

离散主义与后古典经济学丛书

大数据 ——

本质上是社会经济的离散化解构与全息化重构，
表现为行业间海量数据的跨界融合，
行业内海量数据的深度挖掘。

BIG DATA
ECONOMICS

徐晋 著

大数据 经济学



上海交通大学出版社
SHANGHAI JIAO TONG UNIVERSITY PRESS

**BIG DATA
ECONOMICS**

徐晋 著

大数据 经济学



上海交通大学出版社
SHANGHAI JIAO TONG UNIVERSITY PRESS

内容提要

本书是第一本关于大数据的经济学著作,指出大数据的本质就是社会经济的离散化解构与重构。本书提出离散主义,指出精神产品大生产与生命时间稀缺性,在此基础上讨论大数据经济学的基本范式,即离散化解构、全息化重构、价值关联。本书探讨了数据理性,论述了理性超越、边界可达与蚂蚁效应,研究了数据成本对企业边界的影响、数据资源的权属。本书最后讨论了大数据平台的构建模式,对人的行为异化等进行了分析,并对未来进行了展望。

本书适合经济学工作者、企业高管、政府公务员,以及从事大数据领域相关工作的人士阅读参考。

图书在版编目(CIP)数据

大数据经济学 /徐晋著. — 上海:上海交通大学出版社,2014
ISBN 978-7-313-11378-8

I. 大... II. 徐... III. 信息经济学 IV. F062.5

中国版本图书馆 CIP 数据核字(2014)第 103498 号



大数据经济学

著 者:徐 晋

出版发行:上海交通大学出版社

邮政编码:200030

出 版 人:韩建民

印 制:浙江云广印业股份有限公司

开 本:787mm×960mm 1/16

字 数:268 千字

版 次:2014 年 6 月第 1 版

书 号:ISBN 978-7-313-11378-8/F

定 价:49.00 元

地 址:上海市番禺路 951 号

电 话:021-64071208

经 销:全国新华书店

印 张:16.25

印 次:2014 年 6 月第 1 次印刷

版权所有 侵权必究

告读者:如发现本书有印装质量问题请与印刷厂质量科联系

联系电话:0573-86577317

大数据,广义上说是经济社会离散化解构之后的数字化表达,表现为海量数据的价值关联。

宏观而言,是行业之间海量数据的跨界融合;微观而言,是行业内部海量数据的深度挖掘。

——作者

前言 | 经济学变革的突破点

大数据是目前最红火的 IT 名词。很多学者和从事具体实践工作的人士，从各个不同的角度进行分析和研究。全世界每天产生超过 250 兆字节的数据，未来十年，全球数据量还将以每年 40% 的数据增长，而且到了 2020 年全球数据量将达到 35ZB。在大数据、云计算、移动、社交等全球趋势下，我们正进入一个全新的以数据为核心的大数据时代。

大数据已经成为一种新的经济资产类别，就像货币或黄金一样，大数据已经成为科技界和企业界关注的热点。一个国家拥有数据的规模和运用数据的能力将成为综合国力的重要组成部分，对数据的占有和控制也将成为国家间和企业间新的争夺焦点。

目前的研究视角，全部基于技术进步及其在具体行业或者案例中的应用。从技术视角看大数据，其优势在于能够非常直接地联系到具体应用，而且绚丽多姿的商业案例给我们无限的遐想空间。这个视角宣传和研究大数据，可以迅速普及大数据知识，提高政府和企业对大数据的敏感程度，并愿意为此加大投资和政策倾斜。

本书的研究，则是基于经济学的视角。从经济学视角看大数据，其优势在于能够洞悉人类社会在这次科技革命中的未来发展趋势——社会的离散化解构与重构。能够看见未来就能够确定方向，能够确定方向就能够制定战略。

