

基于大数据知识工程的真实世界中医临床研究思路与实践

孙婷¹, 张盼¹, 宋淑洁¹, 李君¹, 李晓东², 沈绍武³, 肖勇³, 田双桂³, 黄纓^{4*}, 毛树松^{2*}

1. 湖北中医药大学中医临床学院, 武汉 430065

2. 湖北省中医院, 武汉 430060

3. 湖北中医药大学信息工程学院, 武汉 430065

4. 湖北中医药大学附属荆州市中医医院, 荆州 434000

摘要 通过回顾中医临床大数据知识工程的起源及其40年发展历程,厘清了各研究阶段的重大事件和研究成果,阐述了基于病案的真实世界中医临床研究思路与方法,强调了从数据驱动向数据与知识双驱动模式转换,推动真实世界中医临床研究模式变革与创新的重要性,明确了模式变革与创新的基础是将中医临床病例数据库知识属性化重构,使隐性知识显性化,实现知识的完整表达,构建中医临床病例知识库。以病例数据知识化、知识编码及知识图谱为示例,重点展示了中医临床大数据知识工程的关键技术和核心环节,充分证实了大数据知识工程是真实世界中医临床研究的必由之路。在国家中医药高质量发展战略指引下,未来应培育中医临床研究新质生产力,形成具有新发展理念中医临床研究新业态为目标,构建大数据知识工程技术平台和中医临床知识仓库,为开展中医临床领域大模型研究提供技术平台和知识资源,促进中医临床智能化和中医药事业高质量发展。

关键词 大数据知识工程;真实世界中医临床研究;数据知识化;知识编码

随着人类社会迅速发展,科学技术日新月异,以大数据为代表的新兴技术将人类从“信息时代”带入了“大数据时代”^[1]。具有独特认识论、方法论和技术体系的大数据知识工程学应运而生,已成为当今科

收稿日期:2024-04-24;修回日期:2024-08-17

基金项目:湖北省医学领军人才培养工程项目(鄂卫通[2019]47号);全国老中医药专家学术经验继承工作项目(国中医药人教函[2022]76号);湖北省自然科学基金一般面上项目(2024AFB983)

作者简介:孙婷,博士研究生,研究方向为名老中医经验传承,电子邮箱:136807166@qq.com;黄纓(通信作者),教授,研究方向为名老中医经验传承,电子邮箱:hy3398@126.com;毛树松(共同通信作者),教授,研究方向为中医药标准化与信息化,电子邮箱:mss_812@126.com

引用格式:孙婷,张盼,宋淑洁,等.基于大数据知识工程的真实世界中医临床研究思路与实践[J].科技导报,2024,42(21):55-65;
doi:10.3981/j.issn.1000-7857.2024.05.00468

学技术的前沿学科,并以迅猛之势引领着各行各业的技术革命,推动着整个人类社会的深刻变革和跨越式发展。大数据知识工程学是大数据科学与知识工程学相互交融的工程学科^[2-4],是处理人类行为活动要素之间关联关系的工程技术学科,主要包括数据采集、数据治理、数据知识化、知识加工、知识建模和知识图谱化,以及人工智能(artificial intelligence, AI)应用系统研发等知识工程技术^[5]。大数据科学的理念与中医药现代化的发展战略及其“传承精华、守正创新”的发展目标高度契合,同时知识工程学的技术方法完全适配真实世界中医临床研究^[6-8],可应对由于数据量巨大、数据结构复杂等带来的数据标准化处理、数据深层次解读等瓶颈问题^[9-11],能满足高质量真实世界中医临床研究的要求。同时,中医病案是真实世界中医临床研究的重要数据资源,王永炎院士指出,“当今各级各类中医师们诊疗疾病时保留的完整的病历资料是宝贵的大数据资料,然而它是象、数、易、气、神混沌一体的非线性数据,通过激活数据学及非线性算法可为梳理、发掘、创新、复兴中医临床医学带来新机遇,进一步完善中医临床医学体系,这也是一种渴求进化的大趋势”^[12]。基于此,在历经40年的中医临床专家系统、中医药标准化和信息化、中医临床数据治理及数据知识化技术等研究的基础上,提出中医临床大数据知识工程(以下简称“工程”)的概念,即以中医学理论为基础,遵循中医临床自身发展规律,在大数据世界观指导下,应用知识工程技术开展基于医案的真实世界中医临床研究,构建标准、理论与方法和技术等三大体系,形成以中医学与大数据知识工程学相融合的新交缘学科,服务于中医临床研究模式的创新与变革,实现中医临床研究从数据驱动向数据与知识双驱动模式的转变,为真实世界中医临床研究提供新思路和新方法,促进其高质量发展。

1 中医临床大数据知识工程的起源与发展历程

1.1 前期研究阶段

1.1.1 中医临床人工智能研究(1979—1984年)

20世纪70年代末至80年代中期,李国平院士

和中医大师黄绳武,倡导建立中医控制论研究室,开展计算机辨证论治、中医临床专家系统研究,历时5年,研发了中医内科、妇科、儿科等多个临床专科的计算机专家诊疗系统软件,并由鲍亦万、吕继端、张大钊等编撰了《控制中医学 中医学证治系统分析》《中医计算机模拟及专家系统概论》等专著,开创了控制中医学研究的先河。但受限于当时的计算机技术水平和信息标准的缺失而未能得到深化与推广应用,突显了开展中医临床信息管理技术与标准研究的必要性和紧迫性。

1.1.2 中医临床信息管理系统研究(1985—1995年)

