

基于知识图谱的课程知识体系构建到学习者画像生成及应用技术研究

邓莉琼¹, 任媛¹, 王杰宇¹, 田辰晨²

¹空军通信士官学校, 辽宁 大连

²大连财经学院, 辽宁 大连

收稿日期: 2021年10月1日; 录用日期: 2021年10月29日; 发布日期: 2021年11月5日

摘要

信息化教学飞速发展, 如何将人工智能技术应用于教育领域中, 强化信息化教学手段是急需解决的问题。本文面向教育领域, 首先提出了一种基于知识图谱的从课程知识体系建立、学习者画像生成到自适应学习推荐的完整技术框架, 并以计算机网络等课程为例对知识图谱的构造技术进行了示例; 其次, 在自适应学习推荐中利用RippleNet算法实现了学习推荐, 并通过实验验证了该算法具有较高的准确度; 最后, 总结归纳了教育知识图谱的主要应用方法, 对教育知识图谱的构建及应用具有一定的指导意义。

关键词

知识图谱, 教育, 学习者画像, 自适应学习

From the Construction of Knowledge System to the Generation of Learner Portrait Based on Knowledge Graph and Research on Its Application Technology

Liqiong Deng¹, Yuan Ren¹, Jieyu Wang¹, Chenchen Tian²

¹Air Force Communication NCO Academy, Dalian Liaoning

²Dalian Financial and Economic College, Dalian Liaoning

Received: Oct. 1st, 2021; accepted: Oct. 29th, 2021; published: Nov. 5th, 2021

Abstract

With the rapid development of information teaching, how to apply artificial intelligence technology to the field of education and strengthen information teaching methods are urgent problems to

文章引用: 邓莉琼, 任媛, 王杰宇, 田辰晨. 基于知识图谱的课程知识体系构建到学习者画像生成及应用技术研究[J]. 计算机科学与应用, 2021, 11(11): 2591-2598. DOI: 10.12677/csa.2021.1111262

be solved. Aiming at the field of education, this paper first proposes a complete technical framework based on knowledge graph from the establishment of knowledge system, the generation of learner portraits to adaptive learning recommendation, and takes courses such as computer networks as examples to show the construction technology of knowledge graph. Secondly, the RippleNet algorithm is used to realize the learning recommendation in the adaptive learning recommendation, and the experiment verifies that the algorithm has high accuracy. Finally, the application methods of educational knowledge map are summarized, which has certain guiding significance for the construction and application of educational knowledge graph.

Keywords

Knowledge Graph, Education, Learner Portrait, Adaptive Learning

Copyright © 2021 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

未来将是知识互联的阶段，利用知识图谱能够实现对各领域数据中知识语义的解释和表达。在教学领域中，每门课程的信息中不仅包含其自身的语义知识点，还包含各类相关的信息化教学资源，诸如网络课程、虚拟仿真、微课视频等。如何帮助学生明确学习需求，在众多的知识点中规划合理的学习路径，提高对教学资源的利用率，解决信息化教学资源的自适应问题具有重要的研究价值。

教育知识图谱是近些年将知识图谱技术应用于教育信息化的新研究方向。它指的是教育领域的知识域可视化，实际是对教学领域之中一系列概念及其相互之间关系的形式化表达。基于知识图谱的教育把目光转移到知识点上，从大量无序的信息资源中重构知识之间的联接，有效组织各学科的知识体系而试图通过更底层的知识点掌握情况，帮助理清知识脉络。

然而目前教育知识图谱技术的研究主要还集中在知识实体的构建上，对于教育知识图谱从构建再到最后的应用尚未有完整的技术框架描述；在教育知识图谱的自适应推荐中，目前的推荐算法都难于发现学习者潜在学习兴趣；在教育应用中还停留于技术层面，没有对如何应用进行具体描述。本文针对以上问题，首先提出了教育知识图谱构建及应用的完整技术框架，基于 TextRank 算法实现对课程的知识体系构建，然后利用学习标签挖掘方法实现学习者画像生成，最后基于 RippleNet 算法实现教育图谱中的自适应学习推荐，并描述了知识图谱技术在实现个性化学习中的具体应用方法。