从经济学视角研究大数据，能够让我们豁然开朗，很多之前想不到或者没有重视的经济学问题会迎刃而解。很多经济学家一直想找到经济学变革的突破点，这个突破点在哪里？这个突破点就在大数据所带来的社会解构。这个突破点带来的变革就是本书提出的经济学研究前提的重大变革：社会大生产重心转移到精神产品生产、生命时间成为稀缺资源。在这两个重大变革基础上，本书提出人们的理性从有限理性变为数据理性，并因此可以被超越与逼近；成本从过去的物质化成本，更多表现为数据成本；企业边界从过去的显性边界，变为隐性的、跨界的、离散边界；社会发展从过去的简单发展模型，变革为依托于

制度解构的离散发展模式；社会发生异化，个人行为、企业私权和政府行政，都在因为社会的解构与重构，而相应发生变革，但是法律和制度急需完善却又无从应付。

经济学的变革，须上升到形而上的高度，有自己的哲学基础。这个基础既要包括方法论基础，又必须包括价值观基础。本书提出新的方法论基础——离散主义。由于大数据本质上是信息科技带来的社会影响，而信息科技的重要数据基础就是离散数学。离散的观点与过去连续的观点，在思维模式上有本质的不同。基于离散主义的方法论，我们可以很容易认识到社会发展的基本形态，以及未来可能发生变革的主要趋势。本书认为人本主义可以作为新经济学的价值观基础。这种提法，主要因为当社会化大生产转变为数字化的精神产品大生产时，人们所支付的成本就不仅仅是购买产品的货币，更是使用产品所消费的时间。所以，生命时间稀缺是未来经济关心的核心，也是未来产品设计与销售的核心。未来经济的发展，必然回归到“人是万物的尺度”这个最终出发点上。

本书虽然是一本关于大数据的经济学著作，论述的是大数据经济的理论前提、哲学方法论基础、微观经济性质、社会经济变革。但本书主要目的，是基于以上突破点和变革，尝试提出经济学发展的新变革和新思想，归纳起来称为“后古典主义”经济学（具体可以参考“离散主义与后古典经济学”一文，《当代经济科学》，2014年第2期）。本书的很多观点，都是一种理论创新的尝试。考虑到体系的构建非一日之功，在很多良师益友的敦促下，本书完成得实在有点匆忙，书中值得商榷之处很多。如果体系能够成立，那必然是源于很多前辈、同仁的指点教导，是大家的智慧汇总。本书肯定会有错误、缺漏，包括引用不当之处，这些概由本人承担。愿听取诸位学者和实业界人士批评、指正、教诲，期待能够在今后修订版时更上一层楼。

徐 晋

2014年3月31日

目 录

第1章 导论	1
1.1 大数据现象与研究现状	1
1.1.1 大数据现象	1
1.1.2 研究现状	2
阅读材料:大数据改变世界——数字化的你永存于世	4
1.2 数据、信息与大数据的定义	5
1.2.1 数据与信息的定义	5
1.2.2 大数据的本质与定义	7
1.3 大数据的产生与特征	8
1.3.1 大数据的产生	8
1.3.2 大数据的特征	9
相关知识:数据的表达	11
1.4 大数据经济学的提出	11
1.4.1 大数据经济学的定义	11
1.4.2 本书的逻辑框架	13
阅读材料:如何避免失败——像 Google 一样掌控“大数据”	15
第2章 大数据经济带来的变革	17
2.1 信息变革:从信息不对称到数据不对称	17
2.1.1 信息不对称的提出	17
2.1.2 大数据下的信息不对称变革	18
2.2 成本变革:从传统成本到数据成本	20
2.2.1 成本与交易成本	20
2.2.2 数据成本的凸显	22
2.3 思想变革:个体精准性、局部全样本与整体离散化	23