本阶段以中医临床信息编码和信息管理技术研究为目标,开展中医病证分类编码技术和中医病案首页中西医双重诊断研究,为国家中医药管理局制定具有中西医疾病分类编码的“中医病案首页”提供依据。在此基础上,研制中医病案首页信息管理系统,中医病案统计信息管理系统,中医医院科研、人事、护理等信息管理系统,并在全国推广应用,编撰了《中医病证分类编码》《护理信息学概论 计算机在护理中的应用》和《汉英中医药分类词典 计算机词库范本》等专著。

1.2 工程预研究阶段

1.2.1 中医药信息标准体系构建研究(1996—2015年)

本阶段以中医药信息标准制修订研究为目标,完成中医药信息标准体系的顶层设计,编制《中医药信息标准体系表》,并由国家中医药管理局国中医药办〔2013〕41号文予以发布。该体系表将信息标准划分为基础标准、技术标准、管理标准、工作标准等4个类目和25个分类目,并将现行的230余项相关信息类国家标准、行业标准,以及110余项团体标准,进行分类整理与汇编,初步建立中医药信息标准体系。其中,制修订了包括国家标准《中医病证分类与代码》《全国可运输产品(中药部分)分类与代码》和临床症状、舌象、脉象、针灸、中药等中医诊疗信息分类代码团体标准等共计30余项,覆盖中医病案中的全部诊疗信息,为中医临床数据知识化提供标准依据。

1.2.2 临床数据治理及知识关联方法研究(2016—2020年)

在建立中医药信息标准体系的基础上,由广东省中医院、上海龙华医院、湖北省中医院等10个单位组成了研究团队,遴选了8个专科专病,针对临床数据治理和知识关联方法等关键技术开展临床应用研究,总结并编撰了《中医临床数据结构化与知识关联方法学概论》和《中医临床数据标准体系及其标准汇编》,为形成中医临床大数据知识工程研究思路与方法奠定基础。

1.3 工程研究阶段(2021年至今)

本阶段界定了中医临床大数据知识工程的概念和目标,并对“工程”进行总体规划与设计,开展了“工程”的理论与方法、标准和技术等3大体系构建研究,其主要研究进展简介如下。

1) 以刘保延提出的真实世界中医临床研究范式^[13]为指导,创立了基于大数据知识工程的真实世界中医临床研究新模式^[14]:基于临床数据的“是什么”的研究和临床循证的“为什么”的研究,即以中医临床海量数据中的“事实与规律”替代先验性科研假设开展中医临床循证研究。新模式流程见图1。

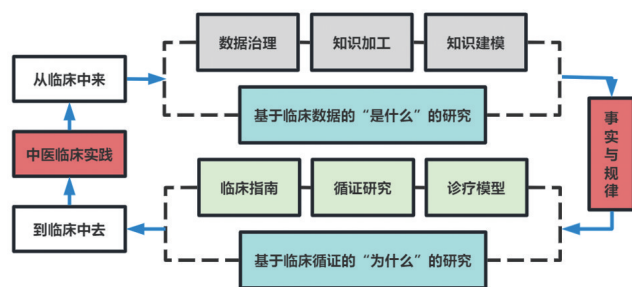


图1 真实世界中医临床研究新模式流程

2) 通过构建中医临床诊疗知识本体,融合中医学、大数据科学和知识工程学等3个学科理论,建立了中医临床大数据知识工程的理论和方法学体系^[15]。

3) 以中医临床信息分类与代码标准为依据,研制了《中医临床基本知识编码库》,制定了将中医临床病例数据库进行知识属性化重构,形成中医临床病例知识库的技术规范^[16-17]。

4) 以人工智能自然语言处理技术,研制了基于病案的中医临床病例数据库与知识库自动生成

系统,解决了海量病案数据生成病例数据库与知识库的瓶颈问题,完成了中医临床大数据知识工程技术平台的方案设计^[18]。

5) 应用大数据知识工程技术方法开展了中医肝病、脾胃病、小儿肾病、名老中医临证经验传承等中医临床大数据知识工程专题研究,探索中医临床精准化诊疗(证、治、效)模型构建和知识图谱绘制的技术方法。

2 基于病案的真实世界中医临床研究思路与方法

2.1 研究思路

以《“十四五”中医药发展规划》为指导,以中医临床精准化诊疗(证、治、效)模型研究为目标,践行真实世界中医临床研究新模式和实施中医临床大数据知识工程,组织开展基于海量病案数据的真实世界中医临床研究。遴选中医临床优势病种,收集整理该病种的原始病案资料,通过对其临床数据的规范化、数字化、结构化、知识化等数据治理过程,形成中医临床病例数据库与知识库。以中医临床病例知识库为对象,探索该病种临床知识分布规律及知识关联关系,构建中医临床精准化诊疗(证、治、效)模型,以经临床循证研究验证与优化的中医临床诊疗模型为据,制定该优势病种中医临床诊疗指南,反馈应用于指导临床实践,促进中医临床诊疗的高质量发展。

2.2 研究方法

2.2.1 标准操作流程

依据中医临床大数据知识工程规划和设计的要求^[1],基于病案的真实世界中医临床研究标准操作流程(standard operating procedure, SOP),包括原始病案资料收集、数据规范化、数字化、结构化、知识化、模型化及其临床验证与优化和反馈控制等闭环管理环节,其标准操作流程见图2。

