

面向继续医学教育的多模态围产保健知识图谱构建研究*

吴萌¹ 杨林¹ 沈柳¹ 张素菡² 王敏¹ 孙振凤³ 徐晓巍¹ 刘娜娜²
王亚新² 侯丽¹ 李姣¹ 马良坤²

(¹ 中国医学科学院/北京协和医学院医学信息研究所 北京 100020

² 中国医学科学院北京协和医院 北京 100730

³ 北京市大兴区妇幼保健院 北京 102699)

[摘要] **目的/意义** 将多模态知识图谱应用于围产保健继续医学教育,以提高医护人员围产保健服务质量。**方法/过程** 结合北京协和医院围产保健继续教育项目学员需求和课程主要特点构建顶层本体,对本体类、关系和属性进行设计和定义,利用不同粒度教学资源表示教学内容,利用 Protégé 软件构建围产保健知识图谱概念层和实例层。基于团队自主研发的医学知识服务平台 MedKaaS,实现多模态实体进一步交互式构建,以及围产保健课程知识的可视化、检索和问答。**结果/结论** 该多模态知识图谱实现了围产保健课程内容的语义化整合,有助于学员直观和有针对性地学习围产保健知识。

[关键词] 知识图谱;围产保健;医学教育

[中图分类号] R-058 **[文献标识码]** A **[DOI]** 10.3969/j.issn.1673-6036.2024.04.002

Construction of Multimodal Perinatal Health Care Knowledge Graph towards Continuing Medical Education

WU Meng¹, YANG Lin¹, SHEN Liu¹, ZHANG Suhan², WANG Min¹, SUN Zhenfeng³, XU Xiaowei¹, LIU Nana², WANG Yaxin², HOU Li¹, LI Jiao¹, MA Liangkun²

¹Institute of Medical Information, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100020, China; ²Peking Union Medical College Hospital, Chinese Academy of Medical Sciences, Beijing 100730, China; ³Beijing Daxing District Maternal and Child Health Hospital, Beijing 102699, China

[Abstract] **Purpose/Significance** Using multimodal knowledge graph in perinatal health care continuing medical education is conducive to improving the quality of perinatal health care service for medical staff. **Method/Process** The top-level ontology is constructed based on the needs of students of perinatal health care continuing education project in Peking Union Medical College Hospital and the main characteristics of the courses, the classes, relationships and attributes are designed and defined, and different granularity teaching resources are used to represent the teaching content. The Protégé software is used to construct the instances of perinatal health care knowledge graph. Based on MedKaaS developed by our team, the further interactive construction of multi-modal entities, visualization, retrieval and question

[修回日期] 2024-02-22

[作者简介] 吴萌, 硕士, 助理研究员, 发表论著 6 篇; 通信作者: 李姣, 研究员, 博士生导师。

[基金项目] 北京市科技计划项目(项目编号: Z231100004623010); 国家社会科学基金青年项目(项目编号: 22CTQ024); 中国医学科学院医学与健康科技创新工程重大协同创新项目(项目编号: 2021-12M-1-056, 2021-12M-1-023)。

and answer of perinatal health care course knowledge are realized. **Result/Conclusion** The multimodal knowledge graph realizes the semantic integration of perinatal health care course content, and helps students learn perinatal health care knowledge intuitively and optionally.

[**Keywords**] knowledge graph; perinatal health care; medical education

1 引言

2023年3月中共中央办公厅、国务院办公厅印发《关于进一步完善医疗卫生服务体系的意见》，强调要对孕产妇等重点人群开展针对性的健康促进和预防保健服务^[1]。加强围产保健工作，对降低围产期并发症发生率、保障母婴安全具有重要意义^[2]。针对围产保健相关专业知识和临床技能持续更新迭代^[3]，但我国当前部分区域的围产保健服务仍存在诸多不足^[4-5]。2023年北京协和医院围产保健团队面向医护人员开展国家级继续教育项目“多学科在围产保健管理中的应用实践学习班”，旨在转变医护人员的传统围产保健管理理念，提高其专业知识和临床技能，提升全国范围内围产期女性及家庭的规范化健康服务。该培训课程结合课件教学、门诊观摩、体验式教学、实操指导等多种教学方式，利用音乐、视频、图片等教学资源，围绕营养、运动、心理等多个维度，讲授围产保健知识、技能、经验和案例。根据调查反馈，学员迫切期望有一套可以随时随地复习、巩固相关知识与技能的集成在线学习资源，以便将所学知识快速应用于所属医疗卫生机构。如何以结构化方式组织包含多种学习资源的围产保健课程，帮助围产保健相关从业人员进一步获取专业知识和相关技能，实现个性化学习，是本研究要解决的问题。

