

基于知识森林的多轮对话导学模型

肖鑫园, 唐九阳

引用本文

肖鑫园, 唐九阳. [基于知识森林的多轮对话导学模型](#)[J]. 计算机科学, 2025, 52(4): 249-254.

XIAO Xinyuan, TANG Jiuyang. [Multi-turn Dialogue Tutoring Model Based on Knowledge Fores](#)[J].

Computer Science, 2025, 52(4): 249-254.

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)

[k-DmeansWM:一种基于P2P网络的分布式聚类算法](#)

k-DmeansWM:An Effective Distributed Clustering Algorithm Based on P2P

计算机科学, 2010, 37(1): 39-41.

[面向异构数据集成的数据源能力描述框架](#)

计算机科学, 2004, 31(4): 170-172.

[多维可视化技术综述](#)

计算机科学, 2008, 35(11): 1-7.

[基于相关性和冗余度的联合特征选择方法](#)

Joint Feature Selection Method Based on Relevance and Redundancy

计算机科学, 2012, 39(4): 181-184.

[一种基于双层栅格划分的无线传感器网络目标定位方法](#)

Target Localization Based on Double-level Grid Division in Wireless Sensor Networks

计算机科学, 2012, 39(6): 25-29.

基于知识森林的多轮对话导学模型

肖鑫园 唐九阳

国防科技大学系统工程学院 长沙 410000

国防科技大学大数据与决策实验室 长沙 410000

(19174926443@163.com)

摘要 在智能导学场景中,传统知识图谱的智能导学算法无法表征知识主题间的认知关系,且传统推荐方法缺少对学生和知识体系的动态交互设计,这两种问题会导致学生在学习过程中面临知识迷航。基于知识森林的多轮对话导学模型从知识森林模型的角度出发,提出一种涵盖知识前后序结构、知识中心度与难度,动态可交互的导学模型,并围绕中心度与先验知识定量评估了学生在学习过程中面临的知识迷航代价。为验证算法的有效性,对随机学习策略的知识迷航问题进行量化分析,在学科数据集 GeoQSP, HisQSP, PhyQSP 和 DSAQSP 上进行了多轮对话导学模型与随机学习策略的对比实验。实验结果表明,基于知识森林的多轮对话导学模型可以较大程度缓解因缺乏知识主题间的认知关系和动态交互能力导致的知识迷航问题。

关键词: 知识森林;智能导学;知识迷航

中图分类号 TP391

Multi-turn Dialogue Tutoring Model Based on Knowledge Forest

XIAO Xinyuan and TANG Jiuyang

School of Systems Engineering, National University of Defense Technology, Changsha 410000, China

Big Data and Decision Lab, National University of Defense Technology, Changsha 410000, China

Abstract In the intelligent tutoring scenario, the intelligent tutoring algorithm of traditional knowledge graph cannot represent the cognitive relationship between knowledge topics, and the traditional recommendation method lacks the dynamic interaction design between students and the knowledge system. These two problems will lead to students being lost in the learning process. Multi-turn dialogue tutoring model based on knowledge forest is from the perspective of the knowledge forest model. This paper proposes a tutoring model that covers the sequence structure of knowledge, knowledge centrality and difficulty, and dynamic interactive. It quantitatively evaluates the cost of knowledge lost faced by students in the learning process around the centrality and prior knowledge. In order to verify the effectiveness of the algorithm, this paper quantitatively analyzes the knowledge disorientation problem of random learning strategy, and carries out comparative experiments between multi-turn dialogue tutoring model and random learning strategy on subject datasets GeoQSP, HisQSP, PhyQSP and DSAQSP. The experimental results show that the multi-turn dialogue tutoring model based on knowledge forest can greatly alleviate the problem of knowledge lost caused by the lack of cognitive relationship between knowledge topics and the lack of dynamic interaction ability.

Keywords Knowledge forest, Intelligent tutoring, Knowledge disorientation

教育是国家的根本大计。21世纪以来,信息的爆炸性增长为传统教育行业带来了颠覆性的机遇和挑战。互联网信息技术的革新,极大程度上便利了学生的在线教育,各项信息技术和知识工程技术的成熟为学生带来了大量的学习信息资源,但是自学的效率和效果难以得到保证。自学效率低下,很大一部分原因是无序化,碎片化的信息也带来了信息过载^[1]和知识迷航问题,并且学生在接触一个陌生的学习领域时也会面临缺乏知识拓扑结构带来的学习迷航问题。对于学生而言,不能很好地理解知识地图的拓扑结构,就不能将知识体系

以低迷航成本的方式归纳为掌握的知识。知识工程教育领域中场景应用的要求是精准和规范,规范化的知识是人类处理信息和理解世界的基础^[2]。但是当前现存的信息大多为半结构化和非结构化信息,它们的存在给知识梳理、实体对齐和智能导学等领域带来了挑战。

知识图谱自出现以来,逐渐在教育领域占据了主导地位,越来越多的人认识到知识图谱在丰富个性化学习、概念映射和教育内容推荐系统方面的重要作用^[3]。在信息组织领域,以领域为中心的学科知识图谱提供了为复杂知识体系进行

到稿日期:2024-07-22 返修日期:2024-09-08

基金项目:国家重点研发计划(2020AAA0108800)

This work was supported by the National Key Research and Development Program of China (2020AAA0108800).

