

大数据背景下的地域性 中医学学术流派研究*

郜 密^{1,2}, 王振国^{1**}, 张丰聪¹

(1. 山东中医药大学中医文献研究所 济南 250355; 2. 安徽中医药大学中医临床学院 合肥 230038)

摘 要:随着大数据时代的来临,大数据的理念与技术应用于中医药领域的研究正在蓬勃发展。中医学学术流派是中医学学术发展与理论创新的重要形式,是中医学发展史上的一个鲜明特色。当前中医学学术流派数据具有异构性、广分布、多样性、海量等特点。本文通过分析大数据时代带来的影响,提出将大数据分析技术应用于中医学学术流派研究的思路和方法,期望能为地域性中医学学术流派大数据环境提供一个坚实的支撑平台,推动地域性中医学学术流派的传承与发展。

关键词:大数据 hadoop技术 地域性中医学学术流派

doi:10.11842/wst.2018.01.005 中图分类号:R22 文献标识码:A

近年来,随着“大数据时代”的来临,大数据的理念与技术应用于中医药领域的研究逐步开展。在中国,海量的中医药领域数据日益剧增^[1],大数据技术的逐渐成熟也为中医药领域的数据分析带来了新的思路,如何充分利用好积累的数据和信息,并把数据和信息的分析提升到前所未有的高度,是大数据时代的重大挑战。

中医学学术流派是中医学学术发展与理论创新的重要形式,是中医学发展史上的一个鲜明特色。不同的地域形成了不同的学术流派,其间的相互争鸣与渗透,又进一步促进了中医学学术的发展,最终形成了中医学“一源多流”的发展格局。

当前,中医学学术流派数据具有异构性、广分布、多样性、海量性等特点。本文在既往研究的基础上,以地域性中医流派为切入点,探讨在大数据背景下地域性中医学学术流派研究的思路和方法。

1 大数据时代为中医药发展提供了新的机遇和挑战

在信息日益普遍化的今天,正在兴起的大数据技术革命为中医药走向现代化提供了有效的技术手段^[2,3]。大数据,即通过高速捕捉、发现或分析,从大容量数据中获取价值的一种新的技术架构,具有大量(volume)、高速(velocity)、多样(variety)、价值(value)等“4V”特点^[4]。大数据技术革命给中医的现代化带来了前所未有的机遇,中医药信息具有模糊性和经验性的特点,有了大数据技术,中医药信息的数据化和现代化就有了走向现实的可能性^[5]。如何从凌乱、无模式和复杂的大数据中挖掘出有用的知识,需要使用到多种方法来实现数据的挖掘和处理,包括:机器学习与数据挖掘算法(关联规则、决策树、模糊理论、粗糙集、人工神经网络及遗传算法等)、预测性分析、可视化分析、语义引擎、知识计算等^[6-10]。

大数据时代技术的逐渐成熟也为中医药领域的数据分析带来了新的思路,已经广泛运用到中医药文献

收稿日期:2017-06-09

修回日期:2017-10-29

* 中国博士后科学基金面上资助项目(2016M592240):地域性中医学学术流派评价的方法学研究,负责人:郜密;安徽中医药大学校级科研项目(2016ts001):基于文献计量学的新安医学文献评价研究,负责人:郜密。

** 通讯作者:王振国,教授,主要研究方向:中医文献及中医流派研究。

数据、中医诊断客观化数据、中药及复方数据、中医临床数据等领域的相关研究。

中医学在其历史悠久的发展过程中,积累了海量的中医药文献资源。如何有效提取非结构化文本中的模式知识,是当前中医药文献数据中需要处理的重要问题。如有研究建立的古今名医辨治数据库,将文献分为古代、近现代、现代三个部分,古代和近现代部分为固态数据库,现代研究进展部分借鉴云数据库,建立动态数据库,实现数据库的开放式优化^[11]。而在中医诊治过程中,一方面依靠传统的望、闻、问、切等手段来获取疾病资料,同时随着各种新兴触感器,如舌诊仪、脉诊仪、可穿戴设备等出现和应用,以及医院信息系统的普及,极大丰富了中医诊疗数据,如有学者^[12]利用人神经网络构建了中医舌诊知识库等。中药及复方数据,更是纷繁复杂,通过分析灵活运用这些庞大、多类别的数据,可以进一步完善中药药性理论、指导中药复方开发^[13]。中医临床数据研究中,有学者正在展开对名老中医经验或相关地域流派的数据挖掘,获得其临床诊疗及用药规律。如有学者对“吴门医派”的相关数据挖掘,对其相关文献、辨治特色、临床经验等方面做了详细的研究^[14]。

