

语义知识图谱与科学知识图谱在中医药领域的应用^{*}

孙肇阳

黄小圆 许溪彬

(天津中医药大学中医学院 天津 301617)

(天津中医药大学针灸推拿学院 天津 301617)

沙福慧 任海燕 邱继文

(天津中医药大学中医学院 天津 301617)

〔摘要〕 介绍语义知识图谱和科学知识图谱的概念、特点及研究现状,分别阐述其在中医药领域应用情况,分析两类知识图谱区别及二者交互对促进中医药领域知识图谱技术发展的作用。

〔关键词〕 语义知识图谱;科学知识图谱;中医药

〔中图分类号〕 R-058 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2021.07.008

Application of Semantic Knowledge Graph and Mapping Knowledge Domain in the Field of Traditional Chinese Medicine

SUN Zhaoyang, School of Traditional Chinese Medicine, Tianjin University of Traditional Chinese Medicine, Tianjin 301617, China; HUANG Xiaoyuan, XU Xibin, School of Acupuncture-Moxibustion and Tuina, Tianjin University of Traditional Chinese Medicine, Tianjin 301617, China; SHA Fuhui, REN Haiyan, QIU Jiwen, School of Traditional Chinese Medicine, Tianjin University of Traditional Chinese Medicine, Tianjin 301617, China

〔Abstract〕 The paper introduces the concept, characteristics and study status of semantic knowledge graph and mapping knowledge domain, expounds the application of semantic knowledge graph and mapping knowledge domain in the field of Traditional Chinese Medicine (TCM), and analyzes the differences between the two types of knowledge graph and the effect of their interaction on promoting the development of knowledge graph technology in the field of TCM.

〔Keywords〕 semantic knowledge graph; mapping knowledge domain; Traditional Chinese Medicine (TCM)

〔收稿日期〕 2020-08-03

〔作者简介〕 孙肇阳,本科生,发表论文1篇;通讯作者:任海燕,博士。

〔基金项目〕 中国工程院2019年重大咨询项目子课题“人工智能在健康管理中的应用战略研究”(项目编号:2019-ZD-6-03)。

1 引言

知识图谱是显示知识发展进程与结构关系的一系列图形组合,中医药领域对于知识图谱的绘制流程、绘制软件及其应用等方面的研究热度逐渐提升^[1],但目前尚无研究从技术与应用角度对中医药

领域内的知识图谱做出明确区分。当前中医药领域知识图谱主要可分为两类,一是以语义知识图谱为代表的大规模语义网络,二是隶属于文献计量学的科学知识图谱。本文探讨两者所涉及的相关技术并论述其在中医药领域的具体应用。

2 语义知识图谱与科学知识图谱技术

2.1 语义知识图谱

2.1.1 概述 为提供海量数据的智能检索服务,Google公司构建统一结构化的语义知识图谱(Semantic Knowledge Graph, SKG)。为表述方便本文将建立在语义网技术基础上所有类似架构的知识库统称为语义知识图谱。SKG由实体和概念以及它们之间的语义关系组成,呈现为三元组形式,其构建流程主要涉及知识抽取、融合、存储3个阶段。在知识抽取阶段的关键技术主要有:实体抽取、关系抽取、属性抽取;在知识融合阶段的关键技术主要有:实体链接、本体构建、知识更新,最后通过表示语言例如RDF架构和图数据库进行存储。目前中医药领域内技术研究主要集中在本体构建和大型中医药数据库构建方面。

2.1.2 本体构建 张德政、谢永红和李曼等^[2]提出基于本体的中医核心知识图谱表示及其构建方法。崔家鹏、王彩霞和袁东超等^[3]研制出基于脾脏象理论的52种语义关系,绘制出脾脏象理论生理、病理、诊疗体系以及“脾主运化”、“脾主统血”等知识本体图谱,构建脏象理论知识体系。李新龙、刘岩和周莉等^[4]通过指南、标准、教材类文献梳理构建中医师辨证论治知识体系的顶层本体。中医药领域本体定义了词汇表中的基本术语及其关系,知识图谱在本体基础上对实体进行丰富和扩充,因此顶层本体构建是中医药知识图谱构建的首要环节。现阶段本体构建过程中对领域术语及其关系规范尚未达成共识,一定程度上影响了其对中医药知识图谱构建的指导意义,不同本体之间有待进一步整合。

