

中医知识图谱构建及应用综述

杨一帆, 刘忠博, 周玉新

(内蒙古民族大学 计算机科学与技术学院, 内蒙古 通辽 028000)

摘要: 如今中医已开始朝着数字化与智能化方向发展, 而知识图谱作为人工智能的重要分支技术正逐步应用于各个领域。在此背景下, 中医知识图谱应运而生。其通过建模关键实体和语义关系全面整合了中医知识, 为医生系统学习以及对患者的辅助诊断起到了积极的推动作用。介绍了中医知识图谱构建所需的方法, 并对多种方法进行了总结与比较。最后, 对中医临床决策诊断、中医智能语义搜索、中医智能问答、中医数据挖掘以及中医知识图谱面临的挑战与未来展望进行讨论。

关键词: 中医知识图谱; 自然语言处理; 知识图谱构建

DOI: 10.11907/rjdk.241436

中图分类号: TP391.1

文献标识码: A

开放科学(资源服务)标识码(OSID):

文章编号: 1672-7800(2025)008-0017-10



Review of the Construction and Application of Traditional Chinese Medicine Knowledge Graph

YANG Yifan, LIU Zhongbo, ZHOU Yuxin

(School of Computer Science and Technology, Inner Mongolia Minzu University, Tongliao 028000, China)

Abstract: Traditional Chinese medicine is gradually entering the era of digitization and intelligence, and knowledge graphs, as an important branch technology of artificial intelligence, are gradually being applied in various fields. In this context, the traditional Chinese medicine knowledge graph emerged, which comprehensively integrates traditional Chinese medicine knowledge by modeling key entities and semantic relationships, playing a positive role in promoting the systematic learning of doctors and assisting patients in diagnosis. This article introduces the methods required for constructing a knowledge graph of traditional Chinese medicine, and summarizes, compares, and applies various methods. Finally, the challenges and prospects encountered in clinical decision-making diagnosis, intelligent semantic search, intelligent Q&A, data mining, and knowledge graph of traditional Chinese medicine will be discussed.

Key Words: traditional Chinese medicine knowledge graph; natural language processing; knowledge graph building

0 引言

随着人工智能和自然语言处理技术的快速发展, 信息也呈几何倍数增长。中医作为中华民族的瑰宝, 相较于西医更注重对身体整体的调养, 具有药物副作用小和治愈后不易复发的特点, 在国外也备受推崇。使用知识图谱构建技术构建一个中医知识图谱对于大众认识与使用中药, 以

及中医从业者进行系统化查询与借鉴可起到重要作用。中医知识图谱中有大量的中医方剂、临床电子病历、中医典籍等中医数据, 好的构建技术可以挖掘中医知识中的隐性联系, 对于中医知识的保留与更新有着重要意义。中医作为我国的传统医学, 包含了病机、病理、病灶、病位、病人体征等不同信息。在以知识关联为主的 Web3.0 背景下, 如何对海量的电子病历、临床方案、中医典籍、中医文献、中医网站等中医信息进行更高效的提取与应用成为研究

收稿日期: 2024-05-09

扫描二维码阅读全文:



基金项目: 内蒙古自治区自然科学基金面上项目(2022MS06028); 内蒙古自治区研究生科研创新项目(KC2024074S); 内蒙古民族大学智慧农牧创新团队项目(2021); 内蒙古民族大学博士科研启动基金项目(BS438)

作者简介: 杨一帆(2000-), 男, 内蒙古民族大学计算机科学与技术学院硕士研究生, 研究方向为知识图谱; 刘忠博(1999-), 男, 内蒙古民族大学计算机科学与技术学院硕士研究生, 研究方向为知识图谱; 周玉新(1977-), 男, 硕士, 内蒙古民族大学计算机科学与技术学院副教授、硕士生导师, 研究方向为数据挖掘。本文通讯作者: 周玉新。

人员需要面对的问题^[1]。

在 AI 研究的早期阶段,人们就已经开始思考如何将知识以结构化的形式表示出来,这直接促成了语义网络的发展。2012 年,Google 正式提出知识图谱(knowledge graph),其本质上就是语义网中的知识库。作为大数据时代的关键技术之一,知识图谱在实现智能化语义检索方面发挥着至关重要的作用。其提供了可由计算机读取和解释的语义结构信息,将信息组织成一种结构化的图谱,其中包含了大量的实体、属性和关系^[2]。该结构使搜索引擎可以更加准确地理解用户的查询意图,并以更深层次的语义理解来匹配相关内容,从而提供更精确、个性化的搜索结果。同时,知识图谱是动态的,可以随着新数据的不断加入和知识的不断更新而持续演进,从而保持时效性和准确性。

在知识图谱领域,由于中医属于中国特有的医学,所以中医知识图谱在国外文献中并没有太多介绍。在国内,有由上海曙光医院建立的中医知识图谱^[3]、中国中医科学院中医药信息研究所建立的中医临床知识图谱,以及王松^[4]建立的周仲瑛辨治肺癌知识图谱,还有一些中医咨询平台如中医通、小鹿中医、中医堂等。

目前现有的知识图谱可以分为通用知识图谱和领域知识图谱。其中,通用知识图谱一般是以自底向上的方式构建,大多是面向全领域的,且主要应用于互联网推荐、搜索和问答中。目前拥有世界上最大的通用知识图谱的两大计算机公司分别是微软和谷歌。而领域知识图谱一般是以自顶向下的方式构建,多面向单一领域,因而覆盖的知识粒度更细、对领域知识处理更细致,且对领域知识的深度检索要求更高。于彤等^[5]创建了包含很多中医内容的中医药大型知识图谱。本文讨论的中医知识图谱属于

领域知识图谱。在医学领域,收集和整理相关信息时必须极其严谨,以确保数据的深度、完整性和准确性。

本文介绍了中医知识图谱的基础构建方法:首先,通过文本挖掘技术从中医经典文献、医案和科研论文中提取中医实体及其关系。其次,利用自然语言处理技术进行实体识别和关系抽取,生成初步的知识图谱。再次,通过专家标注和领域知识校正,提高知识图谱的准确性和可靠性。最后,采用图数据库存储和管理这些数据,进而形成一个系统的中医知识图谱,并对每一步应用的现有技术进行对比和分析。本文归纳了中医知识图谱在临床决策支持诊断、智能语义搜索、中医智能问答、中医数据挖掘中的应用,并总结了中医知识图谱研究现状和未来的发展方向。

1 中医知识图谱构建

中医知识图谱属于领域知识图谱,相较于通用知识图谱的构建难度更大。本文介绍了中医知识图谱里的中医数据来源、中医知识表示、中医知识抽取、中医知识融合、中医知识存储、中医知识推理、中医图谱质量评估方法共 7 个模块。首先将不同类型的中医数据进行知识抽取,然后将抽取的异构知识进行融合,再使用三元组的形式将处理好的知识进行表示,最后将处理好的中医信息放到知识库中进行知识存储。对于因信息不全导致中医知识图谱不够完善的问题,可以利用知识推理对不完善的地方进行知识补全。对于已建立好的中医知识库,可以采用中医图谱质量评估方法来提高中医知识图谱质量。中医知识图谱构建具体过程如图 1 所示。