2.3.1	统计、整体、个体与抽样	23
2.3.2	局部全样本与个体精准性	24
2.3.3	整体离散化	26
第3章	哲学方法论:离散主义	28
3.1	离散数学概论	28
3.2	离散数学与现代信息技术	30
3.3	离散主义思想与离散经济	32
3.3.1	离散主义思想	32
3.3.2	离散经济	34
3.4	基于离散主义的应用经济学	37
3.4.1	网络经济学	37
3.4.2	平台经济学	38
3.4.3	大数据经济学	39
	阅读材料:互联网思维与拓扑学	40
第4章	理论基础:精神产品生产与生命时间稀缺	43
4.1	物质资料大生产与工业革命	43
4.1.1	社会化大生产	43
4.1.2	三次工业革命	44
4.2	马洛斯需求层次理论与产品供给	46
4.3	社会化大生产的重心转移:精神产品与数字化生存	48
4.3.1	精神产品生产及其特点	48
4.3.2	数字化生存	50
	阅读材料:虚拟货币——比特币	52
4.4	传统经济学前提:外在物质资源稀缺	53
4.5	眼球、注意力及其经济价值	55
4.5.1	注意力经济与眼球经济的提出	55
4.5.2	注意力经济与互联网	56
	阅读材料:社交网络引爆“眼球经济”	58
4.6	资源稀缺性的重大变化:内在生命时间稀缺	59

4.6.1 人是万物的尺度	59
4.6.2 理论分析	60
阅读材料:《时间规划局》	61
第5章 基本范式:离散化解构、全息化重构与价值关联	63
5.1 经济的离散化解构与全息化重构	63
5.2 解构与数字化表达	65
5.2.1 供求关系与商品属性的数字化表达	65
5.2.2 社会商业空间的数字化表达	66
5.2.3 信息不对称的数字化表达	68
5.2.4 社会利益关系的网格化表达	69
5.3 全息化重构	70
5.3.1 重构的经济前提与技术前提	70
5.3.2 重构的基本原则与基本原理	74
5.3.3 全息化重构的表现形式	77
5.3.4 全息化重组具有三个重要维度	79
5.4 价值关联	81
5.4.1 数据关联定义与核心	82
5.4.2 价值关联模式	83
阅读材料:互联网的解构与重构	84
第6章 数据理性	89
6.1 理性与理性人	89
6.1.1 理性人假设	89
6.1.2 理性人假设的批判	91
阅读材料:诺贝尔奖得主西蒙与有限理性	94
6.2 数据理性:边界可达与理性超越	95
6.2.1 数据理性的定义	95
6.2.2 数据理性的特点之一:边界可达	97
6.2.3 数据理性的特点之二:理性超越	98
6.3 数据理性的量变:羊群效应	99

6.4	数据理性的质变:蚂蚁效应	101
	阅读材料:大数据将改变人类解决问题的方式	103
第7章	市场、数据成本与新企业边界	105
7.1	大数据经济下的市场	105
7.1.1	交易成本构成结构	105
7.1.2	价格发现手段和机制	107
7.1.3	消费行为与决策模式	109
7.2	市场结构变化	111
7.2.1	市场要素离散化	112
7.2.2	信息壁垒数据化	113
7.2.3	经济关系网格化	115
7.2.4	企业行为跨界化	116
7.2.5	交易模式平台化	118
7.3	数据成本与企业边界	119
7.3.1	数据成本与数据不对称	119
7.3.2	企业边界的含义及主导因素	120
7.3.3	企业边界变化的表现	123
	阅读材料:大数据发展十大趋势分析	125
第8章	大数据的权属	128
8.1	数据权属与数据垄断	128
8.1.1	数据权的定义与属性	128
8.1.2	数据主权	131
8.1.3	数据垄断	132
8.2	隐私与安全	134
8.2.1	个人数据与创富手段	134
8.2.2	数据隐私:不只是技术问题	137
8.2.3	个人隐私的保护:方式与建议	139
8.3	网络中立原则	141
8.3.1	定义与发起	141