2.1.3 数据库构建 相关研究^[5-8]提出以中医药学语言系统(Traditional Chinese Medicine Language

System, TCMLS)为框架,从数据来源、研究内容、图形化展示等方面探讨大型中医药知识图谱构建方法,建立中医药知识图谱网站:<http://www.tcm-kb.cn/kg/index.php>。郭文龙^[9]等经过分词和手工整理得到各类型方剂7710首,中药8957味,实现方剂组成可视化和知识图谱构建。针对构建中医药知识图谱中的文本抽取、关系数据转换、数据融合技术以及自动化构建方法、标准化流程,不少作者^[9-12]都进行了详细论述。大型知识图谱数据库构建是知识问答系统、辅助决策系统等具体应用实施的前提,如何保证知识图谱质量及应用效果是尚待解决的问题。

2.2 科学知识图谱

科学知识图谱(Mapping Knowledge Domain, MKD)起源于1955年加菲尔德^[13]提出的将引文索引应用于检索文献的思想。1965年普赖斯指出引证网络类似于当代科学发展的“地形图”,自此分析引文网络开始成为研究当代科学发展脉络的常用方法。2005年陈悦和刘则渊^[14]引入MKD概念,其绘制的一般流程为:数据检索——数据预处理——构建知识单元——数据分析——可视化与解读,核心技术包括引文分析、关键词共现、社会网络分析等,用于图形化展示各领域学科结构、各学科研究内容、学科间关系,识别和分析学科发展新趋势以及预测前沿,目前已成为国内热门文献研究方式。

3 SKG与MKD在中医药领域中的应用

3.1 SKG在中医药领域的应用

3.1.1 经方挖掘 尹丹、周璐和周雨孜等^[15]以中医经方为研究对象,利用拓扑有向图认知结构模拟“方证对应”过程,进行单个医案的经方用药分析及以方测证等功能的智能推理。赵凯、王华星和施娜等^[16]使用Neo4j图数据库构建基于《伤寒论》桂枝汤类方的小型语义知识图谱。运用图数据库中的查询语言模拟人联想思维进行检索,构建经方知识检索框架,是经方研究新方式。但经方文本语言难以精确结构化,其症状规律缺乏标准。例如将“啬

“恶寒，淅淅恶风，翕翕发热”规范为“恶寒，恶风，发热”并不符合文献本意。此外《伤寒论》原文省略表达较多，如第52条“脉浮而数者，可发汗，宜麻黄汤”，症状描述仅有脉浮数与常识不符，需要结合其他条文整体理解。因此如何规范整理经方原文是亟待解决的问题。

3.1.2 辅助诊疗 张莹莹^[17]初步构建中医药语义知识图谱并应用于舌诊系统中。聂莉莉、李传富和许晓倩等^[18]借助自然语言处理方法在传统两层模型（疾病-症候）基础上增加特征属性，构建基于“疾病-症候-特征”3层结构模型的医学诊断知识图谱。贾李蓉、刘丽红和刘静等^[19]基于中医药学语言系统设计构建知识问答系统，为用户提供辅助诊疗建议。董丽丽、程炯和张翔等^[20]创造性地提出融合医学知识图谱与深度学习的疾病诊断方法。单纯应用知识图谱查询语言模拟人的思维对疾病进行诊疗难以保证其准确率，而董丽丽提出的融合知识图谱与深度学习疾病诊断方法，较其他传统疾病诊断方法取得更好的分类性能。将知识图谱技术与人工智能相关技术相结合从而提高不同类型疾病的诊断准确率是知识图谱技术在辅助诊疗领域未来主要发展方向。

