵌入表示并进行知识状态的预测. 通过在真实公开数据集上的性能对比和模型消融实验, 并可视化 SPKT 模型的知识追踪效果, 证明了 SPKT 在预测性能上优于现有的知识追踪模型.

**关键词:** 知识追踪; 异构图网络; 注意力机制; 学习能力; 问题重要性

**基金项目:** 国家自然科学基金(No.62277046, No.61977061)

**中图分类号:** TP301 **文献标识码:** A **文章编号:** 0372-2112(2023)12-3549-08

**电子学报 URL:** <http://www.ejournal.org.cn>

**DOI:** 10.12263/DZXB.20221427

## Student-Problem Association Based Heterogeneous Graph Knowledge Tracing Model

YAN Qiu-yan, SI Yu-qing, YUAN Guan, WANG Zhi-xiao

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

**Abstract:** Knowledge tracing aims to assess learners' learning status and predict their future performance based on previous answers. However, most of the existing knowledge tracing models only focus on the relationship between problems or skills, and ignore the structural relationship between students and problems. Therefore, we propose a student-problem association based heterogeneous graph knowledge tracing model (SPKT). The model integrates the learning ability of students and the importance of problems in knowledge tracing, and uses graph attention network to learn the interaction between students' problems, so as to obtain the embedded representation of students and problems, meanwhile predicting the learners' status. It is proved that SPKT is superior to the existing knowledge tracking model in terms of prediction performance through a large number of experiments and data visualization.

**Key words:** knowledge tracing; heterogeneous graph network; attention mechanism; learning ability; importance of problems

**Foundation Item(s):** National Natural Science Foundation of China (No.62277046, No.61977061)

## 1 引言

知识追踪的主要任务是根据已有的答题情况对学生知识状态进行建模, 从而预测学生对技能的掌握程度<sup>[1]</sup>. 目前的主流方法大致可分为传统方法<sup>[2]</sup>和深度学习方法<sup>[3]</sup>, 由于传统方法人工成本高昂, 当前的知识追踪模型多以深度学习方法为主. 深度学习方法又可按模型时空关系分为两类, 一类基于答题时序以循环神经网络 RNN(Recurrent Neural Network)为基础, 其对技能的结构信息建模过于简略. 一类基于技能结构以

图神经网络 GNN(Graph Neural Network)为主. 基于图神经网络的知识追踪模型可以更好地表达知识、问题间的结构并利用边缘信息扩展模型, 因此在图的基础上进行知识追踪具有更广阔的研究前景<sup>[4-6]</sup>.

本文中设定有三种类型的节点, 即学生节点、问题节点和技能节点. 一个问题可包含几个技能, 一个技能可涉及多个问题. 为了更好地表达学生和题目、技能之间的关系, 本文针对学生学习能力和题目重要性提出两个经验假设: 假设 1: 两个学生具有相似的答题序列往往意味着他们具有相似的学习过程, 而答对问题数量更多的学生, 则具有

更强的学习能力. 假设2:在问题集合中,一个问题所涉及的技能越多,且与其相似的问题越少,则回答这个问题所带来的收益越大,对于学生的重要性也越大.

本文将学生、问题看作两类节点  $V = \{U, Q\}$ ,  $U$  为学生集合,  $Q$  为问题集合. 学生与问题间的关系看作两类边  $E = \{T, F\}$ ,  $T$  为正确边集合,  $F$  为错误边集合. 基于上述两个假设,本文提出了基于学生-问题关联的异构图知识追踪模型,将知识追踪定义为在异构图上的元组分类问题. 本文的主要贡献包括:(1)基于假设1,利用问题与技能间和学生与问题间的结构关系,结合学生答题表现和学生节点的度,分别计算学生的技能掌握程度和学习能力,二者融合后建立学生节点的特征嵌入;(2)基于假设2,认为问题的重要性与其包含技能的丰富度成正比,与问题之间的区分度成反比,据此通过随机游走策略配合 PageRank 算法<sup>[7]</sup>,计算不同问题的重要性,建立问题节点的特征嵌入;(3)引入多头注意力机制,将融合学生能力和问题重要性后的学生节点嵌入和问题节点嵌入,通过多头注意力机制重新融合,提取学生-问题关联中有价值的信息,提高知识追踪的预测精度;(4)在四个真实公开数据集上进行了对比实验和消融实验,并可视化 SPKT 模型的知识追踪效果,验证了本文提出的 SPKT 模型的有效性及其先进性.

## 2 相关工作

### 2.1 知识追踪

知识追踪研究可分为传统方法和深度学习方法两大类. 传统方法分为贝叶斯知识追踪 BKT (Bayesian Knowledge Tracing)<sup>[2]</sup>和基于逻辑回归模型的知识追踪 IRT (Item Response Theory)<sup>[8]</sup>. 这两类模型属于高度结构化的模型,需要较高人工成本,深度知识追踪模型 DKT (Deep Knowledge Tracing)<sup>[3]</sup>的出现使得知识追踪迈入了深度学习领域. DKT 将 RNN 应用到知识追踪任务中,取得了比传统方法更好的预测性能,但它忽略了知识结构的影响,因此 Chen<sup>[9]</sup>将技能间前驱后继关系建模为有序对,用作知识追踪模型中的约束. 此外,PEBG (Pre-training Embeddings via Bipartite Graph)<sup>[10]</sup>模型对每个问题进行预训练嵌入,在获得的嵌入上训练深层知识追踪模型. 基于练习感知的知识追踪模型 EKT (Exercise-aware Knowledge Tracing for student performance prediction)<sup>[11]</sup>添加了问题的技能属性,使用内存网络衡量学习者在学习每个问题时,对技能掌握程度的影响. 融合认知诊断理论和多头注意力机制的预测模型 C&RM-MAKT (Cognitive & Response Model-Multi-head Attention Knowledge Tracing)<sup>[12]</sup>使用多头注意力机制计算出历史序列数据对预测结果的影响程度. 上述方法多将答题记录简化为时间序列,不能挖掘

