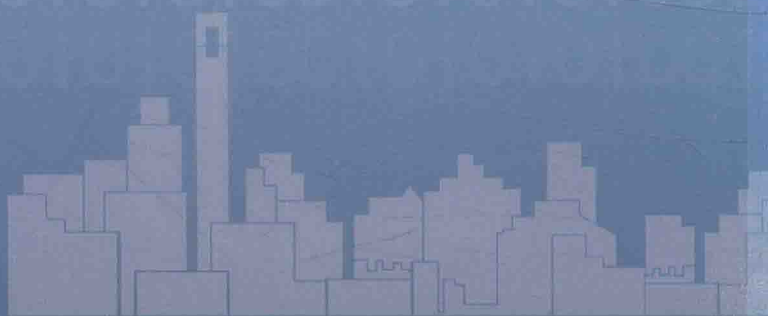


全球大数据产业 技术创新态势及 相关政策研究

党倩娜 曹 磊 罗天雨 著



上海科学普及出版社

全球大数据产业技术创新态势及相关政策研究

党倩娜 曹 磊 罗天雨 著

上海科学普及出版社

图书在版编目(CIP)数据

全球大数据产业技术创新态势及相关政策研究/
党倩娜,曹磊,罗天雨著. --上海:上海科学普及出版
社,2015.12

ISBN 978-7-5427-6610-6

I. ①全… II. ①党…②曹…③罗… III. ①数据处理—
高技术产业—技术革新—研究—中国 IV. ①F279.244.4

中国版本图书馆 CIP 数据核字(2015)第 301909 号

责任编辑 史炎均

全球大数据产业技术创新态势及相关政策研究

党倩娜 曹磊 罗天雨 著

上海科学普及出版社出版发行

(上海中山北路 832 号 邮政编码 200070)

<http://www.pspsh.com>

各地新华书店经销 虎彩印艺股份有限公司印刷

开本 787×1092 1/16 印张 14 字数 280 000

2015 年 12 月第 1 版 2015 年 12 月第 1 次印刷

ISBN 978-7-5427-6610-6 定价: 48.00 元

本书如有缺页、错装或坏损等严重质量问题

请向出版社联系调换



上海科技
发展基金会

上海科技发展基金会(www.sstdf.org)的宗旨是促进科学技术的繁荣和发展,促进科学技术的普及和推广,促进科技人才的成长和提高,为推动科技进步,提高广大人民群众的科学文化水平作贡献。本书受“上海科技发展基金会”资助出版。

“上海市科协资助青年科技人才出版科技著作 晨光计划”出版说明

“上海市科协资助青年科技人才出版科技著作晨光计划”(以下简称“晨光计划”)由上海市科协、上海科技发展基金会联合主办,上海科学普及出版社有限责任公司协办。“晨光计划”旨在支持和鼓励上海青年科技人才著书立说,加快科学研究和传播,促进青年科技人才成长,切实推动建设具有全球影响力的科技创新中心。“晨光计划”专门资助上海青年科技人才出版自然科学领域的优秀首部原创性学术或科普著作,原则上每年资助10人,每人资助一种著作1500册的出版费用(每人资助额不超过10万元)。申请人经市科协所属学会、协会、研究会,区县科协,园区科协等基层科协,高等院校、科研院所、企业等有关单位推荐,或经本人所在单位同意后直接向上海市科协提出资助申请,申请资料可在上海市科协网站(www.sast.gov.cn)“通知公告”栏下载。

编辑委员会

主 编：党倩娜

撰稿人：党倩娜 曹 磊 罗天雨

曹 可 陈薇娜

人类的信息化进程已经从数字化进入数据化的新阶段,世界正在进入“大数据”(Big Data)时代,数据正在成为重要的资源要素,大数据是智慧的源泉。数据能力成为一个国家或区域新的竞争力。

近年来,上海科学技术情报研究所在国内比较早地关注了大数据这一最新的趋势。2010年,从追踪“智慧地球”入手进行研究。自2011年起,开始承担并完成了上海市经信委、上海科技发展基金、国家社会科学基金等多项重要课题,对大数据产业化、数据资产、数据安全等方面进行深入研究,已经取得了一系列成果。同时,上海科技情报所的大数据研究团队先后在《科学时报》、《文汇报》、《科技日报》和国内一些专业刊物上陆续公开发表了一系列文章。

