

美国大数据发展及应用现状研究

王同涛¹, 蒋德明², 向远博¹

(1. 中国科学技术交流中心, 北京 100045;

2. 河北省张家口市第十九中学, 河北张家口 075000)

摘要: 大数据技术和产业是未来国家竞争的焦点之一。本文系统梳理了美国政府发展大数据技术及其应用的经验, 重点介绍了美国政府如何推动大数据在公共服务和民生领域的应用, 并提出了我国发展大数据技术和应用领域的政策建议。

关键词: 美国; 大数据; 数据开发; 数据应用

中图分类号: G327.712 **文献标识码:** A **DOI:** 10.3772/j.issn.1009-8623.2018.06.012

大数据 (Big Data) 一般是指无法在可容忍的时间内用传统 IT 技术和硬件工具对其进行感知、获取、管理、处理和服务的数据集合^[1]。人、机、物三元世界的高度融合引发了数据规模的爆炸式增长和数据模式的高度复杂化, 世界已进入了网络化的大数据时代^[2]。根据监测统计, 2017 年全球的数据总量为 21.6ZB (1ZB 等于 10 万亿亿字节), 目前全球数据的增长速度在每年 40% 左右, 预计到 2020 年全球的数据总量将达到 40ZB。著名市场咨询公司 SNS 研究报告预测, 至 2020 年全球大数据市场规模将达到 800 亿美元, 年均实现 15.37% 的增长。

大数据是信息化的又一重要阶段, 全球数据的爆发增长、海量集聚, 对经济发展、社会治理等都产生了非常大的影响。大数据技术和应用正成为全球新一轮技术和产业竞争的竞技场, 世界主要大国纷纷推出大数据产业有关的发展战略。美国是最早开展大数据研发和应用的, 其采取多种政策举措, 积极推进国家大数据战略, 取得了大数据领域的世界领先地位。

1 美国发展大数据做法及主要经验

美国将大数据看作新时代的石油, 将其放在极

重要的战略位置, 并采取了行之有效的措施推进大数据产业的发展, 使美国成为世界大数据产业的发祥地和产业中心。大数据之所以能被迅速、广泛应用, 与美国高度重视大数据价值、积极推动数据开放等政策密切相关。

1.1 将大数据开发和应用作为国家战略, 出台政策引导

美国在推进大数据应用上形成了从发展战略、法律框架、资助体系到行动计划的完整布局, 已先后实施多项政策行动。2009 年, 美国国家科学技术委员会 (NSTC) 发布了首个大数据相关报告《开发数字数据的威力》。2012 年 3 月, 白宫发布《大数据研究和发展计划》, 并成立“大数据高级指导小组”^[3]。2013 年 11 月, 白宫推出“数据-知识-行动”计划, 进一步细化了大数据改造国家治理模式、促进前沿创新、提振经济增长的路径, 这是美国向数字治国、数字经济、数字城市、数字国防转型的重要举措。2014 年 5 月, 美国总统办公室提交《大数据: 把握机遇, 维护价值》政策报告, 强调政府部门和私人部门紧密合作, 利用大数据最大限度地促进增长, 减少风险^[4,5]。2016 年 5 月, 白宫发布《联邦大数据研发战略计划》, 在已有基