通信作者:唐九阳(jiuyang_tang@nudt.edu.cn)

导航的可能^[4-5]。知识图谱促进了人们对学科知识体系的深入理解,也有助于教育工作者有效预测在理解知识结构和策划教学内容时面临的挑战^[6]。

个性化教育是一种逐渐兴起的教育方法,它针对每个具体的学生,根据其优势、先验知识、技能和目标需求等因素,设计出一个符合具体场景的教育方式。Qi等^[7]提出了一种基于兴趣相似度和知识关联度的高效个性化推荐算法,但其缺乏在多轮交互模式下的有效论证。Wang^[8]从利用用户信息协同构建知识图谱的角度构建了推荐系统,但是缺乏对知识拓扑关系的建模。Cheng等^[9]提出了一种基于本体的个体排序方法,但是缺乏对学生知识迷航程度的量化。Fu等^[10]提出的基于相似度的推荐方法收集与目标用户相似的用户的评价数据,以预测其偏好特征,在局部缓解知识迷航方面取得了一定的效果。

郑庆华团队提出的知识森林^[11]由于依赖关系明晰,具备良好的可编辑性和可扩展性,在智能导学领域有着良好的应用前景。相比传统知识图谱,知识森林中主题分面树,知识主题等高度结构化的知识构建方式可以缓解传统知识图谱导学面临的认知无序和学习迷航问题。

针对传统知识图谱的智能导学算法无法表征知识间的认知关系,且传统推荐方法缺少对学生和知识体系的动态交互设计这两大问题,文中提出了基于知识森林的多轮对话导学模型。该模型有效解决了上述两个问题,缓解了知识迷航,在有较强逻辑性和知识依赖性的学科场景下取得了较好的效果。

1 相关工作

个性化导学推荐技术的目的,是使学生更容易且更全面地掌握对应的学科知识体系。已有的个性化学习推荐分为3类方法:基于学习者特征、基于知识特征以及基于认知关系特征。

基于学习者特征的方法,主要对学习者的知识体系与习得状态进行建模,从满足学习者的需求出发,进行知识推荐。Lin等^[12]提出的混合决策树学习路径,使学习者获得了高于平均水平的创造力得分。Chun等^[13]提出的学习对象管理模型(LOMM)利用决策树和Apriori算法生成预测模型,对学习者的学习进行分类,以此对学习者的个体特征进行自适应优化学习。Cheng等^[14]提出了一种结合学生概念模型和变化挖掘机制的方法,从学生的历史评价数据中分析学生的学习问题,从教师实践的角度进行技术辅助。

基于知识特征的方法,更加关注知识本身的系统性与连贯性,从知识拓扑结构的角度生成学习者的学习路径。Chu等^[15]利用本体技术实现了一个概念图学习系统,该系统有利于资源之间的共享和重用,是知识地图在降低认知负荷方向上的扩展。Colace等^[16]提出了一种利用贝叶斯网络构建本体的新算法,在课程知识领域的定义、适应学习路径的生成和评估阶段进行了显著的改进。Teddy等^[17]提出的知识集成自编码器将外部知识源融入学习过程中,更加精准地捕获了知识之间的底层结构和关系,将

知识特征进行了更加精细的建模。

基于认知关系特征的方法,主要解决用户与知识学习过程之间产生的学习迷航问题,是学习者特征与知识特征的进阶方法。Zhao等^[18]提出的基于改进蚁群优化算法的微型学习路径推荐方法能够根据学习者的过渡情况组织或重组学习路径,提高学习效率。Lin等^[19]利用基于注意力的机制与高阶特征交互改进了基于网络的知识共享推荐系统,有效地区分了用户之间的学习差异。Zhang等^[20]在此基础上,提出简化的知识感知注意力网络(SKAN),通过删除两种设计来简化知识感知聚合,进一步对用户交互项目进行聚合,以增强用户表示学习。

2 基于知识森林的多轮对话导学模型

2.1 问题定义

给定学习者知识主题集合为 U , $u_i \in U$ 为每个独立的知识主题,其中 $u_{di} \in U_d$ 为学习者已经习得的先验知识(初始定义为空集);学习者学习的目标知识主题为 U_r 。基于上述定义,每一步导学的目标是根据学习者已习得的先验知识推荐出一个总学习迷航代价最小的路径。

