

中医临床大数据知识工程规划与设计

张 盼 沈绍武* 田双桂 肖 勇 宋淑洁
(湖北中医药大学 湖北 武汉 430065)

摘要: 为解决如何充分利用丰富的临床数据资源,应用大数据知识工程理论和方法创建临床研究新模式和促进临床智慧化等中医临床研究重大问题,文章结合大数据的理念和知识工程的技术方法,在中医学理论指导下,遵循中医临床自身发展规律,对中医临床大数据知识工程进行系统规划和设计,提出“以知识关联方法揭示中医临床事实与规律”和“开展临床循证研究对其进行验证与优化”相结合的中医临床研究新模式,探讨以构建中医临床大数据知识工程的标准、方法和技术等三大体系为主要内容的开创性实施策略,为开展中医临床大数据知识工程研究提供理论基础和技术支撑,服务于中医药传承创新发展和实现医疗大数据治理能力现代化。

关键词: 大数据; 知识工程; 中医临床; 研究模式

DOI 标识: doi: 10.3969/j.issn.1008-0805.2022.03.74

中图分类号: R2-03 **文献标识码:** A **文章编号:** 1008-0805(2022)03-0764-03

云计算、大数据、物联网、人工智能等高新技术应运而生,标志着人类从“信息时代”进入“大数据时代”^[1]。大数据的理念和技术方法正在影响并改变着我们认识和理解世界的方式。随着数据获取愈发便捷,数据以不可估量的速度海量增加。虽然数据的存储处理技术不断的更新和成熟,但在如此庞大、混乱的数据面前,如何对海量的数据资源进行快速挖掘、高效地管理和利用,并发现有价值的知识和规律,是当前亟待解决的问题。同时,经过数十年信息化建设,全国各级中医医院基本实现了临床病历的信息化管理,积累了大量中医临床数据,这些数据是中医医院现代化的宝贵战略资源。如何管理和利用好这些资源,攻克临床数据整合的技术壁垒^[2],将临床医疗业务型数据转化为临床研究型数据,实现临床科研一体化^[3],服务于中医药传承创新发展和实现医疗大数据治理能力现代化,已经成为当前迫切需要解决的重大课题。

1 大数据知识工程研究现状

费根鲍姆(Feigenbaum)在1977年第五届国际人工智能联合会上首次提出“知识工程”^[4],是以知识为对象,研究知识系统的知识表达、知识处理和应用的等方法。随着大数据时代的到来,一门引领高科技前沿学科“大数据知识工程”应运而生,并立即引起了学界的高度重视和反响。中国人民大学、复旦大学、合肥工业大学等纷纷成立知识工程研究实验室^[5],围绕大数据、知识工程等进行知识管理、知识挖掘等研究。如吴信东教授将大数据与知识工程相结合进行研究,提出“大数据知识工程(Knowledge Engineering with Big Data, BigKE)”的HACE定理和框架理论;陆汝钊教授团队致力于知识工程平台开发、推广及应用等方面的研究^[6]。

中医以知识为研究对象起步于20世纪70年代末,1977年我国研制开发了第一个专家系统“中医关幼波肝炎诊断治疗程

序”^[7],湖北中医学院(现湖北中医药大学)成立中医控制论研究室研发黄绳武、吴绍基、吕继端等名老中医专家系统,以计算机技术来解决中医临床问题,但主要集中在名老中医个人经验临床诊疗的知识推理、用药规律等方面。中医特有的思想方法、含义深邃的中医概念以及大量临床的隐性知识都对其知识的完整表达和知识获取与知识利用产生重大影响和难度,大数据知识工程的理论和技术方法为中医药传承创新发展提供了一个全新的思路和途径^[8]。因此,中医临床大数据知识工程研究是当前我们面临的一个崭新课题。

2 中医临床大数据知识工程研究目的及意义

2.1 促进中医药标准化与信息化双向融合,推动中医药标准应用

中医药标准化是推进中医药行业治理体系和治理能力现代化的基础性制度。“十一五”期间,初步建立了中医药标准体系^[9],围绕该标准体系,我国发布了一系列中医药标准,但标准推广工作仍存在一定困难,标准应用方面比较薄弱^[10]。中医药信息化是中医药事业创新发展的技术支撑,“中医药标准化与信息化”创新发展的当务之急是实现双向融合^[11],更好为实现中医医院治理体系和治理能力现代化提供基础保障。对中医临床大数据知识工程的研究,一方面将标准融入信息系统,提升系统的标准化程度和数据的共享性,为中医临床数据的共享提供基础保障;另一方面将信息技术融入中医药标准化,实现中医药标准管理与应用工程化,实质性推进中医药标准的应用和拓宽信息技术的应用领域和途径,提升中医药标准化与信息化水平。