3.1.3 医案挖掘 邓宇、周卫强和张振铭等^[21]基于中医药行业术语规范标准将文本形式的医案转换为符合图谱要求的结构化数据，将中医医案知识结构化、层次化的关联予以呈现。许铮^[22]基于数据挖掘对名中医司徒仪教授治疗内异症痛经的用药规律进行研究。朱丹^[23]基于当代100余位名老中医以及姚乃礼辨治脂肪性肝病的文献、临证医案，总结当代名老中医治疗脂肪性肝病的证治规律，进行共性经验总结和个性经验分析。李新龙^[24-25]团队将辨证论治知识体系中显性知识与隐性知识综合分析，实现“病-证-治-效”等层面表达个体诊疗规律，以失眠症为例展示医师个体化辨证论治知识体系。名老中医经验及医案中包含大量隐性知识，与书本显性知识不同，隐性知识需要通过长期具体诊疗过程体悟积累，这是中医历来讲究师承的重要原因。构建名老中医经验、医案知识图谱，可以将碎片化的个体诊疗经验上升为中医界诊疗共识，是提高中医临证水平、推动中医药规范化和标准化的

关键。

3.2 MKD 在中医药领域的应用

通过检索中国知网（China National Knowledge Infrastructure, CNKI）数据资源系统中与中医药领域相关的 MKD 文献作为文献来源，检索式为：SU = 中医药 and (SU = 知识图谱 or SU = 可视化 or SU = CiteSpace)，检索时间为2020年5月12日，共检出文献169篇。通过人工筛选排除重复及与 SKG 相关论文，获得与 MKD 相关文献共计120篇。将文献数据导入 VOSviewer，加载同义词表从而绘制关键词共现 MKD，见图1。可以看出 MKD 在中医药学中的应用研究主要基于 CNKI、PubMed、Web of Science 等学术数据库，数据来源以期刊文献为主，研究内容主要为中医药领域研究现状及热点分析、学科前沿预测。多采用文献计量学相关方法，主要借助 CiteSpace 软件，少数使用 VOSviewer 软件构建研究者、研究机构、关键词等信息共现图和聚类图，从而展示常见疾病与疗法^[26]、中医经典研究^[27]、基础理论研究^[28]、方剂药物研究^[29-31]等方面发展态势和研究热点。

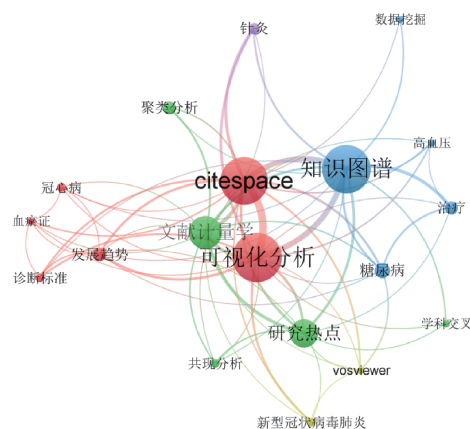


图1 MKD 在中医药领域的应用

4 讨论

4.1 SKG 与 MKD 的区别

对于 SKG 与 MKD 的本质区别，冯新翎、何胜和熊太纯^[32]等认为 MKD 本质是知识管理方法，一般与知识获取、组织、共享和创新密切相关；SKG

本质是知识库，参与知识获取、组织、存储和创新过程。可以将 MDK 理解为科学知识结构、关系及演化过程的可视化呈现手段，而 SKG 则是基于互联

网大数据构建的图数据库，SKG 拓展了原先 MKD 应用领域和构建技术，可视化呈现只是构建知识图谱数据库及其相关应用中微小的一环。

表 1 SKG 与 MKD 在中医药领域应用的区别

维度	SKG	MKD
数据来源	学术数据库	互联网大数据
软件工具	CiteSpace 为主	Neo4j、Protégé、WebOnto、MySQL 等
绘制流程	数据检索、数据预处理、构建知识单元、数据分析、可视化与解读	知识抽取、知识融合、知识存储
关键技术	聚类分析、共现分析	实体抽取、关系抽取、属性抽取、本体构建、实体链接、知识更新等
应用目标	分析中医药领域研究热点及发展趋势	构建问答系统、辅助诊疗系统医案挖掘，经方研究
学科定位	文献计量学	计算机科学
主要作用	可视化分析	数据存储与分析