答题记录中蕴含的结构信息. 受图神经网络的启发 Nakagawa 提出了 GKT<sup>[4]</sup>,但 GKT 忽略了问题间潜在的层次图性质. Tong<sup>[6]</sup>提出的 HGKT 有效的解决了这一问题. Yang<sup>[5]</sup>利用图卷积网络 GCN (Graph Convolutional Network) 来聚合问题嵌入和技能嵌入,提出的 GIKT (A Graph-based Interaction Model for Knowledge Tracing) 探讨了问题和技能间的相关性.

### 2.2 异构网络

异构图神经网络可以模拟现实世界的异构信息网络,在推荐系统、网络挖掘等领域都取得了成果. 随着图表示学习技术的兴起,以图卷积网络 GCN 和 GAT<sup>[13]</sup> (Graph Attention Networks) 为代表的图神经网络被应用于异构网络的表示学习中. Wang 提出一种基于层次注意力的异构图神经网络模型 HAN (Heterogeneous Graph Attention Network)<sup>[14]</sup>,利用语义级别注意力学习每条元路径的权重,并通过相应的聚合操作得到最终的节点表示. 由于大多基于元路径的表示学习方法舍弃了元路径内部的节点信息,因此 Fu<sup>[15]</sup>等提出了基于元路径的异构嵌入模型 MAGNN (Metapath Aggregated Graph Neural Network for Heterogeneous Graph Embedding),使用注意力机制将元路径内部的语义信息纳入考虑,并聚合多条元路径信息得到最终的节点表示. 在知识追踪领域,基于会话序列的异构知识追踪模型 SGKT (Session Graph-based Knowledge Tracing for student performance prediction)<sup>[16]</sup>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. 整体来看,在基于异构图的知识追踪中,多数研究只考虑知识和问题间的关联,忽略了学生节点在结构中的作用.

## 3 方法描述

模型整体框架如图1所示,模型由三个模块组成:嵌入模块,信息聚合传递模块和预测模块. 其中嵌入模块又分为学生节点嵌入和问题节点嵌入. 模型的主要任务是根据学生的历史答题记录追踪学生的知识状态,并预测学生未来的答题表现.

### 3.1 问题定义

假设在线学习系统中,存在一组学生集  $U$ , 一组问题集  $Q$ , 一组技能集  $S$ . 本文引入学生-问题关联,用  $M^{|U| \times |Q|}$  矩阵表示;问题-技能关联,用  $N^{|Q| \times |S|}$  矩阵表示.

$M$  中的第  $i$  行即为学生  $i$  当前时刻的答题序列  $u_i'$ , 学生在做题过程中会产生新的答题记录,不同时刻的  $M$  矩阵并不相同.  $M$  将根据模型训练后学生与问题间的权重进行更新并保存为  $M'$ . 当学生  $i$  在  $t+1$  时刻答对了问题  $j$ , 则证明在  $t+1$  时刻,学生  $i$  对问题  $j$  的掌握程度最好,因此只需将  $t$  时刻训练得到的  $M'$  中的第  $i$  行第  $j$  列置为 1, 其

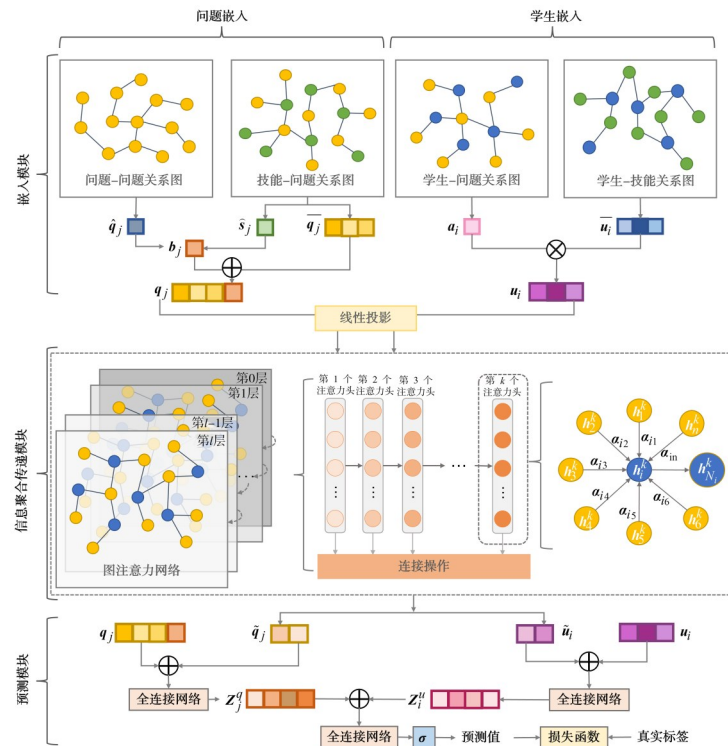


图1 SPKT模型框架图

他位置保持不变,即可获得  $t+1$  时刻的  $M$  矩阵.

