

学校编码：10384
学号：10420121152194

分类号_____密级_____
UDC_____

廈門大學

硕士学位论文

大数据时代的科学研究变革探析

The Transformation of Scientific Research Activities
in the Era of Big Data

刘伟榕

指导教师姓名：贺威 副教授
专业名称：科学技术哲学
论文提交日期：2015年4月
论文答辩时间：2015年5月
学位授予日期：2015年 月

答辩委员会主席：_____

评阅人：_____

2015年4月

厦门大学学位论文原创性声明

本人呈交的学位论文是本人在导师指导下,独立完成的研究成果。本人在论文写作中参考其他个人或集体已经发表的研究成果,均在文中以适当方式明确标明,并符合法律规范和《厦门大学研究生学术活动规范(试行)》。

另外,该学位论文为()课题(组)的研究成果,获得()课题(组)经费或实验室的资助,在()实验室完成。(请在以上括号内填写课题或课题组负责人或实验室名称,未有此项声明内容的,可以不作特别声明。)

声明人(签名):

年 月 日

厦门大学学位论文著作权使用声明

本人同意厦门大学根据《中华人民共和国学位条例暂行实施办法》等规定保留和使用此学位论文，并向主管部门或其指定机构送交学位论文（包括纸质版和电子版），允许学位论文进入厦门大学图书馆及其数据库被查阅、借阅。本人同意厦门大学将学位论文加入全国博士、硕士学位论文共建单位数据库进行检索，将学位论文的标题和摘要汇编出版，采用影印、缩印或者其它方式合理复制学位论文。

本学位论文属于：

1. 经厦门大学保密委员会审查核定的保密学位论文，于 年 月 日解密，解密后适用上述授权。

2. 不保密，适用上述授权。

（请在以上相应括号内打“√”或填上相应内容。保密学位论文应是已经厦门大学保密委员会审定过的学位论文，未经厦门大学保密委员会审定的学位论文均为公开学位论文。此声明栏不填写的，默认为公开学位论文，均适用上述授权。）

声明人（签名）：

年 月 日

论文摘要

大数据是信息时代的必经阶段，是信息社会从“量变”走向“质变”的表征。人类进入了一个对世界进行数据化解构和重构的时代，用数据思维和手段去发现知识和创造价值成为趋势，大数据已然成为全球学术研究的新热点。作为人类文明进步的积极力量，科学研究活动不仅引领着大数据时代，其自身的发展条件、研究方法和范式、进步方式也在酝酿着积极的变革与创新。

大数据时代首先带来了科学研究条件的变革。科学研究活动正在创造出规模庞大、模式复杂、自我繁殖的“数据自然界”。越来越多的新科学发现将依赖于对“数据自然界”的挖掘和利用，包含大数据管理、分析和支撑的新科研工具将成为必需的设施。新的研究手段也带来研究理念的转变，通过数据挖掘寻求科学规律的数据密集型学科领域正在不断涌现，社会科学的定量化研究迎来了新机遇。很多国家都注意到了科学研究条件的变革所蕴含的巨大潜力，纷纷将大数据提高到了国家科技战略的高度。

科学研究条件的变革也引发了科学研究方法的革新。数据不仅是科学研究的强大工具，更成为科学研究的直接对象。数据驱动的科学不需要预先做出理论假设，而是对已有大数据的挖掘和分析，就可以发现蕴藏的知识或规律。以数据为核心的科学研究正在凝聚科学共同体、集结跨学科人才、促成广泛的科学联盟。一种基于数据的、开放协同的科学研究“第四范式”正在形成，已经在天文学、物理学、生命科学等领域引发了新突破。

新条件和新范式促成了科学进步方式的改变。“第四范式”用数据挖掘、机器归纳、人机合作等手段补充了科学知识的生产方式，提升了科学向真理逼近的速率和程度。大数据思维和技术增强了科学“解决问题”的实用性。相关性、复杂性思维补充了单一的因果规律，从而拓展了科学的预测能力；整体性和实践性的研究理念纠正了传统科学对还原论的偏执，提升了地方性、实践性知识的地位；社会科学得到前所未有的数据支撑，进一步实现各学科的交叉融合。科学研究的社会价值正在跃升：科学数据的挖掘利用、科研成果的快速转化、新型科研机构、科技产品社会化生产将带来更多的认识价值、经济价值与人文价值。