2 相关研究现状

知识图谱作为一种表示现实世界知识的重要方式，为知识浏览、检索、可视化和共享提供基础，在当前教育信息化发展中扮演重要角色。多模态知识图谱的提出为图像、声音和视频等现实世界数据

的表示带来解决方案^[6]。在教育领域，知识图谱能够基于教学过程丰富的多源异构数据、资源，表征各学科多层次、多粒度的知识谱系和认知过程，利用知识图谱为教育教学提供创新的支撑工具，已成为当前教育信息化发展的重要内容^[7]。Li Z等^[8]提出多资源的教育知识图谱，并探索多个课程之间的关系。余胜泉等^[9]研发基于育人知识图谱的“AI好老师”智能助理系统，实现人机结合的高效育人。许斌等^[10]利用知识图谱技术，构造出基础教育知识图谱 EduKG，并将其应用于知识搜索、知识快照、知识问答、知识链接。茶思月等^[11]以高中“信息技术”课程为例，收集了图片、音频、视频等形式的关于高中信息技术课程的多模态教学资源，设计基于学科核心素养的多模态教学知识图谱框架。Liu P等^[12]构建了多模态音乐知识图谱，通过对在线教育数据的挖掘提出基于用户兴趣的个性化学习策略。在医学领域，张坤丽等^[13]以产科专业叙词表、临床路径和诊疗规范等为数据来源构建中文产科知识图谱，以产科疾病为核心构建其与症状、检查、药品等实体之间的关系。陆泉等^[14]对武汉大学临床医学专业主干课程的课程资料进行深度挖掘，得到临床医学课程间、知识主题间及课程与知识主题间的关联，构建临床医学课程知识主题库及临床医学课程知识主题图谱。

本研究结合围产保健课程中存在大量图片、视频等多模态教学资源的特点，构建多模态围产保健知识图谱，旨在以结构化的方式组织围产保健课程知识，帮助围产保健相关医护人员快速获取专业知识和相关技能，实现个性化学习，进而助力优质孕产宣教资源下沉惠及更多家庭。

3 多模态围产保健知识图谱构建

3.1 数据来源

本研究的数据资料来源于北京协和医院围产保

健团队提供的 24 个教学单元的课件及相应教学视频、音频。课程结合课件教学、门诊观摩、体验式教学、实操指导等多种方式，围绕营养、运动、心理等多个维度内容，讲授围产保健知识、经验和案例。课程内容覆盖备孕期、妊娠期、分娩期、产后期等多个不同的孕产时期，涉及围产保健、分娩镇痛、母乳喂养、产后康复、婴幼儿养育等多个相关主题。课件内容多以图文并茂、体验实操的形式详细讲解知识点。

3.2 知识图谱构建流程

本研究面向北京协和医院围产保健继续教育课程特点和应用需求，基于本体构建方法、理论构建多模态围产保健知识本体，采用多人协同标注与审核方式提取多模态数据素材，利用医学知识服务平台（medical knowledge as a service, MedKaaS），实现多模态围产保健知识图谱的构建与应用，见图 1。

