

融合课程知识点图谱的知识追踪进行习题推荐

王磊¹, 雷虹², 钱凯³

(1. 陕西省远程教育研究中心, 陕西 西安 710119; 2. 榆林市开放大学, 陕西 榆林 719052;
3. 陕西开放大学新城分校, 陕西 西安 710032)

摘要:为解决个性化习题推荐中因忽略知识点间深层关联而导致的推荐盲区问题,本研究基于“大模型+LangChain”构造课程知识点图谱,并结合知识追踪模型进行习题推荐。首先识别课程知识点并构建关联图谱,明确知识点间逻辑关系;再采用知识追踪算法,动态捕捉学习者知识掌握状态;最后结合知识点图谱与知识追踪结果实现个性化习题推荐。实验结果表明,与基线模型相比,本模型在推荐分值上有所提高,证明了基于大模型构建的课程知识点图谱结合知识追踪的有效性。

关键词:习题推荐;大模型;LangChain;知识追踪;知识点图谱

中图分类号:TP311 **文献标识码:**A

文章编号:1009-3044(2026)02-0049-04

DOI:10.14004/j.cnki.ckt.2026.0054

开放科学(资源服务)标识码(OSID):



0 引言

随着“互联网+教育”及在线学习平台的广泛使用,题库中习题呈现快速增长趋势,如何辅助学习者更高效选取练习题,成为当下教育研究热点。习题练习作为学生学习过程中的重要环节,对于课堂知识的巩固、反思和总结具有重要作用。然而,现有的传统习题推荐方式往往存在局限性。首先,现有习题推荐多依赖于教师的主观经验或教材自带习题,缺失学情分析和对学习个体差异的精准测评。不同学习者在前置知识基础、学习方法和认知能力上存在显著差异,“一刀切”的推荐策略难以满足个性化学习需求,导致一些学习者因习题过难而丧失信心,或部分学习者因习题简单而浪费学习时间。其次,习题库资源虽然丰富,但学习者在海量习题中难以筛选出有针对性的习题,这既增加了学习者的学习负担。在此背景下,智能化的习题推荐技术得到广泛研究,这些技术借助先进的算法模型能够深入挖掘学习者的学习特征,精准分析其知识掌握情况和薄弱知识点,从而为学习者提供个性化、精准化的习题推荐资源。高效的习题推荐方式不仅能够提高学习者的学习效果和学习兴趣,还能为教师提供有价值的教学反馈,提升教学质量。尽管现有研究已取得一定进展,但大多将知识图谱构建与知识追踪视为独立模块,缺乏一个端到端的、能动态利用图谱结构信息来增强知识追踪预测

能力的融合框架。此外,利用大语言模型自动化构建高质量课程知识图谱的研究尚处于起步阶段。因此,本研究针对现有研究中存在的痛点,提出了融合课程知识点图谱与知识追踪的习题推荐方法。

1 研究现状

1.1 基于认知诊断与知识追踪的推荐

在线教育平台积累了大量学生答题数据,针对现有推荐方法未充分考虑学生知识水平动态变化、习题难度与学生能力的匹配度,诸葛斌等^[1]设计了一种基于学生知识追踪的多指标习题推荐算法,从而实现动态精准的个性化推荐。针对习题推荐方法忽视了习题之间的关联关系以及学习者的能力水平,导致习题与学习者失配,浪费了学习者的时间和精力,Li等^[2]通过分析作答序列数据,构建认知诊断模型计算学生对每个知识点的掌握程度和学习进展,同时引入用户对练习难度、类型等的期望,通过模拟退火算法筛选符合用户当前学习阶段和偏好的练习。金天成等^[3]结合认知心理学中关于工作记忆、长时记忆、外部信息以及非连续性认知发展的相关理论,构建了记忆与认知融合的个性化OJ习题推荐模型(MC-DRER)。针对现有习题推荐方法中存在的稀疏数据和忽略群体特征问题,郭英清等^[4]提出了一种结合深度知识追踪(DKT)与矩阵补全的习题推荐算法,旨在提升习题推荐精度。针对当前在线评测(OJ)系统上存有大量习题,学

收稿日期:2025-10-10

基金项目:陕西省教育厅科学研究计划项目资助(25JZ014);陕西省职业技术教育学会2025年度职业教育教学改革研究课题“知识追踪结合大模型构建的课程知识图谱进行习题推荐”,课题编号:2025SZX521,主持人:王磊;2025年度陕西开放大学教育教学改革一般项目“大模型与知识库协同构建课程知识图谱及其在教学中的应用”(sxkd2025yb21)