2.2 变革中医临床研究模式,提升中医临床研究创新能力水平

长期以来,中医临床研究模式是在现代医学的科学观指导下,在理想条件下进行的“以因果关系为目标,以随机对照试验(randomized controlled trial, RCT)为方法”开展临床研究^[12]。随着大数据时代的来临,新的世界观成为当今人类认识世界的主流,即“大数据世界观”。依据中医学的学术特征和临床实践需求,将大数据、人工智能等技术与中医药传承创新相结合,开展真实世界中医临床研究重构临床研究模式,探索发现中医临床诊疗数据中的“事实与规律”,用以开展中医临床循证研究,充分发掘中医临床数据资源利用价值与效率,提升中医临床研究创新能力,为中医药传承创新和智慧便民服务提供数据支撑与技术支持。这种创新型的“认识论和方法论”,即“从狂热的追求事物因果关系中解脱出来,以大数据技术方法研究事物关联关系”是对传统临床研究理念的一次极其深刻的变革。

收稿日期: 2021-10-11; 修订日期: 2021-12-10

基金项目: 湖北省2021年中医药传承与发展资金项目
(HBZY-2021-Z09)

作者简介: 张 盼(1993-),女(汉族),湖北汉川人,湖北中医药大学博士研究生,硕士学位,主要从事中医药标准化与信息化,中医医史文献工作。

* 通讯作者简介: 沈绍武(1962-),男(汉族),湖北武汉人,湖北中医药大学教授,博士研究生导师,硕士学位,主要从事中医药信息化与标准化研究及教学工作。

2.3 促进中医临床大数据治理能力现代化,推进中医临床智能化服务 大数据科学正在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。大数据科学产生了新的认识论、方法论和技术体系,运用大数据知识和技术充分挖掘和利用中医临床数据资源,构建具有数据支撑的临床知识库和开发中医临床智能应用系统,发挥中医药的独特作用,逐步实现中医医院智慧医疗,推动其管理理念、管理模式和管理技术的变革与创新,实现中医医院大数据治理能力现代化,推进中医临床智能化服务^[13]。

3 中医临床大数据知识工程体系构建

本文将大数据知识工程应用于中医临床,对中医临床大数据知识工程进行系统规划和设计,旨在以中医学理论为指导,遵循中医临床自身发展规律,探讨“以知识关联方法揭示中医临床事实与规律”和“开展临床循证研究对其验证与优化”相结合的中医临床研究新模式,将中医临床隐性知识显性化,形成包括显性和隐性知识完整表达的中医临床知识库,开展中医临床知识关联分析,构建精准化中医临床诊断、干预与评价模型,绘制中医临床知识图谱,形成“从临床中来,到临床中去”的闭环反馈控制体系,寻求解决中医传承创新关键问题的方法和途径,探索发现真实世界中医临床诊疗数据中的“事实与规律”,为开展临床循证研究提供思路和目标,加速中医临床知识模型化以及知识库建设,提升中医医院临床研究创新能力以及医疗服务质量和水平,为实现中医医院智慧化及其治理体系与治理能力现代化提供重要技术支撑。

3.1 中医临床大数据知识工程体系框架 中医临床大数据知识工程主要研究内容是整合与利用中医医院现有的临床数据资源,构建中医临床大数据知识工程标准、方法和技术等三大体系,然后选取临床专科专病通过循证研究验证与优化临床模型用以指导临床实践。其中医临床大数据知识工程体系框架图,如图 1 所示。

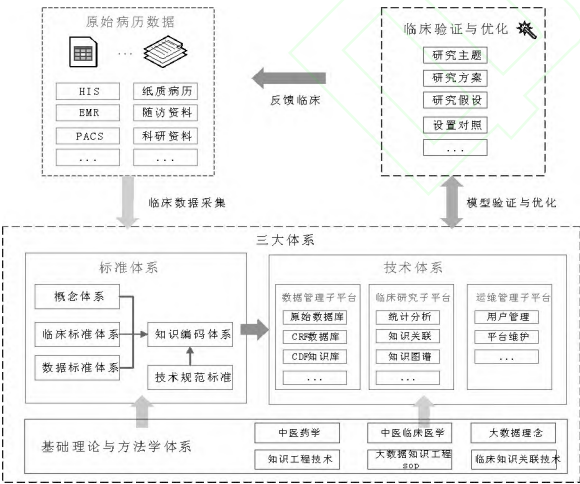


图 1 中医临床大数据知识工程体系框架图

3.2 数据的采集 医院临床数据大多以满足医疗业务为主,难以完全满足临床研究需要。因此,为获取具有中医特点的真实世界临床病例数据,根据中医临床数据标准对中医临床原始资料,包括电子病历、医院信息系统(Hospital Information System,HIS)等资料进行规范化,其主要信息包括患者基本信息、临床诊疗信息、临床治疗信息以及临床评价信息等。并设计统一的中医临床病例报告表(Case Report Form,CRF)构建 CRF 数据库将中医临床数据结构化,便于后续计算机自动抽提和处理,提高数据采集的