2.2 模型框架

通过知识森林学习依赖拓扑关系的分析,本文提出了5个步骤的学习路径构建方法。如图1所示,依据用户的学习需求收集对应的先验知识集,依据知识之间的前后序关系生成最小补偿集,接下来计算每个知识主题的中心度与难度等指标,生成初步的学习序列,当用户知识状态改变时,从新的先验知识集运行新的循环,直至学习序列为空集时,运算结束,输出历史推荐序列。

1)先验知识集生成。先验知识集是学习者在学习目标知识主题之前已经掌握的知识主题的集合。可以在分析个人学习特征的基础上,根据相应学科的学习依赖拓扑图,从在线学习者的学习日志和记录中提取已经掌握的知识主题集。

2)最小补偿集生成。依据构建好的知识主题学习依赖拓扑图,按照认知依赖关系的逆序,递归地找出目标节点的先序知识闭包。在此之后,以学习者先验知识为最小补偿集的下限,学习目标知识主题的先验知识为最小补偿集的上限,生成对应的最小补偿集。

3)知识主题的中心度、难度与学习迷航度的计算。在用户学习知识时,不同的知识主题与其他的知识主题的关联紧密度不一样,我们将之定义为“中心度”。在用户学习一个新知识时,知识的不相关性会给用户的学习效率带来挑战。学习路径推荐的总目标,就是在满足必要知识遍历的条件下,使得用户的学习迷航总代价降到最小。

4)学习序列生成。在完成步骤1)–3)的基础上,从目标知识主题开始,在最小补偿集中,沿着学习依赖关系的逆序,依据知识主题难度的降序,按照宽度优先遍历的原则生成学习路径推荐。

5)依据用户反馈掌握的新的知识主题,将新掌握的知识主题加入先验知识集合 U_d 。重复步骤2)–4),直到先验知识集 U_d 最终包含目标知识主题 U_r 。

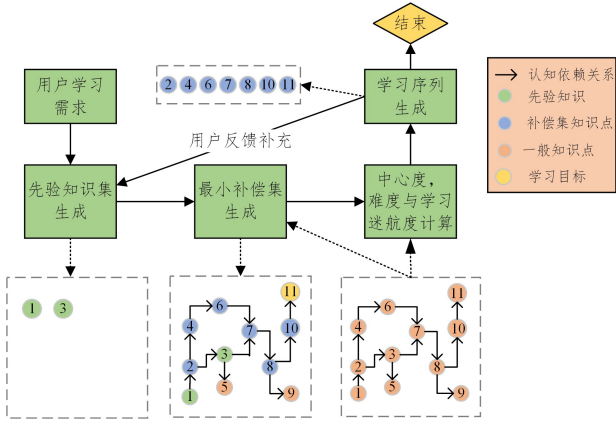


图1 基于知识森林的多轮对话导学模型

Fig.1 Multi-turn dialogue tutoring model based on knowledge forest

2.3 最小补偿集生成

设学习依赖矩阵为 $G_{[k \times k]}$, $g_{ij} = 0, 1$, 其中 0 代表知识主题 R_i 不是 R_j 的先决条件, 1 代表知识主题 R_i 是 R_j 的先决条件。知识森林的学习依赖矩阵表述了不同知识主题之间的学习依赖关系。

在定义学习依赖矩阵的基础上, 新定义学习逆依赖矩阵 $G_r = G^T$, 其中 G_r 为 G 的转置矩阵。在学习依赖逆矩阵中, 对学习依赖关系进行了逆定义, 以方便后续运算。

2.4 基于马尔可夫链的中心度计算

借鉴 Zheng 等^[11]的方法, 本文利用马尔可夫链的学习迁移概率矩阵来对其进行量化。在学习概率矩阵中, 若 p_{ij} 代表第 i 行第 j 列的元素, 则说明掌握知识主题 i 后, 掌握知识主题 j 的概率为 p_{ij} 。

定义学习迁移概率矩阵为:

$$P = \begin{bmatrix} \mathbf{R} & \mathbf{B} \\ \mathbf{O} & \mathbf{I} \end{bmatrix} \quad (1)$$

其中, \mathbf{R} 是 $|V_1| \times |V_1|$ 阶矩阵; \mathbf{B} 是 $|V_1| \times |V_2|$ 阶矩阵; \mathbf{I} 是 $|V_2| \times |V_2|$ 阶单位矩阵, \mathbf{O} 是 $|V_2| \times |V_1|$ 阶零矩阵。在式(1)定义中, 为了方便计算, 对知识主题进行了分组排序, 使得 \mathbf{O} 是 $|V_2| \times |V_1|$ 阶零矩阵, 即将 $|V_2|$ 组零概率的知识主题学习组合排列在矩阵下方的相邻位置。

