

Research Status and Development Trend of Big Data in Traditional Chinese Medicine

Xinping Zhang Zhengxi Li* Yi Sun Yachuan Du Shichen Shen Ting Li

Chengdu University of Traditional Chinese Medicine, Chengdu, Sichuan, 611137, China

Abstract

Objective: Mainly to summarize, summarize and analyze the application of big data in traditional Chinese medicine. **Methods:** Mainly through reviewing the literature and journals in recent years. **Results:** Through research literature and other data, it is found that the application of big data in traditional Chinese medicine is not comprehensive and extensive. **Conclusion:** The development of big data in traditional Chinese medicine is still very large, and development is on the rise. We should increase publicity and research.

Keywords

traditional Chinese medicine; applications of big data in traditional Chinese medicine; future development

中医药大数据研究现状和发展趋势

张心萍 李正熙* 孙毅 杜雅川 沈诗晨 李婷

成都中医药大学, 中国·四川 成都 611137

摘要

目的: 主要对大数据在中医药上的应用进行一定归纳、总结以及分析。**方法:** 主要通过回顾近几年的文献及期刊。**结果:** 通过研究文献及其他资料发现中医药大数据在中国的应用并不全面和广泛。**结论:** 中国中医药大数据的发展空间还非常大, 发展处于上升趋势, 我们应加大宣传和研究。

关键词

中医药; 中医药大数据应用; 未来发展

1 引言

“大数据”是需要新处理模式才能具有更强的决策力、洞察发现力和流程优化能力来适应海量、高增长率和多样化的信息资产。而中医药完全具有大数据的特征容量 (Volume) 种类 (Variety) 速度 (Velocity) 可变性 (Variability) 真实性 (Veracity) 复杂性 (Complexity) 价值 (value)。中医药与大数据的结合可以更方便的进行对中医药庞大数据的管理和分类, 便于对中医药的研究和传承, 如果缺少大数据对中医药的应用, 只会导致珍贵资源的继续流失, 以及中医药的隐藏财富无法被发掘。

2 过去研究中医药的方法

中医药 (traditional Chinese medicine) 在中国历史悠久, 是中国大部分人的首选治疗方式, 中医药在最初是通过医者

的临床实践总结经验, 自己感悟升华, 得到自己的学术思想和学术结论, 最后再到临床指导实践, 大部分是通过师徒关系来传承, 但是有俗语: 凡事留一手, 导致大量的珍贵中医药资料遗失, 并且以往主要以纸张来记载和保存, 纸张保存的最大弊端在于, 纸张的保管时间极短, 并且容易损坏, 书籍与书籍之间不便建立联系, 给中医药的研究造成了很大的阻碍, 并且医者之间不能交流, 大部分只能闭门造车, 使得中医药的发展十分的缓慢, 当时他们也并未意识到大数据的重要性。大量的资料显示, 中国大部分的中医学术资料保存情况十分的惨淡, 大部分的文献内容只能通过猜测最后再通过各种实验得以证实或被推翻。传统中医对医学的实验比较传统和保守, 因为数据不能得以共通, 所以个人医者会单独的做实验, 导致中医的创新研发不能得以推广也极易被迫淘汰和掩埋。中医药大数据的出现, 不仅为中医药提供了发现新知的的方法, 而且促进了原有的中医药研究方法的重新整合,