然而,在当前对大数据的诸多研究中仍存在诸多的问题,尤其是在关于大数据的技术创新态势方面,如大数据技术究竟发展到哪一阶段?它的过去、今天和未来有着怎样的演化趋势?哪些机构、哪些区域处于领先地位,形成了怎样的创新网络?其产业体系框架该如何架构,对商业模式形成了怎样的影响?作为新一代战略性新兴产业具有怎样的创新特性?全球主要国家的相关政策及其政策支持模式有哪些?上海目前处于什么状态,应当怎样发展?对于这些问题的思考将对当前国家尤其是上海地区相应政策的制定具有系统性的、重要的意义,要想解决这些问题,势必要求我们对大数据产业及其技术创新发展态势进行全面的分析。

对此,上海图书馆上海科技情报研究所认真地进行了探索和研究。2013年,申请了上海市科技发展基金重点课题,试图通过利用知识图谱、专利地图等新兴可视化的研究方法,将定量分析和定性分析结合起来,通过知识、技术和产业这三维视角的结合,全面、客观地展现大数据领域的创新演化路径、前沿方向、创新群体及网络,分析其主要驱动因素和创新特性,研究全球产业链、竞争格局

及产业政策的支持模式,希望为社会各界及时了解最新态势,把握方向,为政府制定相关发展思路与政策提供一定的参考与支撑。

本书是我们系列研究的一个阶段性成果。作为年轻的研究团队,在研究分析和资料引用中难免存在纰漏和错误,不足之处希望社会各界给予指导。

上海图书馆副馆长

上海科学技术情报研究所副所长

陈 超

第一章 绪论	1
1.1 大数据的研究背景	1
1.2 大数据的定义及相关概念	1
1.3 大数据产业的国内外研究概况	6
1.4 研究总体思路与主要内容	9
1.5 大数据产业创新发展趋势分析框架	11
1.6 主要分析设计及数据库的初步构建	16
第二章 大数据产业技术创新总体概况	20
2.1 技术创新演化历程	20
2.2 研究机构和创新群体分布	31
2.3 研究重点与前沿分布	40
第三章 大数据产业链与竞争格局分析	56
3.1 大数据产业链初步形成	56
3.2 大数据产业与战略性新兴产业关联分析	64
3.3 大数据产业竞争格局分析	71
第四章 大数据对商业模式影响分析	83
4.1 大数据思维——产业发展模式的颠覆	83
4.2 大数据资源——平台生态系统的构建和整合	87
4.3 大数据工具——企业变革力量	90

第五章 大数据产业创新趋势与特点	93
5.1 关于大数据的几个重要认识	93
5.2 数据驱动型商业模式创新	94
5.3 大数据产业创新特性	95
第六章 主要国家和地区大数据领域相关政策概况与分析	99
6.1 美国	99
6.2 欧盟	130
6.3 日本	142
6.4 韩国	146
6.5 新加坡	148
6.6 主要国家和地区的比较分析	151
第七章 中国大数据领域相关政策分析	156
7.1 中国大数据领域创新发展概况	156
7.2 国家和部分省市大数据领域相关的政策概况	163
7.3 部分省市大数据发展的优先领域	165
7.4 部分省市大数据发展的主要政策措施	166
7.5 国内大数据产业发展的三种模式	171
第八章 上海大数据产业发展现状与相关政策分析	172
8.1 上海大数据领域创新发展概况	172
8.2 上海大数据产业链发展概况	173
8.3 上海大数据产业重点领域进展	174
8.4 上海主要区(县)大数据相关产业园区发展重点	177
8.5 上海与主要国家或地区的大数据相关政策对比分析	180
8.6 上海大数据产业发展的优劣势分析	190
第九章 上海大数据产业发展的思路与建议	192
9.1 大数据产业的特点	192
9.2 大数据产业发展的总体思路与政策框架	193
9.3 大数据产业发展的政策与措施建议	194
第十章 结论与建议	202
10.1 有关结论	202
10.2 下一步研究的建议	203
参考文献	205

