

大数据环境下的情报学

□ 贺德方 / 中国科学技术信息研究所 北京 100038

摘要: 大数据已经成为国际国内研究关注的热点话题和重要方向。文章从大数据的内涵出发,阐释了大数据问题的内在价值与现实意义,分析了大数据在数据质量、对技术的要求、信息服务以及信息安全、隐私保护等方面给传统情报学带来的挑战,也分析了大数据的发展给情报学学科发展和情报事业发展带来的新机遇。最后文章也建议情报学与情报工作应利用既有优势和成果积极参与到大数据的研究和发展中去,以大数据的研究为契机实现情报学的大发展。

关键词: 大数据, 情报学, 情报事业

DOI: 10.3772/j.issn.1673—2286.2012.11.002

随着信息通讯技术的不断发展、科学与整个社会对于计算的依赖,人们在日常工作和生活中使用计算机、手机以及其他各类技术设备的过程中,都产生出大量的数据,其规模和复杂性超越了传统技术手段所能处理的范围,这种情况催生了大数据(Big Data)相关的研究和发展,世界范围内IBM、Oracle等IT企业巨头纷纷投身于大数据业务。今年3月,美国奥巴马政府宣布启动“大数据研究和发展计划”,六个联邦政府部门和机构将投资2亿美元,用于提高从大量数据中访问、组织、收集、发现信息的工具和技术水平,这也意味着大数据技术从市场行为正式上升为美国的国家科技战略^[1]。

大数据的快速发展反映的是量化的思维方式,这与我国情报界这些年一直倡导的“事实数据+工具方法+专家智慧”的情报研究思路与情报实践方法也不谋而合。尽管大数据问题的出现会给传统的情报学研究带来不少新问题、新挑战,但大数据实质上是对数据资源的开发利用,这也给情报学的发展带来了前所未有的战略机遇。

1 大数据的内涵与价值

1.1 大数据的内涵

一般认为,大数据是指数据规模与复杂性超过了传统关系数据库处理能力的数据^[2,3],其特征可以概括为“3V”:即数量(Volume)、速度(Velocity)和多样性(Variety)^[4]。Volume指大数据容量和规模远超过传统数据;Velocity指速度,Oracle公司将其解释为数据产生的速度极快^[5],IBM公司则将其理解为需要快速地对数据进行处理^[6];Variety是指数据的类型多样,大数据包括不同结构、不同来源、不同媒体形态的各种数据。除3V之外,IBM公司和Oracle公司还指出大数据的特征还应该包括第四个V:价值(Value),意指大数据蕴含着重要的经济价值和社会价值。

大数据的4V特征给数据获取、数据监护^[7](Data Curation, 又译为数据监管^[8])、数据存储、数据搜索、数据共享、数据分析和数据可视化等领域的既有技术方法带来了极大挑战,因此大数据发展的最初阶段集中表现为大数据相关技术的研发和进步。

1.2 大数据的意义与价值

大数据的引人注目既是因为其规模、复杂性和速度超越了技术处理能力,迫使人们不断研发相关的新技术,更是因为随着量的积累,数据的价值产生了质变。

美国管理学家与统计学家戴明曾指出“除了上帝,任何人都必须用数据来说话”。这反映出以美国为典型的西方国家崇尚定量的思维方式,体现了尊重数据、利用数据、以数据说话的思维和行为范式。大数据的研究与发展集中体现了将数据作为重要的战略资源进行开发利用的思想,是定量思维方式与以数据说话行为范式的集中体现,因而具有重要的战略指导意义。

同时,大数据的应用与实践也具有显著的现实经济价值与社会价值。正如麦肯锡公司所指出的:大数据可以在任何一个行业内创造更多价值。在医疗行业,如果美国医疗业全面使用大数据,每年能多创造3000亿美元的价值;在个人位置信息领域,利用大数据技术可以为服务商带来1000亿美元的利润,并为消费者和商业用户带来7000亿美元的价值^[3]。2008年Nature刊发一篇文章显示:Google搜索引擎利用看似无规律的巨量用户查询数据能够提前1—2周预测流感的爆发^[9]。今年, Twitter基于其每天上亿条的信息,发布了针对2012年美国大选的“Twitter 政治指数”,评估选民对两位候选人的满意度,以此得出的评估结果与传统大规模民意调查得出的结果基本一致。

2 大数据给情报学提出的挑战

情报学是以数据、信息、知识、情报为主要研究对象的学科,大数据问题的出现也给情报学提出了诸多挑战,体现在以下几个方面:

(1) 相比于传统的信息内容,特别是相对规范的科技信息,大数据更零散、更原始、更复杂。这些巨量、快速、多样的数据存在明显的碎片化特征;不少数据仅仅停留在原始数据层面,并未成为真正的信息;大量数据是真假不分的杂乱数据,数据背后反映的是复杂的现实世界无显著规律的行为。

(2) 大数据对相关技术提出了新的要求。技术是大数据问题中的关键问题,并将一直影响着大数据问题的研究。对大数据的有效存储、管理和使用是实现大数据目标的基本要求。在此基础上,传统的信息获取、存

储管理、分类、索引、检索、聚类、人机交互技术等信息分析处理技术方法如何顺利转移到对大数据的管理和利用上来是对情报学提出的重要挑战。

(3) 用户需求更复杂,对服务价值的期望值更高:大数据4V中的价值特性使得用户对基于大数据服务的价值有更高的期望值,但同时用户的需求也更加复杂。实现从数据到用户价值的转化则需要情报人员一方面能够更深刻地了解用户需求,另一方面需要更加系统性地了解数据及其特征,建立高效可控的数据处理流程,掌握多种数据分析方法,需要情报人员具备更加全面、综合的素质与能力。

(4) 信息安全和公民隐私问题更加突出:在大数据时代,信息安全和隐私问题又将成为情报学研究的新挑战。随着全球各国对大数据获取、处理分析能力的提高,大量零散数据中蕴含的有价值信息能够得以发现并利用,这种情况下,信息安全也从传统计算机网络安全、保密管理等可控安全管理变为无法确知安全隐患的不可控的安全管理,而同时信息公开、数据公开的呼声都越来越强烈,数据开放与信息安全之间的矛盾更加突出。同样,大数据中包含大量用户信息,这也使得对大数据的开发利用很容易侵犯公民的隐私,恶意利用公民隐私的技术门槛大大降低,公民隐私保护也将成为极具挑战的研究问题。

3 大数据给情报学带来新机遇

大数据尽管起初作为技术概念被提出,因为初期发展集中体现为数据存储管理技术的发展和提升,但大数据问题的实质是对数据的管理与开发利用,与传统情报学关注的信息资源管理的开发利用具有广泛的共通性。因此大数据问题的兴起给近20年来备受挑战的情报学带来了新的发展机遇。

(1) 大数据的价值链与传统情报工作的价值链一致:对大数据的发掘利用涉及包括数据获取、数据监护、数据存储、数据检索、数据共享、数据分析、数据展示等多个环节在内的价值链条,这一链条与信息管理和传统情报工作的信息价值链基本吻合。情报学领域可以将包括信息采集、信息整序、信息组织、信息检索、信息分析、信息可视化等方面成熟的理论方法和技术应用到大数据的工作中,在促进大数据研究发展的同时,扩大传统情报服务范围。

(2) 大数据工作与“事实数据+工具方法+专家智

慧”的情报方法有效契合：科学研究的定量方法在大数据领域充分体现，当工具和方法作为人类智慧的结晶还必须与人类的智慧本身有效结合，偏离定性和人类洞察力的纯粹事实数据不会具有真正的意义。传统情报学更加侧重于各类科技文献资源和科学数据，大数据的兴起和发展能够丰富传统情报研究中事实数据的来源、内容与媒体形态，不同的事实数据互相补充、相互印证，能够促进传统情报工作水平和情报产品质量的提升。

(3) 大数据的推进能够为传统情报学提供有力的数据处理分析工具和方法。信息技术和分析算法尽管是情报学一直关注的重要研究内容，但也一直是情报学学科发展中的软肋。大数据环境下，各国政府和市场主体将更加重视对大数据相关技术的研发投入，也将吸引大批优秀人才进入本领域，进行实现相关工具方法的快速进步和发展，有力弥补本学科技术能力不足的长期劣势。

(4) 大数据的发展能够更凸显情报学的应用价值，推动本学科的快速发展：大数据的价值不仅在于能够进一步加大情报工作对我国科技战略和科技发展工作的支撑作用，使情报工作与科研工作更加紧密结合，大数据的研究和发展还将成为公共管理、商业机构运营等各类活动的定量决策支撑。在此过程中，理解业务需求、熟悉相关技术方法、了解数据资源的情报人员也将扮演越来越重要的角色。学科人才的旺盛市场需求反过来也会促进情报学教学科研工作的全面提升和发展，进而实现情报学的整体发展。以美国为例，包括约翰·霍普金斯大学、加州大学圣迭戈分校、伊利诺伊大学、密歇根大学、康奈尔大学和麻省理工大学在内的不少高校已经设置了专门硕士、博士学位，以培养能够按照数据的生命周期全过程对数据进行管理的专业人才^[10]。