从而为中医药的研究发展提供了新的机遇。^[1]如图1所示。

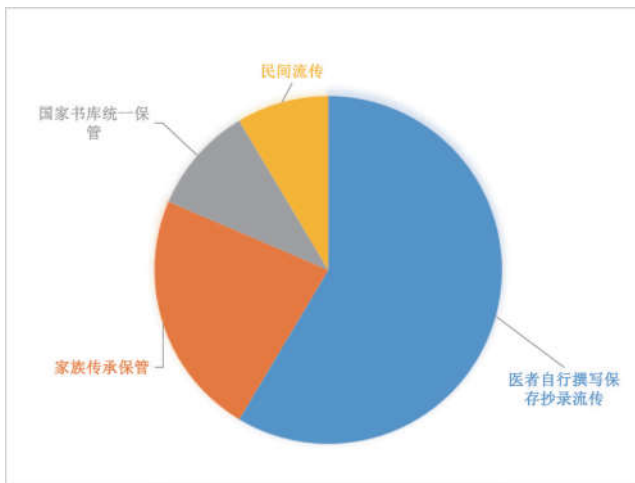


图1 中医药资料保存方式

3 中医药大数据的发展历程

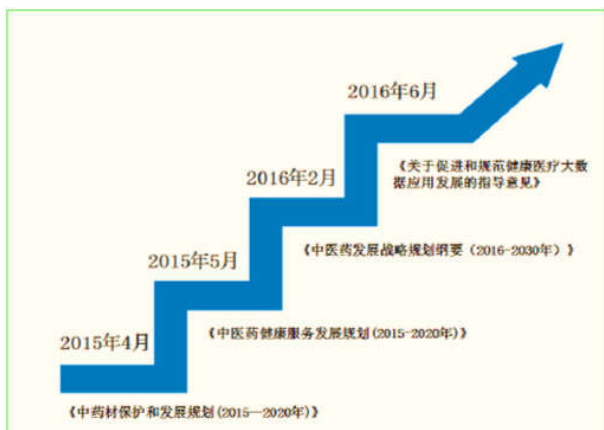


图2 中医药大数据的发展历程

2002年8月apache.org的开源项目NUTCH最早引用“大数据”这个术语，互联网慢慢的变得广泛，人们也开始在网上获取资源和信息，互联网时代带来了“大数据”的理念，大数据在中医药领域中的应用也逐渐显现，越用越广，中医药也利用大数据应用开始展现它隐藏的价值，中医药信息学、中医药预防疾病、中医诊断疾病、中医治疗疾病、中医方药知识发现、中药开发等相继用到大数据的挖掘和分析方法，中医药大数据发展分为：手工1.0、数字2.0、智能3.0、一体4.0四个阶段。手工1.0为问诊过程通过望闻问切来感知；数字2.0为采集一些数字化的信息；智能3.0为采用个体化、非常态数据自动检测；一体4.0为采用个体化、非常态数据自动检测。中医药信息学、中医药预防疾病、中医诊断疾病、中医治疗疾病、中医方药知识发现、中药开发等相继用到大数据的挖

掘和分析方法。如图2所示。

4 中医药大数据的现状热点以及中医药大数据的应用

对于海量并瞬息万变的大数据来说，存储必然不是首要目标，更重要的是如何从这些数据中获取包括商业价值在内的红利，才是真正的意义所在。中医三千多年的历史，积累了大量的数据，数据库的建立使得大量的中医数据信息转化为可检索的电子信息形式，以极快的速度发现所需要的信息。中医药大数据可以协助医生更加了解病人的病况，以及去预测病人对于某些疾病的易感染性和某些隐藏疾病的突发性和多种疾病的并发性和对病人的不利条件，以此来减少治疗过程的时间和花费，以免造成药不对症或用药无效或用药过敏的尴尬情况，并实施更精准的治疗方案。强有力的避免了医生和患者以及患者家属之间可能发生的矛盾和冲突。

中医药领域具有典型的大数据特征：数量巨大、种类丰富、处理速度快、数据价值极高。如中国中医科学院2012年4所医院全年的门诊达到698万；在中医院，每个患者不但要经过辩证论治的个体化诊疗，还会经过各种理化检测进行疾病及其预后的诊断，所以不光有病历资料中包含的信息，还会有生化检验，多种影像或病理切片检查的生物学信息。中医药材经过统计超过12800种，通过书本记载不会出现误用、少用甚至将某种药材彻底遗忘的情况。但是将大数据引入中医药中后，将各个药材的名称、属性、用法、用量归纳入库，这样对科学家的研究进度会有重大提高，对医生开药方也是变得非常简洁，对中医药材有兴趣研究的相关人士，在他们的学习过程中也更简洁更容易明白。大数据也会整体贯穿与中医药学理论与临床各个领域，强调结构与功能的统一性及完整性，即把研究对象看做有机整体，反对孤立、片面、局部地阐述事物及现象的内在规律及特征。所以在研究对象研究角度及层面认识上，中医药学与大数据研究均立足于整体，均以揭示事物的整体规律为指归^[2]。重新审视我们已固定化的中医药思维，中国在中医药领域的传统思想及学术观点可以简洁概括为，通过医者的临床实践得出个人经验再通过师徒关系将个人经验得以传达及延续。通过大数据将传统的中医药文化现代化、科技化、数字化，把经验与学术知识得以共享和讨论最终得以发展，将中医药发展以及更深入的研究