作者简介:王磊(1984—),男,陕西彬州人,副教授,博士,主要研究方向为教育信息化;雷虹(1983—),女,陕西榆林人,讲师,理学学士,主要研究方向为计算机科学与技术及成人教育;钱凯(1979—),男,陕西西安人,讲师,主要研究方向为信息技术。

本栏目责任编辑:谢媛媛

软件技术

49

生难以根据自己的知识水平和学习需求快速找到合适的习题,金天成等^[5]进一步通过分析习题的最优解来挖掘习题之间的相关性,并比较了学生所提交程序与习题最优解的相似性来评估学生所提交程序的实际正确率。张凯等^[6]针对在线教育突出问题,设计并实现了以知识追踪为核心的智能导学系统,有效解决了学习者知识状态监测与个性化资源推荐难题。

1.2 基于知识图谱的推荐

叶俊民等^[7]提出了一种知识图谱增强的在线评测系统习题推荐算法,旨在通过构建习题知识图谱,结合习题难度和学习者的能力水平,从而实现个性化习题推荐。钱浩文等^[8]以“学习者—知识点掌握缺陷”为背景,提出一种基于树状关系图的个性化习题资源推荐方法,旨在解决现有推荐系统缺乏对知识点关联性分析以及无法精准定位学生薄弱环节的问题。

1.3 融合模型研究

聂庭焜等^[9]提出了一种分层次学习注意力权重的双路注意力推荐模型(SEFM),旨在通过结合因子分解机(FM)与压缩激励注意力网络(SENET) 2个模块提升习题推荐效果和个性化匹配度。面对众多的习题资源,学习者很难从海量的习题资源中找到真正适合自己的习题集来进行训练,导致出现学习效率低下和学习自主性不高等问题。唐燕等^[10]提出了个性化习题推荐算法,对个性化习题推荐算法的发展现状进行研究,介绍常用的个性化习题推荐算法以及5个算法评价指标并给出未来的研究热点和方向。针对传统习题推荐系统挖掘学生特征不足以及知识掌握关联性弱的问题,诸葛斌等^[11]通过融合知识追踪机与协同过滤算法,构建多特征融合的个性化习题推荐框架。

以上研究存在因忽略知识点间深层关联而导致的推荐知识点盲区问题。因此,本研究基于“大模型+LangChain”框架识别课程知识点并构建课程知识点图谱,明确知识点间逻辑关系,使用知识追踪算法跟踪学习者知识掌握状态,最后结合知识点图谱与知识追踪结果实现个性化习题推荐。

2 习题推荐框架

本研究提出的习题推荐框架分为知识库融合、知识点图谱构建以及结合知识追踪生成薄弱知识点3个模块,其中知识库融合对象包括教材、课程标准、教案以及学术文献。构建的知识点图谱结合薄弱知识点生成推荐习题列表,具体流程如图1所示。

2.1 构建课程知识点图谱

本研究采用基于深度学习以及大模型两种方式构建课程知识点图谱。在构建知识点图谱过程中,利用模型获取知识点,确定各知识点之间的逻辑关系,并在知识点图谱中清晰展现知识点的层级关系和关联结构。设知识点集合为 $K = \{k_1, k_2, \dots, k_n\}$,知识点之间的关系可以用有向图 $G = (V, E)$ 表示,其中顶点集合