2.2.2 操作规定与要求

1) 遴选中医专科专病。

中医专科专病指在中医专科临床具有鲜明特色和突出优势的病种,围绕中医专科优势病种开展

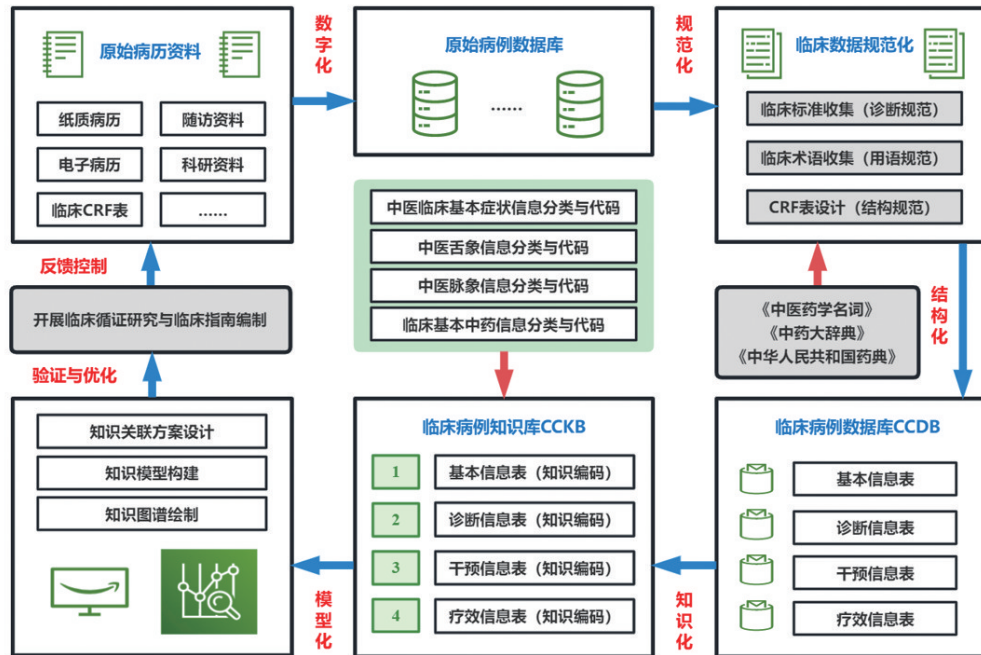


图2 基于病案的真实世界中医临床研究标准操作流程

基于大数据知识工程的真实世界中医临床研究,对规范诊疗行为、提高诊疗质量具有现实意义。该研究首先遴选中医专科优势病种,依据临床科研设计的纳入标准和排除标准,多元化搜集该病种的原始病案资料,将经过筛选和整理的病案数据录入计算机,构建中医原始病案数据库。

2) 中医病案数据治理。

数据治理是在数据科学理论与方法指导下,通过数据的多元化采集、规范化清理、主题化汇聚、集约化存储和知识化处理,形成“数据说话、数据管理和数据决策”的大数据治理能力。中医优势病种的病案数据治理,是基于病案的真实世界中医临床研究的基础性工作,以解决用于开展临床研究的中医病案数据的完整性和规范性等质量问题^[19]。首先依据中医药学名词审定委员会发布的《中医药学名词》和《中华人民共和国药典》,以及中医药学语言系统语义网络框架^[20-21]等标准与研究,分类规范该病种病例数据中的中医药名词术语,例如病名、证候、症状、舌象、脉象、治法、方药等,将不规范的用语表达和诊断描述规范化,使其符合中医临床术语标准的要求;再根据“工程”规范的中医临床信息采集表,包括基本信息、诊断信息、干预信息和疗效信

息等4个子表,设计优势病种中医临床病例报告表(case report form, CRF)和中医临床病例数据库结构,对病例数据进行结构化处理,形成优势病种的临床病例数据库(clinical case data base, CCDB);以《中医临床基本症状信息分类与代码》《中医舌象诊断信息分类与代码》《中医脉象诊断信息分类与代码》和《临床中药基本信息分类与代码》等中医临床知识属性分类与代码标准,对CCDB进行知识属性重构,即将其中的每一个知识实体(概念)均以其名称代码和其属性代码来准确、完整地表达该实体的知识内涵,使其隐含的知识显性化,形成该优势病种的临床病例知识库(clinical case knowledge base, CCKB)。至此,通过构建中医优势病种的CCDB和CCKB,完成病案数据的治理,为开展该优势病种的病例知识加工、知识建模与知识图谱绘制等提供知识资源。

3) 临床病例知识关联与知识建模。

以该优势病种的CCKB为对象,应用知识工程的技术方法,例如决策树、神经网络、关联规则、聚类算法等,对该知识库中的临床病例知识进行关联分析,挖掘该知识库中隐含的临床诊疗“事实与规律”,并将其模型化,建立该优势病种具有临床数据

支撑的中医临床诊疗(证、治、效)模型(简称“模型A”),完成该优势病种临床病例知识关联和临床诊疗模型构建,为进一步开展该病种诊疗模型临床循证研究奠定基础。

4) 诊疗模型临床循证研究。

采用循证医学的方法,对“模型A”开展临床循证研究,以对其进行验证和优化。依据该优势病种的中医临床病例报告表规范,实时、准确、完整采集其临床诊疗信息;采用上述“工程”标准操作规程(流程与规定要求),构建该优势病种循证研究的CCDB和CCKB;以该CCKB为对象,进行知识关联分析,建立基于实时规范采集临床病例数据的中医临床诊疗“模型B”。以“模型B”为据,对“模型A”进行验证与优化,形成该优势病种中医临床诊疗“模型C”,再以“模型C”为据,研制该优势病种的中医临床诊疗指南,用于指导临床实践,最终实现“从临床中来,到临床中去”的具有反馈控制的管理闭环。