2 地域性中医学术流派研究蓬勃发展

在几千年源远流长的历史发展进程中,中医学界涌现出了众多的医学名家和学术流派^[15]。不同中医学学术流派之间各种医学学说的争鸣与交融,不断地补充和深化中医药学的理论体系,也不断地完善和提高中医药学术水平和临床疗效,极大地促进了中医药学术的进步和发展^[16-20]。国家对中医学术流派的研究高度重视,自2006年起,国家中医药管理局设立“中医学术流派研究”科技专项,2008年设立了“中医学术流派研究室”,开展了一系列的研究,2012年12月又从中国500多家单位中遴选公布了第一批64个全国中医学术流派传承工作室,开启了从国家层面上集中人力和物力对代表性中医学术流派进行规范化建设,并对其进行研究、保护和传承的新阶段。

自20世纪七十年代至今,整理、发掘、研究地域性学术流派的趋势已经逐步趋向一个新的高潮^[19]。目前对地域性学术流派的整理研究,主要涉及新安医派、齐鲁医派、海派医派、钱塘医派、孟河医派、岭南医派、永嘉医派、湖湘医派、闽台医派、吴中医派、盱江医派、御医学派等^[25]。而自称流派者,达500多家。近年来每

个流派都对自身的发展历程进行了梳理,从形成因素、名医名著、学术特色、临床经验等不同角度展开了论述,发表了一系列的论文,出版了一系列的著作。

尽管地域性中医学术流派的研究呈现百花齐放的繁荣局面,但仍存在以下几个问题:①现有研究过分强调了个性,而忽略了流派之间的共性,学术流派资源需要深入整合研究;②各个流派的成果都在各个机构分散保存,未能实现资源共享和协作;③各个地域性中医学术流派缺乏统一的评价标准;④各个流派的学术价值需要进一步挖掘提升。

3 大数据背景下的地域性中医学术流派研究思路

3.1 总体思路

在大数据时代,应赋予流派研究以新的意义。可将大数据分析技术(机器学习与数据挖掘算法、预测性分析、可视化分析、语义引擎、知识计算、数据质量和数据管理等)与地域性中医学术流派研究相结合,对当前散在的多类数据源载体中的海量信息和知识进行集中式管理,为中医学术流派大数据环境提供坚实的支撑平台,并提供高效、高精度的大数据分析与挖掘,实现地域性中医学术流派数据的资源共享和知识发现。同时,构建“地域性中医学术流派评价方法学体系”中的一级、二级、三级评价要素,期望形成行业内公认的标准和规范,进而推广至地域性中医学术流派的评价中。

3.2 研究方案

3.2.1 调查梳理

针对当前分布于各个机构、数据库、知识库等多类数据源载体中的有关中医流派的信息和知识进行梳理调研。

3.2.2 平台构建

搭建hadoop平台和相关上层应用系统。

3.2.3 知识管理

设计开发“基于hadoop技术的地域性中医学术流派知识库”数据平台,将相关数据录入并规范统一,数据维护包括对数据的容错处理、容灾备份,保障平台数据的完整和一致性,实现中医学术流派数据的可视化。

3.2.4 数据分析

以组件化的形式集成多种数据计算与处理模型,其中包括mapreduce经典并行计算模型,高效的数据抽取与查询组件Hive和Spark SQL,并且可以集成内存计算模型spark,对地域性中医学术流派大数据进行分析处理。

3.2.5 系统应用

采用模块化形式,如关联计算、知识计算、语义分析、数据挖掘、机器学习、预测分析以及丰富的可视化模块,实现系统应用。提炼不同流派之间的共性规律,挖掘各个流派的特色诊疗技术,为临床诊疗服务。