大数据为科学研究带来了新机遇，然而机遇变为现实并不容易，科研领域广泛变革有赖于一系列挑战的成功解决。在技术层面，大数据技术还未进入成熟期，需要应对科研数据管理、分析与成本上的挑战。在思维层面，科研方法论受到观察渗透理论、信号与噪声的一体性、相关性与因果性复杂关系等的制约。在制度和文化方面，面临着数据开放、隐私保护、反垄断等方面的挑战。

关键词：大数据；科学研究变革；第四范式

Abstract

Big Data is the inevitable result of the computer, the Internet and other information technology development. It symbolizes the information society from the “quantitative” to “qualitative change”. Mankind has entered an era of world data deconstruction and reconstruction. Thinking with data and tools to discover knowledge and create value become a trend. Big Data has become the new hot spot for global academic research. As a positive force for the progress of human civilization, scientific research activities not only leading the Big Data era, its own conditions for development, research methods and paradigms, progressive approach is also brewing a positive change and innovation.

Big Data era has brought changes in the scientific conditions. Research activities have created a large-scale, complex patterns, self-propagating “the Data Nature”. More and more new scientific discovery will depend on the mining and utilization of “the nature of data”. The new research tool contains a large data management, analysis and service support facilities will be required. The new research means research also brought about changes in the concept. Data-intensive science using data mining as the main means is evolving. Quantitative social science research ushered in new opportunities. Many countries have taken note of the enormous potential of change inherent in scientific research conditions, have to raise the height of big data to national science and technology strategy.

Research conditions change also brings innovative research methods. Scientists not only to solve scientific problems by means of data, but also treat the data as an object of scientific research and tools. Data-driven research is not required to make a pre-theoretical assumptions. Scientists through the big data mining and analysis to discover hidden knowledge or the law. Data-centric scientific can gather scientific community, build interdisciplinary talent, led a broad coalition of science. A data-based, open collaborative scientific research “the Fourth Paradigm” is taking shape. It is leading to breakthroughs in astronomy, physics, life sciences and other areas.

New conditions and new paradigms contributed to a change in the way of scientific progress. “the Fourth Paradigm” supplement production of scientific knowledge, enhance the rate and extent of scientific

approximation to the truth. Big Data thinking and technology has enhanced the scientific capacity of “solving problems”. Relevance, complexity thinking complements the single causal law, thereby expanding the scientific prediction capability. Holistic and practical research concept has corrected reductionism, enhance the status of local, practical knowledge. Social science data obtained unprecedented support. Social value of scientific research has been improved.

Change to reach a wide range of scientific advances and breakthroughs are also facing a series of challenges. At the technical level, big data technology has not yet matured, we need to respond to scientific data management, analysis on the challenges and costs. In the thinking level, research methodology is restricted by following three aspects: observation of percolation theory, oneness of signal and noise, complex relationship between correlation and causality. At the institutional and cultural level, we need to achieve a balance between personal interests and the overall interests of science.

Keywords: Big Data; transformation of scientific research; the Fourth Paradigm

目 录

绪 论	7
第一节 研究背景和意义	1
一、大数据时代已经来临.....	1
二、“大数据”成为学术研究热点.....	2
三、科学研究在大数据时代孕育变革.....	3
第二节 研究现状和本文创新点	4
一、研究现状.....	4
二、本文创新点.....	7
第一章 大数据时代的内涵及影响	9
第一节 大数据：信息社会的必然产物	9
第二节 大数据的内涵与技术创新	11
第三节 大数据的时代意义	12
一、人类“信息环境”的改变.....	13
二、世界的数据化解构与重构.....	14
第二章 科学研究条件的变革	16
第一节 科学研究新环境：“数据自然界”	16
第二节 新工具：大数据的管理、分析与服务	18
一、科研数据的传输与存储.....	18
二、科研数据的分析处理工具.....	19
三、新的知识服务模式.....	20
第三节 新领域：以数据为核心的学科群	20
一、数据密集型科学的兴起.....	21
二、社会科学的数据化探索.....	22
第四节 新政策：大数据成为国家战略	23
第三章 科学研究方法的变革	25
第一节 数据成为科学研究的对象	25