2.2.3 中医CCKB构建关键技术示例

中医CCKB是依据中医临床知识编码标准将中医病例数据知识化,并对CCDB进行知识属性化重构而形成的。数据知识化与知识编码是CCKB构建的关键技术,现以某中医医院妇科优势病种“妇人腹痛”临床病案的诊断数据和干预数据的知识化、知识编码和知识图谱为例,分别说明如下。

1) 中医临床诊断数据知识化与编码示例。

中医临床诊断知识依据GB/T 15657—2021《中医病证分类与代码》《国际疾病分类ICD-10》、

T/CIATCM 020—2019《中医基本症状信息分类与代码》、T/CIATCM 010—2019《中医舌象诊断信息分类与代码》和T/CIATCM 011—2019《中医脉象诊断信息分类与代码》等标准进行规范和分类编码。现以妇人腹痛病-湿热瘀结证临床病例的诊断数据知识化与知识编码为例,说明如下。

(1) 妇人腹痛病-湿热瘀结证的诊断知识结构(图3)。

(2) 妇人腹痛病-湿热瘀结证的诊断数据知识化与知识编码(表1)。

2) 中医临床干预数据知识编码及图谱示例。

(1) 中医临床干预数据知识编码示例。

本病例中临床干预是在核心处方“柴枳败酱汤”的基础上进行随证加减,增加了蒲公英30g、乳香20g、没药20g、黄柏9g、苡米30g等中药。现仅以该核心处方为例进行临床干预知识编码,其随证加减用药的知识编码以此类推。首先对该核心处方中的中药名称进行规范,如“红藤”规范为“大血藤”,“败酱”规范为“败酱草”等,再依据《中华人民共和国药典》2020年版和GB/T 15657—2021《中医病证分类与代码》、T/CIATCM 024—2019《临床中药基本信息分类与代码》等中药临床知识属性分类编码标准对该处方知识进行编码示例(表2)。

(2) 中医临床干预数据知识图谱示例。

知识图谱是大数据时代最重要的知识表示形式^[22],能够将中医药知识内涵通过数据挖掘、信息处理、知识计量和图形绘制而呈现^[23-24],已逐渐被

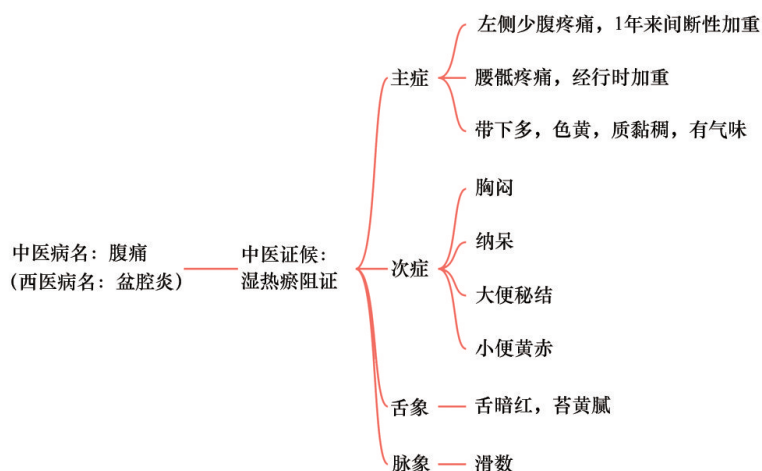


图3 妇人腹痛病-湿热瘀结证临床诊断知识结构

表1 中医临床诊断数据知识化与知识编码示例

分类	临床诊断描述	诊断术语规范	诊断知识编码
中医病名	妇人腹痛病	腹痛	A17.36
中医证候	湿热瘀结证	湿热瘀阻证	B02.05.04.03.06.01
西医病名	盆腔炎	盆腔炎	N73.902
症状	主症:左侧少腹疼痛2年,1年来间断性加重,隐隐作痛,痛连腰骶,经行时加重,带下多,色黄,质黏稠,有气味	左侧少腹疼痛,1年来间断性加重	ZZ004103.A040602Z3R1121
		腰骶疼痛,经行时加重	ZZ004103.A040701R403
		带下多,色黄,质黏稠,有气味	ZZ1121090102.D303H09J1
	次症:平素月经规律, lmp19930823, 胸闷纳呆,大便秘结,小便黄赤	胸闷	ZZ052104.A0402T03
		纳呆	ZZ08010602
		大便秘结	ZZ09510701.K1
舌象	舌暗红,苔黄腻	小便黄赤	ZZ0962030402.D906
脉象	滑数	滑数	SZZ1300000000.T04002200 MZ41620000

表2 柴枳败酱汤知识编码示例

中药名称	剂量/g	配伍(Y Y02)	四气(Y X01)	五味(Y X02)	归经(Y X04)	功效(Y X06)
大血藤	30	君药(Y Y0201001)	平(Y X0110)	苦(Y X0202)	大肠,肝经 (Y X0409, X0401)	清热解毒(C02.06.02) 活血(C02.08) 祛风止痛 (C02.14.01.05.02)
败酱草	30	君药(Y Y0201001)	凉(Y X0108)	辛,苦(Y X0204, Y X0202)	胃,大肠,肝经 (Y X0407, Y X0409, Y X0401)	解毒排脓 (C02.21.03.01.03) 祛瘀止痛 (C02.08.06.06) 活血消痈(C02.21.02)
三棱	12	臣药(Y Y0201002)	平(Y X0110)	辛,苦(Y X0204, Y X0202)	肝,脾经 (Y X0401, Y X0403)	破血行气 (C02.07.03.12) 消积止痛(C02.07)
莪术	12	臣药(Y Y0201002)	温(Y X0106)	辛,苦(Y X0204, Y X0202)	肝,脾经 (Y X0401, Y X0403)	破血行气 (C02.07.03.12) 消积止痛(C02.07)
柴胡	9	佐药(Y Y0201003)	微寒 (Y X0103)	辛,苦(Y X0204, Y X0202)	肝,胆,肺经 (Y X0401, Y X0406, Y X0404)	疏散退热(C02.01) 疏肝解郁 (C02.07.01.05) 升举阳气(C02.11)
枳实	9	佐药(Y Y0201003)	微寒 (Y X0103)	辛,苦,酸(Y X0204, Y X0202, Y X0201)	脾,胃经 (Y X0403, Y X0407)	破气消积 (C02.07.03.13) 化痰散痞(C02.15)
赤芍	15	佐药(Y Y0201003)	微寒 (Y X0103)	苦(Y X0202)	肝经(Y X0401)	清热凉血 (C02.06.03.05) 散瘀止痛 (C02.08.06.06)

