

医学知识图谱应用研究进展^{*}

胡红娟

周 阳

(南华大学附属第一医院护理部 衡阳 421001)

(南华大学护理学院 衡阳 421001)

匡泽民

谭 玲

(首都医科大学附属北京安贞医院高血压科 北京 100029)

(北京邮电大学计算机学院 北京 100876)

〔摘要〕 介绍医学知识图谱相关概念, 详细阐述医学知识图谱应用情况, 包括临床决策支持、文献可视化、智慧搜索、智能问答等方面, 提出问题与展望, 为相关研究提供参考。

〔关键词〕 医学知识图谱; 应用; 临床决策支持; 文献可视化; 智能问答; 智慧搜索

〔中图分类号〕 R-058 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2022.05.006

Research Progress of Application of Medical Knowledge Graph HU Hongjuan, Nursing Department, The First Affiliated Hospital of University of South China, Hengyang 421001, China; ZHOU Yang, School of Nursing, University of South China, Hengyang 421001, China; KUANG Zemin, Department of Hypertension, Beijing Anzhen Hospital, Beijing 100029, China; TAN Ling, School of Computing, Beijing University of Posts and Telecommunications, Beijing 100876, China

〔Abstract〕 The paper introduces the concepts related to medical knowledge graph, expounds the application of medical knowledge graph in detail, including clinical decision support, literature visualization, intelligent search, intelligent question and answer, etc., puts forward questions and prospects, and provides references for related study.

〔Keywords〕 medical knowledge graph; application; clinical decision support; literature visualization; intelligent question answering; intelligent search

1 引言

生机构与互联网合作, 加强区域医疗卫生信息资源整合及临床、科研数据整合共享和应用, 支持研发医疗健康相关的人工智能技术。近年来各类知识图谱不断涌现^[1-3], 逐渐融入到医学领域, 医学知识

《“健康中国 2030”纲要》提出, 鼓励医疗卫

〔收稿日期〕 2021-08-03

〔作者简介〕 胡红娟, 副教授, 硕士生导师, 发表论文 13 篇, 参编教材 1 部; 通信作者: 匡泽民, 主任医师, 硕士生导师, 发表论文 50 余篇, 参编著作 7 部, 获软件著作权 6 项。

〔基金项目〕 北京市自然科学基金面上专项“高血压患者产生型健康医疗数据的表示学习研究”(项目编号: M22009); 湖南省教育厅科学研究项目“基于 BCW + BCT 的‘互联网 + 护理’干预模式在农村老年慢病患者中的构建与评价”(项目编号: 19A426); 湖南省科技创新计划项目“基于临床指南与大数据的中青年高血压运动决策支持系统的构建及推广应用”(项目编号: 2020SK51810)。

图谱成为人工智能辅助医疗系统的基本组成部分^[4]。知识图谱将医学知识与互联网充分结合,探索科研数据整合共享和应用的方法,为人工智能研发提供依据。2017 年我国学科目录进行调整,首次出现知识图谱学科方向。目前医学是知识图谱应用最广的领域之一,医学知识图谱应用也是目前智能大数据的前沿研究问题。

有学者^[5]提出狭义的知识图谱是一种大规模的语义网络,包含实体、概念及其之间的各种语义关系。但与传统语义网络相比,知识图谱具有规模巨大、语义丰富、质量精良、结构友好等特性。广义知识图谱则是指一种技术体系,指大数据时代知识工程的一系列代表性技术的总和。总的来说,知识图谱实质是建立在语义网络上的知识系统^[6]。医学知识图谱将知识图谱与医学知识结合,将互联网技术与医学情境融合,实现医学数据的自动化与智能化处理。构建医学知识图谱可为临床辅助决策、文献可视化分析、智能问答以及智慧搜索提供支持。

2 医学知识图谱应用

2.1 临床决策支持

临床决策支持系统 (Clinical Decision Support System, CDSS) 使用个体化数据对患者进行特定评估并向临床医师提出决策建议。医学知识图谱的诞生为 CDSS 提供了深层关系发现与推理能力。目前研究人员积极探索临床环境中决策支持的潜力^[7],与知识图谱结合开发适用于医疗环境的临床决策支持系统。Zhao C、Jiang J 和 Guan Y 等^[8]开发电子病历 (Electronic Medical Record, EMR) 驱动的 CDSS。该系统利用电子病历中的医学知识网络表示医学知识,构建由症状、检查、检查结果、疾病和治疗 5 类实体组成的知识网络,将患者现状输入后可获得关于疾病测试、初始诊断及治疗相应建议。在知识推理方面,由于贝叶斯网络受到推理及学习的计算复杂性限制,该系统选择马尔可夫随机场 (Markov Random Field, MRF) 执行推理任务。Goodwin T 和 Harabagiu S^[9]设计一个用于回答医学问题的 CDSS,利用 MiMic - III 提供一个公开的电