第二节 数据分析替代理论假说.....	26
第三节 以数据为核心的科研合作.....	28
一、“数据众包”集结跨学科人才.....	28
二、数据聚合凝聚科学共同体.....	29
三、技术协作促进广泛的科学联盟.....	29
第四节 科学研究酝酿“第四范式”.....	31
第四章 科学进步方式的变革.....	33
第一节 数据驱动提高科学“逼真度”.....	33
一、数据挖掘加快共性规律显现.....	33
二、大数据归纳法提高知识可靠性.....	33
三、人机合作实现知识自动化.....	34
第二节 数据思维促进科学的多样性.....	35
一、提升现有科学的预测力.....	36
二、促进社会科学的发展.....	37
三、地方性知识获得承认.....	37
四、促进学科融合与交叉.....	38
第三节 智能化的科学价值实现方式.....	39
一、科学数据的快速增值.....	39
二、知识管理与应用的智能化.....	40
三、新型科研机构的涌现.....	41
四、科技产品的社会化生产.....	41
第五章 科学研究变革实现的挑战.....	43
第一节 大数据技术尚未进入成熟期.....	43
一、数据管理的技术难题.....	44
二、数据处理与分析的复杂挑战.....	44
三、大数据独有的高成本问题.....	45
第二节 “第四范式”的方法论困境.....	45
一、数据处理中的“观察渗透理论”.....	45
二、信号与噪声的一体性.....	46

三、追求相关关系的局限性.....	47
第三节 科学数据权益与伦理争议.....	48
一、科研资源的垄断隐患.....	48
二、数据权益与隐私纠纷.....	49
三、数据开放政策落地受阻.....	49
结 语.....	51
参考文献.....	52
致 谢.....	55

厦门大学博硕士学位论文摘要库

Contents

Introduction	1
Section 1 Background and Significance.....	1
Section 2 Research situation and innovations.....	4
Chapter 1 Meaning and Influence of Big Data era	9
Section 1 Big Data: the inevitable product of the information era.....	9
Section 2 Connotation and technological innovation of Big Data.....	11
Section 3 Meaning of Big Data era.....	12
Chapter 2 Changes of scientific research conditions	16
Section 1 New research environment: “the Data Nature”.....	16
Section 2 New tools: Big Data management, analysis and service.....	18
Section 3 New areas: data-centric group of subjects.....	20
Section 4 New policy: Big Data has become national strategies.....	23
Chapter 3 Changes in scientific method	25
Section 1 Data being the object of scientific research.....	25
Section 2 Data analysis alternative hypothesis.....	26
Section 3 Research cooperation in data-centric.....	28
Section 4 The fourth paradigm of scientific research.....	31
Chapter 4 Changes in the way of scientific progress	33
Section 1 Data-driven improving science "fidelity".....	33
Section 2 Data idea promoting the diversity of science.....	35
Section 3 Intelligent way to achieve scientific value.....	39
Chapter 5 Challenge to achieve scientific revolution	43
Section 1 Big Data technology is not yet matured.....	43
Section 2 methodological dilemma of “the Fourth Paradigm”.....	44
Section 3 Scientific data rights and ethical controversy.....	48
Epilogue	51
Bibliography	52
Acknowledgements	55

绪 论

人类正在进入大数据时代,推动这个时代到来的科学研究活动也在酝酿着积极的变革与创新。本论文综合最新的大数据研究、前沿的科研案例及科学哲学理论,尝试对科学研究活动正在发生与将要发生的变革进行探讨。