表2 柴枳败酱汤知识编码示例(续)

中药名称	剂量/g	配伍(Y Y02)	四气(Y X01)	五味(Y X02)	归经(Y X04)	功效(Y X06)
白芍	15	佐药 (Y Y0201003)	微寒(Y X0103)	苦,酸 (Y X0202, Y X0201)	肝,脾经(Y X0401, Y X0403)	养血调经(C02.11.02.03.06) 敛阴止汗(C02.20.01.01.03) 柔肝止痛(C02.11) 平抑肝阳(C02.05)
丹参	20	佐药 (Y Y0201003)	微寒(Y X0103)	苦(Y X0202)	肝,心经(Y X0401, Y X0402)	活血祛瘀(C02.08.02) 通经止痛(C02.08.06.03) 清心除烦(C02.06.04.01) 凉血消痈(C02.08.08.01)
香附	12	佐药 (Y Y0201003)	平(Y X0110)	辛,微苦 (Y X0204, Y X0202)	肝,脾,三焦经 (Y X0401, Y X0403, Y X0411)	疏肝解郁(C02.07.01.05) 理气宽中(C02.07) 调经止痛(C02.08.06.03)
大黄	9	佐药 (Y Y0201003)	寒(Y X0101)	苦(Y X0202)	脾,胃,大肠,肝,心 包经(Y X0403, Y X0407, Y X0409, Y X0401, Y X0412)	泻下攻积(C02.04) 清热泻火(C02.06.01) 凉血解毒(C02.06.02.02) 逐瘀通经(C02.08) 利湿退黄(C02.09.04)
甘草	6	使药 (Y Y0201004)	平(Y X0110)	甘(Y X0203)	心,肺,脾, 胃经(Y X0402, Y X0404, Y X0403, Y X0407)	补脾益气(C02.11.01.03.01) 清热解毒(C02.06.02) 祛痰止咳(C02.15.01.01.01) 缓急止痛(C02.05)
牛膝	12	使药 (Y Y0201004)	平(Y X0110)	苦,甘 (Y X0202, Y X0203)	肝,肾经(Y X0401, Y X0405)	逐瘀通经(C02.08) 引血下行(C02.09) 补肝肾(C02.11.03.04.03) 强筋骨(C02.23.10)

应用于中医药领域的智能搜索、深度问答、知识推荐、辅助决策等方面^[25-27]。现以表2中中药处方“柴枳败酱汤”的临床药用知识关联关系绘制图谱来呈现其知识内涵(图4~图6)。

3 中医临床大数据知识工程的临床价值与深化发展

中医临床大数据知识工程将现代大数据、知识工程、人工智能等新兴前沿技术与临床中医学深度融合,是新兴前沿技术与临床中医学的“融合剂”,也是中医临床标准化和信息化双向融合的成功范例,具有高科技、高效能、高质量等临床价值。其一,搭建高效能中医临床研究技术平台。“工程”为