第一作者简介: 王同涛 (1980—), 男, 工学博士, 副研究员, 主要研究方向为科技管理、科技政策。

通讯作者简介: 向远博 (1984—), 男, 硕士, 助理研究员, 主要研究方向为科技管理和科技政策。邮箱: xiangyb@castec.org.cn。

收稿日期: 2018-04-20

础上提出美国下一步的大数据发展战略。通过一系列战略性引导,在全社会形成了开发和应用大数据的热潮^[6-8]。

1.2 政府部门资助,推动大数据技术的研发与应用

奥巴马政府在2012年3月推出的“大数据研究与开发计划”是继“信息高速公路”之后的又一重大科技部署,提出将利用大数据技术在科学研究、环境保护、生物医药研究、教育以及国家安全等领域实现突破。该计划涉及美国国家科学基金会、国防部、能源部等6个联邦政府部门,投资2亿多美元用于相关工具与技术的开发。主要包括:美国国家科学基金会和国立卫生研究院负责大数据科学和工程的核心方法和技术研究,具体项目包括管理、分析、可视化,以及从大量的多样化数据集中提取有用信息的核心科学技术;能源部投入2500万美元用于建立可扩展的数据管理、分析和可视化研究所;国防部高级研究计划署的项目则集中在大数据辅助决策,包括情报、侦查、网络间谍等方面,通过集成传感器、感知能力和决策支持建立真正的自治系统,实现操作和决策的自动化;美国地质勘探局通过给科学家提供深入分析的场所和时间、最高水平的计算能力和理解大数据集的协作工具,促进地理系统科学的创新。

除了上述6个部门,还有多个部门正从联邦层面开展大数据相关计划,涉及面广,研发种类很多。例如,国土安全部项目主要推进可视化数据分析,应用领域主要为自然灾害、恐怖事件、边境安全、网络威胁等。美国国立卫生研究院开展了世界最大的遗传变异研究数据集——国际千人基因组项目(数据已达约200TB),由亚马逊网站免费云服务(AWS)提供相关支持。此外,美国科技政策办公室积极推动公共组织和私营部门建立大数据方面的伙伴关系。

1.3 引导建立多元化的大数据应用格局

美国政府鼓励企业开展大数据应用活动。互联网企业重视大数据并最早将其商业化发展。网飞(Netflix)和脸书(Facebook)等大型科技公司运用用户在网络上留下的数字痕迹,通过大数据运算分析用户需求,并依此向用户推荐观看特定电影或联系某人。现在,美国一些大型公司已在支持大数据相关竞赛,并且资助高等院校的大数据研究项目。

IBM、惠普、微软、EMC等IT巨头不断通过并购方式整合大数据技术。

行业协会组织等非营利机构在推动大数据应用中也发挥着重要作用,积极为各方提供公共服务。例如“数据无边界”(Data Without Borders)项目就是通过开展无偿的数据收集、分析以及可视化为非营利组织提供帮助。高等院校注重大数据相关人才的培养。一些大学将培养下一代“大数据科学家”为目标,开始创建有关新课程。

1.4 政府率先开展大规模数据公开行动

通过政府数据公开计划,大力推动政府数据信息公开。2009年,奥巴马在就任总统后的第一个工作日就签发“开放政府”备忘录,推动联邦政府公开大量政府资料,并将数据通过中央信息交换库——Data.gov网站进行公布,供民众和社会进行查阅。目前,Data.gov共开放原始数据集和地理数据集约28.5万个,涵盖了农业、气象、金融、就业、人口统计、教育、医疗、交通、能源等50多个门类。为进一步完善Data.gov平台的功能,政府还在其中加入了数据的分级评定、高级搜索、用户交流以及和社会网站互动等新功能。同时,还构建了开放政府平台(OGPL),将开源的政府平台代码提供出来,并允许任何城市、组织或者政府机构自己创建新的开放站点。联邦政府公布的数据给企业进行新的产品、服务开发提供了优质资源,也给民众带来了气象服务、定位服务等更好的公共服务。2013年5月,奥巴马签署总统行政令,要求在保护隐私安全与机密性的前提下,将数据公开纳入政府的义务范围。随着发布数据的不断增加,联邦政府也在致力于保障所有相关的利益方都可以获得标准格式的数据。如科技政策办公室于2012年拨付2000万美元以支持不同部门发布数据集,并提高其可使用性。

通过“我的大数据”计划等促进个人数据使用。以下是自2010年开始联邦政府开展的一系列主题活动,以方便民众更安全地获取个人数据来办理私人业务,包括:(1)“蓝纽扣”计划。消费者使用“蓝纽扣”获取个人健康信息,以便管理其健康、经济状况,并与信息提供方交换信息。目前,已有超过1.5亿的美国人能够从健康服务企业、医药实验室、零售药房供应商与州免疫信息数据库获得他们所需的个人健康数据。(2)“创建副本”

