

健康医疗大数据发展现状研究^{*}

陈 敏 刘 宁

(华中科技大学同济医学院医药卫生管理学院 武汉 430030)

〔摘要〕 介绍国内外健康医疗大数据发展计划、学术组织、标准化组织、研究领域、研究项目和开放数据资源等相关发展状况,简要阐述存在问题及发展策略,可为健康医疗大数据研究和应用提供参考。

〔关键词〕 健康医疗;大数据;发展现状

〔中图分类号〕 R-056 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2017.07.001

Research on the Current Status of Development for Healthcare Big Data CHEN Min, LIU Ning, School of Medicine and Health Management, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430030, China

〔Abstract〕 The paper introduces related development status of the development planning, academic organizations, standardization organizations, research fields, research projects and open data resources topics about healthcare big data at home and abroad, and briefly expounds existing problems and countermeasures. It could provide a reference for applications and research on healthcare big data.

〔Keywords〕 Health care; Big data; Current status

1 引言

大数据浪潮席卷全球,数据驱动发展成为时代主题,健康医疗已成为大数据应用的重要领域^[1]。健康医疗大数据的应用不仅可为人类带来更好的健康医疗服务,同时利用大数据方法可不断发现新的知识内容,促进医学知识和医学技术的进步。但健康医疗数据类型的复杂和多样性,使得无法利用常规软件工具在合理的时间内摄取、管理并整合成为有价值的信息。

2 相关发展计划

2.1 国外

2012 年联合国发布《大数据促进发展:挑战与机遇》白皮书,以推动各国政府机构、重大行业对大数据技术的研究和应用。自 2012 年开始,美、英、加、澳、日、韩等发达国家相继发布一系列大数据技术研究和发 展计划,大力推进大数据研究和应用。如美国 2012 年发布《大数据研究和发 展计划》,成立“大数据高级指导小组”,2013 年推出“数据—知识—行动”计划,2014 年进一步发布《大数据:把握机遇,维护价值》政策报告,启动“数据开放行动”,陆续公开 50 个门类的政府数据,鼓励商业部门进行开发和创新。欧盟正在力推《数据价值链战略计划》,英国发布《英国数据能力发 展战略规划》,加拿大发布《健康大数据分析白皮书》,澳大利亚发布《公共服务大数据战略》,日本

〔收稿日期〕 2017-07-02

〔作者简介〕 陈敏,教授,发表论文 150 余篇。

〔基金项目〕 国家高技术研究发展计划课题(项目编号:2012AA02A616)。

发布《创建最尖端 IT 国家宣言》，韩国提出《大数据中心战略》。

2.2 国内

与国外部分发达国家相比，我国大数据发展起步较晚。2015 年我国出台《促进大数据发展行动纲要》，提出要构建包括电子健康档案、电子病历的健康医疗服务大数据，建设覆盖公共卫生、医疗服务、医疗保障、药品供应、计划生育和综合管理业务的健康医疗管理和服务大数据应用体系，开展健康医疗大数据创新应用研究。2016 年 4 月国家卫生和计划生育委员会规划司从慧民服务工程、业务协同工程、业务监管工程、平台基础建设工程 4 个方面分列 3 级平台必备 70 项功能，绝大多数均涉及健康医疗大数据应用，健康医疗大数据将成为“十三五”人口健康信息化建设的核心工作。2016 年 6 月《关于促进和规范健康医疗大数据应用发展的指导意见》从夯实应用基础、深化应用、规范和推动“互联网+健康医疗”服务、加强保障体系建设等 4 个方面部署了 14 项健康医疗大数据重点任务和重大工程。此外，一些省市也出台相关专门规划，如《上海推进大数据研究与发展三年行动计划（2013-2015 年）》、《广东省大数据发展规划（2015-2020 年）》、《贵州省大数据产业发展应用规划纲要（2014-2020 年）》、《武汉市大数据产业发展行动计划（2014-2018）》、《厦门市大数据应用与产业发展规划（2015-2020 年）》等。

3 学术组织

3.1 美国大数据转化知识联盟

2012 年美国国立卫生研究院（National Institutes of Health, NIH）成立了大数据转化知识联盟（Big Data to Knowledge, BD2K）^[2]，该组织关注生物医学研究领域的大数据及其科学整合和最大化，设置了 4 个目标：（1）整合电子生物医学数字资源，使其能够可发现、可获取、可引用而被广泛使用。（2）研究和开发生物医学大数据分析方法和软件、工具。（3）加强数据科学的人才培养。（4）支