依据马尔可夫链的基本特性, 知识主题 R_i 的中心度为:

$$d_i = \begin{cases} \sum_{k=1}^{|V_1|} q_{ki}, & i \leq |V_1| \\ \sum_{k=1}^{|V_1|} \sum_{m=1}^{|V_2|} b_{mi}, & |V_1| + 1 \leq i \leq |V_2| \end{cases} \quad (2)$$

其中, q_{ki} 是矩阵 $\mathbf{Q} = (\mathbf{I} - \mathbf{B})^{-1}$ 的第 k 行第 i 列元素; b_{mi} 是矩阵 \mathbf{B} 的第 m 行第 i 列元素。

2.5 知识主题的难度

传统教学场景中, 学生学习知识主题的难度会对学习体验以及学习效率产生不可忽视的影响。传统知识点难度有两种评估方式: 1) 评估该知识点的所有前序知识到该知识点的学习转移概率的最小值; 2) 评估该知识点所对应的所有前序知识到该知识点转移概率的乘积。第一种评估方式忽略了前序知识的数量对难度的影响; 第二种评估方式基于知识点的全独立性假设, 事实上, 知识点之间可能不会全随机独立, 这可能会低估学生学习知识点的概率。因此, 对知识点难度的

评估应考虑到知识点的序贯性与联系性。

对于知识主题 R_i , 定义其难度 D_i 为:

$$D_i = \begin{cases} \frac{\sum_{l=1}^{|V_1|} \exp(d_{li}) \times \ln [p(l_i)^{-1} + 1]}{\theta \times \exp(d_i)}, & |V_i| \neq 0 \\ \frac{1}{\theta}, & |V_i| = 0 \end{cases} \quad (3)$$

其中, d_i 为知识主题 R_i 的中心度; $|V_i|$ 为知识主题 R_i 的前序知识的个数; d_{li} 为知识主题 R_i 的前序知识 R_l 的中心度; θ 为归一化参数; $p(l_i)$ 为在学习迁移概率矩阵中, 掌握知识主题 l 后掌握知识 i 的概率。

定义 1 (知识主题的学习迷航度) 对于知识主题 R_i , 由式(3)可知其难度; 对于学生学习的实际过程, 学习一个知识点所带来的迷航程度是由知识点本身的难度和学生对该知识点的前序知识的掌握情况来共同决定的。

2.6 推荐学习序列生成

对知识主题的中心度和难度等概念进行定义后, 即可对每一时间步的学习序列生成推荐结果。最优推荐学习序列的目标, 是使得用户在能学习到目标知识主题的前提下, 所付出的知识迷航代价最小。

具体策略是: 从目标知识主题 U_i 开始, 在最小补偿集中, 沿着学习依赖关系的逆序, 依据知识主题难度的升序, 按宽度优先的原则生成学习路径推荐。

推荐学习序列生成算法的伪代码如算法 1 所示。

算法 1 推荐学习序列生成算法

Input: 最小补偿集 G_m , 学习依赖逆矩阵 G^T

Output: 一级推荐学习序列 RS_1

1. 选择目标知识主题 U_i 为生成点 P_G
2. for session1 do;
3. 依据生 P_G 以及 G^T , 选择出候选推荐点集 R
4. for $r \in R$, do;
5. 计算候选推荐点 r 对应的中心度 d_r
6. 计算候选推荐点 r 对应的难度 D_r
7. end for
8. 将 r 依据难度 D 的升序加入二级候选推荐序列 RS_2 中
9. f 每个候选推荐点 $r \in R$ 被遍历;
10. 将二级推荐序列 RS_2 加入一级推荐序列 RS_1 中
11. end for
12. else repeat 4-8
13. return RS_1

具体来说, 第一步是依据给定的当前先验知识的最小补偿集和学习依赖关系, 从目标知识主题出发, 依据学习依赖关系的逆序生成下一步的候选知识主题候选推荐集合 R 。在知识主题候选推荐集合 R 中, 分别对每一个候选的知识主题进行评估。计算出难度后, 依据学习难度的降序, 将每一个时间步对应的推荐序列记为 RS_2 。在遍历完整个知识主题候选推荐集合 R 之后, 经过对话会得到一个时间步上的学习路径推荐。同时, 会将本时间步所生成的推荐序列中的知识点加入到 RS_1 中。

定义知识主题 R_i 的学习迷航度 TV_i 为:

$$TV_i = \begin{cases} D_i - \sum_{l=1}^{|D_i|} D_l \times p(R_i | R_l), & |V_i| \neq 0 \\ D_i, & |V_i| = 0 \end{cases} \quad (4)$$