8.3.2 当前的认知与具体方式	142
阅读材料:从 Snapchat 看社交网络未来——自毁型数据崛起 ...	143
第9章 大数据技术进展与应用	147
9.1 大数据技术框架与应用概述	147
9.1.1 大数据技术基本框架	147
9.1.2 大数据技术应用概况	149
阅读材料:大数据时代创新性突破的三大技术方向	151
9.2 数据源与数据存储	152
9.2.1 当前技术进展	152
9.2.2 具体技术应用	153
阅读材料:从数据角度看人类实现远程传送的可能性	154
9.3 数据计算与数据洞察	155
9.3.1 当前技术进展	155
9.3.2 具体技术应用	157
9.4 数据监管与用户接口	158
阅读材料:新一代信息技术勾勒未来“智慧城市”蓝图	159
第10章 大数据与平台	162
10.1 大数据 vs 平台:解构 vs 重构	162
10.1.1 一般意义上的平台与平台现象	162
10.1.2 平台外部性与多属行为	163
10.1.3 平台的分类	167
10.1.4 平台的业务模式	169
10.2 平台组织结构与大数据	171
10.2.1 平台组织结构	171
10.2.2 参与层:原始信息	173
10.2.3 规则层:数据结构	177
10.2.4 数据层:数据解构	178

第 11 章 大数据平台的商业模式	183
11.1 平台商业模式的载体结构:用户—规则—数据	183
11.1.1 平台商业模式的载体结构	183
11.1.2 用户—规则—数据的循环通道	184
11.1.3 平台商业模式载体结构的演化规律	186
11.1.4 平台商业模式载体结构演进的趋势:平台联盟与平台云	187
11.2 大数据下的平台模式解析	188
11.2.1 平台商业模式的盈利模型:市场发展的逻辑	188
11.2.2 要素流平台商业模式	190
阅读材料:从平台经济学角度解读 O2O	191
11.2.3 信息流平台商业模式	192
阅读材料:兴趣平台 Pinterest	194
11.2.4 规则流平台商业模式	195
阅读材料:制定评价规则的成功案例——大众点评	197
11.3 规则流平台专题:赌博平台	198
11.3.1 赌博平台的一般原理	198
11.3.2 赌博平台的一般平台商业模式	199
11.3.3 赌博平台的演化	201
第 12 章 社会异化与监管	204
12.1 异化的定义与机理	204
12.1.1 什么是异化	204
12.1.2 异化的机理及表现	205
阅读材料:缸中之脑与《黑客帝国》	206
12.2 行为异化	207
12.2.1 理性超越与经验超越	208
12.2.2 欲望泛滥与人格释放	209
12.2.3 网络游戏与行为异化	211
阅读材料:“大数据”蒙蔽下的《WOW》的萎靡史	213

12.3 私权异化	215
12.3.1 私权与表现形式	215
12.3.2 私权的扩张与公权的私有化	217
12.3.3 大数据公权的模式	218
12.3.4 公权私权混合的典型——棱镜计划作业	219
 第13章 大数据时代的机遇与展望	221
13.1 机遇:数字化解构的全面深化	221
13.2 三个战略方向:网络、数据、平台	224
13.2.1 网络战略	224
13.2.2 大数据战略	225
13.2.3 平台战略	227
阅读材料:美国政府的大数据计划(部分)	228
 参考文献	231
 索引	239

第1章 导 论

本章主要从经济学角度,给出了“大数据”的定义,并尝试提出大数据经济学。对大数据的认知,既可以基于一般的数据技术及其应用的视角,也可以基于现代经济与社会变革的视角。本章指出,大数据经济的本质,是社会经济的离散化解构与重构,而大数据则是一种经济现象——海量数据的价值关联。最后给出了全书的基本逻辑框架。

1.1 大数据现象与研究现状

1.1.1 大数据现象

全世界每天有超过 250 兆字节的数据产生,未来十年,全球数据量还将以每年 40% 的数据增长,而且到了 2020 年全球数据量将达到 35ZB($1\text{ZB}=10^{21}$ 字节)。通过互联网、物联网、移动终端、社交网络等形形色色的途径,人们将每时每刻面对海量数据。有分析表明,过去两年内产生的数据量,就占了有史以来的 90%。在大数据、云计算、移动、社交等全球趋势下,我们正进入一个全新的以数据为核心的大数据时代。