实现的更彻底更有效。中医药大数据若要应用在临床中则需要通过多次的辩证论证, 实验证明来得以实现完成个人一体化的治疗方案, 并分类管理患者的医疗资料, 保证医生能随时观察患者的病情发展, 形成庞大的患者数据库, 来实现大数据应用于中医药的大价值。中医药大数据使理论变为研究最后应用在患者身上。中医药大数据通过传统理论对数据进行分析, 关注整体健康与疾病, 体现了群体社会的主要特征, 结合大众心理来发掘中医药大数据的深层意义, 以及获取更多多方面的数据和信息。使中医药文化应用的更全面, 让大众从中医药中获益更多。

大数据应用于中医药领域最大的价值是从中医药知识中获得相关联的知识的追求, 并综合和高效的利用中医药的数据和资源, 感知所处环境中的疾病并预测未来, 并判定与调控疾病的发展趋势。^[1]

与西医相较, 中医以临床实践为基础, 以人为核心, 把疾病对人的衣食住行、精神状态、心情的影响的外在表现, 与干预、效果相关联, 形成了辩证论治个性化诊疗的体系。

利用大数据的手段揭示辩证论治个性化背后隐含的深刻规律, 甚至可以利用可穿戴设备监测实时的个体数据, 利用可视化方法来勾画每一个人的健康知识图谱。

亚洲相比于欧洲大数据比较落后, 更多的还处于探索学习阶段, 现今中医药大数据可以来源于网络数据、电子信息、书本文案等非专业中医药资料, 但从这些中医药大数据总结, 得出规律和中医药理论。目前中国中医药大数据除了文献期刊数据以外, 尚缺乏大规模且非常规范的数据库, 即使现在的中医药诊断的总数据十分的庞大, 但需要大范围, 多次数, 多人员的信息数据库建设, 才得以保存且高效的利用中医药大数据, 由此可见, 中国现在对中医药大数据的建设还有待提高并需要更多此方面的人才来共同加入, 推进中医药大数据的发展, 现在中医药大数据分布的比较零散, 专业介绍和归纳中医药大数据的图书或文档比较少, 中医药大数据的资料和信息都被零散的分布在各个医学领域, 建立独有的资料库和数据库是首要任务。并且未来亚洲会成为大数据的主战场, 因为世界总人口的 60% 都居住在亚洲, 亚洲的上网人数也在直线上升。

2007 年中国卫生部信息中心对 3765 所医院临床信息系统应用比例调查结果显示: LIS 系统应用率为 26.43%, PACS

系统的应用率仅为 9.00%^[5]。这表明中国的临床医疗信息技术还处于初始发展阶段, 各界人士对于这方面的认知急需提高, 加大宣传力度, 和各医院的配合, 争取实现医疗信息数据系统一体化。与此同时, 中医药大数据的获取也并非易事, 个人个体难以获得中医药大数据的详细资料, 以致难以达到大数据的应用, 信息数据及时公开十分必要, 并且公开要确保完整性, 可靠性, 安全性, 一致性, 避免个人个体获取到错误信息错误数据, 进行深入研究, 得到错误结论。并且要完全做到患者的个人信息保密问题, 防止造成患者个人信息泄露, 为患者造成不必要的困扰。

中医药大数据的研究现状主要是通过文献分析研究, 万方数据库检索历年中医药大数据方面的文献, 得出中医药大数据相关文献 294 篇, 文献量逐年增加, 文献类型中应用研究类文献占 59.1%。研究领域中证候研究与方剂研究文献分别占 29.39% 和 26.22%; 数据挖掘方法采用关联规则、频数分析、聚类分析和人工神经网络的文献占 76.36%; 数据挖掘软件使用商业化数据挖掘软件的文献占 61.05%。详见表 1。