其中, $p(R_i | R_j)$ 代表掌握知识主题 R_j 后掌握知识主题 R_i 的概率, D_i 代表知识主题 R_i 的难度。

在定义完知识主题的学习迷航度之后, 可以对学生在每一个学习步(每一个习得的先验知识和知识森林的知识依赖关系进行交互的实时状态)中的学习状态给予客观的评估。但是, 要完成对整个学习过程的评估, 需要对全学习过程的迷航程度进行定义。

定义知识迷航度总和 KT 为:

$$KT = \sum_{i=1}^{|V_e|} TV_i \quad (5)$$

其中, $|V_e|$ 为每一个已执行的学习知识主题的集合; TV_i 为知识主题 R_i 的学习迷航度。

至此, 可以对一个学生在知识森林内学习不同知识主题的过程进行客观的建模, 同时也可以推荐出一个使得学习效率最高即知识迷航代价最小的学习策略。

为了将基于知识森林的多轮对话导学模型与没有知识森林做逻辑支撑的学习场景作对比, 同样对其进行建模, 将其定义为随机学习策略。3.4 节将通过实验将随机学习策略与基于知识森林的多轮对话导学模型进行对比。

2.7 随机学习策略

在用户进行学习的过程中, 有两种学习方式: 确定性学习和随机性学习。在上文介绍的多轮对话导学模型(确定性学习)中, 依据每一时间步的用户反馈输入、模型的初始输入(先验知识) U_d 以及对应学科的学习依赖矩阵 G , 可以确定地推算出下一步的学习策略。接下来, 将通过随机学习过程进行建模, 使用模式化、规律化的数学表达对其进行展示。

首先, 用户在每一个时间步有两种选择: 进行一步随机跳跃学习或者进行一步随机继承学习。

2.7.1 随机跳跃学习

随机跳跃学习可以理解为用户不满足现有的先验知识框架, 尝试在框架之外进行突破性的学习。学习时, 在集合 $U - U_d$ 中, 通过等概率抽样选取一个知识主题 R_i 进行学习, 并将学习到的知识主题添加到先验知识集中。

2.7.2 随机继承学习

随机继承学习可以理解为用户对现有知识体系有一定了解, 因此选择对先验知识节点的后继节点依据一定的随机规律进行继承。知识贡献度 T_i 可以很好地衡量未知序关系和全局关系条件下知识点的重要性, 因此, 学习可以依据知识贡献度 T_i 的值大小来进行一个有权重的随机抽样, 权重与知识贡献度成正比。

定义 2(知识贡献度 T_i) 在知识森林中, 各个知识主题之间相互影响, 但是此类影响力会随着跳数的增加而衰减。其计算方法如式(6)所示:

$$\begin{aligned} F_{suc}(i) &= \sum_{j \in pre(suc(i))} d(j) \\ T(k) &= F_{suc_1}(i) + \frac{1}{2} F_{suc_2}(i) \\ S(j) &= \sum_{k \in suc(j) + pre(j)} T(k) \\ W(i) &= \sum_{i \in suc(i) + pre(i)} S(j) \end{aligned} \quad (6)$$

其中, $pre(i)$ 代表入度, 即知识点实体 i 的一跳前向知识主题, 它衡量的是实体 i 的受益程度; $suc(i)$ 代表出度, 即知识点实体 i 的一跳后向知识主题, 它衡量的是实体 i 的贡献程度;

F_{suc_1} 代表一阶知识贡献度, F_{suc_2} 代表二阶知识贡献度, $T(k)$ 为一阶知识和二阶知识贡献度的加权和, 即知识主题 k 的知识贡献度。

定义知识掌握度 LD 为:

$$LD = \frac{\sum_{k=1}^{k \in U_d} d_k}{\sum_{r=1}^{r \in U} d_r} \quad (7)$$

LD 为已获得先验知识的中心度之和与学科总知识中心度之和的比值, 反映了学生对整个知识森林体系的掌握程度。

2.7.3 选择策略

选择策略即每一时间步进行随机跳跃学习和随机继承学习的概率。它基于学生对知识森林体系的掌握程度来确定。在某一时间步中, 若进行随机跳跃学习的概率为 P_x , 随机继承学习的概率为 P_{1-x} , 则有:

$$P_x + P_{1-x} = 1 \quad (8)$$

除此之外, 学生学习时, 随机跳跃学习的概率会随着掌握程度变化。由于学习增量对于知识森林的习得效益边际递减, 因此将刻画变化的函数定义为选择策略函数 $P_x = f(\beta)$ 。

$$P_x = \ln \left(\frac{1 + \frac{1}{e^\beta - 1}}{LD + \frac{1}{e^\beta - 1}} \right) \quad (9)$$