持建立数据生态系统，建立卓越的数据科学研究中心。BD2K 下设执行委员会、多方咨询委员会、NIH 科学数据委员会和项目管理工作组。其中项目管理工作组包括 BD2K 中心（内含 12 个研究中心）、临床数据使用研究小组、共享和可操作环境建设小组、生物医学软件资源索引小组、生物医学数据发现索引小组、基于社区的数据和元数据小组、NIH 标准信息资源小组、生物医学数据资源库商业可持续小组、生物医学大数据软件开发小组、培训小组等。

3.2 英国医疗卫生信息与发现中心

2013 年 5 月英国牛津大学成立了首个综合运用大数据技术的医药卫生科研机构——李嘉诚卫生信息与发现中心。该中心有靶标发现研究所和大数据研究所两个机构，旨在通过收集、存储和分析大量医疗信息，确定新药物的研发方向，探索特定疾病的新疗法。

3.3 澳大利亚卓越研究中心

澳大利亚研究基金会（Australian Research Council, ARC）通过卓越研究中心项目对大数据研究进行资助，包括在墨尔本大学投入 2 000 万美元进行数学和统计领域大数据、大模型、新见解卓越研究中心建设，用于开展包括健康在内的社会问题新数据模型研究；资助生物信息学卓越研究中心，研究领域包括生物大数据及其应用。

3.4 开放医疗与健康联盟

2015 年 5 月开放医疗与健康联盟（Open Medical and Healthcare Alliance, OMAHA）在杭州成立。OMAHA 是一个推动医疗、健康产业信息开放、共享、开源的民间非盈利性组织，由国内优秀的相关健康信息技术及健康医疗应用开发商、各类医疗及健康服务机构、医疗器械及健康类消费电子产品生产商、运营商、互联网企业、媒体等相关产业链机构和个人组成。OMAHA 致力于推动产业开放个人的医疗和健康数据，定义个人健康档案的标准格式，探索相关信息标准的协作和落地，从而提高居民对自身健康医疗数据使用的完整性、可及性和可用性^[3]。

3.5 中国卫生信息学会

中国卫生信息学会为配合国家健康医疗大数据发展,在原有 16 个专委会的基础上,又新成了 15 个健康医疗大数据相关专委会。

3.6 健康医疗大数据研究院

2015 年 5 月国家卫生和计划生育委员会统计信息中心、电子科技大学联合发起成立健康医疗大数据研究院。将运用大数据挖掘与分析等关键技术,对居民电子健康档案、电子病历、全员人口数据库以及医疗卫生机构运行中产生的大量诊疗数据进行挖掘,在突发公共事件、流行性疾病爆发、健康服务业发展、人口流动等领域提供分析和预警,同时提供科学管理和决策的数据基础和信息依据。

3.7 中国通信学会大数据专家委员会

2012 年 10 月中国通信学会大数据专家委员会在北京成立,是我国首个专门研究大数据应用和发展的学术咨询组织。该组织宗旨是跟踪大数据的最新进展,探讨大数据发展与应用的重点问题,搭建学术性、行业性高端平台,促进国内外通信企业、监管部门、研究机构、学术机构的交流与合作,推动我国大数据的科研与发展。其中健康医疗大数据也是其重要研究内容。

3.8 中国计算机学会大数据专家委员会

2012 年 10 月中国计算机学会大数据专家委员会 (CCF Task Force on Big Data, CCF TFBD) 在大连成立。其宗旨是探讨大数据的核心科学与技术问题,推动大数据学科方向的建设与发展。构建面向大数据产学研用的学术交流、技术合作与数据共享平台,对相关政府部门提供战略性的意见与建议。CCF TFBD 下设 5 个工作组,分别为会议组织 (学术会议、技术会议) 工作组、学术交流工作组、产学研用合作工作组、开源社区与大数据共享联盟工作组、战略材料编写工作组。此外,还有上海大数据联盟、中国大数据技术大会 (Big Data Technology Conference, BDTC) 等, BDTC 是目前国内最具

影响、规模最大的大数据领域盛会。大会的前身是 Hadoop 中国云计算大会 (Hadoop in China, HiC)。

3.9 中国医学科学院生物医学大数据中心

在国家大数据战略背景下,中国医学科学院/北京协和医学院 (以下简称“院校”) 组织成立了生物医学大数据中心,中心联合了院校所属医院、研究机构在医疗信息化和生命组学研究当中所积累的资源 and 人才,打造生物医学大数据的技术支撑平台和功能应用平台。

