

后浪

大数据经济

大数据时代，互联网加法如何做？

谢文 著

大数据纳入中国国家行动方略，大数据产业起飞在即

互联网数朝元老、中国的“凯文·凯利”
百度深入解读大数据的大机会

陈彤

小米科技副总裁、新浪网前总编辑

胡舒立

财新传媒总编辑

王巍

中国金融博物馆理事长

鼎力推荐



北京联合出版公司
Beijing United Publishing Co., Ltd.

 北京联合出版公司
Beijing United Publishing Co., Ltd.

，数据

大数据经济 / 谢文著. —北京：北京联合出版公司，2015.10

ISBN 978-7-5502-6358-1

I. ①大… II. ①谢… III. ①信息经济—研究 IV. ①F062.5

中国版本图书馆CIP数据核字（2015）第236513号

Copyright © 2015 POST WAVE PUBLISHING CONSULTING (Beijing) Co., Ltd.

本书版权属于后浪出版咨询(北京)有限责任公司

大数据经济

著 者：谢 文

选题策划：后浪出版公司

出版统筹：吴兴元

特约编辑：高龙柱

责任编辑：张 萌

封面设计：李海超

营销推广：ONEBOOK

装帧制造：墨白空间

北京联合出版公司出版

（北京市西城区德外大街83号楼9层 100088）

北京京都六环印刷厂印刷 新华书店经销

字数120千字 720×1030毫米 1/16 14.5印张 插页2

2016年1月第1版 2016年1月第1次印刷

ISBN 978-7-5502-6358-1

定价：32.00元

后浪出版咨询(北京)有限责任公司常年法律顾问：北京大成律师事务所 周天晖 copyright@hinabook.com

未经许可，不得以任何方式复制或抄袭本书部分或全部内容

版权所有，侵权必究

本书若有质量问题，请与本公司图书销售中心联系调换。电话：010-64010019

目 录

① 一切都是数据，数据就是一切 1

关于大数据，你知道的都不对 7

数据的由来——从三千年前说起 14

数据的进化——从数据到大数据 20

大数据时代的下一个“倒霉蛋” 25

数据已成为生活必需品？ 29

② 大数据时代的游戏规则 33

得数据者得天下 39

扎克伯格定律 42

数据大≠大数据 45

大数据商业模式 48

产业链的颠覆和重组 51

数据服务业才是未来 54

大数据蓝海的历史使命 57

大数据时代的标准之争 60

创新的大数据商业模式 63

看得见的大数据未来 67

《大数据时代》解读 70

大数据变革前夜 73

③ 创造性毁灭：新时代的生存法则 81

创造性毁灭还是从0到1? 87

大数据战争的产业制高点 90

假如微博成为主流媒体 93

假如微博也是搜索 96

移动互联网未来的格局 99

即将引爆的电子商务革命 102

创新的企业机制 105

十亿美金的教训：技巧有余，眼光不足 108

互联网的溢出效应 111

百度是个什么样的公司 114

一软到底还是软硬兼施 117

谷歌转型的挑战 120

乔布斯真正的伟大之处 123

网络业内的资本狂潮 126

一个时代的终结 131

历史的转折点 134

④ 互联网的加法怎么做 137

互联网的加法怎么做 143

真假“互联网+”的辨识标准 146

互联网思维还是思维互联网? 149

攻防转换中的互联网金融 152

互联网金融还是大数据金融? 155

通用电气的工业互联网 158

新挑战与旧思维 162

真假产业革命 166

唯有创新能救夕阳产业 169

打车软件的道理与利益 172

迎接大数据制造业的到来 175

⑤ 互联网时代的中国特色 179

互联网时代的《仿造西洋火车无利多害折》 185

三条战线的创新之战 188

网络业的亚健康状态 197

电信业的体制之困 199

宽带战略的产业效益 203

微博是块试金石 206

虚拟世界雾霾重重 209

虚拟世界的隐私权博弈 212

失衡的数据 215

互联网的“瞻对”问题 218

无可奈何的反垄断 221

后 记 224

出版后记 225

1

一切都是数据，数据就是一切

2

3

4

5



大数据是什么？从哪里来的？有什么用？讨论大数据难免要围绕着这些问题展开。但至今，对这些基本问题还没有什么共识达成。当然，这也没有严重影响大数据的发展。在什么领域里阻力小、见效快、收益大，这些领域里的大数据创新就数量多、影响大。

微观派愿意就具体的技术问题讨论大数据，例如数据大了怎么存储最好，数据种类多了如何统一，数据增长快了怎么能比较省钱地应对。

中观派愿意讨论利用大数据的好处，例如如何通过数据挖掘提供产品和服务的营收，如何利用用户行为数据增加产品和服务的针对性。

宏观派愿意讨论大数据对社会的影响，尤其是负面影响，例如大数据如何威胁个人隐私、企业利益和国家安全。

三者一致的地方在于都认为大数据是个可以控制、可以利用、可以取舍的技术现象。对大数据，用也好，不用也好，用一部分舍弃一部分也好。总之，大数据只是一种工具。

但万一不仅仅如此呢？

如果世界上的万事万物都正在快速地被人类数据化,存储在计算机里,流动于互联网中,万物皆数,万物互联,那么就会形成一个与现实世界平行的数据世界。人类在数据世界里完成的社会行为在比例上不断增加,在内容上不断丰富,那么,该如何理解这种变化呢?

如果不断增多的人类制造物都开始采用数据化生产方式,新的原材料开始用数据化的方式生成,新的需求通过数据化的方式获得,商业销售和货币交换都以数据化的方式进行,那么,大数据还只是网络业少数人讨论的技术问题吗?