计划。在 2014 年,美国国税局建立了名为“Get transcript”的共享数据库,纳税人在此可以获得个人近 3 年的纳税记录,居民可以方便地下载纳税申报单,更加便捷地进行抵押、贷款等活动。(3)“绿纽扣”计划。美国政府与电力行业在 2012 年合作推出“绿纽扣”计划,为家庭与企业提供能源使用信息,目前已为 5 900 万家庭与企业提供服务,并帮助他们节约能源。(4)“我的学生数据”计划。美国教育部将助学金免费申请表与联邦助学情况的信息进行共享,这些信息囊括了借贷、补助金、注册与超额偿付等,使学生与资助人能够上网下载所需信息资源。

1.5 重视数据应用与隐私保护的平衡

美国在数据保护方面的总体思路是既要规避大数据技术应用可能带来的种种风险,又要为未来的技术与商业模式创新留有足够的发展空间。美国政府认为目前大数据应用中最严峻的挑战是如何保护隐私,并且正在不断修改相关法律法规以加强隐私保护,提出的改进重点在于改进消费者隐私权法案、通过有关国家数据外泄立法、保护非美籍人士隐私、规范在校学生数据采集使用、修正电子通信隐私法等。

美国关注大数据结果的使用,而非数据的收集和分析过程。由于大数据技术的负面影响主要体现在对分析结果的滥用上,美国不断加强对大数据应用的政策监管,并发展隐私保护第三方中介机构,加强信息处理透明度,保护个人对于数据的知情权与维护权,推动机构对大数据的算法和应用保持透明。

2 美国大数据的应用

由于政府不遗余力的推动和引导,美国从政府到企业,从公共服务领域到教育、科技、商业领域,都在因大数据的应用而发生深刻变革,社会各界都在主动开展大数据的应用和推广。大型科技公司是美国大数据技术研发和商业应用主体,本文主要介绍政府推动大数据在公共服务和民生领域的应用。

2.1 大数据在医疗卫生领域的应用

美国通过制定各种政策和倡议支持健康医疗领域大数据的使用,直接改善了医疗费用和质量,

甚至是整个健康医疗生态系统。经过整合和挖掘的医疗大数据对个体医生、康保中心、大型医院、责任医疗组织和研究机构具有显著作用。麦肯锡估计,医疗大数据的利用每年可为美国节省 3 000 亿到 4 500 亿美元的医疗开支^[9,10]。

开展健康医疗相关立法,使公共数据更易获得。早在 1996 年,克林顿政府设立了著名的健康保险携带和责任法案,通过规定医疗信息的标准和格式来推动医疗行政流程成本的下降。后来联邦又陆续发布了一系列法律:2009 年的开放政府法案和美国卫生和人类服务部基于健康数据计划下的一系列后续措施,致力于整合各个机构的医疗数据;2010 年 3 月开始生效的合理医疗费用法案授权美国卫生和人类服务部来发布相关数据,用以提高医疗健康市场和保险市场的透明度;促进经济和临床保健的卫生信息技术法案(HITECH)批准了 400 亿美元资金补贴,鼓励服务提供者使用电子病历系统(EMRS),目标是到 2019 年把目前 70% 的使用率提高到 90%。法案还批准了 20 亿美元用于支持电子病历系统人员培训和相关基础设施建设^[11,12]。

通过医疗信息交换共享及用户使用促进医疗数据共享。美国的医疗保险、医疗补助服务中心创立了信息产品和数据分析办公室,监督和协助各部门数据的上传和发布。时任美国总统奥巴马于 2015 年 1 月公布了其在生命科学领域力推的新项目——精准医疗计划,计划将在癌症和糖尿病等疾病治疗领域发力,将健康个性化信息提供给所有人。从 2016 年开始,美国政府将划拨 2.15 亿美元经费用于精准医疗项目,除支持新方法开发外,还会进行 100 万例以上美国国内的病例研究,通过移动医疗设备追踪志愿者的相关信息,如生物标本、基因信息、磁共振扫描图像等电子医疗数据以及生活数据;参与者可控制本人信息共享程度,既获得本人健康数据信息,又助力医疗大数据信息的形成。