4 标准化组织

4.1 ISO/IEC JTC1 WG9

2014 年 11 月 ISO/IEC JTC1 成立负责大数据国际标准化的大数据工作组 WG9。ISO/IEC JTC1 WG9 工作重点包括:开发大数据基础性标准,包括参考架构和术语;识别大数据标准化需求;同大数据相关的 JTC1 其他工作组保持联络关系;同 JTC1 外其他大数据相关标准组织保持联络关系。目前, WG9 正在研制信息技术大数据概述和术语、信息技术大数据参考架构两项国际标准。

4.2 ISO/IEC JTC1/SC32

2012 年 ISO/IEC JTC1/SC32 成立下一代分析技术与大数据研究组。2014 年 6 月开始启动 SQL 对多维数组的支持、SQL 对 JSON 的支持、数据集注册元模型和数据源注册元模型 4 项大数据标准项目。

4.3 ITU-T

ITU-T 大数据标准化工作主要由研究组 SG13 负责,具体包括该研究组下设的 Q2 (第 2 课题组) 的研究项目“针对大数据的物联网具体需求和能力要求”、Q17 的研究项目“基于云计算的大数据需求和能力”、Q18 的研究项目“大数据业务的功能架构”等。

4.4 美国国家标准及技术研究所 (National Institute of Standards and Technology, NIST)

2013 年 6 月美国国家标准及技术研究所 (Na-

tional Institute of Standards and Technology, NIST) 成立大数据公共工作组, 工作范围包括建立来自于产业界、学术界和政府的公共环境, 形成达成共识的定义、术语、安全参考体系结构和技术路线图, 提出数据分析技术应满足的互操作、可移植性、可用性和扩展性需求和安全有效地支持大数据应用的技术基础设施。该工作组下设术语和定义、用例和需求、安全和隐私、参考体系结构和技术路线图 5 个分组, 目前已完成大数据定义、大数据分类、大数据用例和需求、大数据安全和隐私需求、大数据参考架构、大数据技术路线图等标准文档。

4.5 全国信息技术标准化技术委员会

2014 年 12 月全国信息技术标准化技术委员会 (SAC TC28) 正式成立大数据标准工作组。工作组主要负责制定和完善我国大数据领域标准体系, 组织开展大数据相关标准和技术研究, 承担国家和行业标准制修订计划任务, 进行标准实施的宣传与推广, 对口 ISO/IEC JTC1 WG9 大数据工作组。SAC TC28 的大数据标准工作组下设大数据总体专题组、大数据国际专题组、大数据技术专题组、大数据产品和平台专题组、大数据安全专题组、工业大数据专题组和电子商务大数据专题组 7 个专题组。

5 研究与应用领域

健康医疗领域研究和应用主要包括: (1) 健康医疗大数据分析研究方法研究。包括信息抽取、知识组织、大数据实时处理、统计分析方法与工具、社交媒体健康信息传播模型、多变量动态网络分析方法、健康医疗数据的可视化方法、生物学大数据的深度整合方法等。(2) 智能健康管理。包括患病风险预测、自我健康管理、健康评估预警、康复跟踪、健康处方制定、健康异常提醒、远程健康监测设备研发等。(3) 医疗服务。包括临床决策支持、诊疗方案成本效益分析、比较效果研究、临床质量分析、合理用药分析、药物不良反应分析、医疗不良事件分析、医疗器械安全性分析与评价、患者行为预测等。(4) 疾病控制与应急管理。包括传

染病监测与预警、病原体快速筛检、慢性病发病趋势、疾病干预效果模拟与分析等。(5) 卫生综合管理。包括医疗资源分配、绩效分析、公众健康监测、疾病负担分析、医疗服务定价、费用控制等^[4]。(6) 医疗保障监管。包括医疗欺诈自动识别、医保控费、医疗保险定价等。(7) 精准医疗与医药研发。包括基因组学、蛋白质组学、转录组学、结构基因组学、功能基因组学、药物靶标发现、新药研发等。

6 研究项目

6.1 谷歌大数据疾病预防项目

2014 年 7 月谷歌启动大数据疾病预防 Baseline Study 项目, 将匿名收集 175 人基因和分子信息, 之后将扩大到数千人的相关数据。该项目希望帮助研究人员更早地发现心脏病和癌症的各种迹象, 寻找这些信息中隐藏的“生物标签”, 进而推广预防措施。例如可发现能帮助人们分解高脂肪食物的生物标签, 拥有这些生物标签的人可延缓罹患高胆固醇和心脏病, 没有生物标签的人则可能更早患上心脏疾病。研究人员可通过检查了解哪些人缺乏这类标签并帮助他们纠正习惯, 或开发新的治疗方法帮助其更好地分解高脂肪食物。