表 1 中医药大数据研究现状

中医药大数据相关文献	中医药大数据相关文献目前有 249 篇, 并且文献量逐年增加
应用研究类文献	应用研究类文献占 59.1%
研究领域中证候研究	研究领域中证候研究占 29.39%
研究领域中方剂研究	研究领域中方剂研究占 26.22%
数据挖掘方法采用关联规则、频数分析、聚类分析和人工神经网络	数据挖掘方法采用关联规则、频数分析、聚类分析和人工神经网络占 76.36%
数据挖掘软件使用商业化数据挖掘软件的文献	数据挖掘软件使用商业化数据挖掘软件的文献占 61.05%

所以目前中医药大数据的应用主要体现在临床决策支持方面, 大数据分析和数据挖掘技术适合处理中医药数据, 但目前数据挖掘技术还未能改变中医药数据难以处理的大格局。

5 大数据作用于中医药的效果

在中国中医药几千年的发展史上, 前人给我们留下大量的理论知识和经验, 但是因为当时只能通过古籍记载, 大量的信息缺失以及许多古籍的重要部分受损导致整个古籍不能使用, 但是我们可以通过在与之相关联的古籍中经过揣摩加以推测得出大体信息, 但因为书本之间的关联不容易被发现, 所以大量的珍贵资料被迫沉沦在历史长河中。大数据作为重要的战略资源代入中医药后, 会通过数据搜集、计算、分类详细的分析出书本的关系, 后人也相应得到了大量的中医药

珍贵资料。规律是通过挖掘所储存的古籍图书数据，从大量模糊、不确定的数据中后的具体规律，形成更加具体、准确的知识。^[4]中国大量中医药古籍的记载形式都是通过文本形式，传播形式有限，利用率低，易损坏，大数据时代来临，将资料数据整理归类，给中国中医药古籍带来了给多的利用机会和利用途径。^[4]通过中医药大数据的应用，以现代科学诠释中医药学理论，是中医药面向国际化的重要手段，是中医药现代化的重要任务，通过大数据的分析整理，挖掘并预测事物及现象的整体规律是中医药学创新的根本要求。^[2]现在大数据时代对数据的搜集和更新要求越来越高，随着技术的更新，我们现在可以做到古籍资料和科研文献数据与大数据融合，科研资料数据占据了大数据的大部分，在数字的工作条件下，科研人员更倾向于使用和阅读数据资料，古籍资料的数据是中医药领域的基本数据，将各个类别的数据分类并融合，可以做到资源的高效利用。如图3所示。



图3 大数据作用于中医药的效果

在中医药大数据应用后，临床业务水平（临床决策支持，如用药分析、药品不良反应、疾病并发症、治疗效果相关性分析）、医疗质量控制（疾病诊断与预测、临床实验数据的分析与处理等方面）、工作效率（辅助临床医疗工作、规范性用药、临床路径的优化）、公共管理（管理绩效分析；流行病、急病等预防干预及措施评价；公众健康监测）、健康服务（危及健康因素的监控与预警、网络平台、社区服务）

等方面都有明显提高。

通过对医生的经验和知识规划的形式采集，建立知识库，用符号推理的方式进行医疗诊断，但随着规则库的增大，搜索空间也急剧增加，推理的匹配尝试的效率会相应降低，若采用贝叶斯网络解决医学诊断过程中建模的不确定性，来精确的预测以及得出相应得对策。

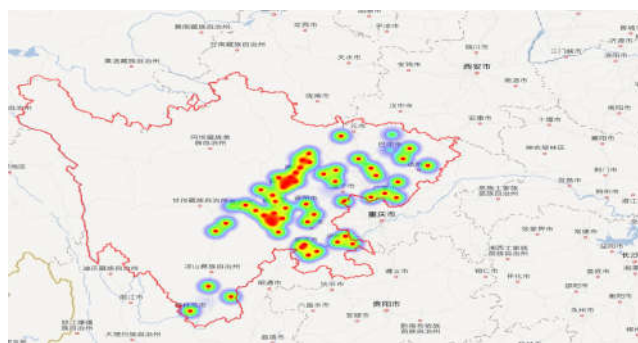


图4 可视化图

通过如此的可视化图（图4）来提前确定一定规模的未知疾病，为疫情控制尽量的争取更多的时间，并且更大程度上的判断人员流向，控制疫情，方便治疗药物和疫苗的迅速研发，在此基础上建立有关疫情的复杂动态网络的传播动力学。