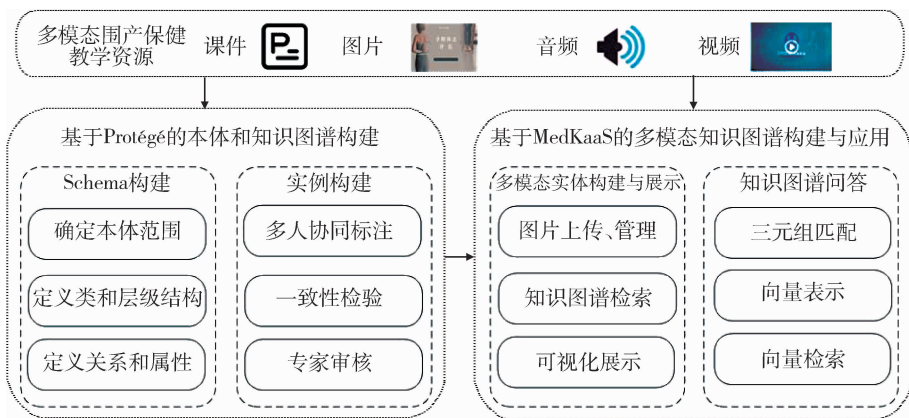


图 1 多模态围产保健知识图谱构建流程

3.3 本体构建

采用本体设计和构建围产保健知识图谱的顶层架构。本体作为一种基于语义的知识表示模型，定义和约束知识图谱中的概念、属性和关系。围产保健本体构建的主要流程包括确定本体范围、构建语义类型、构建关系和属性、构建实例。

3.3.1 确定本体范围 首先，梳理围产保健课程学员需求和该课程的主要特点，明确本体的描述范围。需求主要包括：针对性回顾围产保健知识，选择性学习感兴趣的知识点；通过对图片和视频的直观学习，将围产保健知识快速应用于实际医疗服务工作中。课程特点包括：采用多种教学方式，利用图片、视频等教学资料加强学员对相关技能和实践操作的掌握；拥有多学科交叉融合的教学内容，涉及营养学、运动学、心理学、医学等多学科知识。

3.3.2 构建语义类型 教育知识图谱主要分为静

态知识图谱和动态事理图谱。前者以教学过程中涉及的元素为实体节点、以教学元素间的逻辑关系为边。后者以教学事件或教学活动为表征对象，以逻辑事理关系（顺承、因果）等为边^[7]。本研究以围产保健教学过程中教学资源 and 教学元素之间的时间和逻辑关系为关注点，构建静态多模态知识图谱。结合本体描述范围和应用场景，以及相关领域课程类知识图谱框架，设计多模态围产保健知识图谱的顶层本体结构。其中设计顶层大类 5 个：教学资源、人、教学工具、时期和主题。在教学资源大类中，利用课程、教学单元、教学章节和知识点 4 种不同粒度的教学资源拆分和组织教学内容。利用包含关系表示这 4 种不同粒度教学资源之间的关联。利用教学工具，大类组织和表示多模态教学工具，如视频、图片和课件等。参考围产的不同时期以及围产保健教学课件，提炼和详细分类课程涉及的主题，见图 2。



图2 围产保健知识图谱顶层大类

3.3.3 构建关系和属性 定义类之间的关系，将不同粒度教学资源关联，见表1。其中“后续教学章节”“后续教学单元”“后续知识点”表示教学章节、教学单元和知识点之间的连接顺序，便于学员在学习过程中串联知识，保留教学过程中的逻辑顺序。利用“有主讲教师”“有主题”“涉及时期”对教学单元详细描述，便于学员个性化筛选学习。利用“有课件”“有视频”和“有图片”将教学资源与多模态教学工具关联，提供丰富直观的教学知识。明确关系的定义域和值域。定义类的属性为围产保健课程知识提供更全面的信息，见表2。因不同类拥有不同属性，所以须明确属性和其定义域。首先定义课程和教学单元的教学目标、时长、开始时间、结束时间和教学地点等属性，其次定义知识点和图片的主要内容属性，并定义人员的属性。

表1 围产保健知识图谱关系

关系名称	关系定义域	关系值域
包含	无	无
后续教学章节	教学章节	教学章节
后续教学单元	教学单元	教学单元
后续知识点	知识点	知识点
有主讲教师	教学单元	主讲教师
有主题	教学单元	主题
涉及时期	教学单元	时期
有视频	教学单元、教学章节、知识点	视频
有课件	教学单元	课件
有图片	知识点	图片