第一节 研究背景和意义

随着信息技术的持续飞速发展,数据爆炸现象越来越显著,世界数据化的广度和深度在不断加大。在此背景下,大数据已成为商业投资、国家战略、学界研究的热点。科学研究活动作为推动人类文明进步的积极力量,在研究条件、研究方法等方面正孕育着变革,值得我们关注和研究。

一、大数据时代已经来临

2010 年是人类数据历史上的里程碑,这一年全世界的数据总量跃进 ZB 时代,达到了 1.2ZB (相当于 1.2×10^{14} GB)。从印刷术发明至今,人类创造出的所有印刷品中包含的数据量仅为 200PB (相当于 2×10^8 GB)。信息时代只用了几十年就超过了印刷时代用千年时间积累的数据量,人们面临的数据现象有了质的变化:一是数据体量大且呈指数增长,如图 0-1 所示,2012 年全球数据量为 2.8ZB,预计 2015 年接近 8.6ZB,年增长率接近 45%^①;二是数据进出的高速性,对数据的快速处理能力提出了极高要求;三是数据类型繁多,非结构化数据成为主体,在数据总量中的比重超过 75%。^②

数据规模和复杂度的改变催生了高性能云存储体系、开源云计算框架、非结构化计算模型、数据挖掘等相关数据技术的创新,使得人们不仅能够驾驭大数据集,更能通过数据分析获取新的知识和价值。大数据技术及其应用的巨大潜力,使得大数据方法论、价值论广泛传播。2012 年 2 月 11 日《纽约时报》率先发文称“大数据时代”已经到来,社会各个领域所积累的数据量正在从量变转向质变,由数据驱动的知识发现和决策辅助正在科学、商业、政治等领域引爆变革。^③

^① IDC. The Digital Universe in 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East [EB/OL]. [2012-12]. <http://www.emc.com/collateral/analyst-reports/idc-the-digital-universe-in-2020.pdf>.

^② 李国杰. 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]. 中国科学院院刊, 2012 (6): 647-657.

^③ Lohr S. The Age of Big Data [EB/OL]. [2012-02-12]. <http://nytimes.com/2012/02/12/Sunday-review/big-datas-impact-in-the-world.html>.

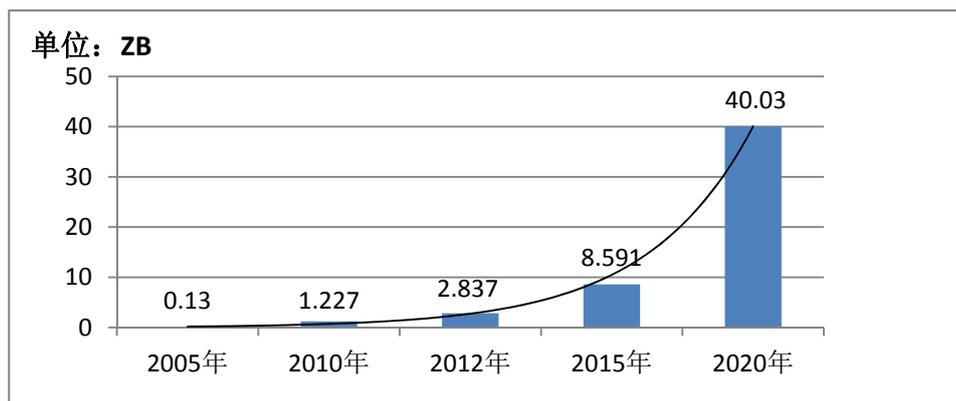


图 0-1 全球数据量现状及预测

大数据将是推动社会变革的重要力量，这已经成为产业界和学术界的共识。据调查，在2015年6月之前，全球将有逾六成的大型企业投资大数据项目。^①各国政府都已在关注和布局大数据：美国在2012年3月发布了《大数据研究与发展计划》、澳大利亚在2013年8月发布《公共服务大数据战略》、英国在2013年10月发布《英国数据能力战略》、法国在2013年7月发布《法国政府大数据五项支持计划》。^②