工作效率。

3.3 标准体系建设 标准化在中医药发展中起着基础性、战略性、全局性作用,标准体系是一定范围内标准按其内在联系形成的科学的有机整体。立足于当前中医药创新发展的新形势和新需求,以中医学理论为基础,遵循中医临床自身的发展规律,构建标准体系框架,完成其标准体系的顶层设计,以此指导相关标准的制修订,建立健全中医临床大数据知识工程标准体系,为中医临床大数据知识工程提供规范化指导和标准支撑。

3.3.1 建立概念体系 以临床中医学、大数据科学和知识工程学等三个学科的基本概念为基础,以中医临床知识体系为纲,设计中医临床大数据知识工程概念体系框架,系统收集整理相关概念,按照概念间的内涵关系,形成中医临床大数据知识工程名词术语库,形成一个有机联系的概念共同体,为中医临床大数据知识工程提供用语规范化指导和标准。

3.3.2 构建临床基础知识编码库 通过建立临床标准体系和数据标准体系,将临床数据结构化和数据化,即将真实世界临床普通文本格式的病历资料转变成高度结构化电子病历,形成结构化 CRF 数据库,再以临床基础知识编码为据,对其进行知识属性化重构。将临床隐性知识显性化,进而形成数据化中医临床病例(Case Data Form,CDF)知识库,为构建精准化中医临床诊疗知识模型,揭示临床的事实与规律提供数据资源。

3.3.3 建立技术规范体系 技术规范体系主要对知识处理流程(Standard Operation Procedure,SOP)和知识库构建技术进行规范。中医临床大数据知识工程知识处理涉及知识的表达、获取、组织和知识的模型化、图谱化以及智能化应用等多个环节,基于真实世界中医临床知识的特性,为保障工程的有序和有效开展,必须构建一个标准化中医临床大数据知识工程的知识处理流程,对不同处理节点进行规范,最终形成一个环环相扣、承上启下的知识处理闭环,作为指导构建中医临床知识库的依据。

3.4 理论与方法体系建设 理论和方法学体系是学科建设的基础,中医临床大数据知识工程是中医临床医学、大数据科学和知识工程学相互交融所形成的一门新学科,是根据该学科的研究领域、研究目标和行为特征有机整合而形成,包括体系框架、知识类型、知识特点以及知识表达、知识获取、知识组织、知识关联、知识模型、知识图谱等中医临床大数据知识处理的理论与方法,为中医临床大数据知识工程提供理论与方法学支撑。

3.4.1 建立理论体系 将中医药学、中医临床医学以及大数据理念等多学科的深度融合,建立符合中医特有的思想方法、含义深邃的中医概念的中医临床大数据知识工程理论体系,突破传统的以专家系统进行知识获取和知识再工程的瓶颈,解决中医临床数据资源利用的问题,为建立中医临床研究新模式和新途径打下坚实的理论基础。

3.4.2 建立方法体系 方法学(Methodology)是以方法为研究对象,专门探索方法的一般结构、发展趋势和方向,以及各种方法的相互关系的专门学问。根据理论体系,从知识工程技术、大数据知识工程 SOP 和临床知识关联技术等不同方法进行系统梳理,构建中医临床大数据知识工程方法学体系的整体框架。其中,SOP 是整个方法学体系中的核心重点及重要环节之一。

3.5 技术体系建设 技术体系建设主要以真实世界中医临床数据为研究对象,依据大数据知识工程的知识处理 SOP 规范要求,构建中医临床大数据知识工程技术平台,该平台从“以数据为中心”迁移到“以知识为中心”,涵盖数据库、自然语言处理、知识图谱、统计分析、关联分析、可视化等核心技术,包含资源管理、数据

管理、知识管理、知识处理、知识服务和运维管理等功能,为开展中医临床大数据知识工程研究提供工作平台和技术支撑,为临床专业人员提供中医临床知识服务,为智慧中医奠定坚实基础。