在  $N$  中,当问题  $i$  涉及技能  $j$ ,则将矩阵的第  $i$  行,第  $j$  列置为 1.  $N$  中的第  $i$  行表示问题  $i$  的初始嵌入  $\bar{q}_i$ ,由于问题与技能之间的关系是静态的,因此  $N$  在所有时刻  $t$  下都是固定的.在二者基础上,本文对学生的学习和问题的重要性进行定义.

**定义 1** 学生能力系数 根据学生  $i$  的答题序列  $u_i'$  对学生的答题能力进行定义,表示为能力系数  $a_i$ ,具体定义见式(1).

**定义 2** 问题重要性系数 根据问题与技能间的关系,综合考虑问题的丰富度和区分度,利用  $N$  矩阵对问题  $j$  的重要性进行定义,表示为  $b_j$ ,具体定义见式(4)~(6).

### 3.2 学生节点嵌入模块

首先引入学生能力系数  $a_i$  作为学生的属性,能力系数由图 1 中信息聚合传递模块里的图注意力网络计算获得,如式(1)所示.其中,  $A_i$  表示学生  $i$  答题情况的集合,  $|A_i|$  表示学生  $i$  回答问题的总次数,  $r_{it}$  为答题情况,表示学生  $i$  在第  $t$  次的答题情况,若回答正确,则  $r_{it}$  的值为 1,错误则为 0.

$$a_i = \frac{\sum_{t \in A_i} r_{it}}{|A_i|} \quad (1)$$

其次,学生  $i$  的知识关联  $\bar{u}_i$  的计算方式如式(2)所示.其中,  $u_i'$  表示学生的答题序列,将学生的答题序列

与  $N$  矩阵做乘积得到学生的知识关联.  $\bar{u}_i$  是维度为  $1 \times S$  的向量,  $S$  表示技能数量,每列代表一个技能,每列的数值越大,则证明学生  $i$  正确回答与该技能相关问题的数量越多.例如当学生甲正确回答了问题  $c$  和问题  $d$ ,问题  $c$  和问题  $d$  均涉及技能  $A$ ,按式(2)计算,即可知学生甲在技能  $A$  上获得学习收益两次.

$$\bar{u}_i = u_i' \cdot N \quad (2)$$

最后,将能力系数  $a_i$  嵌入到学生的知识关联  $\bar{u}_i$  中,获得学生的嵌入表示  $u_i$ ,其计算过程如式(3)所示.例如存在甲乙两位学生,且学生甲的学习能力明显高于学生乙,那么当两人在某个技能上学习的次数一致时,学生甲的掌握程度应高于学生乙,因此本文将系数  $a_i$  与学生的知识关联  $\bar{u}_i$  做乘积.

$$u_i = a_i \cdot \bar{u}_i \quad (3)$$

### 3.3 问题节点嵌入模块

本文采用计算网络节点重要度的 PageRank 算法,度量某节点在网络中的辐射能力,分别定义了问题节点的丰富度和区分度.针对问题  $j$  直接关联的某个技能节点,该技能节点在技能-技能关联图中辐射的技能越多 ( $PR(s_k)$  越大),在技能-问题关系图中直接关联的问题越少 ( $L(s_k)$  越小),则说明此技能节点可传递给问题  $j$  的信息越多,因此,问题  $j$  在所有与其直接关联的技能节点上获得信息总量定义为问题的丰富度,计算过程如式(4)所示.同理,对于问题  $j$  直接关联的问题节点,

若该节点在问题-问题关联图辐射能力越强( $\text{PR}(\mathbf{q}_i)$ 越大),而该节点关联的问题越少( $L(\mathbf{q}_i)$ 越小),则说明此问题节点与问题 $j$ 的共享越多,问题 $j$ 与其他问题的区分度越低,因此定义式(5)为问题节点的区分度.

$$\hat{\mathbf{s}}_j = \sum_{k \in Z_j} \frac{\text{PR}(\mathbf{s}_k)}{L(\mathbf{s}_k)} \quad (4)$$

式(4)中, $\hat{\mathbf{s}}_j$ 表示问题节点 $j$ 的丰富度, $Z_j$ 代表问题 $j$ 包含的技能节点的集合, $\text{PR}(\mathbf{s}_k)$ 是技能节点 $\mathbf{s}_k$ 在技能-技能关联图中的辐射能力, $L(\mathbf{s}_k)$ 表示技能节点 $\mathbf{s}_k$ 在技能-问题图的出度,即为技能节点 $\mathbf{s}_k$ 直接关联的问题节点数量.

$$\hat{\mathbf{q}}_j = \sum_{i \in R_j} \frac{\text{PR}(\mathbf{q}_i)}{L(\mathbf{q}_i)} \quad (5)$$

式(5)中, $\hat{\mathbf{q}}_j$ 表示问题节点 $j$ 的区分度, $R_j$ 为与问题 $j$ 相关的问题集合, $\text{PR}(\mathbf{q}_i)$ 是问题节点 $\mathbf{q}_i$ 在问题-问题图中辐射节点的能力, $L(\mathbf{q}_i)$ 表示问题节点 $\mathbf{q}_i$ 的出度,即为问题节点 $\mathbf{q}_i$ 相关的问题节点数量.