图 目 录

图 1.1	本研究主要技术路线图	11
图 1.2	麦肯锡技术-经济研究框架	12
图 1.3	世界经济论坛数据生态系统分析框架	12
图 1.4	OECD 数据价值链研究框架	13
图 1.5	大数据技术创新生态系统示意图	15
图 1.6	大数据技术创新态势分析框架	15
图 2.1	1994~2014 年大数据文献总量分布趋势	22
图 2.2	1994~2013 年大数据研究热点演变趋势	23
图 2.3	大数据领域重点文献研究演变情况	24
图 2.4	大数据领域技术创新演化图	30
图 2.5	论文作者前 10 位国家分布	31
图 2.6	大数据文献作者国家地区及其技术领域分布	32
图 2.7	大数据文献数量前 25 位研究机构情况	33
图 2.8	大数据文献研究机构合作分布	34
图 2.9	中国科学院大数据文献合作发表情况	34
图 2.10	大数据企业分布图	37
图 2.11	1994~2014 年大数据相关文献数量前 25 位研究方向领域分布	41
图 2.12	大数据相关文献研究学科和重点领域分布	45
图 2.13	大数据领域 Mapreduce 技术引用文献研究热点分布	46
图 2.14	大数据相关专利群分布图	47
图 2.15	部分大数据企业专利分布图	48
图 2.16	大数据核心文献关键词分布图	54
图 3.1	大数据产业链基本架构	57
图 3.2	Bloomberg 大数据产业地图(2.0 版)	57
图 3.3	Dvae Feinleib 产业地图	61
图 3.4	新一代信息技术产业与大数据产业链关联示意图	65
图 3.5	节能环保产业与大数据产业链关联示意图	66
图 3.6	生物产业与大数据产业链关联示意图	68
图 3.7	高端装备制造业与大数据产业链关联示意图	69

图 3.8	新能源产业与大数据产业链关联示意图	70
图 3.9	部分企业专利分类布局	78
图 6.1	美国大数据文献技术领域分布	100
图 6.2	美国大数据相关政策体系	101
图 6.3	英格兰大数据文献技术领域分布	133
图 6.4	德国大数据文献技术领域分布	134
图 6.5	欧盟开放数据战略时间路径图	138
图 6.6	数据开放对于政府、企业、个人以及科学研究的利弊	140
图 6.7	日本大数据文献技术领域分布	143
图 6.8	日本活力 ICT 战略示意图	144
图 6.9	日本新 IT 国家战略五大智能信息通讯技术示意图	145
图 6.10	韩国大数据文献技术领域分布	146
图 6.11	新加坡典型的企业数据分析中心	149
图 7.1	1994~2013 年大数据文献总量和中美两国文献总量增长趋势	156
图 7.2	中美两国文献每年引文数分布趋势	157
图 7.3	中国大数据文献技术领域分布	159
图 7.4	中国科学院大数据文献合作发表情况	161
图 8.1	上海部分大数据相关产业园区分布图	177
图 8.2	广义大数据产业链与部分地区政策分布示意图	189
图 9.1	上海大数据产业发展的总体思路	194
图 9.2	上海“哑铃型+支撑垫”的政策框架体系	194

表 目 录

表 1.1	大数据技术与传统数据处理技术对照表	3
表 2.1	大数据产业领域部分公司典型产品或技术一览表	27
表 2.2	资助 5 篇以上文献的机构分布	35
表 2.3	部分新兴大数据企业列表	39
表 2.4	被引次数总量最高前 10 篇文献	42
表 2.5	大数据领域研究重点方向	49
表 2.6	被引次数总量最高前 30 篇文献	51
表 2.7	2012 年和 2013 年主要关键词表	53
表 3.1	大数据产业链各环节企业	58
表 3.2	大数据产业链各环节产品	62
表 3.3	部分典型企业专利申请概况	73
表 3.4	IBM 近年大数据领域主要产品	79
表 3.5	谷歌公司主要产品和服务	81
表 6.1	2014 年 NITRD 成员及参与部门及缩写	102
表 6.2	美国大数据政策文件汇总	103
表 6.3	2012~2014 年大数据研发高级领导小组说明	106
表 6.4	“大数据研发倡议”项目列表	110
表 6.5	“大数据研发倡议”中未涉及的“大数据跨联邦政府”项目清单	111
表 6.6	“从数据到知识到行动”项目列表	117
表 6.7	“从数据到知识到行动”新项目列表	118
表 6.8	2012~2014 年 BCC 提案征集资助项目数和基金额度	123
表 6.9	欧盟地区各主要国家和地区发表论文 10 篇以上排名	131
表 6.10	资助 5 篇以上文献的欧洲地区机构分布	132
表 6.11	欧盟数字价值链战略主要内容	136
表 6.12	欧洲地区部分国家 2003/98/EC 指令推进情况	139
表 6.13	日本部分大数据专利权人排名	143
表 6.14	新加坡大数据主要技术实施时间表	150
表 6.15	部分国家和地区主要政策措施一览表	152
表 7.1	中国部分研究机构论文数量及排名	158