3.2.6 评价体系

研究中将遴选主要评价指标,采用专家访谈以及问卷调查的方法,拟定一级、二级、三级评价要素,构建“地域性中医学术流派评价方法学体系”。

3.3 平台架构

3.3.1 数据存储层

数据存储层是整个平台的基础,主要功能是通过分布于各个流派数据库、文献知识库、web 站点以及各个机构中的多源数据进行抽取,并统一转化存储进hadoop集群中。基于hadoop的分布式文件系统hdfs,以及构建于之上的hbase分布式数据库对数据进行统一

组织,并提供高容错和灵活的拓展性,是平台的中枢。

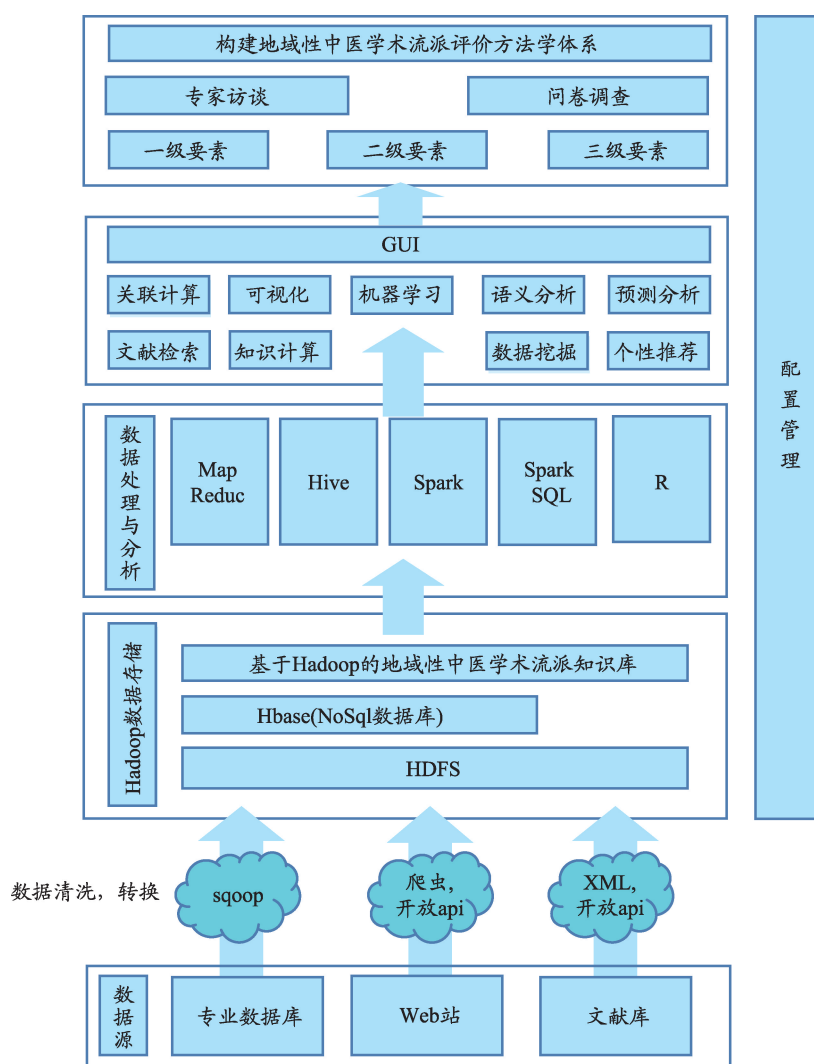
3.3.2 数据处理分析层

数据的处理与分析在平台中起着管道的作用,主要为上层的业务应用提供支撑。以组件化的形式集成多种数据计算与处理模型。其中包括mapreduce经典并行计算模型,高效的数据抽取与查询组件Hive 和 Spark SQL,并且可以集成内存计算模型spark。是连接数据持久层和业务应用层的纽带,并行计算模型保障了对中医学术流派大数据快速,高效的处理能力。

3.3.3 业务应用层

业务应用层是平台的关键,采用模块化形式包括各类应用系统,数据挖掘与机器学习引擎,以及丰富的可视化模块,这些模块可以根据需求灵活组织和拆分,实现中医学术流派数据的集成和分析。

3.3.4 平台架构图



4 结语

大数据时代的来临,带来了挑战,也带来了机遇,为中医药领域的研究打开了新的视角。地域性中医学流派是中医学术发展与理论创新的重要形式,是中医发展史上的一个鲜明特色。然而当前地域性中医学流派数据具有异构性、广分布、多样性、海量等特点。如果将大数据分析技术应用于地域性中医学流派