二、“大数据”成为学术研究热点

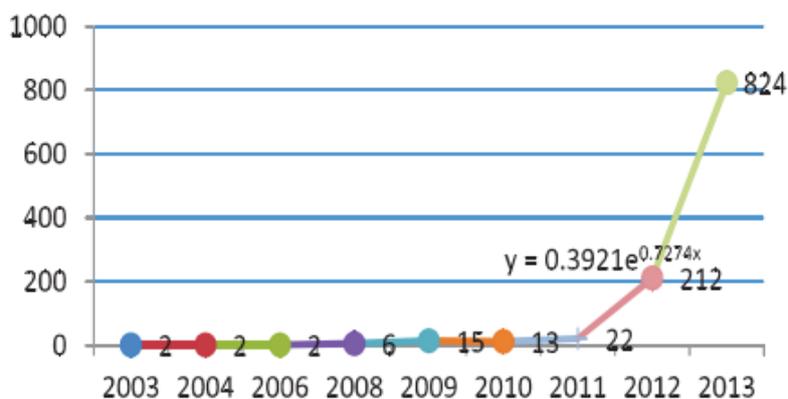
“大数据”已是国际顶级学术刊物的研究热点。2008年 *Nature* 推出专刊“Big Data”^③、2011年 *Science* 出版专刊“Dealing with data”^④，2012年 *ERCIM News*（欧洲信息学与数学研究协会会刊）也出版专刊“Big Data”。在顶级刊物的带领下，越来越多的学者投身研究大数据。从2003-2014年10月20日，全球著名的引文数据库“Web of Science”中以“Big Data”为主题的文献共有1556篇。如图0-2所示，2012年伊始，与大数据相关的研究成果数量迅速攀升，从2011年的22篇猛增到2012年的212篇，再到2013年的824篇。

^① Gartner. Gartner Survey Reveals That 64 Percent of Organizations Have Invested or Plan to Invest in Big Data in 2013. [EB/OL]. [2013-10-11]http://www.gartner.com/newsroom/id/2593815.

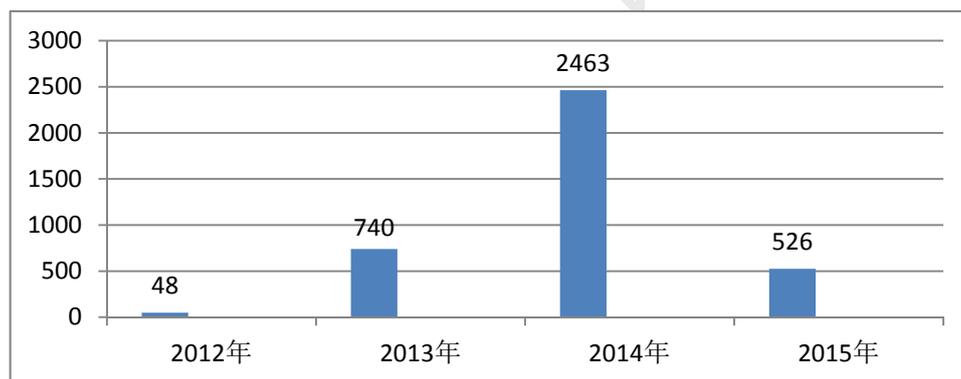
^② 张勇进, 王璟璇. 主要发达国家大数据政策比较研究[J]. 中国行政管理, 2014 (12): 113-117.

^③ Nature. Big Data[EB/OL]. [2015-03-03]. http://www.nature.com/news/specials/biadata/index.html.

^④ Science. Special online collection: Dealing with data[EB/OL]. [2015-03-03]. http://www.sciencemag.org/site/special/data.