4.2 SKG 与 MKD 交互

目前中医药领域已有 SKG 和 MKD 交互，即依托科研文献发现实体关系，实现从 MKD 计量关联到 SKG 语义关联转换。如李新龙等通过筛选指南、标准、教材类文献，借助预先设计的标准化数据提取表，人工提取文献所涉及中医辨证论治知识元素及各元素间联系，从而形成辨证论治顶层本体。借助学术文献数据，通过 MKD 计量关联向 SKG 语义关联转换，相较传统利用结构化大数据构建 SKG 的方式进一步拓展了知识深度与准确度，有助于中医理论体系深度构建并形成更加专业化、标准化的中医药知识图谱。

5 结语

本文总结归纳了语义知识图谱与科学知识图谱概念、特点，分析二者在中医药领域中的应用以及技术、应用层面的区别，提出由科学知识图谱计量关联向语义知识图谱语义关联的转换是构建中医药理论体系知识图谱的重要方式。

参考文献

1 李新龙, 刘岩, 何丽云, 等. 知识图谱研究概况及其在中医药领域的应用 [J]. 中国中医药信息杂志, 2017, 24 (7): 129 – 132.

2 张德政, 谢永红, 李曼, 等. 基于本体的中医知识图谱构建 [J]. 情报工程, 2017, 3 (1): 35 – 42.

3 崔家鹏, 王彩霞, 袁东超, 等. 基于本体的脾脏象理论知识体系构建研究 [J]. 中华中医药学刊, 2018, 36 (2): 388 – 391.

4 李新龙, 刘岩, 周莉, 等. 基于指南类文献的辨证论治知识体系顶层本体的构建 [J]. 中医杂志, 2018, 59 (13): 1154 – 1159.

5 贾李蓉, 刘静, 于彤, 等. 中医药知识图谱构建 [J]. 医学信息学杂志, 2015, 36 (8): 51 – 53.

6 于彤, 李敬华, 于琦, 等. 中医养生知识图谱的构建与应用 [J]. 中国数字医学, 2017, 12 (12): 64 – 66.

7 于彤, 李敬华, 朱玲, 等. 中医临床知识图谱的构建与应用 [J]. 科技新时代, 2017 (4): 51 – 54.

8 于彤, 刘静, 贾李蓉, 等. 大型中医药知识图谱构建研究 [J]. 中国数字医学, 2015, 10 (3): 80 – 82.

9 郭文龙. 中医方剂知识图谱构建研究与实现 [D]. 兰州: 兰州大学, 2019.

10 罗计根. 面向中医领域知识图谱构建的关键技术研究及应用 [D]. 南昌: 江西中医药大学, 2019.

11 阮彤, 孙程琳, 王昊奋, 等. 中医药知识图谱构建与应用 [J]. 医学信息学杂志, 2016, 37 (4): 8 – 13.

12 肖猛. 面向中医证候的健康领域知识图谱构建与应用研究 [D]. 长春: 吉林大学, 2019.

13 Eugene G. Citation Indexes for Science. A New Dimension in Documentation through Association of Ideas [J]. International Journal of Epidemiology, 2006, 35 (5): 1123 – 1127.