4 结语

大数据发展需要大视野，情报学的发展同样需要大视野。

虽然在我国政府层面尚未正式使用大数据这一概念，但国家有关部门和地方政府以及各类企事业单位都在积极关注大数据研究和发展。在科技部发布的《国家“十二五”科学和技术发展规划》中，已经将海量数据存储服务系统、海量数据处理等列入关键技术攻关或重点前沿研究。在工信部今年发布的《物联网“十二五”发展规划》上，已经把海量数据存储、数据挖掘、图像视频智能分析、信息感知技术、信息传输技术、信息安全技术等大数据相关技术列入到项目规划中。教育部也在积极支持大数据相关的研究工作，目前已经支持在中国人民大学建设面向海量数据管理的研究和应用平台。广东省经济和信息化委员会同有关部门起草的《广东省实施大数据战略工作方案》将报送给省政府，借鉴其他国家利用数据的工作机制和先进经验，在政府工作中注重数据梳理，同时更好地实现政务公开，满足公众的知情权。各高校、地方政府、产业园等机构也在积极关注大数据的研究与进展。大数据已经成为国内产业界、学术界和政府部门广泛关注的热点问题。

在这种环境下，情报学和情报工作需要发挥、利用学科的既有优势与既有成果积极参与到大数据的发展应用过程中，并积极从大数据方面的成果中汲取对传统情报工作有益的思路、成果和方法，实现守住传统阵地与开辟全新领域的有机结合，充分发挥情报学与情报工作在信息收集、处理与分析利用方面的既有优势。只有这样，才能实现大数据时代中情报学的大发展、大繁荣！

参考文献

- [1] The White House. Big Data is a Big Deal [EB/OL].(2012-03-29)[2012-11-16].<http://www.whitehouse.gov/blog/2012/03/29/big-data-big-deal>
- [2] Wikipedia. Big Data [EB/OL].(2012-11-15)[2012-11-16].http://en.wikipedia.org/wiki/Big_data
- [3] Manyika J, Chui M, Bughin J, et al. Big Data: The next frontier for innovation, competition, and productivity[R]. McKinsey Global Institute, 2011
- [4] Gartner. Big Data [EB/OL]. [2012-11-16]. <http://www.gartner.com/it-glossary/big-data/>
- [5] Oracle. Oracle: Big Data for the Enterprise [EB/OL]. (2012-01)[2012-11-16]. <http://www.oracle.com/us/products/database/big-data-for-enterprise-519135.pdf>
- [6] Zikopoulos P C, DeRoos D, Parasuraman K, et al. Harness the Power of Big Data—The IBM Big Data Platform[M]. McGraw-Hill Osborne Media, 2012: 9-15
- [7] 杨鹤林. 从数据监护看美国高校图书馆的机构库建设新思路——来自DataStaR的启示[J]. 大学图书馆学报, 2012, 30(2): 23-28, 73
- [8] 沈婷婷, 卢志国. 数据监管在我国高校图书馆的应用展望[J]. 图书情报工作, 2012, 56(7): 54-57, 87.
- [9] Ginsberg J, Mohebbi M H, Patel R S, et al. Detecting influenza epidemics using search engine query data[J]. Nature, 457(7232): 1012-1014.
- [10] Walters T O. Data Curation Program Development in U.S. Universities: The Georgia Institute of Technology Example[J]. The International Journal of Digital Curation, 2009(4) 3: 83-92

作者简介

贺德方, 中国科学技术信息研究所所长, 研究员。E-mail: hedf@istic.ac.cn

Informatics under the Environment of Big Data

He Defang / Institute of Scientific and Technical Information of China, Beijing, 100038

Abstract: Big Data has become a hot research topic and an important research direction in China and abroad. This article, from the viewpoint of the connotation of Big Data, explains its potential value and practical significance. Analyzes the challenge brought by Big Data to Informatics in data quality, technology requirements, information service and information security, as well as privacy protection. Also, analyzes new opportunities that the development of Big Data brings to informatics and Intelligence work. Finally, it suggests that our researches of informatics and Intelligence work should devote to study and development of Big Data with its existing advantages and results, takes Big Data study as an opportunity to realize the great development.

Keywords: Big Data, Informatics, Intelligence work

(收稿日期: 2012-11-16)