表2 围产保健知识图谱属性（部分）

属性名称	属性定义域
名称	所有类
教学目标	课程、教学单元
时长	课程、教学单元
开始时间	课程、教学单元
结束时间	课程、教学单元
教学地点	课程、教学单元
顺序	课程、教学单元、教学章节、知识点、教学工具
主要内容	知识点
图片内容	图片
姓名	人
职务	人
单位	人
介绍	人

3.3.4 构建实例 根据以上定义的本体逻辑框架，利用 Protégé 构建围产保健本体。组织 5 名医学信息学专业学生提取 24 个教学单元的课件知识内容，并利用 Protégé 以实例的形式填充围产保健本体，实现围产保健知识图谱的构建。基于 Protégé 完成本体的一致性检验，并邀请 2 名相关领域专家审核构建结果。最终构建的围产保健教学知识图谱顶层本体模型，见图 3。其中，顶层本体包含类 40 个，关系 10 个，属性 13 个。基于该本体构建多模态围产保健知识图谱，包含课程 2 门，教学单元 24 个，教学章节 81 个，知识点 436 个，图片 695 个，视频 26 个，音频 5 个。围产保健知识图谱构建过程充分参考领域专家建议，基于本体的顶层逻辑框架，通过本体一致性检验及专家进一步审核，保证一致性、逻辑性、准确性和覆盖程度。利用知识图谱对围产保健课程的教学内容进行结构化、语义化组织，便于学员筛选感兴趣的主题、时期等相应的教学单元，满足后续的语义搜索、问答等需求，在案例研究中验证知识图谱的应用能力。在实例构建过程中，由于一些围产保健课程的教学形式多样，其课件存在结构化不强、知识点不突出等情况，为课件内容提取带来困难。针对此类课件，通过进一步组内交流讨论和专家咨询，实现准确提取。此外，针对课件中包含实际案例和患者真实数据，仅提取必要的案例数据，并对患者姓名、地址、联系方式和照片等进行脱敏处理。