子病历集合,自动生成一个庞大的医学知识图谱,通过概率推理方法获得问题答案,自动发现与主题相关的问题答案可能性,生成包含答案的相关科学文章。Sheng M、Hu Q 和 Zhang Y 等^[10]提出一种基于知识图谱的数据密集型临床决策支持平台 IDS,该平台围绕患者就诊过程提供一系列 CDSS 服务,即查询、检查、诊断、药物、治疗及预后等。IDS 从医学样本库中提取三元组形成知识图谱,然后通过规则检测构建新的三元组丰富医学样本库。目前该平台已应用于乡村医疗,有利于降低误诊率。朱超宇和刘雷^[11]构建一个知识图谱辅助的肺部影像诊断系统,其结合指南和文献中肺癌及肺结核的相关知识构建知识图谱,医生诊断时将患者影像特征与知识图谱中的影像特征相结合进行分析,得到更加全面的结果。研究显示 CDSS 主要来源于临床数据,其中 41.66% 来自数据集存储库或临床中心,33.33% 来自电子健康档案 (Electronic Health Record, EHR)^[12]。近年来由于大数据和自然语言处理技术的快速发展,从电子病历中自动挖掘知识成为新的研究趋势^[13-14]。基于知识图谱的 CDSS 充分利用电子病历,但不同医院平台的电子病历数据录入有自己的语言,使得跨越平台开发 CDSS 出现障碍,这就要求电子健康记录中数据创建语言的一致性^[15]。临床指南、科研文献也是知识的重要来源,通过文本计算机化构建知识图谱,更科学地为临床决策提供指导。

2.2 医学文献可视化

目前医学行业积累了规模可观的大数据,但是这些数据并未创造价值,其中一个重要原因就是数据混杂,隐匿于大量文献当中,用户无法精确定位。知识图谱的可视化数据分析、视觉知识图谱构建^[16]、CiteSpace^[17]使用能够帮助用户直观理解医学信息内容,实现大数据的价值变现。例如田元祥、刘莎莎和周新宇等^[18]利用 CiteSpace 构建中医肝郁脾虚证候诊断标准知识图谱;李新宇、李翔和廖林丽等^[19]基于 CiteSpace 构建中医药论治干眼知识图谱;张琪、曹林忠和蒋玮等^[20]基于 CiteSpace 分析股骨头坏死中医药治疗知识图谱等。通过计量

软件的可视化分析可以直观了解中医子领域研究现状,通过关键词聚类分析还可了解该子领域研究集中度。该方法还广泛应用于心血管慢病分析中,王伟帅、李阳兵和刘鑫源等^[21]通过从 CNKI 中检索高原高血压相关文献,利用 CiteSpace 分析直观了解到该领域文献在 2011 发文量最高,军区医院在该领域有突出贡献,研究重点从氧化应激转换到血管紧张素,目前研究热点为高血压用药及护理。知识图谱在文献可视化中的应用使文献阅读变得简单直观,但将多篇文献构建成图谱,可能会缺失其中某些数据的展现。因此利用图谱技术时需加强对文献重点内容的提取能力。