推动医疗健康数据标准化体系建设加速大数据医疗应用。健康数据计划在 HealthData.gov 网站进行标准化格式数据的发布,涵盖各种联邦医疗健康数据库,并包含临床服务提供者质量信息和社区医疗的运营情况,附有医学和科学知识、消费产品有关数据以及政府预算等。另外,为确保数据的可读性及应用程序能够被下载和访问,政府还为开发者

特别提供了编程访问接口等优惠条件。国家健康信息技术协调办公室得到精准医疗项目 500 万美元的资助, 用来制定保护隐私、保证跨系统数据交换安全所需的标准和要求^[13]。

2.2 大数据在农业领域的应用

美国一直以来重视农业数据的收集, 因此数据较为齐全, 也较早开始了农业数据开放和使用。美国农业政策制定者正越来越多的依靠大数据的挖掘和使用来为农业发展做决策。

精准农业是美国农业应用大数据最普遍的领域之一。用来监视和预测土壤及周围空气温度和湿度的传感器是数据采集的重要工具, 遍布田间。控制中心实时收集并分析气候、空气质量、土壤信息、作物成熟度乃至设备和劳动力的成本等多方面的数据, 可使农民在播种、施肥等各环节做出更明智的决策。精准农业可精准控制各种原料的用量, 极大地节约和优化化肥、水、农药等农业要素的投入, 实现农业规模化经营^[14,15]。

孟山都公司将农业大数据应用视为未来发展的长期机会。2012 年初开始, 孟山都公司通过收购精密播种公司开始介入农业大数据, 不断收购农业数据分析相关公司。精密播种公司以生产农业广泛应用的硬件和软件为主, 涉及田间播种空间、深度等方面。孟山都公司于 2013 年 10 月又以 9.3 亿美元收购气候集团, 这是农业大数据领域最大的公司之一, 主要从事气象数据分析。此次收购行动是首宗重大的农业“大数据”收购, 标志着孟山都公司将数据与生物技术解决方案整合作为主要方向。2014 年 2 月, 气候集团收购土壤测试服务公司 Solum, 其系统可应用孟山都公司的主要软件产品来确定土壤实际生产力。“我的智慧农场”

(MySmartFarm) 致力于利用大数据改造农业, 它使用测试收集数据和预测数据综合判断, 可以提高农民对预测或趋势的反应能力。MySmartFarm 监测收获、施肥、疾病、土壤和水分等所有数据, 并与气候数据结合产生新的信息, 用以指导农民做出决策、及时采取行动, 这不仅会节约更多的水和化学药品, 还可使农业更加生态化^[16]。

2.3 大数据在教育体系的应用

教育被公认为大数据的重要应用领域。大数据给教育带来的变化将是革命性的。美国国家教育统

计中心等机构已开始统计和分析学生的学习行为、考试分数以及职业规划等数据, 作为教学改革的重要参考^[17]。

2012 年, 美国教育部参与了一项公共教育大数据计划, 该计划耗资 2 亿美元, 旨在利用大数据分析结果帮助教育改革。美国教育部出资 2 500 万美元, 用于研究和理解学生如何学习。美国教育部门创造了“学习分析系统”来运用大数据: 该系统通过数据挖掘、模化及案例运用, 更好、更精确地分析学生到底是如何学习的。这些有用信息将被提供给教育工作者, 帮助其回答在现实中难以回答的问题。

美国企业已将教育大数据成功进行商业化运作和应用。IBM 发挥其在信息技术与业务解决方案领域的优势, 与亚拉巴马州的莫白儿县开展大数据方面的合作, 改善其公共学区的教育现状。结果显示, 大数据在改进学校工作中可发挥重要作用。合作开始时, 该学区学生成绩不好, 并且辍学率非常高, 已增加到 48%。该县利用此前的学生数据建立辍学指示工具来应对辍学挑战, 并将此作为决策依据。IBM 则利用其技术支持该县重新建立大数据库, 通过大数据分析改善学区内教学状况以提高所有学生的成绩。