大数据这个术语最早期的引用可追溯到 apache org 的开源项目 Nutch。当时,大数据被用来描述为更新网络搜索索引需要同时进行批量处理或分析的大量数据集。2012 年,在瑞士达沃斯世界经济论坛上发布的《大数据 大影响》报告称,大数据已经成为一种新的经济资产类别,就像货币或黄金一样,大数据已经成为科技界和企业界关注的热点。一个国家拥有数据的规模和运用数据的能力将成为综合国力的重要组成部分,对数据的占有和控制也将成为国家间和企业间新的争夺焦点。

美国政府 2012 年 3 月在《大数据研究和发展倡议》中明确指出,将投资 2 亿

美元用于大数据相关领域的研究,重点包括国土安全、教育模式、科学和工程领域创新,从而使得“大数据研究”成为了国家意志。英国政府 2013 年 1 月宣布其投资于大数据领域的资金规模将达到 1.89 亿英镑,而涉及的领域包括对地观测、医疗卫生和节能计算技术等方面。日本政府于 2012 年 7 月提出了以电子政府、电子医疗、防灾等为中心制定新 ICT(信息通讯技术)战略,重点关注大数据的研究和应用。

下面是一组“大数据”现象的典型例子:

- 大型强子对撞机(LHC)在 2010 年总计产生了 13Petabytes 的数据;
- 沃尔玛超市每小时处理超过 1 百万个用户业务,这些数据被导入数据库后将包含 2.5Pdtabytes 的数据量,是美国国会图书馆数据量的 167 倍;
- Facebook 从其用户端处理 400 亿张图片;
- 全球公司的商业数据量,每 1.2 年翻一倍;
- 对人类基因进行解码,最初需要 10 年,而现在仅需要一周时间。

在企业层面,目前大多数公司都缺乏足够的分析能力从大数据当中获得有用的信息。2010 年,麻省理工斯隆管理学院与 IBM 进行了一项针对 3 000 名高管的调查,60%的受访者表示,他们拥有的数据远比他们能有效利用的多。

IBM 最近又发布了一项针对 1 700 多名市场总监(CMO)进行的调查,结果显示,71%的组织还没有准备好迎战大数据。在 IBM 实体分析首席科学家、杰出工程师 Jeff Jonas 看来,“计算速度越来越快,企业却越来越笨……今天很多企业能看懂 7%的企业数据,但这个数字很快会下降到 4%,然后继续螺旋式下降”。

随着谷歌 MapReduce 和 GoogleFile System (GFS)的发布,大数据不再仅用来描述大量的数据,还涵盖了处理数据的速度。从 Gartner 2 最新发布的 2012 年技术成熟度曲线报告中,可以清楚地看到大数据的技术应用的成熟度和价值度都出现了显著的变化,预计将会在未来的 2~5 年出现快速的扩张和普及应用,特别是以数据为中心、统计分析、内向扩展、自动化系统和工作负载管理将成为“第四范式”在当前时代的一种标志性展示。

1.1.2 研究现状

托夫勒早在 1980 年《第三次浪潮》中明确提出,“大数据则是第三次浪潮的华彩乐章”。信息革命、互联网革命、大数据革命的融合正在不断颠覆着人类思

维的模式和范式。

吉姆·格雷(Jim Gray)明确将大数据提升到科学研究范式高度,他描绘了数据密集型科研“第四范式”(the fourth paradigm)的愿景,其关键点在于其研究方式不同于基于数学模型的传统研究方式。

Kirkpatrick R. (2013)利用自己目前所从事的工作经验指出大数据在研发领域所能起到的关键性作用,并指出大数据是发生在 21 世纪的一场革命性变革。Madden S. (2012)分析了大数据发展趋势的缘起,并重点强调大数据和平台之间的关系,指出了由大数据所带来的新的挑战。