健康管理数据来自个人体征数据：血液含氧量、血压、心率、心电图和皮肤温度等；以及个人特征数据：年龄、性别、职业、收入、工作区域、社会关系等。

6 中医药大数据的发展趋势

中医药十分需要大数据应用的智能化数据库管理，大数据作为易用，安全，稳定，高性能的一站式大数据处理平台。带给中医药极多的应用热点：公众健康管理、医疗管理优化——中医标准化、医疗服务质量——病症分析、药物创新研究、新型药物商业模式。

表2 中医药大数据的发展趋势

公众健康管理	5000 亿
医疗管理优化——中医标准化	2000 亿
医疗服务质量——病症分析	3000 亿
药物创新研究	3000 亿
新型药物商业模式	1000 亿
其他	2000 亿
预测显示：中医药大数据应用潜力超过 1.6 万亿元人民币 / 年	

中医药大数据开放电子病历系统数据端口的数据采集，

为中医药提供统一、便利的上传接口，采集医院的临床病历资料数据。中医药电子病历的数据采集采用xml文件形式处理，如上提供统一、便利、兼容的便利接口，支持实时的文件处理情况查询，上传批次管理以及问题数据回滚。电子病历如图5所示。

门诊电子病历明细

科室：门诊医生	就诊时间：2015-09-09 03:40	
姓名：[] 性别：女	年龄：[]	
联系电话：[]	身份证：[]	
单位或家庭住址：[]		
主诉： 四诊信息：咳嗽，呼吸声粗，痰量多，咳痰黄稠，不易咯痰，红舌，黄苔，滑数脉，气味腥臭		
临床诊断：痰热壅肺证 证型：痰热壅肺证 药物过敏史： 治法：清热化痰，活血		
病位：肺、心、胆 病性：咳嗽		
中药处方：宣降止咳汤		
地骨皮15.00g	五味子12.00g	白芍15.00g
款冬花12.00g	细辛5.00g	黄芩15.00g
苏子15.00g	炙麻黄12.00g	桂枝10.00g
干姜10.00g		

图5 电子病历

由此可见数据的真实性、可靠性、完整性的保证度更高，如此建构更专业、更稳定、更灵活的数据采集平台——数据的补正算法等，并支持多终端和多源数据的采集技术（移动、PC、可穿戴设备），形成分层多用户管理的保障技术。

中医药大数据即将带来的不止如上，他还会使得数据储存翻番易安信（EMC）表示，数字世界的规模目前每两年就要翻上一番，待到2020年，每年生成和复制出的数据将达到44ZB，数据湖：能够积累大量数据以供将来分析，云技术：为获得所需技术和服务大开方便之门，人工智能：可以主动简化数据管理，开源技术：对大规模数据处理的价值也越来越受关注。

现如今中医药大数据在大众中熟知度较低，应用比率也较少，相关工作人员需要加急加大对这方面的宣传，各大医院方面也需要配合，若是一直保持传统观念和思想，坚持使用原有的医疗信息管理方式，不但不能增加患者的便利，同时也阻碍了中国中医药技术的发展，技术的发展，信息数据管理是前提，是基础，是核心。其实针对中医药大数据研究对象整体或整体若干类似独立研究来获取更高级别的研究结

论和成果，研究证据即通过“真实世界样本”来反映真实世界总体^[6]，比较效益研究理念注重比较不同临床治疗方案的受益和风险，从而获得较为明智的临床决策等在研究理念上已有契合大数据研究趋势。相对而言，大数据时代研究视角更为宽广，思维方式更为开放，产生的影响也更为深远。中医药大数据在未来数据库完全成型并且精确度有了大幅度的增加后，会对整个中医药有更加深层的影响。中医药会更加的面向国际，有更宽广的发展范围。并且很多人赞同大数据的研究将为中医药发展带来大价值。收集这些庞大、多类别的数据，通过分析将其盘活，可以产生让人意想不到的价值。大数据应用于中医药的临床与科研，将给这个领域带来革命性的变化。

到2020年，电脑在网络流量中的占比将只有29%，而智能手机将占到30%。远程医疗和移动医疗必将影响医疗行业。

物联网，麦肯锡全球研究所预计，到2025年，嵌入式传感器和联网设备对世界经济的贡献将高达6.2万亿美元。数据种类、数据量和处理速度的增长，需要采取新的数据管理方法。