新兴企业“希维塔斯学习”主要通过其最大的跨校学习数据库在高等教育领域帮助学生提高成绩: 分析海量的学习数据, 可以判断学生的分数、出勤率等主要趋势; 综合运用 100 多万名学生的学习记录和 700 万条课程相关记录, 利用软件分析可以对学生辍学及学习成绩不佳进行预警。梦盒学习公司和纽顿公司开发了不同版本的利用大数据的适应性学习系统, 并已发布应用。2012 年, 作为适应性学习领域的领先者, 纽顿公司与世界最大的教育出版公司培生集团共同发布了新的适应性学习产品。这款名为“我的实验室 / 高手掌握”的产品主要由培生集团开发, 目标是全球范围内的数百万名学生, 为他们提供真实的学习数据和个性化的学习服务, 使学校通过数据应用提升学生的学习效果, 进而降低教学成本^[18-20]。

2.4 大数据在政府管理方面的应用

美国越来越多的政府部门依靠大数据分析优化管理方式、提升管理能力。大数据在公共政策、

舆情监控等领域得到广泛应用,其中反恐及犯罪预测等是应用较早和较为成功的领域^[21]。

国土安全部是美国在公共部门中应用大数据较为成功的例子。其组织形式决定了非常有必要在许多政府机构间开展数据共享和整合。国土安全部经过多年探索,在有效整合数据方面取得了很大成绩,也在更好地实现数据共享方面积累了大量经验教训。在运用大数据方面,美国各州和地方执法部门有很多成功案例。2013年,波士顿马拉松赛发生了性质恶劣的连环爆炸案。美国联邦调查局运用大数据帮助加速侦破案件,为此调用了现场附近采集的大量数据,包括通讯记录、监控录像和在志愿者处得到的影像资料,共计约10TB左右,最终经过调查人员比对、查找和分析锁定了犯罪嫌疑人。此外,圣克鲁斯警察局是美国警界应用大数据预测分析最早的几个试点单位之一。该局将城市数据以及社交网络数据进行分析处理,力图发现地区的犯罪趋势和犯罪模式,并以此预测特定区域的犯罪概率等。

基础设施和能源管理也是美国运用大数据的重点领域。赌城拉斯维加斯政府重视大数据在市政基础设施方面的运用,开发了用来全面掌握市政管网有关信息的网络仿真模型。通过模型观察地面和地下的各种管线,管理者可以实时掌握地下关键资产的状况,避免发生被施工活动误挖的情况。加州超过80%的电网由其电网系统运营中心负责管理,肩负向3500万用户每年输送高达2.89亿兆瓦电力的任务,海量的数据提高了管理难度,所以他们利用天气、传感器、计量设备等各种数据源的海量数据来实现全网电力供应和需求平衡,以对潜在危机做出快速响应^[22]。

3 有关建议

大数据不仅具有巨大的商业价值,还在国家治理中发挥重要作用,更是未来国际竞争的战略性资源。我国已将大数据上升为国家战略,加紧部署,大力发展,但目前仍存在数据治理制度不完善、核心技术受制于人、数据安全问题严峻、高端人才短缺等问题。借鉴美国发展大数据的经验,结合我国面临的问题,我国应从以下几方面入手发展大数据产业:

(1) 进一步提高对大数据战略和应用价值的重视程度,通过建立标准和法律法规为大数据发展提供制度保障。要加快落实国家大数据战略,加大战略研究力度,制定大数据技术及产业发展路线图。要完善数据开放的标准和法规,为实现各级信息系统互联互通、资源共享奠定基础。

(2) 重点支持大数据基础理论研究,引导应用研究。政府将支持重点放在基础研究的研发支持,特别是多学科交叉共性理论研究中,引导企业开展大数据技术的开发、研究和应用示范。