图 0-2 国际大数据研究发文量随时间变化曲线^①

我国学者对大数据也表现出了极大的关注和研究热情。在 SCI 来源期刊、EI 来源期刊、核心期刊、CSSCI 期刊等核心期刊中以“大数据”为关键词的论文数在 2014 年达到了 2463 篇（图 0-3），较 2012 年增长了 50 倍！从知识图谱（图 0-4）来看，大数据技术及相应的基础研究已经成为我国学术界的研究热点，聚集了信息科学、网络科学、系统科学、社会科学、心理学、经济学等各领域学者。



（数据来源：笔者以“大数据”为关键词搜索出的论文数，截至 2015 年 3 月 30 日）

图 0-3 CNKI 核心期刊数据库中大数据论文数量

三、科学研究在大数据时代孕育变革

在科学研究的数理方法和实验方法中，数据始终处于基础地位。在科学交流与合作中，数据因其在各自学科体系之内的可通约性而价值倍增。随着科学与技术的界限日趋模糊，数据的基础地位愈发显著。现如今，数据现象的量变引起质变，势必引发科学活动的多方面变革。

在科研工具方面，美国能源部（DOE）投资 2500 万美元成立专门的数据研究所帮助科学家对数据进行有效管理，促进其生物和环境研究计划等研究成果。^②包

^① 王炳立. 基于科学计量学的国际大数据研究可视化分析[J]. 情报杂志, 2015 (2): 131-136.

^② 冯海超. 透视美国大数据爆发全景[J]. 互联网周刊, 2013 (1): 38-41.

含采集、管理与分析工具的大数据处理系统正在环境应用科学、海洋科学、生态科学、物理学、天文学、生物学等领域流行起来。科研数据的采集、传输、存储、管理、分析和可视化能力的升级换代可能成为科学研究必须适应的新事实。

在科学研究方法上，在实验科学（第一范式）、理论科学（第二范式）、计算科学（第三范式）之外已经出现了第四种研究范式，被称为“数据密集型科学”，即用数据挖掘寻找未知规律。全球第一个有效挖掘基因组的工具SPANR系统利用数据挖掘技术发现了与胰腺癌、结肠癌、脊髓性肌萎缩症相关的新突变基因，还发现了之前从未曾被人关注到的39个自闭症基因^①，让人对科学研究变革的潜力充满期待。

大数据时代为科学研究带来了重大机遇。当数据规模和复杂度达到一定阈值时，人们将可以做到许多在小数据基础上无法完成的事情，基于对大数据的分析能更好地理解世界，解决曾经难以解决的或不可能解决的科学问题，做出惊人的科学发现。作为一种新的理念和技术手段，大数据的潜在价值是无限的。因此，我们应以超前的意识去研究新时代科学研究所表现出的新特点、在研究方法和模式上的变化和应对措施。如此一来才能更好地迎接和促进科学研究变革的发生，增进人类福祉。

第二节 研究现状和本文创新点

大数据作为新兴事物，对它的理论研究和案例研究都只处于初创阶段，并无太多可资借鉴的重大成果。目前学术界对新时代科学研究的变革关注较少，在案例搜集、理论解释、系统研究方面都有较大的创新空间。

一、研究现状

学术界对科学研究活动在大数据时代的变革关注较少。从图 0-4 可以看出，目前学界的大数据研究论文主要围绕云计算、数据挖掘、数据分析、hadoop、信息服务、数据处理等主题。王炳立^②对 Web of Science 的 1556 篇国际论文的文献共被引时区知识图谱得出了大数据的研究演进脉络（图 0-5），从中可以看出，国际上对数据科学、知识集成、应用社会科学等研究是近两年才兴起的。

^① 冯卫东. 用大数据解译 DNA 获重大突破——加科学家开发出挖掘疾病遗传根源的数字化工具. [EB/OL]. [2015-03-03]. http://digitalpaper.stdaily.com/http_www.kjrb.com/kjrb/html/2014-12/20/content_287233.htm?div=-1

^② 王炳立. 基于科学计量学的国际大数据研究可视化分析[J]. 情报杂志, 2015 (2): 131-136.

Degree papers are in the "[Xiamen University Electronic Theses and Dissertations Database](#)". Full texts are available in the following ways:

1. If your library is a CALIS member libraries, please log on <http://etd.calis.edu.cn/> and submit requests online, or consult the interlibrary loan department in your library.
2. For users of non-CALIS member libraries, please mail to etd@xmu.edu.cn for delivery details.

廈門大學博碩士論文摘要庫