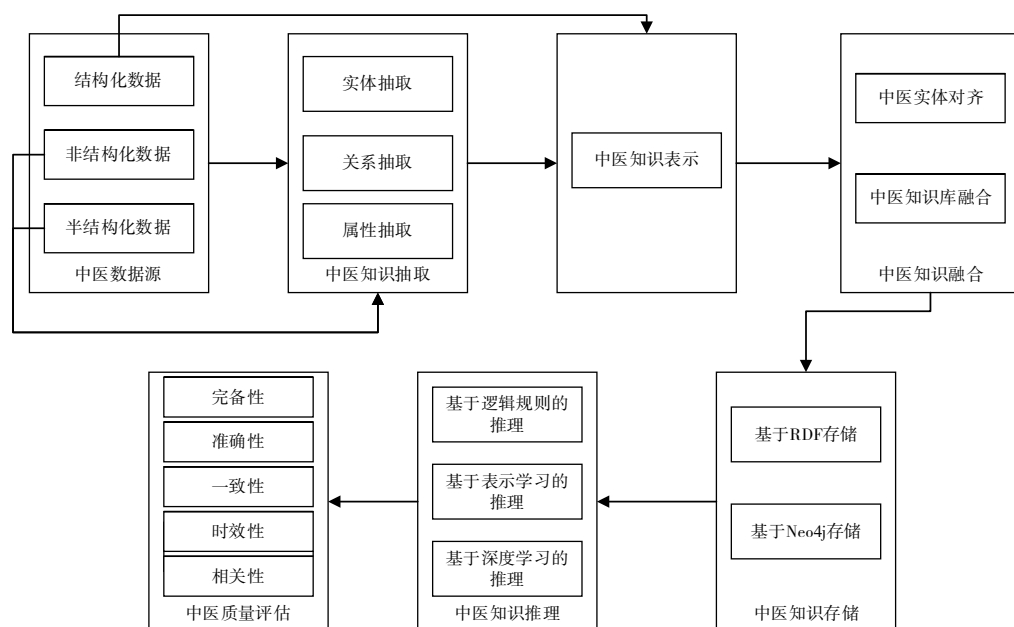


Fig. 1 Construction process of traditional Chinese medicine knowledge graph

图1 中医知识图谱构建过程

1.1 中医数据源

如今随着信息的井喷式发展,中医数据也随之大量增加。当前这些数据源大体可以分为 3 类,分别是:结构化数据、半结构化数据与非结构化数据。结构化数据有明确定义的格式和结构存储,易于组织、存储和分析。例如:患者的基本信息(姓名、体重、年龄)或者生理参数(血压、体温、心率)等都属于结构化数据。半结构化数据大多以 XML (eXtensible Markup Language)、JSON (JavaScript Object Notation) 等格式存储,具有一定的结构,但不符合严格的数据库模式。半结构化数据通常包含标记或标签,便于机器理解和处理,但仍需要一定程度的解析才能进行有效的分析。例如,电子病历中的医嘱信息、检查报告中的描述性文字等都属于半结构化数据。非结构化数据则没有明确的结构和格式,通常以自然语言、图像、音频、视频等形式存在。非结构化数据是医学数据中数量最大、最复杂的一部分,包括病历文本、医生诊断报告、手术记录、医学文献、患者病史描述、医学图片中的文字等。由于非结构化数据的复杂性和多样性,对其的分析和挖掘常常需要借助图像处理、机器学习、自然语言处理等技术手段。中医常用数据库如表 1 所示。

其中,中国中医药数据库官网中的数据相对较全,其他数据源收集的数据均针对不同领域。

1.2 中医知识抽取

知识抽取是构建知识图谱的关键步骤之一。知识抽取是从文本或数据中自动或半自动地提取有价值的知识的过程。在知识图谱构建过程中,对于半结构化数据和非结构化数据需要提取其中的实体、关系和属性。由于人工提取代价太大,所以如何设计出一种可以由计算机自动提取的技术逐渐成为知识图谱构建的重点研究方向。本节主要介绍在中医知识图谱构建过程中,从数据源中自动提取中医实体、中医关系和中医属性。

1.2.1 中医实体抽取

实体抽取又称为命名实体识别(Named Entity Recognition, NER),其目标是从文本中识别和提取出具有特定意

义的实体(如主症、舌象、脉象等),这与最终的中医知识图谱质量以及在获得中医实体时的准确率、召回率、F1 值息息相关。早期的实体抽取采用基于中医词典及规则的方法,该方法依赖于中医领域的专业知识,中医专家通过人工定义的规则与模式匹配的方式构建词典,或者在现有中医词典中抽取中医实体。其不仅要求开发人员具有专业的中医知识,而且需要具备一定的计算机知识。但简单的文本匹配算法通常无法满足实体识别任务的要求。当遇到需要处理中药命名实体中复杂实体的情况时,由于生物命名实体具有多样性和复杂性,常常遵循各种不同的命名规范和形式,因此构建完整的字典会变得十分困难,也无法解决中医中不断出现的新名词问题。为此,有些研究人员开始将中医数据源放入基于机器学习算法与统计模型相结合的大模型中进行实体提取,利用已标注好的语料进行模型训练。常用的相关模型有:线性回归(Linear Regression)、逻辑回归(Logistic Regression)、决策树(Decision Trees)、随机森林(Random Forest)、支持向量机(Support Vector Machine, SVM)、K-近邻(K-Nearest Neighbors, KNN)、朴素贝叶斯(Naive Bayes)和神经网络(Neural Networks)等。机器学习算法可以从数据中学习模式和规律,从而适应不断变化的环境和数据。这一特性使其适用于人工标注专业性不足,且需依赖大规模数据构建复杂关系的任务。

2011 年,Collobert 等^[6]提出一个深层神经网络模型,其生成的结果和运行性能优于传统算法,这也是在深度学习中具有代表性的模型之一。王松等^[7]使用 Bert-BiLSTM-CRF (Bidirectional Encoder Representation from Transformers-Bidirectional Long Short-Term Memory-constant rate factor)对国医大师周仲瑛教授的临床肺癌数据进行知识提取,解决了处理非结构化数据可能需要浪费大量人力以及出现消息遗漏等问题。高晓苑等^[8]在以医案文本为数据集的情况下,将 ERNIE-BiLSTM-CRF 模型分别与 BiLSTM-CRF 和 Bert-BiLSTM-CRF 模型准确率、召回率及 F1 值的变化情况进行比较,最终得出使用在 BERT 模型基础

Table 1 Commonly used databases of traditional Chinese medicine

表 1 中医常用数据库

数据源	简介	数据量
中医古方数据库	方源来自《千金要方》《太平圣惠方》《普济方》等 50 余本经典古籍(数据以 XML 的形式存储)	收录了春秋战国以来的古方 16 万余条
药融云—中医药数据库群	在中成药处方和少数民族药方面做得非常完整,很多查不到的医药信息都可以通过该数据库进行查询(数据以 XML 的形式存储)	成药处方(40 000+)、中医药分类(1 200+)、中医药保护品种(4 000+)、中医病症分类与代码(4 000+)、少数民族药(3 000+)
TCMID 中国中医药数据库	于 2005 年首次发布,其数据采集来源于多个国家(数据以 XML 的形式存储)	提供中医方剂及疾病治疗信息(7 000+)
NPACT 中国中医药数据库	NPACT 是天然存在且基于植物的抗癌化合物—活性—目标数据库,这些化合物具有抗癌活性(数据以 XML 的形式存储)	包含 1 574 个条目
TCM Database	该数据库从前号称是全球最大的计算机药物筛选中国中医药数据库,随着项目扩展升级,现已发展为 TCMBank(数据存储在 MYSQL 中)	该数据库包含来自 37 种中医药成分的 17 032 种中医药化合物
中国中医药数据库官网	由中国中医科学院中医药信息研究所建设,包含大量中医相关数据库(数据以 XML 的形式存储)	采集数据近 7 万条