研究,无疑能够对当前分散的中医流派数据实现资源和知识的共享,实现中医学术流派数据的多维分析及可视化呈现。多学科方法的应用,丰富了地域性中医学流派研究方法论的内容。而在此基础上构建的“地域性中医学流派评价方法学体系”,则可以为地域性中医学流派大数据环境提供一个坚实的支撑平台,从而推动地域性中医学流派的传承与发展。

参考文献

- 1 韩雅丽,付先军,张丰聪,等.文献计量学视角的中医药文献信息化研究现状探讨.世界科学技术—中医药现代化,2015,17(3):427-433.
- 2 黄义强.基于Hadoop的中医数据储存平台设计与开发.江西中医学院学报,2011,23(6):10-14.
- 3 黄欣荣,张艳朋.大数据技术与中医现代化.中医杂志,2014,55(19):1621.
- 4 张华敏,王永炎.高概念大数据时代中医理论研究的机遇.中国中医基础医学杂志,2015,21(1):4-6.
- 5 黄欣荣.复杂性科学与中医.中医杂志,2013,54(19):1621-1626.
- 6 夏于芬,梁光平.大数据背景下的中药现代化.亚太传统医药,2015,11(21):1-3.
- 7 何清,李宁,罗文娟,等.大数据下的机器学习算法综述.模式识别与人工智能,2014,27(4):327-336.
- 8 张艳.大数据背景下的生物医学信息处理.生命科学仪器,2014,12(10):17-20.
- 9 张春丽,成斌.大数据分析技术及其在医药领域中的应用.标记免疫分析与临床,2016,23(3):327-333.
- 10 林子雨,赖永炫,林琛,等.云数据库研究.软件学报,2012,23(5):1148-1166.
- 11 李婧,吴立旗,童文新,等.基于数据挖掘分析《清宫医案集成》止咳方药的应用.中华中医药杂志,2015,30(1):270-273.
- 12 周金海,杨涛,沈大庆,等.基于ANN的中医舌诊八纲辨证知识库构建与应用.计算机应用研究,2010,27(5):1771-1772.
- 13 杨薇,崔英子,杨海森.医疗大数据在中医药研究领域的应用与思考.长春中医药大学学报,2016,32(3):625-627.
- 14 杨税,吴梦婷,李辉.吴门医派历史发展初探.亚太传统医药,2016,12(18):16-17.
- 15 马杰,严世芸.近30年研究历代中医学术争鸣之文献述评.中国中医基础医学杂志,2012,18(8):923.
- 16 王琦.没有新学说就没有新流派.北京中医药大学学报,2011,34(5):293-297.
- 17 中医学术流派研究课题组.争鸣与创新:中医学术流派研究.北京:华夏出版社,2011:5.
- 18 宋咏梅,王振国,刘更生.关于当代中医学术流派评价的几点认识.辽宁中医杂志,2011,38(7):1349-1350.
- 19 洪净,吴厚新.对中医学术流派传承发展中一些关键性问题的思考.中华中医药杂志,2013,28(6):1641-1643.
- 20 黄政德.论医学流派对中医学发展的影响.中医杂志,2000,41(1):7-9.

Research on Regional Academic Schools of Traditional Chinese Medicine under Big Data Background

Gao Luan^{1,2}, Wang Zhenguo¹, Zhang Fengcong¹

(1. Traditional Chinese Medicine Literature Institute, Shandong University of Chinese Medicine, Jinan 250355, China;

2. Clinical College of Chinese Medicine, Anhui University of Traditional Chinese Medicine, Hefei 230038, China)

Abstract: With advent of the big data era, applying ideas and techniques of big data in the research field of traditional Chinese medicine (TCM) is progressing. As a distinctive feature in the history of TCM development, TCM academic school is an important form of academic development and theoretical innovation of TCM. Current data of TCM academic school has the characteristics of heterogeneity, wide distribution, diversity and magnanimity. This paper intended to

apply big data analysis technology to the research of TCM academic school based on analyzing the impact of big data age to promote the inheritance and development of TCM academic schools in the region.

Keywords: Big data, hadoop technology, regional academic schools of traditional Chinese medicine

(责任编辑:张 静,责任译审:王 晶)