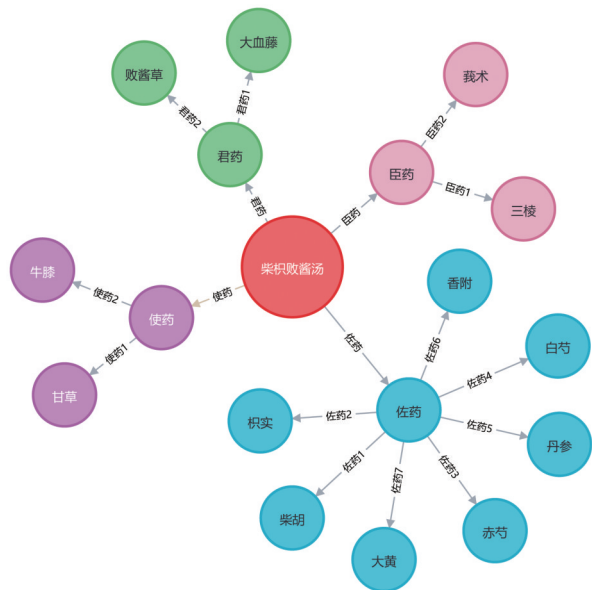


图4 柴枳败酱汤二级知识图谱

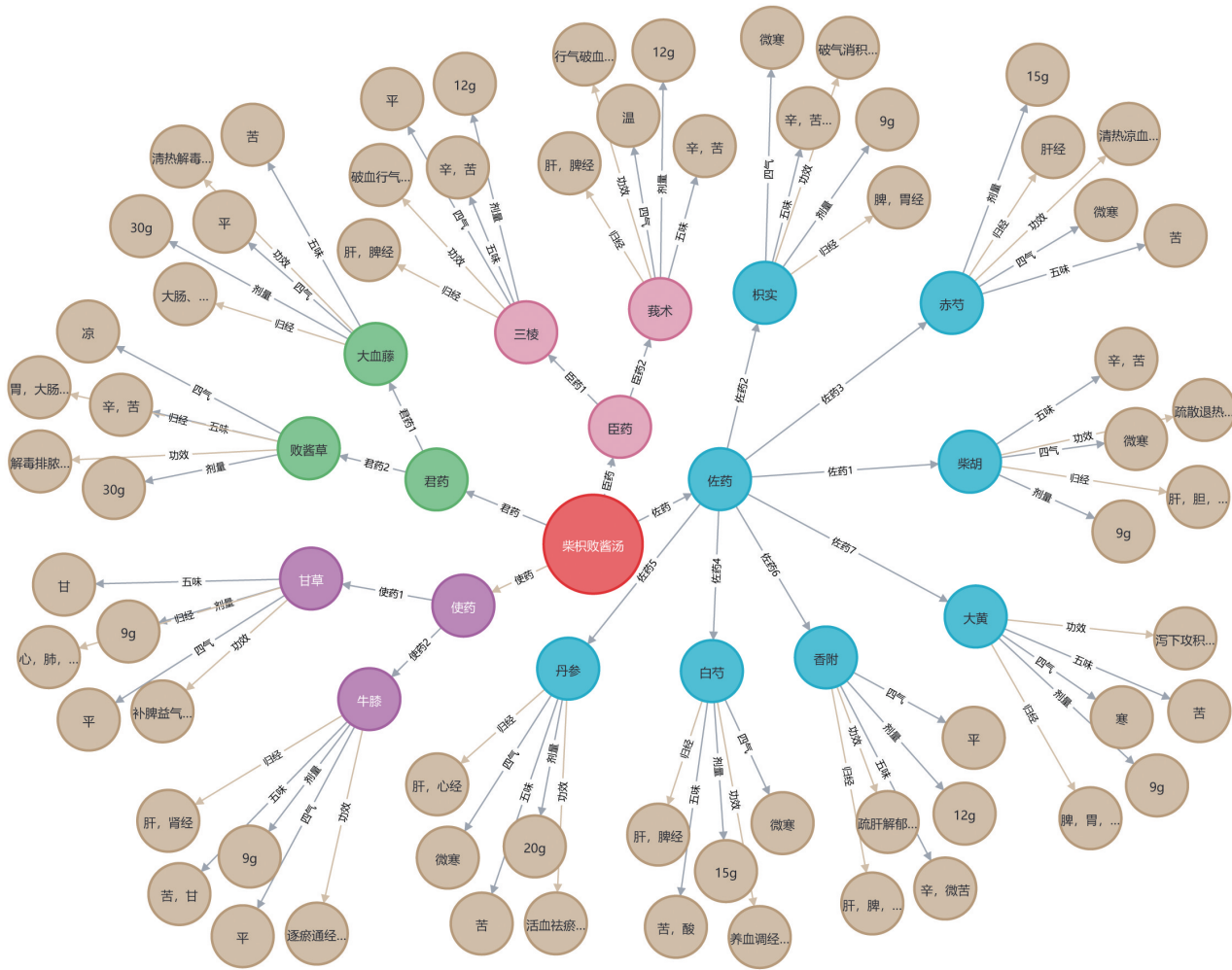
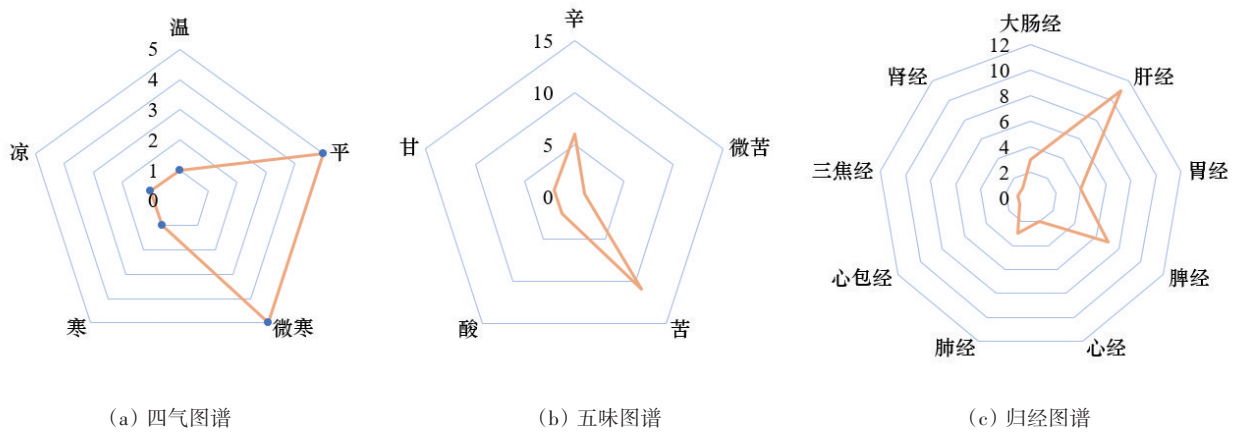