其中, x 为知识掌握度; β 为跳跃参数, 即初始时刻进行随机跳跃学习的概率, 其值的大小与用户的学习习惯有关。

至此, 已经完成了随机学习策略的定义。下一章将在学科数据集上将随机学习策略与多轮对话导学模型进行对比, 以验证多轮对话导学模型的有效性。

3 实验验证及分析

本文选取了自制的 GeoQSP 数据集、HisQSP 数据集、PhyQSP 数据集和 DSAQSP 数据集进行实验; 涵盖的学科知识森林有高中地理、高中历史、高中物理及数据结构与算法等。共进行了 3 组实验: 第一组采取固定先验知识的模式, 对不同跳跃参数 β 下的随机学习和本文所提方法($\beta=0$)进行对比分析; 第二组采用随机先验知识集的模式, 对不同跳跃参数 β 下的两种方法进行分析; 第三组实验对采取和不采取多轮对话策略下每一时间步的平均知识迷航度进行对比分析。为了方便计算, 知识迷航度总和 KT 采用了 \min - \max 进行归一化, 被转化为 KT_r (归一化知识迷航度)。

第一组实验对 GeoQSP, HisQSP, PhyQSP, DSAQSP 这 4 个学科对应数据集下的 KT_r (归一化知识迷航度)进行分析。

如图 2 所示, PhyQSP 数据集下, 随机跳跃学习的知识迷航度总和和对跳跃参数 β 比较敏感, 说明在缺乏学习顺序的拓扑关系的条件下, 学习习惯对学生学习物理科目的影响较大。GeoQSP 和 HisQSP 数据集下, 随机学习策略对应的知识迷航度总和较为集中, 说明地理和历史科目对于学习策略的随机性不太敏感。除此之外, 在 PhyQSP 组中, 参数 $\beta=0.8, \beta=0.75, \beta=0.7$ 对应的知识迷航度总和分布差异较大。从学习代价的角度来看, $\beta=0.8 > \beta=0.75 > \beta=0.7$ 。这说明对于物理这种强逻辑性的科目而言, 是否采用随机跳跃的方式来学习对学习的迷航度和体系性影响较大。在 4 组数据中, 本文

提出的多轮对话学习模型($\beta=0$)组的值分别为 0.3145,0.341,0.3152,0.287,均显著低于不同种的随机学习策略,从而验证

了模型对于减少学生学习时面临的知识迷航问题的有效性,其平均优于随机学习策略一倍以上。

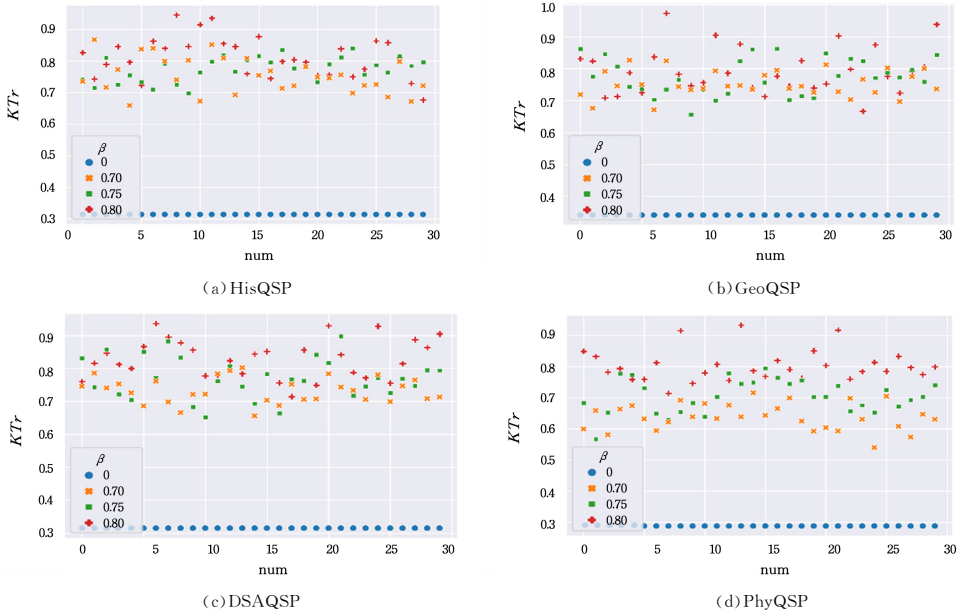


图 2 固定先验知识下导学效果的评估

Fig. 2 Evaluation of tutoring effect with fixed prior knowledge

第二组实验在 GeoQSP, HisQSP, PhyQSP, DSAQSP 这 4 个学科数据集下进行了先验知识随机的归一化知识迷航度对比。同样地,把跳跃参数 β 设置为 0.8,0.75,0.7。在固定学习方式与跳跃参数 β 的条件下,进行了 10 次采样和模拟运算。实验结果如图 3 所示。