2. 相关工作

知识图谱表达了实体、概念及其之间的语义关系，是语义知识表示的方式之一，本质上可以理解为是实现机器语言认知的语义网络知识库。它不仅是知识表示，背后更蕴含了知识。现有的大规模知识图谱，国外主要有 Yago、Dbpedia、Freebase、谷歌知识图谱和 Nell。这些知识图谱中包含的事实数量均是亿级。国内主要有百度知心、搜狗知立方、阿里巴巴知识库等[1]。

本文中所指的教育知识图谱是指采用“节点-线”的图形式，对教育领域的课程数据和学习数据进行语义化组织和表示的方式。以这种结构化的图数据存储方式来表示教育领域知识，能够更加便于计算机“理解”。目前，教育知识图谱主要分为静态知识图谱和动态事理图谱两类。静态知识图谱是以教学资源中涉及的关键元素信息为实体、以教学元素间的关系为边，形成的知识图谱网络；动态事理图谱是

以教学过程中的事件或活动为实体，以事件间的关系为边构成的关系网络。相关的国内外研究有：国内的伟东云教育推出了基于知识图谱的智能化知识导航工具(寻知图[2])；可汗学院将知识图谱作为学科知识可视化的表示工具；Wolfram Research 公司构建的知识搜索引擎采用知识图谱作为底层支撑技术；全球最大的适应性学习平台 Knewton 采用知识图谱作为平台的知识组织工具[3]。

基于知识图谱的学习者画像也是近年来知识图谱在教育领域的应用之一[4]。学习者画像技术通过基于知识图谱叠加学习者的知识掌握状态信息，能够形成学习者的认知图式。传统的学习者画像技术仍然处于基于“打标签”方法，给用户打一组标签，用权重代表用户在各个标签兴趣的强烈程度，存在不完整和不够正确的问题。基于知识图谱的学习者画像可以实现标签的扩展，生成更加饱满的画像。许多研究者从不同角度、不同类型的数据中提出构建基于知识图谱的学习者画像生成方法。例如，加拿大阿萨巴萨卡大学的学习分析研究专家 Siemens 以及哥伦比亚大学的教育数据挖掘专家 Baker，共同提出了构建个性化学习图谱(Personal Learning Graph)的构想，进而构建了基于个性化学习图谱的学习者数字画像[5]。知识图谱不仅可以作为学习者画像的构建技术，还可以利用基于知识图谱的学习者画像进行学习路径的规划。在基于知识图谱的学习者画像技术中，根据学习者的学习需求及其认知风格等个性化特征，在知识图谱中动态规划学习路径，并推荐有针对性的学习资源。这种技术可以根据教学的逻辑顺序和知识的认知规律，依据知识图谱实现了个性化的自适应学习，在一定程度上解决了个性化学习推荐缺乏针对性的弱点。这项技术还处于初期研究阶段，存在用户数据稀疏(Sparsity)，以及冷启动(Cold-start)上网问题，需要更好地开展理论和应用研究。