(3) 加大大数据发展的基础性建设,推动全社会数据开放共享。政府主导建设信息开放平台,推动实现政府信息资源开放共享,引导企业界积极共享数据。同时,不断完善数据采集体系、提升社会信息化程度,针对不同主题建立数据监测系统。

(4) 加强大数据在公共事业领域的应用,带动大数据应用及产业。以政府采购大数据服务、公私合作等方式,将电子政务、教育、医疗、交通等公共事业领域作为重点,带动大数据在全社会的应用。■

参考文献:

- [1] 李国杰,程学旗.大数据研究:未来科技及经济社会发展的重大战略领域——大数据的研究现状与科学思考[J].战略与决策研究,2012,27(6):647-657.
- [2] Steve Lohr. The age of big data[N]. New York Times, 2012-2-11(3).
- [3] The White House. Big Data Research and Development Initiative[R]. Washington DC, 2012.
- [4] The White House. Big Data: Seizing Opportunities Preserving Values[R]. Washington DC, 2014.
- [5] PCAST. Big Data and Privacy: A Technological Perspective[R]. Washington DC, 2014.
- [6] NSTC. Federal Big Data Research and Development Strategic Plan[R]. Washington DC, 2016.
- [7] The Federal Trade Commission. Big Data: A Tool for Inclusion or Exclusion? Understanding the Issues [R]. Washington DC, 2016.
- [8] 乔健. 美国大数据政策的发展趋势[J]. 全球科技经济瞭望, 2016, 31(5): 1-4.
- [9] McKinsey&Company. The Big-data Evolution in US

- Healthcare: Accelerating Value and Innovation[R]. New Jersey, 2013.
- [10] National Institutes of Health. Precision Medicine Initiative[R]. Maryland, 2015.
- [11] BaselKaiyali, David Knott. The "Big Data" Revolution in Healthcare: Accelerating Value and Innovation[R]. New Jersey, 2013.
- [12] 朱彦, 徐俊, 朱玲, 等. 主要发达国家医疗健康大数据政策分析[J]. 中华医学图书情报杂志, 2015, 24(10): 13-17.
- [13] 朱彦, 徐俊, 朱玲, 等. 主要发达国家医疗健康大数据政策分析[C]. 第十届全国医学信息教育可持续发展学术研讨会论文集, 2016: 13-17.
- [14] 佚名. 实地探访美国农业大数据的建设与应用[J]. 农业工程技术, 2016, 36(15): 59-61.
- [15] 牛禄青. 智慧新农业大数据助推农业变革[J]. 新经济导刊, 2015(12): 10-18.
- [16] Megan Stubbs. Big Data in U.S. Agriculture[R]. Washington DC, 2016.
- [17] ArneDuncan. RobustData gives us the roadmap to reform[EB/OL]. [2018-03-20]. <https://www.ed.gov/news/speeches/robust-data-gives-us-roadmap-reform>, 2009-06-08.
- [18] 胡德维. 大数据“革命”教育: 值得深思的话题[N]. 光明日报, 2013-10-19(5).
- [19] 鄂玺如. 大数据给教育带来的思考[J]. 科教导刊, 2014(25): 8.
- [20] Bennett R E. The changing nature of educational assessment[J]. Review of Research in Education, 2015, (39): 370-407.
- [21] LouisaTomar. Big Data in the Public Sector[R]. Washington DC, 2016.
- [22] 孙柏林. “大数据”技术及其在电力行业中的应用[J]. 电气时代, 2013(8): 18-23.

Research on Development and Application Status of Big Data in the US

WANG Tong-tao¹, JIANG De-ming², XIANG Yuan-bo¹

(1. China Science and Technology Exchange Center, Beijing 100045;

2. The Nineteenth Middle School in Zhangjiakou City, Hebei Province, Zhangjiakou, Hebei 075000)

Abstract: Big data technology and industry are one of the focuses of future national competition. This paper systematically introduces the US government's policies and experience in developing big data technologies and their applications, and focuses on the US government's application of big data in public services and people's livelihood, and puts forward policy recommendations for the development of big data technology and application in China.

Key words: US; big data; data development; data application