上将一个短语或实体作为基本单元而改进的 ERNIE 融合 BiLSTM-CRF 模型处理医案文本,效果优于其他两个模型,F1 值显著提高。雷迪等^[9]引入新模型 Lattice-LSTM,在中医药医案数据集下对比了 Lattice-LSTM-CRF 模型与 LSTM-CRF 模型,得出改进后的模型对于中医药领域的实体抽取更具优越性,其对中医药医案中的实体边界有更好的区分效果。但在实体提取时,对于多元化关系的处理能力还有待提高。

1.2.2 中医关系抽取

关系抽取是指在大规模文本数据中自动提取两个或多个实体之间关系的过程。中医领域的实体关系包括病机—症状体征、病机—病理因素、病机—病位、病理因素—中药等。早期用于中医关系抽取的模型是基于模版的关系抽取方法,该模版是由行业专家对将要使用的数据集进行总结归纳得出的。例如从“太子参味道甘、微苦,主治食欲不振、自汗口渴”和“党参味道甘,主治气津两虚、气血双亏”中,可以得到模型[X]的味道是[Y]、主治[Z],或者主治为[X]且味道为[Y]的中药是[Z]。将总结后的模版带入要使用的数据集中进行匹配,即可得到“药名—味道—主治”3个实体中的两个关系。该方法对于抽取特定数据集中实体关系的准确率较高,但是召回率和可移植性较差。应用较多的是基于监督学习的关系抽取方法,其实际上是将之前的关系抽取问题分解为分类问题,可以使用机器学习和深度学习方法对数据进行分类。机器学习需要一些人工标注的数据集来训练其他未标注的数据集,深度学习则可以自动从数据中提取其特点。马月坤等^[10]构建一种针对中医古籍的隐式关系抽取方法,该方法首先定义了8种构式特征与5种构式类型,再使用 CART(Chimeric Antigen Receptor T-Cell Immunotherapy)分类模型进行文本分类,并对其中4类构式制定抽取规则,实现隐式关系抽取。最后将得出的结果在古籍数据集中进行测试,验证了该方法的有效性。潘雪峰^[11]引入新模型 ALBERT-BiLSTM-CRF,在古籍关系数据集上识别实体的精确率为95.65%,召回率为93.97%,F1 值为94.80%,取得了较好效果。其在使用 ALBERT(A Lite BERT)完成词嵌入后,利用 BiLSTM 模型高效解决了梯度爆炸与输出标签之间的相互依赖问题,最后使用 CRF 作为最后的输出层。马郅斌^[12]使用基于正则化的方法改进了 Transformer,减少了模型的复量计算,提升了稳定性。使用相同数据集进行对比实验后,发现改进后的 Transformer 相比原始模型具有更好的效果。然而,目前使用的关系抽取推理规则只适用于特定任务,泛化能力不足。

1.2.3 中医属性抽取

中医知识图谱中的属性抽取是指从文本或其他非结构化数据中提取实体的属性信息,并将其转化为结构化的数据形式,以便构建和更新中医知识图谱。其中,属性是指描述实体的特征、关系或属性的名称或标签。中医知识

图谱中的属性包括中医药的“性、味、归、经”等。属性值是指与实体相关联的特定属性的具体内容或取值,中医知识图谱中的属性值可以是“某一味药用量的具体数值、某一副汤药的炮制时间”。每个实体可以有多个属性,而每个属性都对应着一个或多个属性值。在知识图谱中,属性和属性值的使用使得实体的特征及关系能够以结构化的方式进行存储与管理,从而支持复杂的查询、推理和分析操作。通过定义合适的属性和属性值,可以更好地组织和理解知识图谱中的信息,进而提供更丰富和准确的知识服务。胡定兴^[13]提出 EMR2vec-BLSTM-CRF 模型用于中医临床文本属性抽取,其通过大量中医临床文本语料预训练 Word2vec,生成字向量(CharacterVector of TCM Electronic Medical Record, EMR2vec)作为初始特征,并在 TCMCS-EAE 和 TCMDI 数据集上进行实验验证。结果显示,该模型的 F1 值相较于 BLSTM-CRF 模型有所提升。

综上,传统基于规则和词典的方法在某些特定情况下仍具有优势,如高精度和领域特异性。然而,随着数据复杂性的增加,结合机器学习和深度学习的方法显得尤为重要。这种综合方法能够平衡精确度和适应性,并处理更多样化的数据。深度学习模型如 BERT、ERNIE 和 Lattice-LSTM,在完成中医知识抽取任务时表现出了显著的优越性。尤其是在处理非结构化和复杂的中医文本时,深度学习能够更好地捕捉上下文信息,提高抽取的准确率和召回率。但在训练机器学习和深度学习模型时,高质量的标注数据集至关重要。因此,构建和维护高质量的中医标注数据集应当是一个持续的重点工作。

1.3 中医知识表示

知识表示是为描述世界所作的一组约定,是知识符号化、形式化、模式化的过程^[3]。知识表示学习(Knowledge Representation Learning, KRL)的重点是知识在计算机中的存储方法,不同的表示方式在计算机处理过程中会有不同的存储方式与运行效率。由于中医数据种类众多、格式繁杂,且存储方式不同,因此知识表示较为困难。

目前中医知识表示的主流方式为三元组(实体1—关系—实体2)形式。该知识表示方法与国际医学语义挖掘工具 Sem Rep(Semantic Knowledge Representation)采用的知识表达形式高度一致。Sem Rep 是由美国国立医学图书馆研发并基于 UMLS(Unified Medical Language System)构建的超级叙词表、语义网络、专家词典和词汇工具,能够对科技文献进行分词、语义分析、短语映射、语义谓词归一化、语义约束等处理,最终得到 SPO(Subject-Predication-Object)三元组^[14]。SPO 语义三元组因具有较强的语义表达能力,在中医领域的运用极为广泛。刘灿^[15]采取句法树匹配分析方法构建三元组结构,将不同类型的知识表达为相同的结构。在中医领域,还可以将图算法加入三元组的表示中。但因为图的结构复杂,导致图算法的时间复杂度较高,在一些规模较大的中医知识图谱中会出现计算时间

过长的问题。

随着机器学习的发展,在知识表示中可以将研究目标的语义数据转化为紧凑的低维度向量,从而提高数据的信息密度。按照类型不同,可分为翻译模型、语义模型、神经网络 3 类,如表 2 所示。