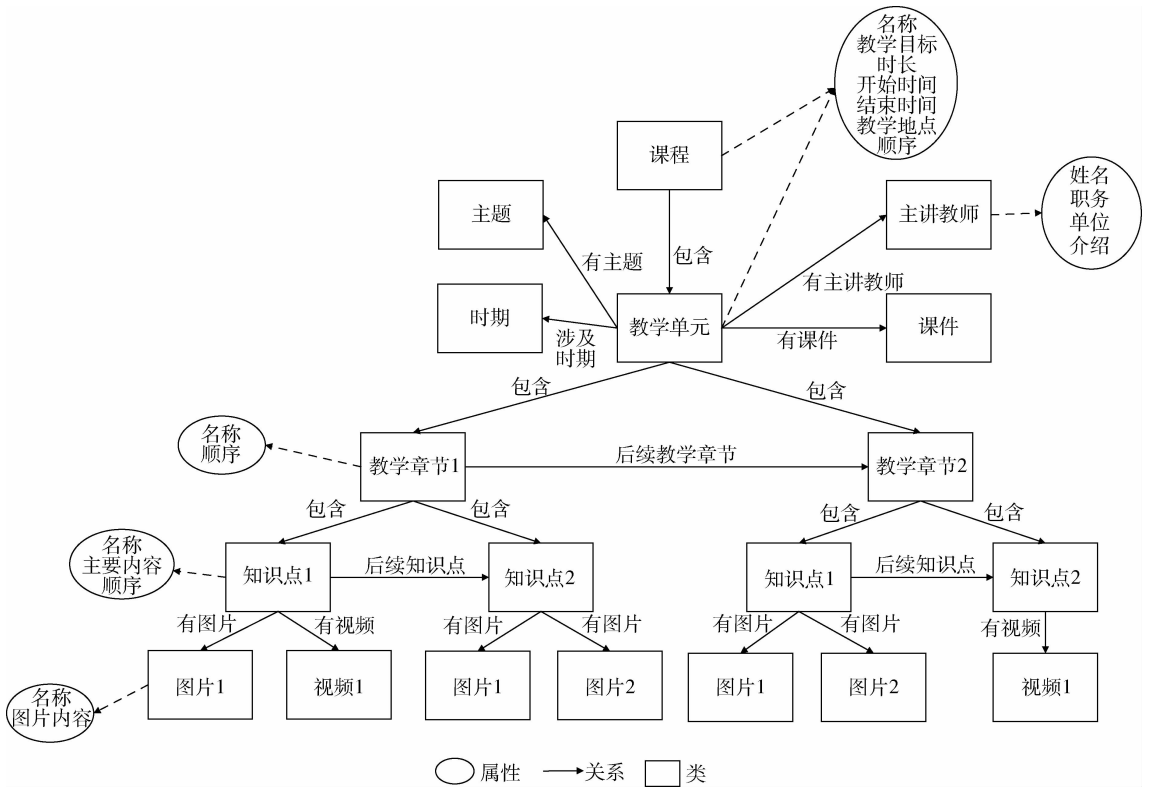


图3 围产保健知识图谱顶层本体结构

构建、浏览和检索。

4 基于 MedKaaS 的多模态围产保健知识图谱构建和应用

4.1 多模态实体构建

利用团队研发的医学知识服务平台 MedKaaS^[15]进行多模态围产保健知识图谱的构建与应用。MedKaaS 是以医学知识为基础、人工智能技术为驱动的医学科技信息知识服务平台。助力消除多模态、异构、分散等特征对海量医学科技资源挖掘和利用的障碍，实现多源异构医学信息资源的语义化表示、整合、存储与利用。MedKaaS 已通过中国电子技术标准化研究院发起基础及医疗领域知识图谱构建平台和应用平台双认证。该平台支持 OWL、RDF、JSON 等本体和知识图谱数据格式的导入，支持实体图片上传、管理和展示的功能。将已构建的围产保健知识图谱导入该平台，利用知识图谱编辑工具构建和展示相应的图片实体，实现多模态围产保健知识图谱的进一步

4.2 应用研究——基于“体态评估、步态指导”案例

基于开源框架 AntV - G6，MedKaaS 平台提供知识图谱的可视化工具，可以展示相应的概念、实例、关系和图片实体，实现多模态知识图谱的可视化。通过可视化工具，学员可以总览相关课程的知识结构、快速定位感兴趣的知识。以“体态评估、步态指导”课程为例，该课程为课程一的一个教学单元，主要讲授孕期体态评估的意义、工具和方法。通过可视化展示的方式，学员可以直观地看到该课程的主题是围产运动，涉及时期是妊娠期，包含“孕期体态评估的意义”“使用工具”“运动处方建议”等教学章节。

此外，该平台使用基于知识图谱和检索结合的问答技术支持问答功能。基于知识图谱的问答技术将问题分词并与知识图谱中的三元组匹配以进行查找。在无法用知识图谱找到答案的情况下，结合基于检索的问答技术。主要流程为将知识图谱中三元组的主语、谓语和宾语进行拼接后实现文本化，利

用 in - batch 负采样优化的向量编码模型 RoBERTa^[16]对文本化的三元组和用户提出的问题进行了向量表示,利用 Elasticsearch 搜索引擎对三元组向量和问题向量构建索引。之后,利用词从大规模图谱三元组文本中检索出语义匹配的候选结果,再使用向量语义相似度进行精确排序,确保检索相关性和效率。基于该知识图谱问答系统,可实现面向围产保健课程知识的初步智能问答,便于学员对授课知识的快速查询和回顾。以“体态评估、步态指导”课程为例,用户在问答对话框中提问“孕期体态评估使用工具包含什么?”时系统会自动回答“评估表格、评估软件和其他工具”。

5 结语

本研究基于 Protégé 本体编辑软件构建多模态围产保健本体,为后续知识图谱的构建提供顶层框架,实现围产保健课程知识点和教学资源的有效整合和语义化表示。利用 MedKaaS 医学知识服务平台进行多模态实体构建和知识图谱应用,实现多模态围产保健知识图谱的交互式、高质量构建。考虑到知识准确性和全面性,采用人工标注的方式进行知识提取,在知识图谱构建效率上有待提高。