图5 柴枳败酱汤三级知识图谱



(a) 四气图谱

(b) 五味图谱

(c) 归经图谱

图6 柴枳败酱汤性味归经图谱

开展中医临床专科专病研究提供数据、知识与知识处理技术以应对一系列数据瓶颈问题,将大幅度提高真实世界中医临床研究的效率与质量,在实现中医临床智慧化与智能化过程中发挥着关键性技术支撑作用,是高效能临床中医学科学研究的技术平台。其二,培养高素质复合型人才梯队。“工程”凝聚医学、理学、工学和管理学等多学科研究人员,组建专业化科研团队,以实现临床中医学智慧化与智能化为目标,充分发挥各专业技术优势,协同开展项目攻关研究,无论是在技术深化与优化,还是在团队协同精神等方面,均有利于高素质复合型人才的培养。其三,推进高质量学科与专科内涵建设。以中医优势病种为抓手,开展基于大数据知识工程的真实世界临床研究,构建中医优势病种精准化“证-治-效”模型,形成中医临床诊疗思维导图,编制中医优势病种诊疗指南,既能提升专科的诊疗水平,又能升华学科的学术内涵,是高质量专科与学科建设的奠基性工作。

习近平总书记在中共中央政治局第十一次集体学习时强调“加快发展新质生产力”,新质生产力在创新中起主导作用,具有高科技、高效能、高质量特征,科技创新能够催生新产业、新模式、新动能,是发展新质生产力的核心要素。实现中医药数字化、工程化和智能化,是引领中医药科技突破与发展和中医药高质量传承与创新的重要路径。“工程”将以培育临床中医学新质生产力为导向,在现有研究基础上稳步实现中医药科技突破与发展,践行中医药高质量传承与创新。其一,依托中医临床海量数据构建“临床中医学知识仓库”。按照“工程”的具体操作流程,批量处理包括标准、文献与医案等海量中医临床数据,利用数据知识化技术,形成多维度、关联性的知识体系,构建临床中医学知识仓库,为开展中医药领域临床研究提供知识资源。其二,应用知识关联分析技术构建精准化中医病证“证-治-效”模型(判别式 AI 小模型)。以中医专科专病为研究对象,应用知识关联分析技术,构建精准化中医病证诊疗模型,即“判别式 AI 小模型”。特别是基于名老中医临证处方决策经验传承开展

的真实世界临床研究,通过知识关联构建知识模型,绘制知识图谱,呈现名老中医的临床诊疗思维导图,是研发名老中医知识推荐系统的核心工作。其三,以临床中医学知识仓库训练优选的开源大语言模型,构建临床中医学领域大模型(生成式 AI 大模型)。当前大语言模型在自然语言处理领域展现了强大的语言理解和生成能力,如何利用其提高诊断准确率、优化治疗方案、提升患者体验是其应用于医学领域亟待解决的问题^[28-29],以临床中医学知识仓库训练优选的开源大语言模型,最终建立临床中医学领域大模型是“工程”未来深化发展的方向。

4 结论

中医临床大数据知识工程学以中医学理论为基础,遵循中医临床自身发展规律,提出了由数据驱动向数据与知识双驱动,变革与创新真实世界中医临床研究模式的主体思想,旨在践行“守正创新,笃行致远”,共促新时代中医药高质量发展^[30]。在“工程”总体规划与设计的指导下,将大数据科学理论、知识工程学理论与中医学理论深度融合,形成了中医临床大数据知识工程的理论体系;通过数据治理的流程规范,制定了构建中医临床病例数据库与中医临床病例知识库的技术规范;利用知识关联分析技术,揭示了真实世界中医临床诊疗数据中的“事实与规律”,示范了精准化“证-治-效”模型构建和知识图谱绘制的技术方法。中医临床大数据知识工程是一个系统工程,具有长期性、复杂性和科学性,需要建立创新型研究范式和多学科协同,是真实世界中医临床研究的必由之路。展望未来,“临床中医学知识仓库”“判别式 AI 小模型”“生成式 AI 大模型”等深化与优化研究将为真实世界中医临床研究和中医药事业高质量发展做出更大的贡献。

参考文献 (References)