3. 教育知识图谱构建及应用的技术框架

基于知识图谱的教育知识图谱构建及应用的技术框架如图 1 所示。主要包括数据采集、图谱构建和个性化学习 3 部分。

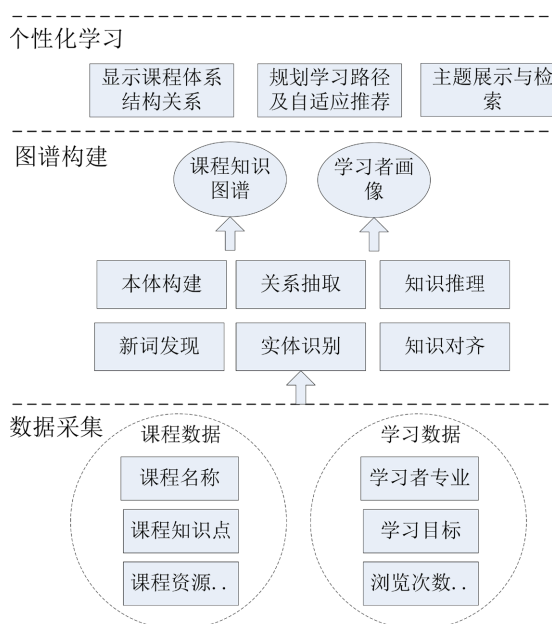


Figure 1. The technical framework for the construction and application of educational knowledge map

图 1. 教育知识图谱构建及应用的技术框架

1) 数据采集主要包含了课程数据和学习数据两大类，其中课程数据指的是诸如课程名称、课程相关资源以及知识点组成等与课程相关的数据。而学习数据指的是学习者在学习中产生的数据，例如专业、

学习目标以及学习行为中产生的浏览次数、答题训练情况等数据。这些数据中既有结构化数据、半结构化数据,也存在非结构化数据。

2) 图谱构建包含了面向课程的课程知识体系图谱以及面向学习者的学习者画像两种知识图谱的构建。知识图谱构建过程是,首先利用新词发现、短语提取等技术进行本体构建,从而形成本体库;然后通过实体识别算法识别实体信息,并利用关系抽取方法抽取实体间关系信息,从而形成资源描述框架(RDF)三元组;最后利用知识对齐与知识推理等更新算法进行知识加工,形成最终的知识库。

3) 个性化学习可以在教育教学领域为学习者提供多种知识服务手段。根据课程知识体系图谱为课程知识表示及知识间关系提供立体的直观展示;根据学习者画像可以为学习者规划针对性的学习路径和自适应的推荐;利用知识图谱的语言可以实现知识中的主题提取展示及检索等服务。

4. 基于知识图谱的课程知识体系构建

基于知识图谱的课程知识体系主要关注包括课程和知识在内的教育实体及其之间的连接关系。课程知识体系本质上是语义网络,是一种基于图的数据结构,用图中的节点表示知识点,边表示知识点与课程之间的关系。基于知识图谱的课程知识体系就是把所有不同课程的知识点连接在一起而得到的一个关系网络,可以帮助教育者从知识点之间的“关系”的角度去分析课程之间关系的能力。

整个构建过程主要包括知识本体构建、实体识别、关系提取、知识对齐及融合几个阶段。知识本体构建是采用本体构建方法描述知识体系的框架;实体识别通过信息抽取算法识别课程知识体系中的实体对象;实体关系提取在实体识别基础上,进一步判断实体之间存在的语义关联关系;知识融合和对齐是对来自多个数据源的实体信息进行融合的过程,消除了知识抽取过程中存在的歧义信息[6]。

课程知识体系中的本体是知识点,因而首先要从课程内容中找到与知识点相关的关键词语从而构建知识点。挖掘每门课程内容中的关键词语是构建课程知识图谱的过程中的第一步,可以从课程简介或者课程体系介绍等文本内容来识别关键字。提取“知识点”本体信息的常用算法有统计机器学习的方法和基于深度学习的方法。基于统计机器学习算法的典型代表是 TextRank 算法[7],通过提取文本中的关键字特征作为关键知识点,主要依赖特征工程;基于深度学习的代表是神经网络算法,其最大优势在于领域特征的自动学习,一定程度上改善了识别效果,但基于深度学习的方法仍处于探索期。由于本文建立的是一个规模较小、范围有限的教育行业知识图谱,实体识别的对象是教学文本数据。这类教学文本资料往往结构统一,大都包含相似的词汇和句型,特征明显且易于提取,因此本文采用基于 TextRank 算法足以满足实体识别的需求。