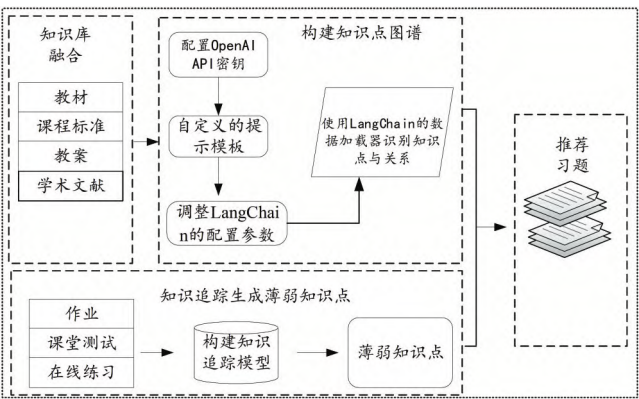


图1 习题推荐框架

$V=K$,边集合 E 表示知识点之间的依赖关系,若 k_i 是 k_j 的前置知识点,则存在有向边 $k_i, k_j \in E$ 。

1) 基于深度学习方法。首先收集课程教材、课件等文本数据并进行清洗和分词等预处理,利用BiLSTM+CRF模型从文本中自动抽取课程知识点实体。其中BiLSTM通过双向长短期记忆网络捕捉文本上下文语义特征,基于双向拼接输出增强全局信息表征;CRF层引入标签转移约束,优化实体标签序列的合理性,解决独立分类假设的局限性。接下来,通过关系抽取模型分析文本上下文,捕捉知识点间的语义关联,本研究定义的知识点关系包括“前置—后续”“包含—被包含”,之后将提取的知识点和关系映射为图结构节点和边,并使用图嵌入技术将图谱映射到低维向量空间,最后结合课程领域知识对自动生成的图谱进行人工校验和优化,确保知识关系的准确性和逻辑性。同时通过持续迭代更新技术使图谱能够动态适应课程内容演变,为个性化学习路径推荐、教学大纲优化等应用提供结构化知识支撑。

2) 基于Deepseek R1+LangChain方法。利用Deepseek R1与LangChain框架构建课程知识图谱,可实现课程知识的自动化抽取与结构化展示。首先,利用LangChain的文档加载器整合课程教案、电子教材等文本数据,并完成数据预处理工作。随后,借助DeepSeek R1的推理能力,设计提示模板引导模型提取知识点三元组,批量处理后生成实体关系列表。接着,利用LangChain框架的工程化能力,将三元组加载进Neo4j图数据库中存储,直观呈现知识点层级与关联。

根据上述2种方法构建知识点图谱,图谱中节点代表不同知识点,边表示它们之间的关系,形成初步的知识点图谱结构。构建过程中需要合理确定节点的粒度和层次,确保知识点图谱既不过于复杂难以理解,也不过于简略而丢失关键信息。构建完图谱后,还需要进行图谱后处理操作,即对生成的图谱进行修复,结合教学实际需求调整节点的位置和边的权重,使图谱更加符合实际教学场景和学生的认知规律。构建初始图谱结构后融合课程教学目标与专业群规则进行图谱结构优化,通过节点布局调整符合认知规

律并实现渐进式学习路径。

2.2 利用知识追踪生成薄弱知识点

1) 知识追踪模型设计。知识追踪算法通过分析学习者历史答题数据来预测未来答题结果,达到动态更新学习者状态的目的,为个性化习题推荐与教学干预提供精准依据。该模块通过在线学习平台、智能教学系统收集学生在学习过程中的学习行为数据,将数据转换为适合知识追踪模型输入的格式,并将学生的答题结果编码为二元变量(正确为1,错误为0),同时记录答题结果对应的知识点。将知识追踪模型预测出的学生知识状态映射到预先构建的知识点图谱上,还需要明确每个掌握程度较低的知识点在图谱中的位置,以及与邻近知识点的关系。同时,进一步考虑知识点之间的依赖关系,分析薄弱知识点后续学习的影响。本研究设计的知识追踪模型使用LSTM神经网络来处理学生的答题序列数据,设输入为学生的答题序列 $X=\{x_1, x_2, \dots, x_T\}$, 隐层状态和输出结果如下:

$$\begin{cases} h_t = \tanh(W_{xh}x_t + W_{hh}h_{t-1} + b_h) \\ y_t = \sigma(W_{hy}h_t + b_y) \end{cases} \quad (1)$$

式中: x_t 包含了第 t 次答题对应的知识点信息和答题结果,网络的隐藏状态 h 编码了学生时刻 t 的知识状态,输出结果 y_t 预测学生下一次答题中的正确率。

2) 知识追踪模型训练。知识追踪分析使用预处理后的数据对选定的知识追踪模型进行训练,通过不断调整模型参数,使模型能够准确预测学生在不同知识点上的掌握情况。最后,利用训练好的模型预测学生知识状态,并根据学生最新的学习数据预测其在各个知识点上的掌握程度。知识追踪模型在训练过程中通过最小化预测误差来训练网络参数:

$$L(\theta) = -\frac{1}{N} \sum_{i=1}^N \sum_{t=1}^T r_{i,t} \log(y_{i,t}) + (1 - r_{i,t}) \log(1 - y_{i,t}) \quad (2)$$