2.3 智能问答

问答系统的构建要求机器能够理解自然语言,需要机器具备强大的背景知识,而知识图谱就是这类背景知识中的重要形式之一。Ruan T、Huang Y 和 Liu X 等^[22]设计一个基于知识图谱的智能问答工具 QAnalysis。用户可用自然语言输入其分析需求,该系统会以图表形式返回答案。该系统实现 3 个方面的改进:不是直接从知识库中提取答案,而是从统计上推导出答案;利用患者图和临床术语图相结合改善以往语义解析易混淆的现状;提高问题回答的精度。Fecho K、Balhoff J 和 Bizon C 等^[23]首先采用知识图谱生产翻译程序,然后基于 3 个翻译推理器 ROBOKOP(用面向知识的路径链接的生物医学对象推理)、RTX(推理工具)和 midiKanren 等生产问答系统,以生物链路模型作为高级本体将领域知识表示为医学实体之间关系的知识图谱,应用 MCAT 问题评估问答系统改进问答系统以提高性能。李贺、刘嘉宇和李世钰等^[24]利用公开数据库构建疾病知识图谱,基于知识图谱确定检查、症状、科室、病因及传染性 5 类问题,通过 AC 多模式匹配算法,结合语义相似度计算进行症状实体识别,封装匹配词转换成查询语言,实现问题回答。同时优化问句分析、信息检索及答案抽取,从而提高医疗问答系统与网络技术、医学知识的融合度。Li C、Hang S 和 Hu X 等^[25]构建老年医疗保健知识图谱,利用爬虫技术从百科网站获取老年健康信息构建本

体库,采用变体的关联规则挖掘算法检测实体之间关系,使用 Limes 算法将实体与 88 446 个实体的关系融合,建立度量表达式最终形成知识图谱,利用知识图谱构建一个历史行为驱动的问答平台 KnowHealth,该平台将健康问题分成 12 类并根据老年人的历史行为回答老年保健相关问题。尽管利用知识图谱在构建医疗问答系统方面开展了相关研究,但受限于现有医学知识图谱的推理能力,知识图谱和问答系统的结合还有一定局限,并且大部分问答系统主要是针对简单问题,对于复杂医疗问题的解决有一定限制^[26]。因此迫切需要推进相关研究的创新。

2.4 智慧搜索

互联网时代医疗信息呈现爆炸式增长,要精准、便捷获取重要信息显得更加困难,因此智能化、情景化、社会化的智慧搜索技术应用成为研究热点。而知识图谱可为有效查询和搜索数据提供更为灵活的数据结果。Hasan S、Rivera D 和 Wu X C 等^[27]提出用于癌症数据的科学数字图书馆框架,核心是知识图谱。该平台利用基于语义网络的知识图谱储存癌症注册表数据,实现特点场景数据查询,例如乳腺癌治疗序列变化的查询。将知识图谱与外部数据集链接,以便于使用多个数据集执行复杂查询。同时该平台还可实现癌症知识可视化,用户通过索引检索和信息检索进行查询和搜索。Shenoi S J、Vi L 和 Sarvesh S 等^[28]开发肿瘤搜索引擎,检索与患者状况、遗传特征、年龄和性别相关的科学文章和临床试验,构建知识图谱,对疾病、基因和人口统计学信息进行扩展查询。Struck A、Walsh B 和 Buchanan A 等^[29]构建一个生物医学证据图谱(Bio-Medical Evidence Graph, BMEG),内容包括患者样本信息、突变、基因表达、药物反应数据、基因组注释和文献分析。BMEG 通过导入和取消几个相关资源来构建连贯图表,将这些复杂数据进行整合和分析,实现癌症生物数据查询。医学知识图谱搜索不仅提供用户网页间超链接的文档关系,还包括不同类型实体间丰富的语义关系。然而受限于医学知识图谱的知识数量和质量,需要不断改进和完善技术框架。

3 结语

知识图谱作为互联网领域的新兴技术，是信息处理和知识组织的强大工具。知识图谱将领域内的复杂知识通过信息抽取、数据挖掘、语义匹配、语义计算和知识推理等过程精准地描述出来，描述知识及其演化过程、发展规律，从而为研究和决策提供准确、可追踪、可解释、可推理的知识数据。但目前知识图谱研究还不够完善和深入，需要在以下方面进行进一步尝试：一是知识图谱应用应标准化，丰富其功能，例如基于高血压指南构建高血压用药知识图谱，实现用药推荐、药物查询、药物问询等功能；二是知识图谱普及应公开化，如临床决策与健康管理相结合，增加其利用率并实现真正落地。