最后,综合考虑问题节点的丰富度和区分度,将问题的重要性定义为 $\mathbf{b}_j$ ,其计算方式见式(6):

$$\mathbf{b}_j = \hat{\mathbf{s}}_j / \hat{\mathbf{q}}_j \quad (6)$$

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,因此综合式(4)~(6),对问题节点的嵌入定义如式(7), $\oplus$ 表示拼接操作.

$$\mathbf{q}_j = \hat{\mathbf{q}}_j \oplus \mathbf{b}_j \quad (7)$$

### 3.4 信息聚合传递模块

由于学生-问题异质图中的问题节点和学生节点分别处于不同的特征空间中,对每种类型为 $\Phi_i$ 的节点,引入节点类型转换矩阵 $\mathbf{W}_{\Phi_i}$ ,将问题、学生节点投影到同一向量空间中进行计算,如式(8)所示.

$$\mathbf{h}_i = \mathbf{W}_{\Phi_i} \cdot \mathbf{F}_i \quad (8)$$

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( $\mathbf{F}_i$ 可以表示学生节点嵌入 $\mathbf{u}_i$ 或问题节点嵌入 $\mathbf{q}_i$ ).在投影后的属性特征基础上,基于GAT网络,采用多头注意力机制聚合异质节点的邻域信息,得到每个节点聚合后的特征表示 $\mathbf{H}_i$ ,但是标准GAT网络仅聚合了异质邻居特征,忽略了中心节点属性特征,因此本文将异质邻居聚合特征 $\mathbf{H}_i$ 与中心节点属性特征 $\mathbf{F}_i$ 拼接,并通过全连接层进行特征融合,如式(9)所示.

$$\mathbf{Z}_i = \sigma(\mathbf{g}(\mathbf{H}_i \oplus \mathbf{F}_i)) \quad (9)$$

式(9)中, $\mathbf{g}(\cdot)$ 表示全连接层.最终分别获得问题的向量嵌入 $\mathbf{Z}_j^q$ ,和学生的向量嵌入 $\mathbf{Z}_i^u$ .

### 3.5 预测模块

为预测学生 $i$ 在下一时刻是否能正确回答问题 $j$ ,

将3.4节中得到的问题和学生嵌入拼接计算出预测概率 $\hat{y}_{ij}$ ,如式(10)所示, $f(\cdot)$ 表示全连接层, $\text{sigmoid}(\cdot)$ 表示非线性激活函数.

$$\hat{y}_{ij} = \text{sigmoid}(f(\mathbf{Z}_i^u \oplus \mathbf{Z}_j^q)) \quad (10)$$

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,如下所示:

$$\mathcal{L} = - \sum_{i,j \in y^+ \cup y^-} (y_{ij} \log \hat{y}_{ij} + (1 - y_{ij}) \log(1 - \hat{y}_{ij})) \quad (11)$$

式(11)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端到端的训练.

## 4 实验

本文实验环境为:Windows11操作系统,CPU型号为Intel Core i7-12700H,核心数为14核,内存为24 GB.显示芯片型号为RTX™3060,显存类型为GDDR6.本文在pytorch框架下实现SPKT模型及其变体模型.

### 4.1 数据集

为了对本文提出的SPKT模型的有效性进行验证,在四个知识追踪方向广泛使用的真实公开数据集上进行了实验.ASSIST2009、ASSIST2012、ASSIST2017均是由ASSISTments在线教育平台收集.其中ASSIST2012中,每个问题只涉及一项技能,但一项技能仍然对应着几个问题.EdNet由于整个数据集太大,本文随机选择了5 000名学生,包含189项技能、12 161个问题和676 974个问题记录.四个数据集的统计数据如表1.

表1 数据集统计

数据集	学生	技能	问题	记录
ASSIST2009	4 151	110	16 891	325 637
ASSIST2012	27 485	265	53 065	2 541 201
ASSIST2017	1 709	102	3 162	942 816
EdNet	5 000	189	12 161	676 974

### 4.2 对比实验

本文选用传统方法BKT和六种深度学习模型作为基线方法进行对比分析,BKT、DKT、GKT、GIKT和SGKT方法在前述部分已经介绍,基于动态键值存储网络的知识追踪模型DKVMN(Dynamic Key-Value Memory Networks for Knowledge Tracing)<sup>[17]</sup>使用存储网络分别存储不同概念的知识状态,而不是使用单个隐藏状态.基于上下文感知的注意力知识追踪模型AKT(Context-Aware Attentive Knowledge Tracing)<sup>[18]</sup>使用了一种新颖的单调注意机制,将学习者未来对评估问题的反应与其过去的反应联系起来;除了问题之间的相似性之外,注意力权重还使用指数衰减和上下文感知的相对距离度量来计算.本节实验对

所有模型均采用5折交叉验证,采用AUC和ACC作为性能评价指标,ACC定义见式(12).

$$ACC = \frac{TP + TN}{TP + FP + FN + TN} \quad (12)$$

评价指标ROC曲线下面积AUC(Area Under Curve)和ROC(Receiver Operating characteristic Curve)相关,其中ROC曲线的横坐标为FPR(False Positive Rate)即负类样本中被判定为正类的比例,纵坐标为TPR(True Positive Rate)即正类样本中被判定为正类的样本.