式中: N 为习题样本数量, θ 为网络参数。式(2)根据学生最新的学习数据计算其在各个知识点上的掌握程度来预测学生知识状态。

2.3 结合课程知识点图谱推荐习题

根据以上步骤得到薄弱知识点后,将薄弱知识点映射至知识图谱中的对应节点。以薄弱节点为起点,利用图遍历算法沿着图谱关联结构向外延伸,挖掘与其紧密相连的知识点,从而构建覆盖强相关和弱相关薄弱知识点集合。设薄弱知识点集合为 E , 对于任意知识点 k_i , 找出薄弱知识点相关知识点集合:

$$N(k_w) = \{k_w | (k_i \rightarrow k_w) \in E\} \quad (3)$$

式中: k_i 代表前置知识点。

设当前知识点为 C , 其前置相关知识点集合为 $S = \{k_1, k_2, \dots, k_n\}$, 每个知识点 k 与 C 的边权值系数为 w_i 。推荐优先级由权值系数决定,权值系数越高,则优先级越高。每个知识点 $k_i \in S$ 的推荐优先级定义为:

$$P(k_i) = w_i \quad (4)$$

对 S 中知识点按优先级降序排列:

$$\text{sort}(S, \text{key} = w_i, \text{reverse} = \text{True}) \quad (5)$$

式中: sort 表示按 w_i 降序排列,从未推荐的知识点中选择权值系数最大的知识点。

根据扩展后的知识点集合,从习题库中筛选匹配题目,优先推荐强相关薄弱的习题,同时融入涉及弱相关知识点。推荐过程中需结合学生答题正确率适当调整相关知识点排序,确保习题难度梯度合理。通过跟踪学生答题表现动态调整推荐策略,若关联知识点掌握情况仍不佳,则进一步细化扩展范围;若已掌握则逐步引入更深层次知识点习题。此方法能系统性地填补知识漏洞,帮助学生构建完整的知识体系,提升学习效率。由于知识点的相关知识点有若干个,需学习知识点与邻居节点边的权值系数,权值系数高的优先推荐,权值系数低的后推荐。

3 实验结果与分析

3.1 实验设计

实验选取 ASSISTments2017 公开数据集, ASSISTments 2017 数据集包含 686 条学生记录, 102 个习题标签, 以及 942 816 条习题记录^[12]。知识追踪过程中分别采用了缺失课程知识点图谱(w/o KG), 传统方式构造的课程知识点图谱(KG{BiLSTM+CRF}), 大模型构造的课程知识点图谱(KG{LLM}) 3 种方式, 对以上 3 种知识追踪方式进行习题推荐对比。

3.2 结果分析

实验结果如表 1 所示。

表 1 实验结果

知识追踪	评价指标/%			
	AUC	P	R	F ₁
w/o KG	81.09	79.65	64.5	71.28
KG{BiLSTM+CRF}	81.37	81.61	65.52	72.63
KG{LLM}	82.51	82.86	66.65	73.89

由实验结果可以看出,无课程知识点图谱时,习题推荐缺乏知识结构支撑,仅能依赖简单规则或用户历史行为,推荐精准度低,学生知识掌握效率受限。结合知识点图谱后准确率得到提升,知识追踪可以动态捕捉学习者知识状态,结合图谱显式知识关联精准定位薄弱点并推荐适配习题,提升推荐精度。传统方式构造的课程知识点图谱通过人工或半自动化梳理知识点关系,虽构建的基础结构丰富,但存在粒度粗、更新慢等问题。

添加传统课程知识点图谱的知识追踪模块推荐效果优于无知识点图谱模型,但知识状态追踪的动态性不足,无法及时适应学生遗忘规律。相较于传统知识图谱构建方法,基于大模型构建知识图谱的推荐模型在 AUC 指标提升了 1.14 个百分点, F1 值提升了 2.61 个百分点,从而验证了 LLM 在捕捉细粒度知识关联上的优势。本研究最终主要采用大模型方法构建课程

知识点图谱,原因是大模型结合强大的模型算力,可自动提取细粒度知识点并构建动态关系网络,支持实时知识追踪,在零样本或者小样本场景下构建效率更高,语义理解能力更强。

3.3 案例分析

以构建某开放大学计算机应用基础、程序设计基础课程知识点图谱为例,发现对于某个学生答错的知识点 A,无课程知识点图谱的知识追踪模型可能推荐了不相关的习题,而融合了课程知识点图谱的模型则能推荐其前置知识点 B 或并列知识点 C 的习题,从而帮助学生巩固基础。基于大模型构造的课程知识点图谱结合知识跟踪方法在开放大学开展教学应用后,学生能更精准获取符合自身学习进度与需求的习题资源,学习效果较之前有所提升,学习积极性明显增强。

4 结束语

本研究提出了融合课程知识点图谱的知识追踪模型进行习题推荐的框架,对比分析了加入课程知识点图谱后的优势,同时比较了采用不同方式构建课程知识点图谱对于习题推荐性能的差异。实验结果表明,添加知识点图谱可以提升知识追踪推荐效果,基于大模型构建的知识点图谱优于传统方式构建的知识点图谱。未来研究可以进一步优化知识点图谱构建方式,融合知识点的应用场景、认知难度等。同时,还可以进一步增强对模型可解释性的研究。唯有让教师清晰理解推荐习题过程的内在逻辑与原理,深刻认识模型给出特定的习题推荐的依据,使学生和教师