1.4 中医知识融合

概念规范化是医学知识融合的核心所在,包括对医学知识进行数据整合、实体对齐、知识消歧和加工的过程,旨在应对医学知识因来源不同而导致的重复、质量参差不齐以及关联不清晰等问题。其目标在于消除多余和错误的知识信息,从而提升医学知识数据内部的结构和逻辑完整性。中医知识融合又可分为实体对齐、知识库融合两部分。

随着实体数量的增加和知识库的不断扩大,对知识库中实体的对齐变得越来越重要。未来知识融合研究的重点之一就是如何高效、准确地进行实体对齐。中医数据在不同的数据源中存在多名称但指代物相同的情况,如:在症状方面,视物欠清、视力模糊、视糊可统一归纳为视物模糊;在病机方面,化疗伤正伤胃、胃虚脾弱、脾胃两伤可统一归纳为脾胃虚弱;在中药方面,炮山甲、炙山甲可统一归纳为穿山甲。由此可见,实体对齐在中医融合中是一个非常重要的步骤。目前对齐算法可分为成对实体对齐和集体实体对齐两种。

(1)成对实体对齐通常发生在不同数据源、不同知识

图谱或不同版本之间,旨在找到其之间的匹配关系,使这些实体可以在不同的上下文中被统一处理。成对实体对齐方法涵盖了基于概率模型和基于机器学习两种主要方法。基于概率模型大多使用 Fellegi 等^[16]提出的用分类问题代替属性相似性评分的实体对齐问题模型。基于机器学习的有监督实体对齐方法包括集成学习框架^[17]、ID3 决策树算法^[18]与 SVM 分类方法^[19]。

(2)集体实体对齐是一种建立在成对实体对齐基础上的方法,其在计算实体相似度时考虑了实体之间的相互关系。该方法包含局部集体实体对齐和全局集体实体对齐两种类型。局部集体实体对齐的召回率和运行速度有一定优越性,但是准确率还有待提升。全局集体实体对齐通过不同匹配结果的相互制约来影响实体间的相似度。在从不同数据源进行实体对齐时,需要分辨其中哪一个数据源的可信度最高,以及相同的数据在各个数据源中出现的频率。阮彤等^[20]在融和多个中医药知识图谱时,针对不同数据源进行了可信度评分,并根据相同数据在不同数据源中出现的次数对数据单元进行排序,以更好地进行实体对齐。

当前大部分中医知识图谱都是围绕某一种疾病或者某一种药物而构建的,例如脾胃病知识库^[21]、中医药知识图谱^[2]等。不同数据库在建立时有不同的需求和框架,使得数据库中的数据呈现多样性。如果想要得到一个大规模的中医数据库,则需要对多种不同的中医数据库进行融

Table 2 Summary of knowledge representation model

表 2 知识表示模型总结

模型	代表模型	优点/缺点	应用
翻译模型	TransE	模型简单、复杂度低,适用于稀疏中医知识库,但无法处理多元关系,仅可有 1-1 关系	由于中医知识图谱中的许多关系是较为简单的单一关系,TransE 可以高效处理简单关系,但对复杂关系的处理能力有限
	TransR	考虑关系的多样性,可提高映射模型的表达能力,但不能区分头尾实体,不适用于大规模知识图谱	中药材往往具有多种功效,不同功效在不同的临床情况下表现不同,TransR 能更好地捕捉这种复杂性
	TransD	加强实体和关系的联系,但空间维数与关系维数不能不同	在中医领域,不同药材和症状的差异较大,TransD 的动态表示有助于捕捉这些异构信息
	PTransE	使用头尾实体的多关系路径,但无法处理异构实体	中医治疗讲究辨证论治,不同路径对应不同治疗方案,PTransE 能捕捉这种复杂的路径关系,提高知识图谱的应用价值
	TransG	考虑实体间的多重语义关系,但使用高斯分布计算繁琐	中药材具有多义性和多重功效等特点,TransG 能有效处理和表示这种复杂性,提高知识表示的准确性
	TransSep	可处理实体异构,但只可在特定领域使用	在中医知识图谱中,需要明确区分不同药材和症状的关系,TransSep 能提高模型的区分能力,减少混淆
语义模型	RESCAL	可以描绘实体关系的二阶联系,但不能计算非对称关系	中医治疗方案中药材间的相互作用复杂,RESCAL 能有效表示这些交互关系,但计算复杂度较高
	DistMult	可以使用少量参数,但只能计算对称关系	中药材的某些功效具有对称性(如互补性和相互增强性),DistMult 可以高效地捕捉和表示这些关系
	双线性隐变量模型	使计算复杂度下降,但使用参数多、复杂度高	中医药方中的药材搭配和整体疗效需要复杂的关系表示,双线性隐变量模型具有更强的表达能力
神经网络	矩阵分解模型	可以有效降维,提升灵活性,但只能使用共现矩阵	中医知识图谱中有大量的药材和配伍关系,矩阵分解模型可以高效处理和简化这些关系,但可能无法捕捉更复杂的关系细节
	CapsE	可以捕捉更多三元组特征,但计算复杂度高	中医知识图谱中的层次结构和复杂交互(如症状与治疗路径的层次关系)需要更复杂的表示,CapsE 拥有更具表现力的结构建模方法,能够精准地捕捉和表达此类复杂语义

合。中医数据库的构建也是一个不断完善、更新的过程。MED-BBK-9K 数据集是腾讯天衍实验室根据真实医疗数据集和互联网中的医疗知识图谱融合而成的,其通过结合下游应用程序的反馈和知识抽取技术提出基于本体指导的知识图谱对齐系统 OntoEA,以减少对齐结果中的类别冲突,提升长尾实体的对齐效果,并顺利实现了 9 162 组实体的一对一高质量匹配^[22]。在医学数据库中,Dieng-Kuntz 等^[23]对医疗数据库进行了转换,形成了医疗本体,并利用半自动语言工具提取了其他文本语料的语义信息。在人工干预过程中,其对本体进行了扩展和改进,并使用启发式规则自动构建了知识的概念层次结构。此外,还可以使用数据联邦的方法,在保持数据源独立的同时,对多个数据源进行统一查询和访问,或者由数据仓库通过 ETL (Extract, Transform, Load) 过程将不同数据源的数据抽取、转换和加载到统一的仓库中,实现对数据集的存储和管理。目前中医领域的数据库融合仍需要大量人工参与,如何建立完备的中医数据库还有待进一步研究。

1.5 中医知识存储

知识存储是指对各种形式的知识和信息组织、保存与管理的过程。其是为了提供一个可靠、安全、高效的平台,用于存储和管理知识资源,以便后续的访问、查询和利用。当前知识存储主要分为基于资源描述框架(Resource Description Framework, RDF)存储和图存储两大类。RDF 是一种表示数据的模型,以三元组的形式来描述数据之间的关系,从而为数据的组织、管理和利用提供了一种强大的方法,有助于实现语义化、智能化的数据处理和应用。基于 RDF 的存储常用于语义网、Linked Data 和知识图谱等应用领域,如 DBpedia、Wikidata 等。基于图的存储是一种用于存储和处理数据的方法,其中数据以图的形式组织和表示。图由节点和边组成,节点表示实体或对象,边表示节点之间的关系或连接,从而为数据的组织、管理与分析提供了强大的工具和平台。常见基于图的存储系统包括图数据库、图计算框架。