但是我们对于大数据要有正确的认识，大数据是一把双刃剑，如果对大数据不能进行正确的利用，不但会增大对中医药发展的负担，还会对中医药的发展形成误导，某大学年度调查统计显示，本校计算机系的女同学中，50%都嫁给了该校的男性老师。该消息一经公布，立即引起校内外的巨大反响，人们对师生恋、校园恋议论纷纷，很多人说是某校的炒作。经几次调查核实，该消息确实是真实有效的。原来该大学计算机系只有两名女生，其中一名女生和计算机系老师相恋结婚，由此得来的50%是真实数据。但是这样的数据即使是真实的，但是极易被炒作，有极大的负面的影响，同样在中医药大数据中思考，中医药大数据如果不能正确的客观的应用，会导致极其错误的评价。同样，大数据的特点在于对未来情况的预测，不少的数据专家认为，当一个企业或项目开始应用大数据时，他们就走上了一条不归路，随着大数据的引用，数据库的建立，数据会越来越庞大，同时也会演变越来越多的枝节，枝节越多，数据分析发生的概率也就越大，如果对数据库不定时的检测和整理，会让人们做出错误决策。上文提到，大数据在中医药的应用，会使中医药大数据开放不少的接口，并牺牲一部分的性能来分析数据，分析数据同时也要在资源平台获取更多的数据，长此以往，大数

据核心系统的业务资源必定会非常紧张，系统的运行风险也随之上升。

所以综上所述，我们在引用大数据的同时分析数据但也不要盲目的相信数据，大数据仅仅是一个参考资源，避免因片面的数据而导致做出错误的决策，并且要做好全面的安全信息防范同时也要加强数据的更新和整理，不要让有用的信息沉淀，也不要反复的使用已经过期的信息，数字信息与时代要做到完美的同步。

表3 中医药大数据应用方向

病人病情数据	病人病情，用药情况记禁忌，病人病情预测，环境不利条件记录
医生处方数据	医生处方记录，医生用药记录，医生临床诊断记录等等
药库数据	药库药物库存情况，出入库情况等等
古籍资料数据	古籍数据资料，古籍研究资料，古籍联系资料等等

7 结语

综上，中医药的历史悠久且中医药大数据的应用影响深刻，现有大量文献表明中国以往中医医者主要通过个人在临床实践中积累经验以及进行实际研究，并用纸质病历将临床诊疗过程记录下来，医者自身感悟总结，最终变成自己的学术观点和学术思想，再回到临床指导实践，这种过程，是一个非常漫长，并以实践去检验真理的过程，导致了中医传承非常困难，成长迟缓，中医药发展不能满足患者病况需求的

情况。中医药大数据的引入将会给中医科研带来翻天覆地的变化和给医者患者极大程度上的便利，若将丰富的中医药数据集合起来，会形成世界上珍贵的医疗历史文化宝库，这个数据资源库将具备独一无二的特色与优势。利用大数据以及中医药自身大数据的特点，会形成中国特有的首选的医疗方式，在国际医疗上也会更加的有优势。中医药大数据带来的大价值不止体现在医生和患者身上，更是在中国医术历史上的传承有了重大的突破。中医药大数据会为中国中医药开启新篇章。

参考文献

- [1] 何伟. 大数据时代与中医学学术创新 [Z]. 中医杂志, 2014, 55(23).
- [2] 解育静. 大数据时代中医药领域面临的机遇与挑战 [A]. 中华医学图书情报杂志, 2015, 24(7).
- [3] 朱毓梅. 大数据时代背景下中医古籍面临的机遇挑战 [J]. 中国中药情报杂志, 2014, 38(3): 11-14.
- [4] 中国医院协会信息管理专业委员会. 中国医院信息化发展研究报告(白皮书)(摘录二) [J]. 中国数字医学, 2008, 3(7): 7-19.
- [5] 任德权. 真实世界研究 GCP 和风险最小化 [J]. 中国中药杂志, 2012, 37(18): 2681-2682.
- [6] 谢雁鸣, 廖星, 王永炎. 将比较效益研究理念引入中医临床评价研究 [J]. 中西医结合学报, 2011, 9(8): 813-818.