表2显示了所有方法在预测学生未来学习情况的AUC

和ACC指标.结果表明SPKT在三个数据集上都达到了最优.但在ASSIST2009数据集上,AKT的表现略优于SPKT,这表明在学生-问题关联密度越高的数据集上,SPKT的表现越好,进一步融合了图结构和注意力机制的知识追踪模型相较于仅依赖注意力机制的模型可以从关联中学习丰富信息.但是由于SPKT是基于学生-问题关联进行知识追踪,因此在学生-问题关联稀疏的数据集上(如ASSIST2009),仍存在改进空间.此外与GIKT,GKT相比,SPKT在四个数据集上均有显著优势,这表明融入学生-问题的异构关联对知识追踪的效果是有正向影响的.

表2 模型预测AUC值对比实验结果

模型	ASSIST2009		ASSIST2012		ASSIST2017		EdNet	
	AUC	ACC	AUC	ACC	AUC	ACC	AUC	ACC
BKT <sup>[2]</sup>	0.651 3	0.647 1	0.623 2	0.654 4	0.611 4	0.632 8	0.627 1	0.615 2
DKT <sup>[3]</sup>	0.766 2	0.743 3	0.730 6	0.726 9	0.727 7	0.703 5	0.764 9	0.718 9
DKVMN <sup>[17]</sup>	0.796 7	0.745 9	0.787 3	0.742 3	0.726 8	0.698 8	0.784 0	0.727 6
AKT <sup>[18]</sup>	<b>0.826 3</b>	0.755 7	0.813 4	0.745 8	0.782 9	0.712 7	0.803 1	0.748 8
GKT <sup>[4]</sup>	0.724 4	0.705 6	0.735 2	0.712 5	0.715 2	0.711 2	0.701 8	0.689 6
GIKT <sup>[5]</sup>	0.789 0	0.738 9	0.775 7	0.739 1	0.761 3	0.735 7	0.753 5	0.726 4
SGKT <sup>[16]</sup>	0.797 5	0.731 2	0.813 5	0.751 2	0.792 4	0.735 6	0.755 0	0.699 5
SPKT	0.825 1	<b>0.784 4</b>	<b>0.815 8</b>	<b>0.761 6</b>	<b>0.802 7</b>	<b>0.750 1</b>	<b>0.811 3</b>	<b>0.774 7</b>

### 4.3 消融实验

为了验证SPKT模型的优势,本节分别针对信息传播层数、注意力头数、学生节点嵌入、问题节点嵌入等内容进行了消融实验分析,以验证SPKT设计的有效性.

#### 4.3.1 传播层数影响

本文将传播层数范围设置在0~4之间,以研究学生-问题间关联对实验效果的影响,结果如表3所示.以学生节点为基点,学生-问题关联阶数等价于网络层数.当层数为0时意味着模型不通过学生-问题关联传递信息,仅使用学生节点自身信息预测学生知识状态及答题结果.当层数为1时,此时使用了一阶的学生-问题关联,即考虑学生自身信息和学生所做题目信息.其他阶数含义依此类推.由表3可知,当网络层数为2、3时,模型表现效果达到峰值.层数过低时,无法充分利用图神经网络的优势捕捉学生-问题间的关联信息.然而,在图结构中,单个节点从图上获取的信息随层数加深呈几何倍增长,当层数超过3时,图中所有节点包含的信息会趋于相同,模型表现呈下降趋势.

#### 4.3.2 学生节点嵌入模块的影响

为了验证SPKT中对学生节点嵌入设计的有效性,本文对模型的四个变体进行了消融实验,四种模型的具体内容为:SPKT-n表示模型未引入学生能力系数且未使用学生知识关联;SPKT-a表示模型仅引入学生能力系数;SPKT-s表示模型仅使用了学生与知识的关联生成学生节点嵌入表示;SPKT-u表示模型引入学生能力系数且使用

表3 传播层数的影响

网络层数	ASSIST2009		ASSIST2012		ASSIST2017		EdNet	
	AUC	ACC	AUC	ACC	AUC	ACC	AUC	ACC
0	0.812 9	0.761 2	0.793 2	0.748 3	0.798 5	0.727 7	0.797 9	0.758 6
1	0.813 3	0.781 4	0.808 1	0.757 7	0.800 3	0.738 2	0.810 7	0.762 8
2	<b>0.825 1</b>	0.784 4	<b>0.815 8</b>	<b>0.761 6</b>	0.801 9	<b>0.750 3</b>	0.811 2	0.772 0
3	0.824 2	<b>0.784 7</b>	0.815 4	0.760 0	<b>0.802 7</b>	0.750 1	<b>0.811 3</b>	<b>0.774 7</b>
4	0.803 9	0.759 8	0.807 7	0.749 9	0.798 8	0.730 3	0.792 6	0.749 0

学生知识关联.

表4结果表明SPKT综合考虑学生的学习能力和学生与知识的关联后,模型性能达到最佳.这说明对学生节点的嵌入设计是有效的.此外在三个数据集上,引入学生能力系数获得了比引入学生与知识的关联更好的效果.这说明学生的学习能力可以为预测学生未来表现提供更多参考.