表 7.2	资助 5 篇以上文献的中国机构分布	159
表 7.3	中美两国论文被引用数据对比	160
表 7.4	中国部分大数据专利权人排名	161
表 7.5	国家和部分省市大数据相关政策	164
表 7.6	部分省市大数据政策的优先领域	165
表 7.7	部分省市大数据政策的主要措施	168
表 8.1	上海部分研究机构论文数量及排名	172
表 8.2	上海大数据产业链企业分布	173
表 8.3	上海市金融信息服务业产业基地入驻项目	175
表 8.4	上海与部分国家和地区主要战略概况	181
表 8.5	上海与部分国家和地区重点发展领域概况	183
表 8.6	上海与部分国家和地区主要措施概况	185
表 9.1	上海大数据产业发展的政策建议	200

第一章 绪 论

1.1 大数据的研究背景

自 2008 年国际金融危机爆发以来,世界各国都在制定新的科技创新战略,以寻求走出危机的根本途径,并以此作为未来可持续发展之路。在新一轮的科技创新浪潮之下,以大数据、云计算和移动互联网等为代表的新兴技术的融合、渗透及其广泛应用,正在引领全球经济社会进行重大的转型,相关领域已成为全球科技创新和经济增长的前沿领域及战略性新兴产业。然而,正是由于它们属于新兴产业,其发展趋势更加难以把握,尤其是在技术创新方面,新的技术和模式频频出现,对于新兴产业的发展具有极大的导向作用。因而,把握住战略性新兴产业的发展趋势,显得尤为重要。

当前,大数据成为社会各界普遍关注的热点。如何实现在大数据驱动下的科技创新,将成为新一轮的热潮。然而,大数据应用范围广,发展速度快,因此,在大数据领域中出现了技术繁杂、创新群体众多、应用模式多样的特点。在这些特点的背后,大数据领域的技术创新究竟经历了怎样的发展历程?其主要驱动因素是什么?目前已经发展到哪一个阶段?还存在哪些前沿领域?像大数据这一类战略性新兴产业和产业究竟具有怎样的创新特性?这一系列的问题亟待进一步深入研究。

1.2 大数据的定义及相关概念

1.2.1 大数据的内涵

大数据(Big Data)一词来源于国外,2001 年即出现在《电脑人工智能》(PCAI)期刊的文章标题中。从字面上来看,大数据表示有很大的数据之意,即该词揭示了具有海量特征的数据。因此长期以来,大数据和海量数据(Mass Data)几乎是指同一事物。

2011 年上半年,国际数据公司(IDC)和麦肯锡全球研究院(MGI)等相继发布

报告,扩展了大数据一词的内涵,使之不仅包括海量(Volume)的特征,还具有快速(Velocity)、多样(Variety)、灵活(Vitality)、价值(Value)和复杂(Complexity)等特点。之后,国际商用机器公司(IBM)、微软公司(Microsoft)、甲骨文公司(Oracle)等一些全球知名IT企业也紧随其后,进一步扩展了大数据一词的内涵。

虽然在这些企业的报告中,并没有对大数据做出确切的定义,但是大数据的4V特征——海量(Volume)、快速(Velocity)、多样(Variety)和价值(Value)则已经成为共识。