均值分布差异较为明显,模型应用效果较好。

第三组实验在 GeoQSP 数据集上,对采取和不采取多轮对话策略下每一时间步的平均知识迷航度进行对比分析。其中,由于现实中可能存在用户依据个人需求对个别知识主题进行无条件优先学习的情形,因此在实验中对推荐的学习序列进行一定概率的随机学习。不采取多轮对话策略组将始终依据第一个时间步推荐的学习序列进行学习;而采取多轮对话策略组在用户进行顺序微调后,根据微调后的认知状态重新进行评估推荐。假设用户对不同单个知识主题顺序微调的概率相互独立,实验结果如图 4 所示。

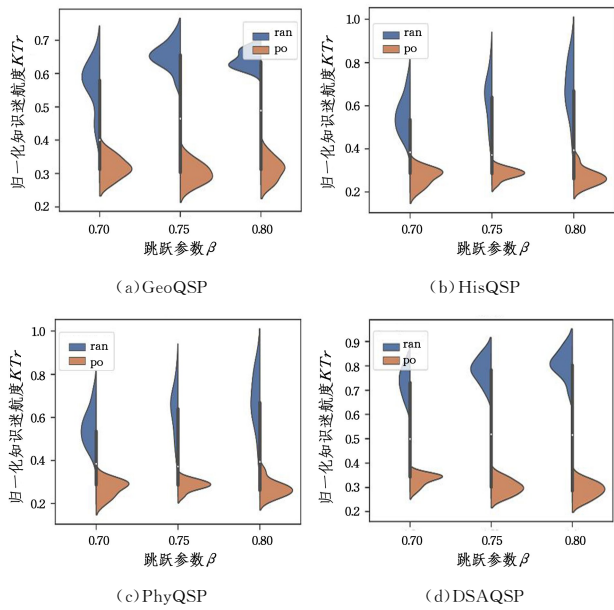


图 3 随机先验知识下导学效果的评估

Fig. 3 Evaluation of tutoring effect with random prior knowledge

从图 3 可以看出,GeoQSP 和 DSAQSP 组对于学习方式是否采用随机学习策略的结果差别较大。采用随机学习策略的知识迷航度均值高于多轮对话学习模型的均值一倍以上。PhyQSP 组虽然表现有明显分层,但是无论是否采用随机学习策略,整体的知识迷航度总和均偏高。相比之下,DSAQSP 组采用随机学习策略的组间均值与采用多轮对话导学模型的

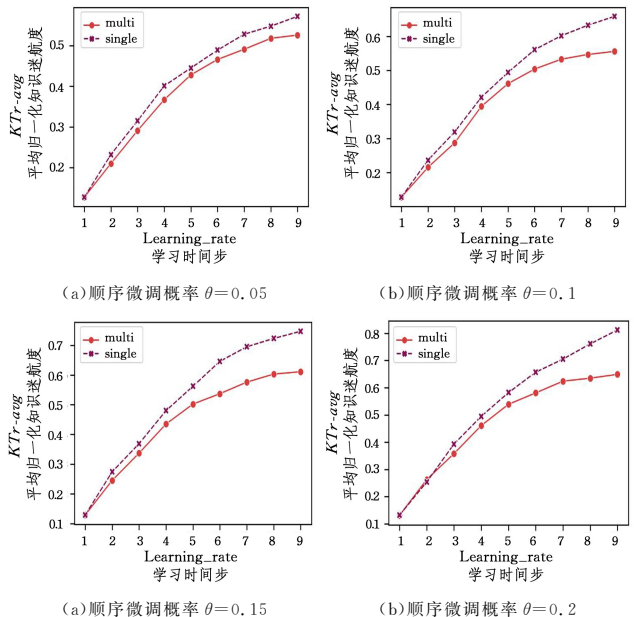


图 4 多轮对话策略的有效性验证

Fig. 4 Validation of multi-turn dialogue strategies

分析实验结果可知,当用户的顺序微调概率越大时,单轮

策略和多轮对话策略平均的归一化知识迷航度也越大,且是否使用多轮对话策略的差异会更加明显。除此之外,在学习时间步较小时,两种方法的差异不大,但是随着学习时间步的增大,采取多轮对话策略会使得知识迷航程度迅速趋于平缓,而单轮策略会削减这种平缓趋势。越到偏后的时间步,多轮对话策略为缓解知识迷航带来的收益越高。