TextRank 算法的主要思想是建立基于词之间邻接网络,并使用 PageRank 计算每个节点的排序值,算法选择排序值数值较大的单词作为关键词。首先将课程文档分成完整的句子 $[S_1, S_2, \dots, S_i]$,对句子 S_i 进行分词和词性标注;分割后从句子中过滤停止词,留下带有指定词性的单词,停止词包含一些常见但无意义的词,如“学时”“课堂”等;将 S_i 分成一组词语 $[t_{i,1}, t_{i,2}, \dots, t_{i,n}]$, $t_{i,n}$ 表示句子中的第 n 个候选词语;算法根据这些词语构建候选关键字网络 $G = [V, E]$,每个候选词语 $t_{i,n}$ 对应一个节点, V 是所有节点的集合, E 则是由代表节点之间共现关系的边组成的集合。共现关系是指一对节点对应的两个词在长度为 K 的文本窗口内共现。根据下列公式1迭代计算各节点的 $Rank(V_i)$ 直到收敛,再选择其中数值较大者作为关键词。

$$Rank(V_i) = (1-d) + d \times \sum_{V_j \in In(V_i)} Rank(V_j) / |Out(V_j)| \quad (1)$$

统计完之后,将提取得到的课程及知识点等关键词作为图谱中的实体,实体的属性包括内容、难度值和学习资源链接等。实体与实体之间的关系可以采用依存句法分析的模板分析方法来实现,由于语言

表达存在句法结构,比如主谓宾、动宾结构、主从句结构等,根据这些结构可以分析出多个词之间关系,比如以动词为起点,构建规则,可对节点的词性与边上的词与词之间依存关系进行限。每匹配一条规则时就生成一个三元组,根据句子语法规则及扩展规则最终得到各个实体之间的关系,例如得到课程包含知识点的关系,并可将知识点的重要程度或难度值作为关系的属性值。最终将得到的所有三元组数据以图数据的方式存储起来[8]。本文将这些三元组实体及实体关系存储为 CSV 格式数据或 SQL 数据库中,可通过 Python 编程进行读取并转换为 Neo4j 的图数据库。本文以《计算机网络技术基础》课程及其相关的两门课程为例,应用知识图谱工具将课程内容数据重构,构建的课程知识体系如图 2 所示,图中黄色实体表示课程、绿色实体表示知识点、红色实体表示参考教材等教学资源。通过将课程、知识点及参考教材通过关系网络的方式连接在一起,显示了课程的知识体系及与其他课程之间的关系。