Begoli E., Horey J. (2012)重新定义了大数据现象,指出其分析海量数据集的一般性流程;并重点分析了大数据的架构和结构特征,最终带来新的机会、新的知识、新的收益。Bakshi K. (2012)以数据量的“大爆炸”为切入点,重点分析了其中增长最为迅速的非结构化数据,指出其在向外扩展,弹性和高可用性的技术特征。McAfee A., Brynjolfsson E. (2012)强调大数据比过去分析的分析工具要强大得多。因此,对于企业和组织而言,必须形成以数据科学家为核心的新的专业组织模式,并将其转化为有用的商业信息。

Borkar V. R., Carey M. J., Li C. (2012)通过对比,将企业传统所沿用的数据仓库和现在的大数据发展结合起来,指出了大数据的发展已经突破了企业发展的边界,变成了一种社会的公共资源。Boyd D., Crawford K. 认为大数据已经引起了各个科学领域专家的重视和研究兴趣。但是,大数据在公共服务领域、意识形态领域、社会关系领域还存在着诸多的疑问。因此重点从六个角度引起对大数据的思考:文化,科技和学术现象,依赖于技术的相互作用,分析和神话,激起广泛的乌托邦和反乌托邦的言辞。Brown B., Chui M., Manyika J. (2011)认为在大数据的发展过程中,必须积极地进行创新,形成个性化定制,尝试新的商业模式,才能真正适应大数据发展的要求。

研究大数据目前存在很多误区,最大的误区就是大数据对社会经济带来的影响,始终停留在技术层面进行讨论。对于“大数据”(Big data)研究机构 Gartner 给出了这样的定义。“大数据”是需要新处理模式才能具有更强的决策力、洞察发现力和流程优化能力的海量、高增长率和多样化的信息资产。

很多观点,仅仅是过去“海量数据”的相关产业观点或技术趋势的“大数据化”翻版而已。因此,当现在人人都在提及大数据时,基本上都仅在技术层面讨论,更多的是从计算机与信息技术的角度分析数据的特征、表达和处理技术,而

没有上升到经济学的高度。

我们说,蒸汽机改变人类历史,是因为它对生产力的释放。同样,我们研究大数据,也必须从社会化大生产的角度、从生产力的高度,定义大数据、分析大数据,而不是就事论事地研究数据挖掘技术。

阅读材料

大数据改变世界——数字化的你永存于世^①

未来一切皆为数据,而且数据还会有生命:只要保留了足够多的数据,完全可以模拟出一个数字化的自己,同样的音容笑貌,同样的思维习惯。

人类历史上,从来没有一个时代像今天这样与数据紧密相连,各种各样的智能终端设备使得数据生成无处不在。数据又是怎样改变我们的现实与未来的呢?

一、商业——个性化的交易

亚马逊总裁杰夫·贝索斯说过:“如果我的网站上有一百万个顾客,我就应该有一百万个商店。”现在的零售网站在挖掘顾客偏好的时候主要有两种方式:一种是基于用户,来判断顾客之间的相似性,比如当你在网上买了一本最新的小说,网站就会自动提醒你买这本小说的顾客中还有 65% 的人买了另外一本。借助“群体的智慧”,让顾客的购买行为来帮助完成“人以群分”。另一种方式是基于商品,通过判断商品之间的关联度来完成推荐,如当你购买一款剃须刀时,网站自然就会向你推荐一款对应的须后水,由此形成“物以类聚”。

不单零售业,大数据也已经开始影响很多传统行业,比如金融和保险,它们最大的挑战就是风险控制。百分点公司与一家银行合作,利用大数据的方式来筛选网上登记的信用卡申请资料。

二、医疗——认识自己

大数据在医疗保健领域的应用更加鼓舞人心。认识自己一直是人类不懈的追求,自从 1953 年詹姆斯·沃森博士和弗朗西斯·克里克博士提出脱氧核糖核酸(DNA)结构之后,基因测序成为人们梦寐以求揭开自身密码的钥匙。但是,基因分析需要超强的计算能力,DNA 分子由排列成约 30 亿个精密序列的

^① 魏一平. 大数据改变世界——数字化的你永存于世[J]. 三联生活周刊, 2013-12-03.