知识存储也是中医知识图谱构建中的重要环节,目前基于图存储的方式 Neo4j 在中医知识图谱构建中逐渐成为主流。Neo4j 是一种流行的图数据库管理系统,专门用于存储和处理图形数据。作为一种 NoSQL 数据库,Neo4j 采用了图形数据库模型,将数据表示为节点、关系和属性的集合,从而可以有效地表示和处理复杂的数据结构与关系。王明强等^[24]在明确应用“疾病—证候—治疗”关系的基础上,采用 Neo4j 进行图数据库构建,使得本体具有较强的扩展性和较高的数据准确性。

1.6 中医知识推理

推理是从已知事实中挖掘出其中隐含的未知信息,而中医知识图谱中知识推理的重点在于依照现有知识,通过对不同推理方法的选择和应用,以达到减少人工参与和更准确地推理出缺失事实的目的。在中医知识图谱中,知识

推理可以更好地收集病患信息、辅助主治医师给出更好的解决方案,以及补全一些中医知识图谱中残缺的部分。中医知识图谱推理方法可以分为:基于逻辑规则的推理、基于图结构的推理、基于表示学习的推理、基于神经网络的推理、基于深度学习的推理。

基于逻辑规则的中医推理机制旨在将中医理论、诊断方法和治疗原则形式化,以便在计算机系统中实现自动化推理。其中,一阶谓词逻辑将命题视为推理的基本单元,命题被定义为含有个体及其所满足或不满足的属性或关系的陈述。如边红^[25]利用中医数据构建了中医胃病的知识图谱,并根据成型的知识图谱构建辅助诊断系统。在专家系统规则库构建时,采用关联规则提取的软集包含度算法,使得专家知识的来源更加可靠,还可以将患者输入的医学信息转变为置信结构,然后使用产生式规则和模糊推理来总结患者所给出的信息,进而向患者提供相应的诊断结果。将基于路径的推理规则纳入中医知识图谱推理过程中,可以显著提高推理性能。如朱吕行^[26]实现了通过学习已知的“疾病—基因”关联信息,为未知的“疾病—基因”关联作出预测。其利用深度游走算法在整个网络中进行游走搜索,获取每个节点在游走路径下的嵌入表示向量,然后利用图卷积神经网络提取的向量特征进行结果预测。

基于表示学习的中医推理主要是学习三元组中实体和关系的低维嵌入表示,然后基于向量化知识表示进行计算和知识图谱推理^[27]。研究者们为了给多跳关系进行建模,提高推理能力,提出很多基于路径的知识表示学习模型。如郑子强^[28]构建的慢性肾脏病知识图谱中,将慢性肾脏病诊疗过程归纳为知识图谱多层实体间的链路预测任务。在表示学习的基础上,使用多维卷积对图谱中实体和关系形成路径中产生的隐含特征进行提取。经过实验验证,发现多个指标均很优异,证明提取的效果非常好。

随着深度学习的广泛应用,知识表示在深度学习中也备受关注,成为一个热门研究领域。基于深度学习的中医推理机制利用深度神经网络对中医知识进行建模,从而实现复杂的自动化推理和预测。目前在中医推理中使用最多的是基于循环神经网络(Recurrent Neural Network, RNN)、卷积神经网络(Convolutional Neural Networks, CNN)和图神经网络(Graph Neural Networks, GNN)的知识推理方法。循环神经网络不仅可以处理可变长度的序列输入,并通过记忆先前的信息来影响后续输出,而且可以通过引入循环连接使网络可以在处理长序列时保留并利用先前的信息,也更容易捕捉到长期的依赖关系,从而提高模型性能。卷积神经网络在设计中使用了卷积层和池化层,使得网络能够对输入数据进行局部感知。通过卷积操作,网络能够有效地捕获输入数据中的局部模式和特征,而通过权值共享机制,网络可以在不同的位置共享参数,从而减少了模型参数量,提高了模型的泛化能力。图神经网络(GNN)具有传递性质,即节点的表示可以通过图的连接传递给其邻居节点,

再传递给邻居的邻居,以此类推。这种传递性质使 GNN 能够在表示学习过程中获得全局信息。如罗生根^[29]通过书籍《中医证候鉴别诊断书》和《中医证候辨证论治辑要》构建中医知识图谱,并提出一种基于路径的推理算法,可推断出两

个没有直接联系的实体的间接关系。如想要查询证候和一味药物有无关系时,首先找到治疗证候时所要使用的方剂,再从方剂中查找是否有要查询的中医药来证明其之间的关系。中医知识推理总结如表 3 所示。

Table 3 Summary of TCM knowledge reasoning
表 3 中医知识推理总结

对比维度	基于逻辑规则的推理	基于表示学习的推理	基于深度学习的推理
数据需求	需要明确的规则和结构化数据	需要知识图谱和嵌入表示的数据	需要大量标注的训练数据
知识表达	使用明确的逻辑规则表达	使用低维向量表示	使用深度神经网络表示
推理能力	能处理明确的逻辑关系和规则	能处理复杂的多样关系	能捕捉复杂模式和高阶关系
可解释性	可直接解释推理过程	通过向量操作解释	需要借助注意力机制等方法
计算复杂度	低,规则执行效率高	中,依赖嵌入向量计算	高,训练和推理需要大量计算资源

1.7 中医图谱质量评估技术

知识图谱质量评估是指通过标准的程序来提高知识图谱质量,以保证获得有更高置信度的知识,从而得到一个更高质量的知识图谱。目前中医知识图谱质量评估主要分为完备性、准确性、一致性、时效性和相关性 5 个维度^[30-33]。完备性是数据质量评估中最重要的维度^[34]。完备性是指知识图谱中所包含的信息是否涵盖了领域内的所有重要实体、关系和属性。如在添加某一味中医药时,也要添加其“味道”“作用”等属性。缺乏完备性可能导致知识图谱在某些情况下无法提供全面、准确的信息,从而影响其在搜索、推理、问答等应用中的效果。准确性是中医知识图谱评估中的关键指标。准确性描述的是知识图谱中数据的正确程度,也可以定义为数据正确反映现实世界事实的程度^[35]。准确性的高低与用户查找知识的正确性密切相关,准确性高的中医知识图谱可使用户在搜索时得到一个正确率高的结果。一致性描述了知识库在特定的知识表示形式和推理机制方面有无逻辑或形式上的矛盾^[35]。如中医中对于一种病灶的操作规范流程,每一次应该是相同的。一致性评估通常需要进行逻辑推理和分析,以发现潜在矛盾和不一致之处。因为知识图谱中一致性的缺失可能导致信息不可信、推理错误等问题,所以确保高度的一致性对于提高中医知识图谱的可靠性和可用性至关重要。时效性指信息的最新程度,知识图谱的时效性可以指知识更新的程度^[36-37]。例如一个患者购买过一副药后应该在系统里及时添加,防止医生为患者重复开药。时效性评估旨在确定知识图谱中的数据是否保持与现实世界同步的状态,以及是否能够及时反映出现实世界的变化。Pipino 等^[38]指出时效性是知识来源和内容评估中不可忽视的因素。相关性评估旨在确定知识图谱中的数据是否与用户查询或应用场景相关,并且是否能够满足用户的信息需求。例如患者在搜索治疗发烧的药物时,相关性强的结果包含发烧时的症状、发烧时该吃的方剂、发烧时的注意事项等,相关性差的结果可能会包含手脚冰凉如何治疗,或者吃什么中药可以提高发质的质量等。相关性强弱决定着用户查询结果是否可以获得所需的信息。在实际构建中医药知识图谱时,质量评估标准无法统一,应当