理解模型为什么会推荐这些习题,这是智能教育系统落地应用的关键。

参考文献:

- [1] 诸葛斌,尹正虎,斯文学,等.基于学生知识追踪的多指标习题推荐算法[J].电信科学,2022,38(9):129-143.
- [2] LI M J, NIU L. Exercise recommendation with preferences and expectations based on ability computation[J]. Computers, Materials & Continua, 2023, 77(1): 263-284.
- [3] 金天成, 梁亮, 肖春芸, 等. 记忆与认知融合的个性化 OJ 习题推荐方法[J]. 计算机学报, 2023, 46(1): 103-124.
- [4] 郭英清, 王敏, 肖明胜. 结合深度知识追踪与矩阵补全的习题推荐方法[J]. 计算机技术与发展, 2023, 33(7): 188-195.
- [5] 金天成, 梁亮, 张伟, 等. 基于深度强化学习与程序分析的 OJ 习题推荐模型[J]. 计算机科学, 2023, 50(8): 58-67.
- [6] 张凯, 方洋洋. 知识追踪驱动的智能导学系统设计与实现[J]. 电脑知识与技术, 2024, 20(12): 31-34.
- [7] 叶俊民, 宋家琦, 张珂, 等. 一种知识图谱增强的在线评测系统习题推荐算法研究[J]. 小型微型计算机系统, 2023, 44(11): 2558-2565.
- [8] 钱浩文, 王忠群, 沈超. 考虑知识点掌握缺陷的个性化习题资源推荐[J]. 安徽工程大学学报, 2023, 38(5): 79-87.
- [9] 聂庭焜, 刘梦赤. 基于双路注意力机制的习题推荐模型研究[J]. 中文信息学报, 2023, 37(7): 152-161.
- [10] 唐燕. 个性化习题推荐算法研究[J]. 信息技术与信息化, 2023(2): 100-103.
- [11] 诸葛斌, 汪盈, 肖梦凡, 等. 基于知识追踪机的多特征融合习题推荐模型[J]. 电信科学, 2024, 40(9): 75-87.
- [12] 段建设, 崔超然, 宋广乐, 等. 基于多尺度注意力融合的知识追踪方法[J]. 南京大学学报(自然科学版), 2021, 57(4): 591-598.

【通联编辑：王力】

(上接第48页)

相信随着人工智能快速发展, AI教育应用愈发深入, AI赋能乡村教师专业发展将迎来更多新技术、新应用、新案例。

参考文献:

- [1] 方红,张天雅.数字化转型赋能乡村教师专业发展:现实阻力与实践路向:基于“AGIL”模型视角的分析[J].中国电化教育,2024(6):61-69.
- [2] 汪基德,李博,朱书慧,等.赋能、负能与使能:生成式人工智能嵌入乡村教师专业发展的多维审思[J].电化教育研究,2024,45(8):80-86,96.
- [3] 郝广龙,黄培森.乡村教师专业发展现代化的向度、困境及其超越[J].当代教育科学,2023(8):52-60.
- [4] 李慧,陈坤.价值认同与实践掣肘:乡村全科教师专业发展困境与突破[J].中小学管理,2023(6):48-52.

- [5] 杨海燕.农村小学全科教师胜任力现状及供需匹配度提升对策[J].教学与管理,2024(20):12-16.
- [6] 杨帆,陈昊璇,朱永新.人工智能助力教师专业发展:价值定位、现实制约与制度建设[J].中国远程教育,2024,44(4):58-68.
- [7] 张铭锐,闫志明,孙铭璐,等.教师知识图谱:人工智能赋能教师专业发展的必由之路[J].现代教育技术,2023,33(8):38-47.
- [8] 刘颖,向磊,杨清涵,等.AIGC赋能乡村教师专业发展:可及类型、行动逻辑与实施路径[J].现代远距离教育,2024(4):84-96.
- [9] 梁华,邵丽,周燕婷,等.高校实验室安全风险管理系统开发与应用[J].计算机时代,2025(1):72-76.
- [10] 刘石奇,刘智,段会敏,等.大模型驱动的教育多智能体系统应用研究:技术架构、发展现状、实践路径与未来展望[J].远程教育杂志,2025,43(1):33-45.

【通联编辑：王力】