结束语 针对现有的导学推荐模型缺乏对知识体系全局前后序关系信息利用的问题,本文提出了基于教育领域知识森林的多轮对话导学模型。知识森林的高度结构化存储,可以为用户的导航学习提供依据。在 GeoQSP, HisQSP, PhyQSP 和 DSAQSP 数据集上,多轮对话导学模型的知识迷航代价总和显著小于采取随机学习策略的知识迷航代价。除此之外,在先验知识集随机的条件下,多轮对话导学模型的知识迷航度总和均有比较集中的分布域,说明模型具有对随机先验知识的鲁棒性。但是,不同学科之间的知识结构特性也有差异,对于知识结构强联系性和弱依赖性的学科,如何更好地应用本模型,需要后续进行探讨。

参考文献

- [1] TARATUKHINA Y V, BART T V, VLASOV V V. Machine learning models of information recommendation system on individualization of education[J]. *Educational Resources and Technology*, 2019, 2(2): 7-14.
- [2] CHEN J Y, HU L Y, WU F. ChatGPT/Generative artificial intelligence promotes the research of teaching mas the core[J]. *Journal of East China Normal University* 2023, 41(7): 177-186.
- [3] ABU-SALIH B, ALOTAIBI S. A systematic literature review of knowledge graph construction and application in education[J]. *Hiyon*, 2024, 10(3): e25383.
- [4] ABU-SALIH B. Domain-specific knowledge graphs: A survey [J]. *Journal of Network & Computer Applications*, 2021, 185: 103076.
- [5] TA CHIEN D C, KHAI T T. Constructing a subject-based ontology through the utilization of a semantic knowledge graph [J]. *International Journal of Information Technology (Singapore)*, 2024, 16(2): 1063-1071.
- [6] CHENG Y, BU X. Research on key technologies of personalized education resource recommendation system based on big data environment[J]. *Journal of Physics: Conference Series*, 2020, 1437(1): 012024.
- [7] WEI Q, YAO X L. Personalized Recommendation of Learning Resources Based on Knowledge Graph[C]// *International Conference on Educational and Information Technology (ICEIT)*. 2022.
- [8] WANG H. Personalized recommendation system based on network feature learning[D]. Shanghai: Shanghai Jiao Tong University, 2018.
- [9] CHENG B Y, ZHANG Y, SHI D X. Ontology-based personalized learning path recommendation for course learning[C]// *International Conference on Information Technology in Medicine and Education (ITME)*. 2018.
- [10] FU R, TIAN M J, TANG Q J. The Design of Personalized Ed-

ucation Resource Recommendation System under Big Data[J]. *Computational Intelligence & Neuroscience*, 2022, 2022: 1-11.

- [11] ZHENG Q, LIU J, WANG B, et al. Knowledge Forest: Theory, Method and Practice[M]. Beijing: Science Press, 2021: 63-64.
- [12] LIN C F, YEH Y C, HUNG Y H. Data mining for providing a personalized learning path in creativity: An application of decision trees[J]. *Computers & Education*, 2013, 68(1): 199-210.
- [13] WANAPU S, FUNG C C, KERD-PRASOP N, et al. An investigation on the correlation of learner styles and learning objects characteristics in a proposed Learning Objects Management Model (LOMM) [J]. *Education and Information Technologies*, 2016, 21(5): 1113-1134.
- [14] CHENG L C, CHU H C, CHENG L C, et al. An innovative approach for assisting teachers in improving instructional strategies via analyzing historical assessment data of students[J]. *International Journal of Distance Education Technologies*, 2015, 13(4): 40.
- [15] CHU K K, LEE C I, TSAI R S. Ontology technology to assist learners' navigation in the concept map learning system (Article) [J]. *Expert Systems with Applications*, 2011, 38(9): 11293-11299.
- [16] COLACE F, DE SANTO M. Ontology for e-learning: a Bayesian approach [J]. *IEEE Transactions on Education*, 2010, 53(2): 223-233.
- [17] LAZEBNIK T, SIMON-KEREN L. Knowledge-integrated autoencoder model [J]. *Expert Systems with Applications*, 2024, 252: 124108.
- [18] ZHAO Q, ZHANG Y, CHEN J. An improved ant colony optimization algorithm for recommendation of micro-learning path [C]// *2016 IEEE International Conference on Computer and Information Technology (CIT)*. 2016.
- [19] LIN J, SUN G, SHEN J, et al. Attention-based high-order feature interactions to enhance the recommender system for web-based knowledge-sharing service [C]// *Web Information Systems Engineering - WISE 2020*. 2020.
- [20] ZHANG H, CHEN Y, LI X, et al. Simplifying knowledge-aware aggregation for knowledge graph collaborative filtering [C]// *19th International Conference on Web Information Systems and Applications (WISA 2022)*. 2022.



XIAO Xinyuan, born in 2001, postgraduate, is a member of CCF (No. O7524G). His main research interests include natural language processing and intelligent question answering.



TANG Jiuyang, born in 1978, Ph.D, professor. His main research interests include smart analytics, big data and social computing.