在构建完成后经由专家或查阅资料对 5 个维度进行评估,或者依照相对应方向中的行业框架进行评估。

2 中医知识图谱的应用

知识图谱在中医领域的应用日益成熟,中医知识的规范化和智能化应用水平也在不断提升。目前中医知识图谱在各领域的应用越来越多,包括医生临床决策支持诊断、智能语义搜索、中医智能问答与中医数据挖掘。

2.1 医生临床决策支持诊断

传统的临床决策是由医生根据患者的综合信息进行疾病诊断,而中医知识图谱不仅可以自动学习知识并对患者的疾病给出对应方案,帮助医生开拓思路,而且在中医知识图谱中有大量处理好的结构化知识可供医生学习。Keen 等^[39]在 20 世纪首次提出决策支持系统 (Decision Support Systems, DSS), 其将 DSS 定义为基于计算机的信息化系统,能够为管理者在制定和解决非结构化或半结构化的管理决策问题时提供支持帮助。将决策支持系统沿用到医学中即得到临床决策支持系统 (Clinical Decision Support Systems, CDSS)。CDSS 的目的是提供个性化、准确、及时的临床建议,帮助医生和其他医疗保健专业人员作出更好的医疗决策。牟丽君^[40]结合人机交互检索设计了中医古方知识图谱的人机交互检索功能,解决了中医药术语使用混乱、标准更新慢等问题,并以检索新冠信息为例,发现人机交互检索的性能优于传统检索方式。尹丹等^[41]为了解决中医药领域的方剂问题构建了一套推理流程,并计算了药物的相似度,设置了药物的功效和禁忌。

2.2 智能语义搜索

基于专业领域知识图谱所构建的语义搜索与传统的网页超链接搜索有所不同,后者仅会给出一些相同案例或者其他社交平台上的大众经验,而前者会结合专业领域知识和知识图谱的特性使用户高效地得到想要的结果及高相关度的智能推荐。中医知识图谱中结合中医古籍、电子病历、文本数据等,在语义搜索时对用户输入的信息提取关键字后,在知识图谱中找到对应的实体及其关系和属性。如于彤等^[42]构建的中医养生知识图谱可以根据用户

提供的体质信息提取用户关键词后,在知识图谱中搜索对应实体的属性及相应关系,为用户推荐贴合自身的养生方法。目前中医领域的智能搜索应用有搜医网、腾讯医典、丁香医生等,如贾李蓉等^[43-44]从2002年即开始构建以大量古籍和611味中药作为数据基础的中医药语言系统,其中包含13万多个概念和30万条术语以及100万条语义关系,通过中医药知识地图系统可视化方式展示用户搜索时的中医药相关概念。同时构建了中医药语言问答系统,根据用户提问为用户推荐相应医案^[45]。

2.3 中医智能问答

智能问答是自然语言处理(Natural Language Processing, NLP)的研究领域之一,其通过一问一答的形式,用户和具有智能问答系统的机器之间进行交互。具有智能问答系统的机器就像一个智者一样与用户友好地交谈,为其提供答案。中医智能问答一方面可以帮助大众普及中医知识,另一方面可以在用户生病时即使足不出户也能得到治疗方案,因此具有较好的应用前景。基于中医知识图谱的智能问答系统可以在用户提供问题后使用自然语言处理技术进行处理,然后在中医知识图谱中进行查询并给出答案。陈程等^[46]对用户问题进行分类并构建了相应模版,从用户提出的问题中搜索关键词,由关键词进入对应模版从而给出答案。郑子强^[28]采用向用户多轮提问的方式,使查找结果更加准确。刘琦^[47]设计了关于脉诊的智能问答,通过对症状节点与脉象节点之间边权的量化推测其中的关联,从而计算健康状态下可能出现的脉象。

2.4 中医数据挖掘

中医数据挖掘是指利用数据挖掘技术和方法对中医相关数据进行分析、挖掘与应用的过程。面对大量、复杂、混乱的中医数据,知识图谱可以更直观、全面地展示其中中药与方剂之间的联系、中药名医的思想等,还能可视化地表达隐式知识。王菁薇等^[48]对中医古籍《伤寒论》的中医药等实体及关系进行抽取。张泠杉等^[49]采用融合知识元理论的信息技术方法,对中医古籍《王旭高医案》中的内容进行梳理,探索其中隐含的逻辑关系。在下游应用中,金连顺等^[50]使用回顾性分析法对张忠德教授的临床病例进行收集,将个性化知识图谱与定性访谈法相结合,挖掘质性肺疾病的临证特征与用药规律。孙明俊等^[51]结合关节影像与病例信息构建类风湿性关节炎知识图谱。刘禹琪^[52]提出链路预测模型SK-Conv E,通过与其他4种经典的链路预测模型进行比较,发现SK-Conv E的预测能力更优。其将SK-Conv E应用于中医名方知识图谱对药物进行重新定位,发现泻肺散、黄连解毒汤、普济消毒饮3个方剂对新冠肺炎有抑制作用。

中医知识图谱的应用相较于一般的电子病历与中医书籍,在提升临床决策的准确性、提高信息检索效率、增强智能问答的响应能力以及挖掘中医数据的深层规律中具有显著优势。为了进一步提升中医知识图谱的应用效果,

可以集成多种技术以提高系统的智能性和实用性。结合本体论和知识图谱技术,既能利用本体的严格逻辑结构,又能发挥知识图谱的灵活性,实现更精确、高效的知识表示和推理。采用大数据处理技术处理中医领域的海量数据,既能应对结构化数据,又能处理复杂的异构数据。此外,可深入推进语义搜索和智能问答技术的应用,通过自然语言处理和语义分析技术增强系统对用户需求的理解能力,提升用户体验。

3 面临挑战与未来展望

知识图谱是语义网与知识库的强力结合,具有较强的关系表达能力和高效的反馈能力。知识图谱对中医数据进行了大整合,将中医数据中不同种类的信息连接在一起形成一种大型关系网络。大数据与中医相结合形成的中医知识图谱已成为目前的研究热点之一。

3.1 知识处理中的算法挑战

在中医知识图谱中,在中医知识抽取、知识对齐、知识融合等方面,可能存在中医文献多为古文或专业术语密集的文本,现有的自然语言处理技术在处理这些文本时面临着词汇歧义、语法复杂等问题,影响了知识处理的准确性。有研究者将中医领域训练词典加入模型,但如何获得准确度高、时间复杂度低的算法仍是值得深入研究的方向。

3.2 中医知识图谱表现单一

中医知识不仅包括文本信息,而且涉及图像、视频、音频等多种信息形式。现有的图谱多以节点和边的形式表示知识,无法有效融合和表示这些多模态信息,影响了知识图谱的全面性和实用性。目前尚未有多模态中医知识图谱方面的研究,因此如何将中医图像、视频等多模态信息有效融合于中医知识图谱,也是下一步研究的方向。

参考文献:

- [1] CHEN M, MAO S W, LIU Y H. Big data: a survey[J]. Mobile Networks and Applications, 2014, 19: 171-209.
- [2] NICKEL M, MURPHY K, TRESP V, et al. A review of relational machine learning for knowledge graphs[J]. Proceedings of the IEEE, 2015, 104(1): 11-33.
- [3] RUAN T, SUN C L, WANG H F, et al. Construction and application of knowledge graph of traditional Chinese medicine[J]. Journal of Medical Informatics, 2016, 37(4): 8-13.
- [4] 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.
- [4] WANG S. Study on construction of knowledge graph for diagnosis and treatment of lung cancer by Professor Zhou Zhongying in traditional Chinese medicine [D]. Nanjing: Nanjing University of Traditional Chinese Medicine, 2022.
- 王松. 国医大师周仲瑛教授辨治肺癌的知识图谱构建研究[D]. 南京: 南京中医药大学, 2022.
- [5] YU T, LIU J, JIA L R, et al. Research on the construction of large-scale knowledge graph of traditional Chinese medicine[J]. China Digital Medi-

- cine, 2015, 10(3):80-82.
- 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.
- [6] COLLOBERT R, WESTON J, BOTTOU L, et al. Natural language processing (almost) from scratch[J]. Journal of Machine Learning Research, 2011, 12: 2493-2537.
- [7] WANG S, LI Z J, YANG T, et al. Study on construction of ontology for diagnosis and treatment of lung cancer by Professor Zhou Zhongying in traditional Chinese medicine[J]. Modernization of Traditional Chinese Medicine and Materia Medica-World Science and Technology, 2022, 24(2): 495-501.
- 王松,李正钧,杨涛,等.国医大师周仲瑛辨治肺癌的中医药本体构建研究[J].世界科学技术-中医药现代化,2022,24(2):495-501.
- [8] GAO X Y, GAO W J, WANG X Y, et al. Method and application of knowledge graph construction of diagnosis and treatment of famous and old Chinese medicine based on medical case text[J]. Modernization of Traditional Chinese Medicine and Materia Medica-World Science and Technology, 2023, 25(9):2967-2974.
- 高晓苑,高文佳,王欣宇,等.基于医案文本的名老中医诊疗知识图谱构建方法及应用[J].世界科学技术-中医药现代化,2023,25(9):2967-2974.
- [9] LEI D. Named entity recognition and relationship extraction for traditional Chinese medicine knowledge graph[D]. Shijiazhuang: Hebei University of Geosciences, 2022.
- 雷迪.面向中医药知识图谱的命名实体识别及关系抽取[D].石家庄:河北地质大学,2022.
- [10] MA Y K, FENG Y C. Research on the method of extracting implicit relations from ancient books of traditional Chinese medicine[J]. Journal of Zhengzhou University(Science Edition), 2024, 56(2):34-42.
- 马月坤,冯焯琛.面向中医古籍的隐式关系抽取方法研究[J].郑州大学学报(理学版),2024,56(2):34-42.
- [11] PAN X F. Research on natural language processing of ancient books of traditional Chinese medicine based on deep learning[D]. Changchun: Changchun University of Traditional Chinese Medicine, 2024.
- 潘雪峰.基于深度学习的中医古籍自然语言处理研究[D].长春:长春中医药大学,2024.
- [12] MA Z B. Research on the optimization of relationship extraction model for traditional Chinese medicine texts[D]. Tangshan: North China University of Science and Technology, 2022.
- 马郅斌.面向中医文本的关系抽取模型优化研究[D].唐山:华北理工大学,2022.
- [13] HU D X. Symptom attribute extraction and medical knowledge graph construction for traditional Chinese medicine electronic medical records[D]. Nanchang: Jiangxi University of Traditional Chinese Medicine, 2023.
- 胡定兴.面向中医电子病历的症状属性抽取与医疗知识图谱构建[D].南昌:江西中医药大学,2023.
- [14] CAI M Z, LI X Y, ZHAO J W, et al. Disease knowledge discovery based on SPO semantic triplet[J]. Data Analysis and Knowledge Discovery, 2022, 6(1):134-144.
- 蔡妙芝,李晓瑛,赵嘉玮,等.基于SPO语义三元组的疾病知识发现[J].数据分析与知识发现,2022,6(1):134-144.
- [15] LIU C. Realization of intelligent prescription auxiliary recommendation system based on knowledge graph[D]. Chongqing: Southwest University, 2024.
- 刘灿.基于知识图谱的药方辅助推荐系统实现[D].重庆:西南大学, 2024.
- [16] FELLEGI I P, SUNTER A B. A theory for record linkage[J]. Journal of the American Statistical Association, 1969, 64: 1183-1210.
- [17] CHEN Z, KALASHNIKOV D V, MEHROTRA S. Exploiting context analysis for combining multiple entity resolution systems[C]//Proceedings of the 2009 ACM SIGMOD International Conference on Management of Data, 2009: 207-218.
- [18] ELFEKY M G, VERYKIOS V S, ELMAGARMID A K. TAILOR: a record linkage toolbox[C]//Proceedings 18th International Conference on Data Engineering, 2002: 17-28.
- [19] CHRISTEN P. Automatic training example selection for scalable unsupervised record linkage[C]//Pacific-Asia Conference on Knowledge Discovery and Data Mining, 2008: 511-518.
- [20] RUAN T, WANG M J, WANG H F, et al. Research on the construction and application of vertical knowledge graph[J]. Knowledge Management Forum, 2016(3):226-234.
- 阮彤,王梦婕,王昊奋,等.垂直知识图谱的构建与应用研究[J].知识管理论坛,2016(3):226-234.
- [21] LI J H, YI X L, YANG D L, et al. Research on the construction of ontological knowledge base of spleen and stomach diseases in traditional Chinese medicine for clinical decision support[J]. Chinese Journal of Medical Innovation, 2014, 11(27):121-125.
- 李敬华,易小烈,杨德利,等.面向临床决策支持的中医脾胃病本体知识库构建研究[J].中国医学创新,2014,11(27):121-125.
- [22] ZHANG Z, CHEN J, CHEN X, et al. An industry evaluation of embedding-based entity alignment[C]//Online: International Conference on Computational Linguistics, 2020.
- [23] DIENG-KUNTZ R, MINIER D, RŮŽIČKA M, et al. Building and using a medical ontology for knowledge management and cooperative work in a health care network[J]. Computers in Biology and Medicine, 2006, 36: 871-892.
- [24] WANG M Q, ZHANG L, CUI Y D, et al. Research on the ontological method of storing TCM dermatological diseases using Neo4j[J]. Modernization of Traditional Chinese Medicine and Materia Medica-World Science and Technology, 2020, 22(8):2914-2921.
- 王明强,张磊,崔一迪,等.利用Neo4j存储中医皮肤病“病-证-治”本体方法的研究[J].世界科学技术-中医药现代化,2020,22(8):2914-2921.
- [25] BIAN H. Research on expert knowledge discovery and reasoning algorithm in medical diagnosis system[D]. Qinhuangdao: Yanshan University, 2017.
- 边红.医学诊断系统中专家知识发现与推理算法研究[D].秦皇岛:燕山大学,2017.
- [26] ZHU L X. Knowledge mining and knowledge discovery for biomedical texts and graphs[D]. Hefei: University of Science and Technology of China, 2019.
- 朱吕行.面向生物医学文本及图谱的知识挖掘与知识发现[D].合肥:中国科学技术大学,2019.
- [27] KROMPASS D, BAIER S, TRESP V. Type-constrained representation learning in knowledge graphs[C]//14th International Semantic Web Conference, 2015: 640-655.
- [28] ZHENG Z Q. Knowledge graph learning and reasoning research for TCM medical cases of chronic kidney disease[D]. Chengdu: University of Electronic Science and Technology of China, 2020.
- 郑子强.面向慢性肾脏病中医医案的知识图谱学习与推理研究[D].成都:电子科技大学,2020.