6.2 湘雅临床大数据系统建设项目

2014 年 1 月中南大学启动“湘雅临床大数据系统建设项目”, 首批共 101 个项目入选, 覆盖 40 余个临床学科。该项目是国内高校首次开展大数据在临床医学领域大范围、系统性的探索和应用。其内容涉及大部分常见疾病和重要疾病, 也包括医疗技术、临床药学、医疗和护理质量及安全等领域。通过临床大数据挖掘和分析, 项目将建立国际先进和国内领先的医疗相关数据运营模式, 促进智慧医疗、精准医疗、医院精细化管理、临床科研、转化医学和基础医学的发展, 为卫生计生行政管理部门提供决策依据。

6.3 贵州首个国家级大数据综合试验区

2016 年 2 月国家发展改革委、工业和信息化部

部、中央网信办发函批复贵州建设首个国家级大数据综合试验区,将围绕数据资源管理与共享开放、数据中心整合、数据资源应用、数据要素流通、大数据产业集聚、大数据国际合作、大数据制度创新等 7 大主要任务开展系统性研究和试验。

7 开放数据资源

数据开放是提高数据利用率,实现信息共享的基础。在医学领域美国政府截至 2016 年 5 月 8 日在国家数据开放平台 HealthData.gov 发布健康医疗类数据集 2 537 项,涵盖医疗保险、公共卫生、医疗服务、流行病学、人口统计、医疗服务费用等主题。2014 年美国食品药品监督管理局公共数据开放项目 Open FDA 正式上线,相关数据集涵盖了药物、设备和食物方面的不良事件、药物标签、执行报告等主题。英国政府专门建立“数据英国”(data.gov.uk)网站,将公众关心的政府开支、财务报告等数据整理汇总并发布在互联网上,对其中的热点议题和重要开支做进一步阐释对公众意见进行反馈。截至 2016 年 5 月 8 日该网站拥有公开数据集 31 150 个、未公开数据集 3 700 个,其中健康相关主题数据集有 1 654 个。澳大利亚政府截至 2016 年 5 月 8 日在 data.gov.au 信息目录开放数据平台公布了 7 938 个数据集,其中有 20 个数据集涉及健康医疗主题。

此外,其他相关机构也发布了大量医学数据

集,其中比较著名的是 MIMIC II 数据库,该数据库包含了波士顿贝斯以色列女执事医疗中心(Beth Israel Deaconess Medical Center, BIDMC) 2001 - 2008 年 ICU 25 000 多名患者的人口学、医疗事件、用药、检查和检验报告、仪器监测等数据,被多国学者用来开展流行病学数据挖掘分析,也被用来开发医学大数据分析平台验证算法。2010 年北京大学成立医学信息学中心,2015 年 5 月被任命为国家医疗数据中心,收集近 400 家医院自 2006 年度以来的 4 800 多万条病案首页数据。2014 年 1 月启动的“湘雅临床大数据系统建设项目”,目前已融合多家医院共 2 000 万患者 100 亿条医疗源数据,可为临床医师和科研人员的医学科学研究提供帮助。

(本文转载自《中国医院管理》2017 年第 37 卷第 2 期,略有改动。)

参考文献

- 1 蔡佳慧,张涛,宗文红. 医疗大数据面临的挑战及思考[J]. 中国卫生信息管理杂志, 2013, (4): 292 - 295.
- 2 NIH. About BD2K [EB/OL]. [2016 - 05 - 04], <http://datascience.nih.gov/bd2k>.
- 3 刘宁,陈敏. 医疗数据开放方法及策略研究[J]. 中国医院管理, 2015, 35 (9): 37 - 39.
- 4 Luo J, Wu M, Gopukumar D, et al. Big Data Application in Biomedical Research and Health Care: A Literature Review [J]. Biomed Inform Insights, 2016, (8): 1 - 10.

《医学信息学杂志》版权声明

(1) 作者所投稿件无“抄袭”、“剽窃”、“一稿两投或多投”等学术不端行为,对于署名无异议,不涉及保密与知识产权的侵权等问题,文责自负。对于因上述问题引起的一切法律纠纷,完全由全体署名作者负责,无需编辑部承担连带责任。(2) 来稿刊用后,该稿包括印刷出版和电子出版在内的出版权、复制权、发行权、汇编权、翻译权及信息网络传播权已经转让给《医学信息学杂志》编辑部。除以纸载体形式出版外,本刊有权以光盘、网络期刊等其他方式刊登文稿,本刊已加入万方数据“数字化期刊群”、重庆维普“中文科技期刊数据库”、清华同方“中国期刊全文数据库”、中邮阅读网。(3) 作者著作权使用费与本刊稿酬一次性给付,不再另行发放。作者如不同意文章入编,投稿时敬请说明。

《医学信息学杂志》编辑部