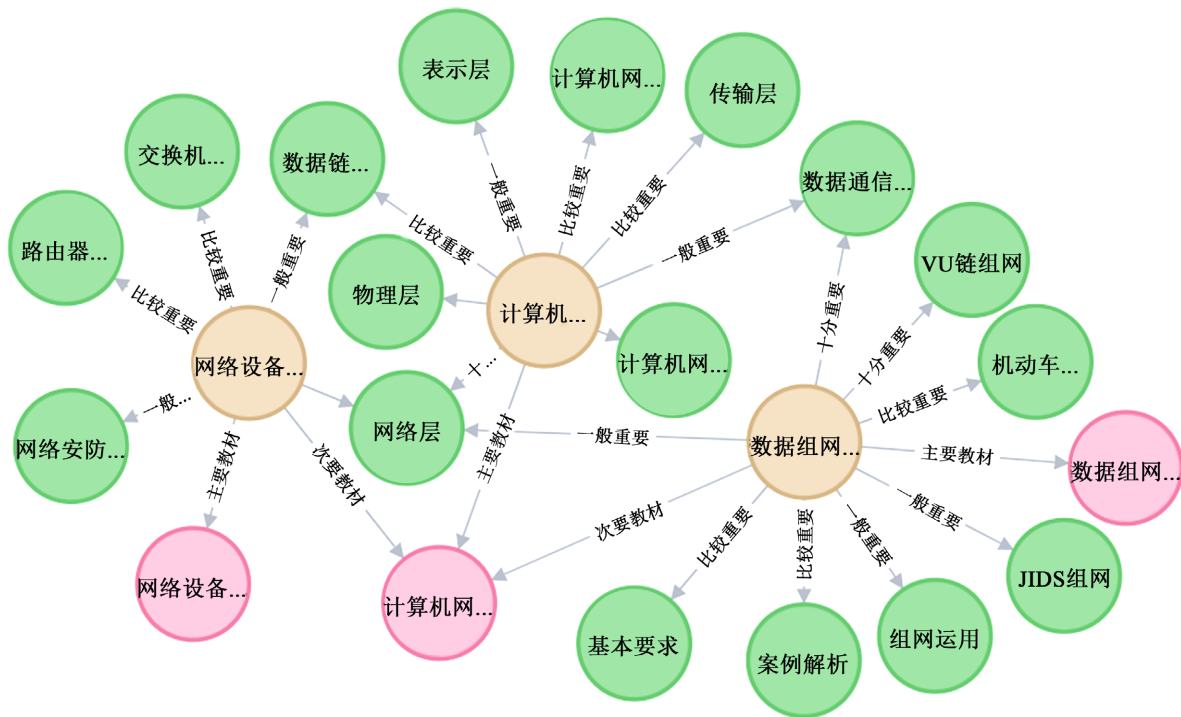


Figure 2. Knowledge system of computer network course
图 2. 计算机网络课程知识体系

5. 基于知识图谱的学习者画像生成技术及自适应学习推荐

学习者画像技术来源于用户画像技术,只不过用户变成了学习者。基于知识图谱的学习者画像是根据学习者的行为和对知识的掌握程度来构建,对学习者数据进行数学建模,将不同数据抽象出的一个标签化的学习者模型。这种技术的核心工作即是给学习者打“标签”,而标签是通过对学生信息分析而来的高度精炼的特征标识,从而对学生进行分类和抽取。基于知识图谱可以实现学习者的标签扩展,一个例子就是 A 的学习目标是计算机网络基础知识, B 的学习目标也是这个知识点,那通过计算机网络基础知识这个实体,将 A 和 B 关联到了一起,那么 A 身上其他的标签也能传播到 B,从而泛化了 B 的标签[9]。

5.1. 基于知识图谱的学习者画像生成

基于知识图谱的学习者画像生成和前面的课程知识体系构建过程类似,只不过数据源和识别的实体

类别发生了变化。学习者数据主要包括学习者本身属性、学习目标、搜索词、在线答题等行为数据。首先对所有原始数据进行字段筛选,并进行初步预处理,获得直接反映学习者属性和行为的字段;然后,基于这些数据构建面向学习者画像的知识图谱。在标签的语义泛化阶段,首先对特定的需求信息进行解析,得到与该需求信息的语义接近的 TopN 词语表示。比如,对于“喜欢动手实践的学习者”需求信息,可以泛化出“实操”、“训练”等相近的词语表示;接着利用知识图谱提供的实体与实体之间的语义相似性和逻辑相关性,通过在词向量空间中计算相关词语之间的欧氏距离,对所有语义泛化后的词语寻找语义相近的实体,基于欧氏距离的计算可以得到词语与实体之间的相关性;最后,得到语义有关的知识实体之后,同样利用向量的相似性计算得到相关实体与已知用户行为标签语义相近的标签表。依据组合计算得到的相似值大小,便可以得到与标签对应用户的相关性的强弱,从而生成可以表示用户特性的用户行为标签关联组合。

5.2. 基于学习者画像的自适应学习

自适应学习是在学习者画像的基础上,通过知识图谱来规划不同的学习路径,为学习者提供个性化难度和个性化节奏的课程和习题,从而提高学习者的学习效率和学习效果。目前将知识图谱融合到推荐系统的算法主要分为两种类别:基于嵌入的方法,通过已知的节点及结构等信息,预测尚未产生连接的两个节点之间产生连接的可能性;基于路径的方法,通过关系路径、图表的类型提取基于元路径的潜在特征来表示节点之间的连接。这两类方法在教育知识图谱中都很难做到发现学习者潜在感兴趣的点,推荐效果有限。因此本文将 RippleNet 算法[10]应用于自适应推荐的学习中,以学习者感兴趣的知识点为种子,在教育知识图谱上,向外一圈一圈的扩散到其他的知识点上,从而发现学习者到潜在感兴趣知识点的可能路径。主要实现过程如下:

- 1) 构建知识点种子集合:以学习者的历史搜索或学习过的知识点作为种子集合;
- 2) 获取第一次 Ripple 表示:构建与学习者相关的第一次偏好传播的 Ripple set,用 (h,r,t) 表示,根据嵌入向量内积,计算知识点 v 与第一层 Ripple set 上的 (h,r) 的归一化相似度,根据归一化后相似度,对第一层 Ripple set 的 (t) 进行加权求和,得到的结果作为这一层的输出。
- 3) 重复第 2 步的扩散过程,将第一层的 Ripple set 的输出作为第二层的输入,先取出第二层的 Ripple set,然后用第二层的 Ripple set 跟 item 的相似度及加权表示作为输出。如此不断往外重复扩散过程。
- 4) 最终的预测,得到了上面多次扩散得到的输出后,把多次的 Ripple 输出累加作为最终某知识点作为学习者潜在感兴趣目标的概率。

下表是以计算机网络课程的教育知识图谱为数据源,比较了不同自适应推荐算法的准确性结果,实验表明如表 1 所示, Ripple 算法在教育知识图谱的推荐中具有较好的推荐效果。

Table 1. Comparison of recommendation algorithms on accuracy of educational knowledge map

表 1. 推荐算法在教育知识图谱准确度上的比较

模型	Ripple	Wide/Deep	SHINE	CKEDKN	DKN
准确度	0.919	0.902	0.781	0.799	0.663

基于知识图谱的学习者画像及自适应学习技术可以为学习者提供个性化的教学引导服务,同时能获得大量学习者的反馈数据,对成绩的提升有不断加强的正面效果。并且在此基础上开发的各类信息化教学资源可以自动根据学习者在学习活动中出现的问题动态调整学习路径,如图 3 所示。从而使学习系统具备自适应性,让“因材施教”得以能够真正实现[11]。