- [1] 毛树松, 沈绍武, 胡芳. 中医临床数据结构化与知识关

- 联方法学概论[M]. 北京: 中国中医药出版社, 2022: 13.
- [2] Wu X D, Chen H H, Wu G Q, et al. Knowledge engineering with big data[J]. IEEE Intelligent Systems, 2015, 30(5): 46-55.
- [3] 陆汝铃. 从大数据到大知识[J]. 重庆邮电大学学报(自然科学版), 2022, 34(6): 921-928.
- [4] 郑庆华. 大数据知识工程[M]. 北京: 科学出版社, 2023.
- [5] 宋淑洁, 田双桂, 张盼, 等. 中医临床研究创新发展的需求分析与对策[J]. 时珍国医国药, 2022, 33(5): 1267-1269.
- [6] Corrigan-Curay J, Sacks L, Woodcock J. Real-world evidence and real-world data for evaluating drug safety and effectiveness[J]. JAMA, 2018, 320(9): 867.
- [7] Basch E, Schrag D. The evolving uses of "real-world" data[J]. JAMA, 2019, 321(14): 1359.
- [8] Concato J, Corrigan-Curay J. Real-world evidence—Where are we now?[J]. New England Journal of Medicine, 2022, 386(18): 1680-1682.
- [9] 李蔓, 邵明义, 符宇, 等. 基于临床数据开展中药新药真实世界研究的策略与问题探讨[J]. 中药药理与临床, 2020, 36(5): 219-223.
- [10] 张雨楠, 李易真, 窦智丽, 等. 真实世界针灸大数据研究现状、问题与解决路径[J]. 中华中医药杂志, 2023, 38(2): 695-698.
- [11] 李宛亭, 乔佳慧, 孟令全, 等. 真实世界数据存在的问题与质量提升对策研究[J]. 中国新药杂志, 2021, 30(13): 1160-1163.
- [12] 王永炎, 王忠, 王燕平. 中医临床医学迎接大数据科学时代的刍议[J]. 中医杂志, 2022, 63(1): 2-4.
- [13] 刘保延. 真实世界的中医临床科研范式[J]. 中医杂志, 2013, 54(6): 451-455.
- [14] 宋淑洁, 毛树松, 张盼, 等. 基于大数据知识工程的中医临床研究新模式构建研究[J]. 时珍国医国药, 2022, 33(9): 2297-2299.
- [15] 陈俞清, 朱瑶, 张雪荣, 等. 中医临床大数据知识工程基本理论与方法学体系[J]. 时珍国医国药, 2022, 33(6): 1531-1532.
- [16] 田双桂, 桑宇慧, 付文娇, 等. 中医临床知识属性分类与代码标准体系构建研究[J]. 时珍国医国药, 2022, 33(10): 2554-2557.
- [17] 张盼, 毛树松, 林玲, 等. 中医临床知识组织与知识编码研究[J]. 时珍国医国药, 2023, 34(4): 1024-1026.
- [18] 肖勇, 张魁, 沈绍武, 等. 中医临床大数据知识工程技术平台架构与功能设计[J]. 时珍国医国药, 2023, 34(6): 1518-1520.
- [19] 周雪忠, 王世华, 张迪, 等. 构建中医药特色真实世界临床研究新模式的实践与思考[J]. 科技导报, 2023, 41(14): 22-31.
- [20] 于彤, 崔蒙, 张竹绿. 从中医药学语言系统到简单知识组织系统本体的转换研究[J]. 中国中医药信息杂志, 2014, 21(12): 38-41.
- [21] 中华人民共和国国家市场监督管理总局, 中国国家标准化管理委员会. 健康信息学 中医药学语言系统语义网络框架: GB/T 38324—2019[S]. 北京: 中国标准出版社, 2019.
- [22] 牟梓君. 小儿脑瘫中医诊疗知识图谱构建及其隐性知识显性化研究[D]. 北京: 中国中医科学院, 2021.
- [23] 赵汉青. 中医药知识图谱研究述评[J]. 中国数字医学, 2023, 18(8): 102-107.
- [24] 张翔, 杨伟杰, 刘文文, 等. 人工智能时代知识图谱表示学习方法体系[J]. 科技导报, 2021, 39(22): 94-110.
- [25] Guo Z H, Liu Q P, Zou B J. Research on knowledge reasoning of TCM based on knowledge graphs[J]. Digital Chinese Medicine, 2022, 5(4): 386-393.
- [26] Yang R, Ye Q, Cheng C L, et al. Decision-making system for the diagnosis of syndrome based on traditional Chinese medicine knowledge graph[J]. Evidence-Based Complementary and Alternative Medicine, 2022, 2022: 8693937.
- [27] Hou M H, Jin J X, Wei C F. Research and practice on the construction of Chinese medicine knowledge intelligent question answering system based on knowledge graph[C]//Proceedings of Second International Symposium on Computer Applications and Information Systems (ISCAIS 2023). Washington D C: SPIE, 2023.
- [28] 胡振生, 杨瑞, 朱嘉豪, 等. 大语言模型在医学领域的研究与应用发展[J]. 人工智能, 2023, 10(4): 10-19.
- [29] 邓佳文, 任福继. 2023年生成式AI大模型发展热点回眸[J]. 科技导报, 2024, 42(1): 266-285.
- [30] 张伯礼. 守正创新, 笃行致远共促新时代中医药高质量发展[J]. 科技导报, 2023, 41(14): 1.

Thinking and practice of real world clinical study of traditional Chinese medicine based on big data knowledge engineering

SUN Ting¹, ZHANG Pan¹, SONG Shujie¹, LI Jun¹, LI Xiaodong², SHEN Shaowu³, XIAO Yong³, TIAN Shuanggui³, HUANG Ying^{4*}, MAO Shusong^{2*}

1. Clinical College of Traditional Chinese Medicine, Hubei University of Chinese Medicine, Wuhan 430065, China
2. Hubei Provincial Hospital of Traditional Chinese Medicine, Wuhan 430060, China
3. College of Information Engineering, Hubei University of Chinese Medicine, Wuhan 430065, China
4. Jingzhou Hospital of Traditional Chinese Medicine Affiliated to Hubei University of Chinese Medicine, Jingzhou 434000, China

Abstract Through a systematic review of the origin and 40 years of development of Traditional Chinese Medicine(TCM) big data knowledge engineering, we clarify the major events and research achievements in each stage of the research, expound the ideas and methods of real world TCM clinical study based on medical records, emphasize the importance of transforming from data driven mode to data and knowledge driven mode to promote the innovation of the real world TCM clinical study model, and argue that the basis of model transformation and innovation is to reconstruct the knowledge attributes of TCM clinical case data base, make tacit knowledge explicit, realize full expression of knowledge, and build a TCM clinical case knowledge base. Taking case data knowledge, knowledge coding and knowledge mapping as examples, we highlight the key technologies and core links of TCM big data knowledge engineering, and fully confirm that big data knowledge engineering is the only way for real world TCM clinical study. Under the guidance of the national high-quality development strategy of TCM, with the goal of cultivating new quality productivity and forming a new business form of clinical research of TCM with new development concepts, we will build a big data knowledge engineering technology platform and a warehouse of TCM clinical knowledge, provide a technical platform and knowledge resources for carrying out large-scale model research in the field of TCM, and promote the clinical intelligence and high-quality development of TCM.

Keywords big data knowledge engineering; real world clinical research of traditional Chinese medicine; data knowledgeization; knowledge encoding ●



(责任编辑 徐丽娇)