#### 4.3.3 问题节点嵌入模块的影响

为了评估SPKT中对问题节点嵌入的设计,对SPKT模型的四个变体进行了消融实验,模型间的具体区别如下:SPKT-f表示模型不添加问题的额外属性;SPKT-q表示模型仅添加了问题的区分度作为问题属性;SPKT-k表示模型仅使用问题的丰富度作为问题属性;SPKT-b表示同时使用问题丰富度和区分度的比值作为问题属性.

表5说明问题的丰富度和区分度均对SPKT的性能有影响.在三个数据集中,问题的丰富度对预测结果的影响

更大,将问题间的区分度纳入考虑同样对模型性能也有一定提升.但在 ASSIST2012数据集下,问题区分度对模型性能提升略高于问题丰富度,这主要是由于 ASSIST2012为

单技能数据集,每个问题只涉及一个技能.综合来看,将问题丰富度和区分度的比值作为问题的重要性属性,对模型的性能提升最为显著,这说明 SPKT改进是有意义的.

表4 学生学习能力嵌入的影响

模型	ASSIST2009		ASSIST2012		ASSIST2017		EdNet	
	AUC	ACC	AUC	ACC	AUC	ACC	AUC	ACC
SPKT-n	0.802 9	0.758 9	0.810 2	0.731 4	0.801 7	0.731 0	0.797 9	0.746 4
SPKT-a	0.815 7	0.759 3	0.813 1	0.740 6	0.802 1	0.737 7	0.809 7	0.753 2
SPKT-s	0.813 0	0.761 9	0.812 6	0.748 4	0.799 8	0.742 7	0.810 4	0.758 2
SPKT-u	<b>0.825 1</b>	<b>0.784 4</b>	<b>0.815 8</b>	<b>0.761 6</b>	<b>0.802 7</b>	<b>0.750 1</b>	<b>0.811 3</b>	<b>0.774 7</b>

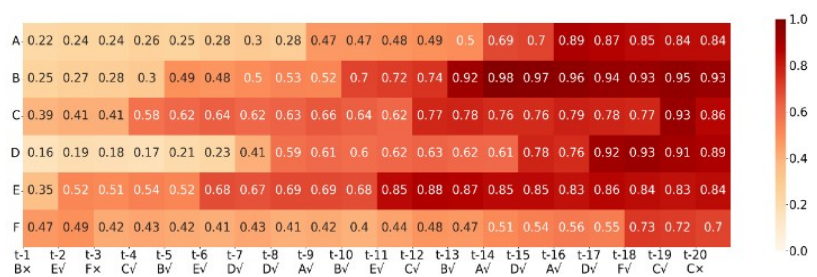
表5 问题重要性嵌入的影响

模型	ASSIST2009		ASSIST2012		ASSIST2017		EdNet	
	AUC	ACC	AUC	ACC	AUC	ACC	AUC	ACC
SPKT-f	0.804 2	0.767 0	0.810 7	0.749 2	0.785 8	0.735 5	0.803 8	0.747 7
SPKT-q	0.814 2	0.769 1	0.813 2	0.748 5	0.794 9	0.741 5	0.795 3	0.759 2
SPKT-k	0.823 0	0.774 2	0.812 6	0.749 4	0.795 8	0.743 8	0.810 3	0.758 5
SPKT-b	<b>0.825 1</b>	<b>0.784 4</b>	<b>0.815 8</b>	<b>0.761 6</b>	<b>0.802 7</b>	<b>0.750 1</b>	<b>0.811 3</b>	<b>0.774 7</b>

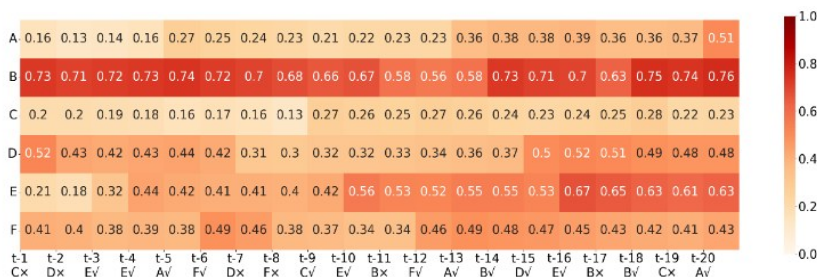
#### 4.4 实验结果可视化分析

为了说明 SPKT 模型对学生节点改进的有效性,本文从 ASSIST2009 中针对六个技能,选取了两名学生的答题序列及知识状态进行展示,如图2所示.其中纵坐标为技能编号,横坐标表示学生每个时刻的答题情况,如“B×”表示学生回答了一道和技能B相关的题目且回答错误.热力图中色块颜色的深浅代表学生对知识点的掌握程度,色块中的数值表示答题后学生的知识状态.从图中可以看出,当学生回答正确一道题目时,学

生对该技能的掌握程度会增加,当学生回答错误一道题目,证明其对该技能的掌握程度有所衰减.不同学生回答同一道题目的技能状态增幅有明显区别,其中学生甲的学习能力系数高于学生乙,即学生甲比学生乙拥有更强的学习能力.学生甲乙均回答正确三个技能A上的问题,但学生甲的技能掌握程度增长了0.61,学生乙只增长了0.35.学生乙回答了四道与技能E相关的题目且结果全部为正确,其在技能E上的掌握情况增长了0.49,而学生甲仅回答正确三道与技能E相关的题目,就获得了0.5的技能掌握程度提升.由此可见,增加



(a) 学生甲知识状态热力图



(b) 学生乙知识状态热力图

图2 SPKT模型学生知识状态热力图

了学生的能力系数对不同学生的学习能力进行区分, 可以更好的反映现实学习规律.