3.5.1 数据管理子平台 数据管理子平台主要是对多源异构数据库进行管理,核心是通过自动提取技术、知识库自动编码技术快捷、方便的将临床电子病历数据进行转换和存储,提高临床科研人员数据处理的效率和数据的可获得性。通过该平台,将临床医疗业务型数据转化为分析型数据,使其隐性知识显性化,深化中医临床知识表达层次,丰富中医临床知识的内涵,其数据库类型包括原始电子病历数据库、标准库、CRF 数据库、CDF 知识库等。

3.5.2 临床研究子平台 临床研究子平台主要是基于数据管理子平台建立的 CRF 数据库、CDF 知识库,从不同维度、不同主题进行分析和建模。根据中医临床知识关联分析的需求进行组织,服务于不同专病专科的中医临床知识主题特征数据提取和分析应用,如证候、症状、舌诊、脉诊以及中药等之间的关联分析,构建精准化中医临床诊断、干预与评价模型,绘制中医临床知识图谱。

3.5.3 运维管理子平台 运维管理子平台主要进行系统维护和用户管理等。平台的运维与安全管理涉及硬件和软件两个方面,主要实现应用(标准、临床数据等)数据库、数据分析软件、数据可视化软件、主题数据抽取软件等的管理与维护;网络、服务器、接入端等硬件设备的管理与维护。此外,需要建立完善的平台安全体系保障平台的正常运行。

3.6 临床验证与优化 基于大数据理念,应用知识工程技术方法,开展真实世界中医临床知识关联分析研究,构建中医临床诊疗模型,再以随机双盲对照方法开展临床循证研究对模型进行验证与优化,最终形成精准化诊疗模型,用以指导临床实践。同时,对中医临床大数据知识工程技术平台应用进行完善和优化。依据中医临床大数据知识工程规划和设计,建立中医临床研究示范单位,开展专科专病相关标准、知识编码、知识库和临床智能应用系统等研究,探索在全国开展中医临床大数据知识工程研究的途径与方法,并提供可借鉴的成功案例。

4 总结与展望

中医临床大数据知识工程是一个系统工程,具有长期性和复杂性,需要建立创新型研究范式和多学科协同。不同于传统的循证医学研究模式,该工程其核心是将大数据、人工智能等技术与中医药传承创新相结合,构建知识编码体系,建立中医临床病例

知识库,将中医临床知识中所蕴含的隐性知识显性化、数据化;以临床专病专科为具体实例,深化中医临床知识表达层次,挖掘中医临床知识内涵,构建中医临床知识模型,揭示真实世界中医临床诊疗数据中的“事实与规律”,用以开展中医临床循证研究,为形成精准化中医临床诊断、干预与评价模型和构建临床知识库与智能应用系统提供重要的数据资源。前期已初步完成对中医临床数据结构化与知识关联方法学的探索,以大数据世界观指导构建中医临床研究新模式,下一步将围绕中医临床大数据知识工程三大体系,建立中医临床大数据知识工程技术平台,研发中医临床 AI 系统,推进技术成果产业化,提升中医临床研究创新能力,服务于中医药传承创新发展和临床智慧化。

参考文献:

- [1] Lynch C. Big data: How do your data grow[J]. Nature, 2008, 455(7209): 28.
- [2] 曾宪涛,朱风雷,任学群,等.基于临床科研一体化技术的临床研究[J].中国循证心血管医学杂志,2017,9(10):1156.
- [3] 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.
- [4] 高建.基于 Catia 知识工程及二次开发的快速建模技术[J].工程与试验,2019,59(3):8.
- [5] 吴信东,陈欢欢,刘均,等.大数据知识工程基础理论及其应用研究[J].中国计算机学会通讯,2016,12(11):68.
- [6] 吴信东.从大数据到大知识: HACE + BigKE[J]. 计算机科学,2016,43(7):3.
- [7] 张德政,哈爽,刘欣,等.中医药领域人工智能的研究与发展[J].情报工程,2018,4(1):13.
- [8] 赵敏.新时代中医药治理新思维[J].时珍国医国药,2021,32(4):1009.
- [9] 宇文亚,韩学杰,谢雁鸣,等.中医药标准推广评价初探[J].中医杂志,2012,53(18):1609.
- [10] 宇文亚,韩学杰,史楠楠,等.中医药标准化的现状分析与思考[J].世界科学技术-中医药现代化,2011,13(3):445.
- [11] 常凯,付文娇.中医药标准推广应用策略研究[J].中华中医药杂志,2017,32(3):1156.
- [12] 张润顺,王映辉,刘保延,等.基于共享系统的真实世界中医临床研究范式初步实施方案的设计[J].中医杂志,2014,55(18):1551.
- [13] 肖勇,沈绍武,孙静,等.后疫情时代中医药信息化建设与发展的思考[J].时珍国医国药,2020,31(12):3055.