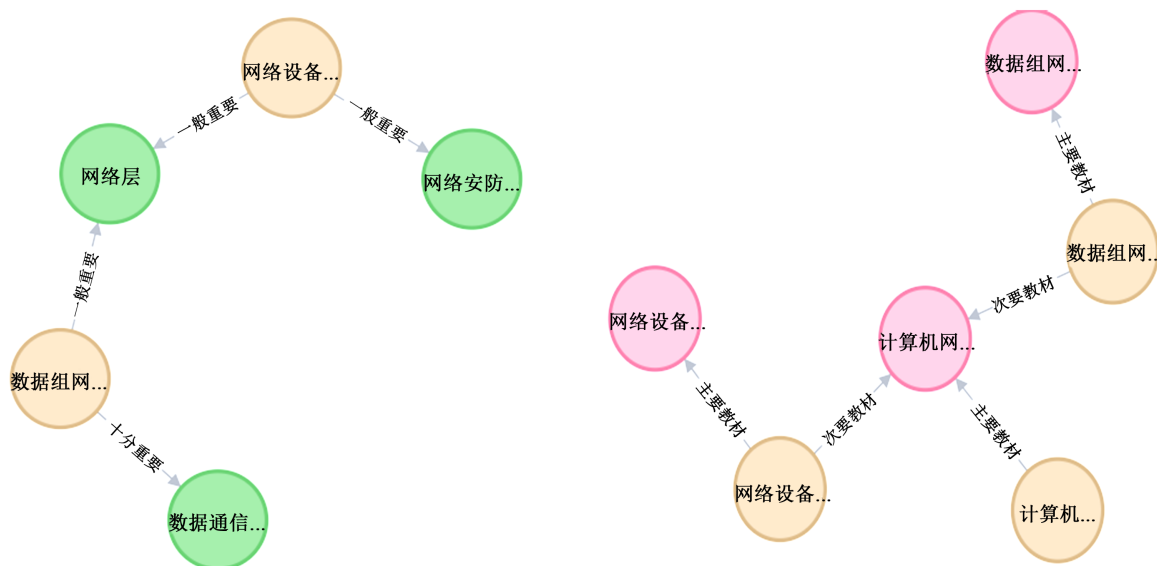


Figure 3. Learning path planning example

图 3. 学习路径规划示例

6. 教育知识图谱应用方法归纳

通过针对具体课程的教育知识图谱进行知识体系构建、学习画像生成再到自适应学习推荐，本文将知识图谱技术在教育中应用方法归纳如下表 2 所示：

Table 2. Application methods of educational knowledge map

表 2. 教育知识图谱应用方法归纳

应用领域	应用方法
知识体系可视化	构建以课程、知识点等为实体的体系架构，以图数据库的方式显示各课程之间的关联性，分析课程的相似度
教育资源聚合	针对不同的教育相关资源，基于知识图谱中的语义关联方法进行管理和融合
学习标签构建	基于知识图片的学习者画像技术，通过收集及分析所有学习者相关行为数据，泛化了学习者标签，实现精确定位
适应性学习推荐	将知识体系知识图谱与学习者画像结合起来，推荐针对性的个性化学习资源和学习路径

7. 结论

基于知识图谱的知识结构可视化，已成为智能化教学平台的基本组件。教育知识图谱的构建及应用是多学科的结合，需要知识图谱、自然语言处理、机器学习和数据挖掘等方面的知识融合。目前该领域的研究尚处于起步阶段，还有很多开放性问题需要将技术基础理论和教育教学理论融合起来共同解决，未来的教育知识图谱将会有极大的发展前景。

参考文献

- [1] 李振, 周东岱, 王勇. “人工智能+”视域下的教育知识图谱: 内涵、技术框架与应用研究[J]. 远程教育杂志, 2019, 37(4): 42-53.

- [2] 伟东云教育. 伟东云知识地图[EB/OL]. <http://www.xunzhitu.com/>, 2019-06-10.
- [3] 吴卓青. 基于 MOOC 数据的教育知识图谱构建与学生群体划分方法[D]: [硕士学位论文]. 大连: 大连理工大学, 2021.
- [4] 赵军. 知识图谱[M]. 北京: 高等教育出版社, 2018: 1-20.
- [5] Chen, P.H., Lu, Y., *et al.* (2018) KnowEdu: A System to Construct Knowledge Graph for Education. *IEEE Access*, **6**, 31553-31563. <https://doi.org/10.1109/ACCESS.2018.2839607>
- [6] 赵瑜, 陈志坤, 杨春. 基于开源数据的军事领域知识图谱构建方法[J]. 指挥信息系统与技术, 2019(3): 64-69.
- [7] 陈曦, 梅广, 张金金, 等. 融合知识图谱和协同过滤的学生成绩预测方法[J]. 计算机应用, 2020(2): 595-601.
- [8] 朱国丞. 基于大数据平台的知识图谱存储访问系统的设计与实现[D]: [硕士学位论文]. 南京: 东南大学, 2018.
- [9] 陈敏. 教育知识图谱内涵及应用场景探析[J]. 教育传播与技术, 2020(3): 61-65.
- [10] Luo, Y.T., Sha, B.C. and Xu, T. (2021) A Recommended Method Based on the Weighted RippleNet Network Mode. *Journal of Physics: Conference Series*, **2025**, Article ID: 012011. <https://doi.org/10.1088/1742-6596/2025/1/012011>
- [11] 李勇宏. 大规模在线内容推荐系统关键算法的设计与实现[D]: [硕士学位论文]. 北京: 北京交通大学, 2018.