#### 4.5 聚类分析

本节主要用以验证加入学生-问题关联, 是否对问题节点的表示也可产生一定积极影响. 由于 GIKT 和 SGKT 与本文模型均为基于异构图的知识追踪模型, 并且三者均在问题节点的嵌入表示上进行了讨论, 因此我们选用 GIKT 和 SPKT 作为聚类实验的对比模型. 实验将技能编号作为问题节点的标签, 利用 KMeans 算法对问题节点进行聚类, 将簇数设置为每个数据集中技能节点的数目. 采用标准化互信息(Normalized Mutual Information, NMI)和调整兰德系数(Adjusted Rand Index, ARI)作为聚类评价指标, 定义见式(13)、式(14)和式(15). 由于 KMeans 的性能受初始质心的影响, 为了避免实验结果的偶然性, 进行 10 次重复实验, 计算实验均值, 表 6 给出了聚类的实验结果.

$$NMI(U, V) = 2 \frac{MI(U, V)}{H(U) + H(V)} \quad (13)$$

其中 MI 为变量间的互信息值,  $H$  为变量的熵值.

兰德系数 RI(Rand Index)为

$$RI = \frac{TP + TN}{TP + FP + TN + FN} = \frac{TP + TN}{C_N^2} \quad (14)$$

调整兰德系数 ARI 为

$$ARI = \frac{RI - E[RI]}{\max(RI) - E[RI]} \quad (15)$$

从表 6 可以看出, 本文的模型 SPKT 在聚类任务上具有较优的性能. 比其他模型的标准互信息(NMI)高出 0.38%~6.21%, 在调整兰德系数(ARI)上, 相较其他模型高出 0.8%~14.94%. SPKT 与其他基线模型相比, 模型在考虑了学生问题间关联和学生答题的能力系数后, 问题节点在技能上的区分能力更好. 因此, 加入学生-问题间关联和学生答题能力对问题的分类分级存在积极影响.

为了更清晰地表现问题节点的聚类效果, 选用 ASSIST2012 数据集中的十个技能以及与这些技能具有关联关系的问题进行可视化展示. 本文分别将三个模型学习到的问题节点特征向量通过 t-SNE 算法进行降维, 并投影到二维空间中, 进行可视化比较. 针对每个问题节点表示向量, 根据所属技能编号进行不同着色. 从图 3 中可以看出, 相比 GIKT, GKT 模型, SPKT 可以更好地将不同技能的问题映射到不同的区域, 不同技能类别的题目之间界限更为清晰. 此外, 在同一个技能下问题聚集的区域中, 问题节点的重叠度不高, 在聚类的基础上更好地保留了问题间的区分度.

表 6 问题节点聚类结果

模型	ASSIST2009		ASSIST2012		ASSIST2017		EdNet	
	NMI	ARI	NMI	ARI	NMI	ARI	NMI	ARI
GIKT <sup>[5]</sup>	0.685 5	0.443 2	0.689 8	0.465 6	0.686 2	0.434 7	0.446 9	0.150 7
SGKT <sup>[16]</sup>	0.638 1	0.421 2	0.644 1	0.371 4	0.728 2	0.486 0	0.399 5	0.125 4
SPKT	<b>0.699 6</b>	<b>0.450 3</b>	<b>0.725 2</b>	<b>0.520 8</b>	<b>0.748 3</b>	<b>0.578 0</b>	<b>0.450 7</b>	<b>0.158 7</b>

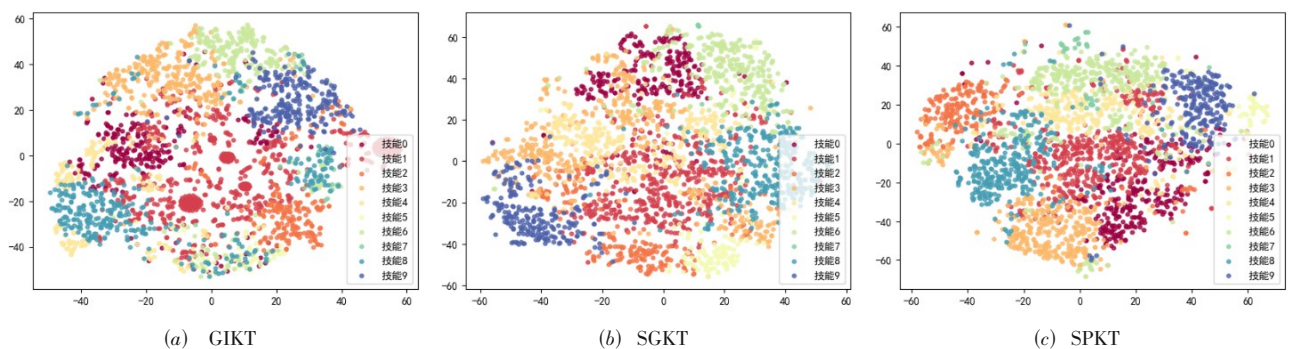


图 3 问题节点向量表示的聚类可视化效果图

## 5 总结与未来工作

本文提出了一种基于学生-问题关联的异构图知识追踪模型 SPKT, 根据学生已有的知识状态, 预测学生隐藏的答题表现. SPKT 模型改进了现有的知识追踪模型, 融合了现实教学场景下学生的学习能力和问题的重要性, 在表达了知识追踪图结构的基础上, 增加了注意力机制.