14 陈悦, 刘则渊. 悄然兴起的科学知识图谱 [J]. 科学学

- 研究, 2005 (2): 149 - 154.
- 15 尹丹, 周璐, 周雨玫, 等. 中医经方知识图谱“图搜索模式”设计研究 [J]. 中国中医药信息杂志, 2019, 26 (8): 94 - 98.
- 16 赵凯, 王华星, 施娜, 等. 基于 Neo4j 桂枝汤类方知识图谱的研究与实现 [J]. 世界中医药, 2019, 14 (10): 2636 - 2639.
- 17 张莹莹. 基于知识图谱的舌像诊疗系统研究与构建 [D]. 成都: 电子科技大学, 2019.
- 18 聂莉莉, 李传富, 许晓倩, 等. 人工智能在医学诊断知识图谱构建中的应用研究 [J]. 医学信息学杂志, 2018, 39 (6): 7 - 12.
- 19 贾李蓉, 刘丽红, 刘静, 等. 基于中医药学语言系统的知识问答系统的设计与构建 [J]. 中华医学图书情报杂志, 2019, 28 (5): 11 - 14.
- 20 董丽丽, 程炯, 张翔, 等. 融合知识图谱与深度学习的疾病诊断方法研究 [J]. 计算机科学与探索, 2020, 14 (5): 815 - 824.
- 21 邓宇, 周卫强, 张振铭, 等. 基于名老中医医案的知识图谱构建 [J]. 湖南中医杂志, 2019, 35 (7): 186 - 187.
- 22 许铮. 名中医司徒仪教授治疗子宫内膜异位症痛经的数据挖掘与实验研究 [D]. 广州: 广州中医药大学, 2018.
- 23 朱丹. 名老中医治疗脂肪性肝病的证治规律研究及知识图谱构建探索 [D]. 北京: 中国中医科学院, 2019.
- 24 李新龙. 中医师辨证论治失眠症知识图谱构建方法研究 [D]. 北京: 中国中医科学院, 2018.
- 25 何丽云, 李新龙, 刘岩, 等. 中医师辨证论治知识图谱构建的思路与方法 [J]. 中医杂志, 2017, 58 (19): 1650 - 1653.
- 26 刘莎莎, 孟洁, 董玉山. 基于 CiteSpace 软件的肝郁脾虚证方药研究的可视化分析 [J]. 中国中医药现代远程教育, 2019, 17 (14): 44 - 47.
- 27 陈陵芳, 邓晴, 晏峻峰. 近 10 年《素问·上古天真论》相关研究知识图谱分析 [J]. 中医杂志, 2017, 58 (23): 2054 - 2058.
- 28 崔德华, 张惠敏, 黄蓉, 等. 基于 CiteSpace 的中医体质学说知识图谱可视化分析 [J]. 中华中医药杂志, 2019, 34 (2): 598 - 601.
- 29 孟祥龙, 樊洲, 肖洋, 等. 中药炮制学学科研究前沿、热点及发展趋势——基于知识图谱的“中药炮制”可视化研究 [J]. 世界科学技术 - 中医药现代化, 2015, 17 (7): 1514 - 1524.
- 30 郭栋, 童元元, 黄生权, 等. 基于数据挖掘的枸杞研究热点分析 [J]. 中国中医药信息杂志, 2016, 23 (9): 48 - 51.
- 31 林骞, 高铸桦, 徐浩. 基于 CiteSpace 软件组方用药规律文献的可视化分析研究 [J]. 中西医结合心脑血管病杂志, 2019, 17 (19): 2901 - 2905.
- 32 冯新翎, 何胜, 熊太纯, 等. “科学知识图谱”与“Google 知识图谱”比较分析——基于知识管理理论视角 [J]. 情报杂志, 2017, 36 (1): 149 - 153.

2021 年《医学信息学杂志》征订启事

《医学信息学杂志》是国内医学信息领域创刊最早的医学信息学方面的国家级期刊。主管：国家卫生健康委员会；主办：中国医学科学院；承办：中国医学科学院医学信息研究所。中国科技核心期刊（中国科技论文统计源期刊），RCCSE 中国核心学术期刊（武汉大学中国科学评价研究中心，Research Center for Chinese Science Evaluation），美国《化学文摘》、《乌利希期刊指南》及 WHO 西太区医学索引（WPRIM）收录，并收录于国内 3 大数据库。主要栏目：专论，医学信息技术，医学信息研究，医学信息组织与利用，医学信息教育，动态等。读者对象：医学信息领域专家学者、管理者、实践者，高等院校相关专业的师生及广大医教研人员。

2021 年《医学信息学杂志》国内外公开发行，每册定价：15 元（月刊），全年 180 元。邮发代号：2 - 664，全国各地邮局均可订阅。也可到编辑部订购：北京市朝阳区雅宝路 3 号（100020）医科院信息所《医学信息学杂志》编辑部；电话：010 - 52328672，52328686，52328687，52328670。

《医学信息学杂志》编辑部