参考文献

- 1 Chung - Lin Huang C H, Chung - Chi Huang. Study on Customized Knowledge Graph of Student Pilot Learning in FITS Training [J]. Journal of Intelligent & Fuzzy Systems, 2021 (40): 7969 - 7979.
- 2 Esfahani A N, Moghaddam N B, Maleki A, et al. The Knowledge Map of Energy Security [J]. Energy Reports, 2021 (7): 3570 - 3589.
- 3 李嘉明, 赵阔, 屈挺, 等. 基于知识图谱的区块链物联网领域研究分析 [J]. 计算机科学, 2021, 48 (S1): 563 - 567.
- 4 Li L, Wang P, Yan J, et al. Real - World Data Medical Knowledge Graph: Construction and Applications [J]. Artif Intell Med, 2020 (103): 101817.
- 5 肖仰华, 徐波, 林欣, 等. 知识图谱 [M]. 北京: 电子工业出版社, 2020.
- 6 Yu T, Li J, Yu Q, et al. Knowledge Graph for TCM Health Preservation: Design, Construction, and Applications [J]. Artif Intell Med, 2017 (77): 48 - 52.
- 7 Shortliffe E H, Sepúlveda M J. Clinical Decision Support in the Era of Artificial Intelligence [J]. JAMA: the Journal of the American Medical Association, 2018, 320 (21): 2199 - 2200.
- 8 Zhao C, Jiang J, Guan Y, et al. EMR - based Medical Knowledge Representation and Inference via Markov Random

- Fields and Distributed Representation Learning [J]. Artif Intell Med, 2018 (87): 49 - 59.
- 9 Goodwin T, Harabagiu S. Medical Question Answering for Clinical Decision Support [C]. Indianapolis: ACM International Conference on Information and Knowledge Management, 2016.
- 10 Sheng M, Hu Q, Zhang Y, et al. A Data - Intensive CDSS Platform Based on Knowledge Graph [M]. Cham: Springer International Publishing, 2018: 146 - 155.
- 11 朱超宇, 刘雷. 基于知识图谱的医学决策支持应用综述 [J]. 数据分析与知识发现, 2020, 4 (12): 26 - 32.
- 12 Souza - Pereira L, Pombo N, Ouhbi S, et al. Clinical Decision Support Systems for Chronic Diseases: A Systematic Literature Review [J]. Computer Methods and Programs in Biomedicine, 2020 (195): 105565.
- 13 Li F, Jin Y, Liu W, et al. Fine - Tuning Bidirectional Encoder Representations From Transformers (BERT) - Based Models on Large - Scale Electronic Health Record Notes: An Empirical Study [J]. JMIR medical informatics, 2019, 7 (3): e14830.
- 14 Ji B, Liu R, Li S, et al. A Hybrid Approach for Named Entity Recognition in Chinese Electronic Medical Record [J]. BMC medical Informatics and Decision Making, 2019, 19 (2): 64.
- 15 Sarkar U, Samal L. How Effective are Clinical Decision Support Systems? [J]. BMJ, 2020 (370): m3499.
- 16 Li X, Liu H, Zhao X, et al. Automatic Approach for Constructing a Knowledge Graph of Knee Osteoarthritis in Chinese [J]. Health Information Science and Systems, 2020, 8 (1): 12.
- 17 付雷杰, 曹岩, 白瑀, 等. 国内垂直领域知识图谱发展现状与展望 [J]. 计算机应用研究, 2021, 38 (11): 3201 - 3214.
- 18 田元祥, 刘莎莎, 周新宇, 等. 基于 CiteSpace 的肝郁脾虚证候诊断标准知识图谱可视化分析 [J]. 世界科学技术 - 中医药现代化, 2021, 23 (3): 953 - 960.
- 19 李新宇, 李翔, 廖林丽, 等. 基于 CiteSpace 的中医药论治干眼可视化分析 [J]. 湖南中医杂志, 2021, 37 (4): 147 - 149.
- 20 张琪, 曹林忠, 蒋玮, 等. 基于“citespace”分析中医药治疗股骨头坏死的研究热点与发展趋势 [J]. 实用中医内科杂志, 2021, 35 (9): 26 - 28, 149 - 151.

(下转第 39 页)