本研究基于围产保健课程构建多模态教学知识图谱,丰富了传统围产保健教学模式,利用课程、教学单元、教学章节、知识点 4 种不同粒度教学资源对教学内容进行有效串联和组织,实现了围产保健课程知识的可视化、检索和问答。学员基于知识图谱可以快捷地获取教学内容、知识组织关系和与知识点相关的各类多模态教学资源,完成对围产保健课程知识点的梳理和回顾。同时,利用多模态形式的教育资源为学员带来更清晰直观的学习感受,使教学知识更加丰富,学生的学习质量和效率得以提高。此外,围产保健多模态知识图谱可为面向课程的智能问答、信息检索和数据的可视化交互等下游教育应用提供有力的支撑。

利益声明: 所有作者均声明不存在利益冲突。

参考文献

- 1 关于进一步完善医疗卫生服务体系的意见 [EB/OL]. [2024 - 01 - 10]. https://www.gov.cn/zhengce/2023-03/23/content_5748063.htm.
- 2 王国华,杨棋英. 孕产期保健对高龄产妇分娩结局及新生儿并发症的影响 [J]. 中国妇幼保健, 2020, 35 (2): 240 - 242.
- 3 刘婷婷,范翠芳. 心理因素在先兆流产中的研究进展 [J]. 中国优生与遗传杂志, 2022, 30 (12): 2298 - 2302.
- 4 郑晓菲,尹凡芑,王欢,等. 以围产医学中心建设助推区域妇幼保健服务体系高质量发展——基于重庆市妇幼保健院的管理实践经验 [J]. 现代医药卫生, 2022, 38 (8): 1420 - 1423.
- 5 唐思婕. 公共孕产服务对农村新生儿超重的影响研究 [D]. 杭州:浙江大学, 2023.
- 6 ZHU X, LI Z, WANG X, et al. Multi - modal knowledge graph construction and application: a survey [EB/OL]. [2024 - 01 - 10]. <https://arxiv.org/pdf/2202.05786.pdf>.
- 7 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.
- 8 LI Z, CHENG L, ZHANG C, et al. Multi - source education knowledge graph construction and fusion for college curricula [EB/OL]. [2024 - 01 - 10]. https://www.researchgate.net/publication/228857266_A_survey_of_ontology_evaluation_techniques.
- 9 余胜泉,彭燕,卢宇. 基于人工智能的育人助理系统——“AI好老师”的体系结构与功能 [J]. 开放教育研究, 2019, 25 (1): 25 - 36.
- 10 许斌,苏伟杰,刘阳. 基础教育知识图谱赋能智慧教育 [J]. 人工智能, 2019, 10 (3): 37 - 43.
- 11 茶思月,张秀,李红. 基于学科核心素养的多模态教学知识图谱构建研究 [J]. 现代信息科技, 2023, 7 (1): 173 - 177, 181.
- 12 LIU P, CAO Y, WANG L. A multimodal fusion online music education system for universities [EB/OL]. [2024 - 01 - 10]. <https://www.hindawi.com/journals/cin/2022/6529110/>.
- 13 张坤丽,胡晨馨,宋玉,等. 基于多源数据的中文产科知识图谱构建 [J]. 郑州大学学报(理学版), 2023, 55 (1): 8 - 14.
- 14 陆泉,谢祎玉,陈静,等. 临床医学课程知识主题图谱构建研究 [J]. 图书情报工作, 2019, 63 (9): 101 - 108.
- 15 XU X, WANG X, WU M, et al. Development of an interactive medical knowledge graph based tool set [EB/OL]. [2024 - 01 - 10]. <https://www.sciencedirect.com/science/article/pii/S1877050923007834>.
- 16 LIU Y, OTT M, GOYAL N, et al. RoBERTa: a robustly optimized BERT pretraining approach [EB/OL]. [2024 - 01 - 10]. <https://arxiv.org/pdf/1907.11692.pdf>.