在真实公开数据集上的实验结果表明 SPKT 模型取得了较好的性能. 在未来研究中, 为了进一步发挥 SPKT 模型的优势, 将在技能、问题节点上融合多模态, 获得更有层次的问题、技能节点表示; 在学生与问题间的关联上补充更多信息, 例如学生做题的状态等, 充分发挥异构图的潜能使知识追踪更切合现实场景, 而非仅依靠对错这一单一指标判断学生是否掌握相应技能.

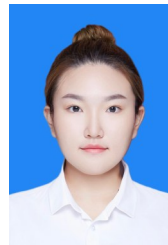
## 参考文献

- [1] 张暖, 江波. 学习者知识追踪研究进展综述[J]. 计算机科学, 2021, 48(4): 213-222.  
ZHANG N, JIANG B. Review progress of learner knowledge tracing[J]. Computer Science, 2021, 48(4): 213-222. (in Chinese)
- [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] PIECH C, BASSEN J, HUANG J, et al. Deep knowledge tracing[EB/OL]. (2022). <https://arxiv.org/abs/1506.05908>.
- [4] 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. Piscataway: IEEE, 2019: 156-163.
- [5] YANG Y, SHEN J, QU Y R, et al. GIKT: A graph-based interaction model for knowledge tracing[C]//European Conference on Machine Learning and Knowledge Discovery in Databases. Cham: Springer, 2021: 299-315.
- [6] TONG H, WANG Z, ZHOU Y, et al. HGKT: Introducing hierarchical exercise graph for knowledge tracing[EB/OL]. (2006). <https://arxiv.org/abs/2006.16915>.
- [7] BRIN S. The PageRank citation ranking: Bringing order to the web[J]. Proceedings of ASIS, 1998, 98: 161-172.
- [8] CEN H, KOEDINGER K, JUNKER B. Comparing two IRT models for conjunctive skills[C]//International Conference on Intelligent Tutoring Systems. Berlin Heidelberg: Springer, 2008: 796-798.
- [9] CHEN P, LU Y, ZHENG V W, et al. Prerequisite-driven deep knowledge tracing[C]//2018 IEEE International Conference on Data Mining. Piscataway: IEEE, 2018: 39-48.
- [10] LIU Y, YANG Y, CHEN X, et al. Improving knowledge tracing via pre-training question embeddings[EB/OL]. (2012). <https://arxiv.org/abs/2012.05031>.
- [11] LIU Q, HUANG Z Y, 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.
- [12] 王炼红, 罗志辉, 林飞鹏, 等. 采用多头注意力机制的 C&RM-MAKT 预测算法[J]. 电子学报, 2023, 51(5): 1215-1222.  
WANG L H, LUO Z H, LIN F P, et al. C&RM-MAKT prediction algorithm using multi-head attention mechanism[J]. Acta Electronica Sinica, 2023, 51(5): 1215-1222. (in Chinese)
- [13] VELICKOVIC P, CUCURULL G, CASANOVA A, et al. Graph attention networks[EB/OL]. (2022). <https://arxiv.org/abs/1710.10903>.
- [14] WANG X, JI H, SHI C, et al. Heterogeneous graph attention network[C]//The World Wide Web Conference. New York: ACM, 2019: 2022-2032.
- [15] FU X, ZHANG J, MENG Z, et al. MAGNN: Metapath aggregated graph neural network for heterogeneous graph embedding[C]//Proceedings of The Web Conference 2020. New York: ACM, 2020: 2331-2341.
- [16] 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.
- [17] ZHANG J, SHI X, KING I, et al. Dynamic key-value memory networks for knowledge tracing[C]//Proceedings of the 26th International Conference on World Wide Web. New York: ACM, 2017: 765-774.
- [18] GHOSH A, HEFFERNAN N, LAN A S. Context-aware attentive knowledge tracing[C]//Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York: ACM, 2020: 2330-2339.

## 作者简介



闫秋艳 女, 1978年8月出生于江苏省徐州市. 现为中国矿业大学计算机科学与技术学院副教授、博士生导师. 研究方向为教育大数据挖掘、机器学习.  
E-mail: yanqy@cumt.edu.cn



司雨晴 女, 1997年8月出生于江苏省徐州市. 本科毕业于中国矿业大学信息安全专业, 硕士就读于中国矿业大学计算机应用技术专业.  
E-mail: siyq@cumt.edu.cn



袁冠(通讯作者) 男, 1982年生, 江苏睢宁人. 现任中国矿业大学计算机科学与技术学院教授. 主要研究方向包括时空大数据技术以及计算智能.  
E-mail: yuanguan@cumt.edu.cn