- [29] LUO J G. Research and application of key technologies for the construction of knowledge graph in the field of traditional Chinese medicine[D]. Nanchang: Jiangxi University of Traditional Chinese Medicine, 2019.
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.
- [30] WANG R Y, STRONG D M. Beyond accuracy: what data quality means to data consumers [J]. Journal of Management Information Systems, 1996, 12(4): 5-33.
- [31] ZAVERI A, RULA A, MAURINO A, et al. Quality assessment for linked data: a survey[J]. Semantic Web, 2016, 7(1): 63-93.
- [32] FÄRBER M, RETTINGER A. Which knowledge graph is best for me? [DB/OL]. <https://arxiv.org/abs/1809.11099>.
- [33] LIANG J Q, ZHANG S, XIAO Y H. How to keep a knowledge base synchronized with its encyclopedia source[C]//Proceedings of the 26th International Joint Conference on Artificial Intelligence, 2017: 3749-3755.
- [34] ISSA S, ADEKUNLE O, HAMDI F, et al. Knowledge graph completeness: a systematic literature review[J]. IEEE Access, 2021, 9: 31322-31339.
- [35] ZAVERI A, RULA A, MAURINO A, et al. Quality assessment methodologies for linked open data [J]. Submitted to Semantic Web Journal, 2013(1): 1-5.
- [36] BIZER C. Quality-driven information filtering in the context of web-based information systems[M]. Saarbrücken: VDM Verlag, 2007.
- [37] FÄRBER M, BARTSCHERER F, MENNE C, et al. Linked data quality of DBpedia, freebase, openCyc, wikidata, and YAGO [J]. Semantic Web, 2018, 9(1): 77-129.
- [38] PIPINO L L, LEE Y W, WANG R Y. Data quality assessment[J]. Communications of the ACM, 2002, 45(4): 211-218.
- [39] KEEN P G W, SCOTT M M S. Decision support systems: an organizational perspective [EB/OL]. <http://www.dssresources.com/books/contents/keen78.html>.
- [40] MOU L J. Research on the construction and application of knowledge graph of ancient Chinese medicine formulas [D]. Shanghai: East China Normal University, 2024.
牟丽君. 中医古方知识图谱的构建与应用研究[D]. 上海:华东师范大学, 2024.
- [41] YIN D, ZHOU L, ZHOU Y M, et al. Research on the design of "graph search mode" of knowledge graph of traditional Chinese medicine formula [J]. China Journal of Traditional Chinese Medicine Information, 2019, 26(8): 94-98.
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.
- [42] YU T, LI J H, YU Q, et al. Construction and application of knowledge graph of traditional Chinese medicine health preservation[J]. China Digital Medicine, 2017, 12(12): 64-66.
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.
- [43] JIA L R, YU T, CUI M, et al. Research progress on language system in traditional Chinese medicine[J]. China Digital Medicine, 2014, 9(10): 57-59, 62.
贾李蓉, 于彤, 崔蒙, 等. 中医药学语言系统研究进展[J]. 中国数字医学, 2014, 9(10): 57-59, 62.
- [44] JIA L R, LIU J, YU T, et al. Construction of knowledge graph of traditional Chinese medicine[J]. Journal of Medical Informatics, 2015, 36(8): 51-53, 59.
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, 59.
- [45] JIA L R, GAO B, LIU J, et al. Design of knowledge question answering system based on language system of traditional Chinese medicine [J]. Journal of Medical Informatics, 2021, 42(6): 52-55.
贾李蓉, 高博, 刘静, 等. 基于中医药学语言系统的知识问答系统设计[J]. 医学信息学杂志, 2021, 42(6): 52-55.
- [46] CHEN C, ZHAI J, QIN J Y, et al. Research on intelligent question answering technology based on traditional Chinese medicine knowledge graph[J]. China New Communications, 2018, 20(2): 204-207.
陈程, 翟洁, 秦锦玉, 等. 基于中医药知识图谱的智能问答技术研究[J]. 中国新通信, 2018, 20(2): 204-207.
- [47] LIU Q. Research on knowledge extraction of scientific and technological literature on pulse diagnosis of traditional Chinese medicine [D]. Changsha: Hunan University of Traditional Chinese Medicine, 2022.
刘琦. 中医脉诊科技文献的知识抽取研究[D]. 长沙:湖南中医药大学, 2022.
- [48] WANG J W, XIAO L, YAN J F. Research on construction of knowledge graph of treatise on febrile diseases based on Neo4j [J]. Computer and Digital Engineering, 2021, 49(2): 264-267, 396.
王菁薇, 肖莉, 晏峻峰. 基于Neo4j的《伤寒论》知识图谱构建研究[J]. 计算机与数字工程, 2021, 49(2): 264-267, 396.
- [49] ZHANG L S, WANG F L, XING C L, et al. An exploration of the logical data and knowledge graph of Wang Xugao's case records based on knowledge element indexing [J]. Journal of Nanjing University of Traditional Chinese Medicine, 2021, 37(4): 592-596.
张冷杉, 王凤兰, 邢琛林, 等. 基于知识元标引的《王旭高医案》逻辑数据及知识图谱探析[J]. 南京中医药大学学报, 2021, 37(4): 592-596.
- [50] JIN L S, ZHANG T, HE W Y, et al. An analysis of clinical characteristics and prescription patterns of Professor Zhang Zhongde's treatment of interstitial lung disease based on knowledge mapping and qualitative interview [J]. Modernization of Traditional Chinese Medicine and Materia Medica-World Science and Technology, 2021, 23(8): 2838-2848.
金连顺, 张瞳, 何伟炎, 等. 基于知识图谱构建和定性访谈法探析张忠德教授辨治间质性肺病临床特征与方药规律[J]. 世界科学技术-中医药现代化, 2021, 23(8): 2838-2848.
- [51] SUN M J, ZHANG D, ZHENG M Z, et al. Traditional Chinese medicine aided diagnosis and treatment system for rheumatoid arthritis based on artificial intelligence [J]. Pattern Recognition and Artificial Intelligence, 2021, 34(4): 343-352.
孙明俊, 张丹, 郑明智, 等. 基于人工智能的类风湿性关节炎中医辅助诊疗系统[J]. 模式识别与人工智能, 2021, 34(4): 343-352.
- [52] LIU Y Q. Research and application of knowledge graph and link prediction model for famous prescriptions of traditional Chinese medicine [D]. Changchun: Northeast Normal University, 2021.
刘禹琪. 中医名方知识图谱构建与链路预测模型的研究及应用[D]. 长春:东北师范大学, 2021.

(责任编辑:黄健)