海量 麦肯锡全球研究院认为,大数据是指容量大小超出常用软件处理能力的数据集合体^①。随着科学技术的发展和进步,被视为大数据的数据集合体也会同步增长,对它的定义并未对大数据兆级或千兆级的容量大小给出限定。此外,每个行业部门对大数据的定义也会有所不同,所以这里主要是基于特定行业所使用的数据处理分析软件和数据容量的大小。许多行业中的大数据容量从数十兆级字节到千兆级字节不等。由此可知,大数据的定义并不落在绝对的“大”字上,大数据中的“大”是一个相对的概念,它是相对于数据所在行业的处理能力而言的“大”。

快速 指数据生成的频率高。据中国互联网数据中心(DCCI)统计,每1秒钟,会有60张Instagram^②手机软件拍摄的照片被上传;每1分钟,会有60小时视频被传到Youtube上;每1天里,搜索引擎产生的日志数量是35TB^③。这些数据生成的速度之快,大大超过了传统信息的产生。

多样 指数据的类型,不仅包括传统的结构化数据,还包括图片、声音、视频等非结构化(或复杂结构化)数据。随着各种新媒体和新型传感器的诞生,这些非结构化数据还会不断出现新的格式。统计显示,全世界结构化数据的增长率大约为32%,而非结构化数据的增长率为63%;至2012年,非结构化数据占有比例将达到互联网整个数据量的75%以上^④,类似的估计还有一些,具体比率不一,但都认同非结构化数据在增长速度上远大于结构化数据。

价值 通常指数据利用所产生的商业价值或经济与社会效益。据麦肯锡全球研究院测算,在美国医疗行业,大数据每年可以创造3000亿美元的价值;在

① McKinsey Global Institute. Big data: The next frontier for innovation, competition, and productivity[EB/OL]. 2011[2013-09-16]. http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation.

② Instagram是一款运行在iOS和Android平台上的应用程序,以一种快速、美妙和有趣的方式将您的随时抓拍下的图片分享彼此。

③ 中新网. DCCI发布:大数据时代互联网营销的5个转变[EB/OL]. 2012-07-27[2014-11-03]. <http://finance.chinanews.com/it/2012/07-27/4063734.shtml>.

④ 中国计算机报. 非结构化数据来袭50%~75%数据来自人与人互动[EB/OL]. 2011-07-13[2014-11-03]. <http://tech.sina.com.cn/i/2011-07-13/11545774485.shtml>.

欧洲公共管理部门,大数据每年有相当于2 500亿欧元的价值(比希腊的GDP还高);利用全球个人位置数据,可以创造出6 000亿美元的消费者价值;大数据还能使零售商的经营利润增长60%^①。另外,甲骨文公司(Oracle)更进一步明确了大数据的价值非大数据本身所具有,恰恰相反,其认为原始的大数据往往具有价值密度低的特点^②。大数据需要经过处理,将数据转换成可用的形式或派生出的可用部分,再对其进行分析,才能创造出价值。

综上所述,尽管大数据的概念目前在学术界还没有统一的定义,但是人们普遍认为,它是指在特定行业中超出常规处理能力、实时生成、类型多样化的数据集合体,具有海量(Volume)、快速(Velocity)、多样(Variety)和价值(Value)几大特征。同时,与大数据(Big Data)有关的英文词语,还有Mass Data、Massive Data、Large Scale Data、Mega Data等。

1.2.2 大数据的外延

大数据已不仅仅局限于阐述大数据现象和大数据本身的特征,还会联系到与大数据相关的生态圈、技术以及应用领域。

(1) 大数据技术

大数据技术是指在大数据生态圈中开发利用大数据的各种方法和工具。因为在大数据的定义中指出,大数据已经超出常规的处理能力,所以大数据技术是区别于传统的对数据处理的技术。大数据技术与传统数据处理技术的对比如表1.1所示。

表 1.1 大数据技术与传统数据处理技术对照表

项目	大数据前沿技术	传统数据处理技术
采集	传感器、社交网络自动生成	人工输入
存储	虚拟化存储	磁盘
传输	无线移动	有线
处理	OLTP, OLAP	单机
分析	大规模数据挖掘算法	小规模数据统计、挖掘
解读	可视化	图表、文字

资料来源: IDC、麦肯锡等各大报告

① McKinsey Global Institute. Big data: The next frontier for innovation, competition, and productivity[EB/OL]. 2011 [2013 - 09 - 16]. http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation.

② Oracle: 企业大数据[R]. Oracle, 2012. 1.