类效果,为更好地辅助医生增强医疗临床应用,提高我国医疗水平提供一定技术方法支撑。

参考文献

- 1 世界卫生组织. 乳腺癌成全球最常见癌症 [EB/OL]. [2021-06-10]. <https://baijiahao.baidu.com/s?id=1690756461465319744&wfr=spider&for=pc>.
 - 2 吴辰文, 李长生, 王伟, 等. 一种改进的 SVM 算法在乳腺癌诊断方面的应用 [J]. 计算机工程与科学, 2017, 39 (3): 562-566.
 - 3 刘军, 彭慧娟, 黄斌, 等. 基于 BP-GamysBoost 的乳腺癌诊断模型 [J]. 计算机与现代化, 2021 (4): 8-14.
 - 4 于凌涛, 夏永强, 闫昱晟, 等. 利用卷积神经网络分类乳腺癌病理图像 [J]. 哈尔滨工程大学学报, 2021, 42 (4): 567-573.
 - 5 冯照石, 范祺. 双重加权朴素贝叶斯算法预测乳腺癌复发率 [J]. 牡丹江师范学院学报 (自然科学版), 2021 (2): 11-15.
 - 6 余莹, 樊重俊, 朱人杰, 等. 基于系统聚类和 SVM 模型的乳腺癌诊断研究 [J]. 智能计算机与应用, 2020, 10 (11): 97-100, 105.
 - 7 常炳国, 李玉琴, 冯智超, 等. 基于主成分机器学习算法的慢性肝病的智能预测新方法 [J]. 计算机科学, 2017, 44 (S2): 65-67, 91.
 - 8 董虎胜. 主成分分析与线性判别分析两种数据降维算法的对比研究 [J]. 现代计算机 (专业版), 2016 (29): 36-40.
 - 9 Belhumeur P N, Hespanha J P. Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997, 19 (7): 711-720.
 - 10 Abuomman A A, Reaz M B I. Ensemble of Binary SVM Classifiers Based on PCA and LDA Feature Extraction for Intrusion Detection [C]. Xi'an: Proc. of the Advanced Information Management, Communicates, Electronic and Automation Control, 2017: 636-640.
 - 11 Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System [EB/OL]. [2019-03-30]. <https://blog.csdn.net/u014033218/article/details/88921277>.
 - 12 CSDN. 机器学习算法之 Xgboost 算法 [EB/OL]. [2021-06-12]. https://blog.csdn.net/qq_20412595/article/details/82621744.
 - 13 余天阳. 基于 XGBoost 的固态白酒发酵产量预测 [J]. 计算机与数字工程, 2020, 48 (5): 1233-1237, 1251.
 - 14 李占山, 刘兆赓. 基于 XGBoost 的特征选择算法 [J]. 通信学报, 2019, 40 (10): 101-108.
 - 15 邓卓, 苏秉华, 张凯. 基于集成学习的乳腺癌分类研究 [J]. 中国医疗设备, 2020, 35 (12): 59-62.
-
- (上接第 33 页)
- 21 王伟帅, 李阳兵, 刘鑫源, 等. 基于 CiteSpace 对中国高原高血压研究热点与演变的知识图谱分析 [J]. 中国高原医学与生物学杂志, 2020, 41 (3): 156-164.
 - 22 Ruan T, Huang Y, Liu X, et al. QAnalysis: A Question - Answer Driven Analytic Tool on Knowledge Graphs for Leveraging Electronic Medical Records for Clinical Research [J]. BMC Med Inform Decis Mak, 2019, 19 (1): 82.
 - 23 Fecho K, Balhoff J, Bizon C, et al. Application of MCAT Questions as a Testing Tool and Evaluation Metric for Knowledge Graph - based Reasoning Systems [J]. Clinical and Translational Science, 2021, 14 (5): 1719-1724.
 - 24 李贺, 刘嘉宇, 李世钰, 等. 基于疾病知识图谱的自动问答系统优化研究 [J]. 数据分析与知识发现, 2021, 5 (5): 115-126.
 - 25 Li C, Hang S, Hu X, et al. KnowHealth: A Knowledge Graph based Question - Answer Platform for Elderly People [C]. Weihai: 12th EAI International Conference on Mobile Multimedia Communications, Mobimedia, 2019.
 - 26 Zhou X, Wu B, Zhou Q. A Depth Evidence Score Fusion Algorithm for Chinese Medical Intelligence Question Answering System [J]. J Healthc Eng, 2018, 1 (3): 1205354.
 - 27 Hasan S, Rivera D, Wu X C, et al. Knowledge Graph - Enabled Cancer Data Analytics [J]. IEEE J Biomed Health Inform, 2020, 24 (7): 1952-1967.
 - 28 Shenoi S J, Vi L, Sarvesh S, et al. Developing a Search Engine for Precision Medicine [J]. AMIA Summits on Translational Science Proceedings, 2020 (2020): 579-588.
 - 29 Struck A, Walsh B, Buchanan A, et al. Exploring Integrative Analysis Using the BioMedical Evidence Graph [J]. JCO Clin Cancer Inform, 2020 (4): 147-159.