如果社会管理、国家安全、公共卫生和交通、教育和医疗都在迅速采用数据化的方式,那么,一个社会该以什么样的态度对待大数据现象,以什么样的姿态面对大数据浪潮的冲击?

如果社会交往、新闻资讯、文化娱乐、思想传播都更主要地通过数据化形式进行,那么,是抗拒或封杀这样的传播方式还是努力去适应这样的传播方式?

如果人们居住的房屋、驾驶的车辆、使用的器具、穿戴的衣物都变成数据化终端,那么,人类该怎样去适应这样的生活方式,该形成怎样的社会习俗?

仅仅把大数据看作一种技术现象、经济现象、社会现象或政治现象是不够的。从个人层面讲,轻视大数据就容易在各种生活选择中不知所措,诸如上学、就业、居家、社交都有可能陷入困境。从企业层面讲,看不到大数据的趋势,轻视各行各业走向数据化的趋势,以为

随便做做就算互联网+了，都可能是首先被淘汰被取代的企业。从社会层面讲，文化、道德、习俗如果比较保守，比较内向，比较排外，那就会在大数据浪潮面前不知所措，消极被动地应对危机。从国家层面讲，如果抓不住大数据的机遇，消极抵抗大数据浪潮，就会像许多处于农业社会的国家面对工业革命的冲击却错失良机，再想赶超就需要数百年的努力。

大数据是一种世界观，大数据是一种历史观，大数据是一种价值观，大数据是一种方法论。大数据其中的技术问题自有专业人士应对，商业问题自有企业家们操心，但由此引发的社会文化、道德、习俗的变迁，国家兴衰与全球范围的竞争，每个人都很难不去面对，不去思考，不去选择。

如果把人类走向信息化社会的努力分为三个阶段或三个时代的话，可以分为计算机时代、互联网时代和大数据时代，理由在于不同时代的驱动力不同。计算机解决的是数据计算问题，互联网解决的是数据传输问题，大数据则是在此基础上直奔主题，用数据化的方式解决人类生存发展的各种问题。

苏联解体后专家们统计，美国与苏联相比，20世纪80年代末计算机拥有量是25比1。当苏联还痴迷于原子弹、航母、导弹、坦克的数量时，却不知道或不理解人类已经开始走向信息化社会了，一个国家的国力已经不再仅仅以军事力量去衡量，而更主要的是以信息生产能力去衡量。

和苏联相仿，中国几乎完全错过了计算机时代，直到20世纪70年代末的改革开放才急起直追。今天虽然在计算机应用的深度和广度上还落后于发达国家，但基本上算是齐头并进了。在互联网兴起时，中国落后美国10年以上。但同样感谢改革开放的国策，从20世纪90年代中期开始奋起，今天也算是第二互联网大国。在2010年左右世界开始进入大数据时代，中国第一次有了和发达国家同时出发的历史机遇。但是，种种迹象表明，中国走向大数据时代的决心不那么大，步伐不那么快，基础性建设不那么多。如果蹉跎十来年，就又会与发达国家拉开整整一个时代的差距。

所以，认识大数据，思考大数据，努力大数据，就应该成为此时此刻的一个重要话题。

关于大数据，你知道的都不对

一个概念，无论它可以抽象到多么高深的程度，其形成、演变、推广的过程往往却很实在、具体，充斥着不同社会力量的博弈。这个概念的对错与否、生命力的短长、对社会的影响往往不取决于概念本身，而在于它的社会价值。

例如，PM2.5是一个衡量空气污染的指标概念，是描述客观存在的一种尺度。但是，这个概念在中国的落地生根却经过了两三年惊心动魄、迂回曲折的艰难历程。这个概念长期被拒绝在中国采用，理由是不合中国国情，不能反映出环境保护方面取得的伟大成就。然后，当亿万百姓可以通过网络获取这一指标的实时报道时，又被说成是外部势力居心叵测的挑衅。终于，现在PM2.5检测体系开始在全国逐步建立，大众的兴奋度却逐步降低，因为据说不经过二三十年，中国是很难达到联合国规定的空气优良标准的。

再例如，基尼系数是世界各国广泛用来衡量社会发展水平和社会不平等程度的一个客观指标概念。但是，这个曾经被中国学界广泛使用的概念近年来却无法获取权威的全国性数据，据说是因为基尼系数七八年前就达

到了0.45的水平，这被认为是一个社会贫富差距过大、继续增加就会导致社会动荡的临界点。一些学人体谅苦心，跳出来说普适性的基尼系数计算方式不适合中国国情，需要建立具有中国特色的基尼系数。因为城乡二元化，所以应该分别计算城乡基尼系数；因为沿海内地发展水平差异巨大，所以应该分别计算沿海地区和内陆地区的基尼系数；因为城市地区有户口的居民与新迁移进城的无户口居民之间生活水平差异巨大，所以应该分别计算正式居民与非正式居民的基尼系数，甚至干脆将非正式居民排除于统计体系之外。于是，社会贫富差距就成了一个只能泛泛而谈而无法实际度量的东西。

还例如，世界多数国家普遍采用的时区制，在中国从来就没有实行过，据说是怕影响国家统一。夏令时制曾经试行过两三年又被取消，据说是因为既麻烦又节电效果不明显。邮政编码先是被大力推行，然后又被取消，然后又被推行，理由先是推行成本过高，后是不推行成本过高。相反，有些概念的命运没有这么坎坷，一旦引进国门就大行其道。纳米是个只有极少数材料物理科学界专家才明白的概念，但今天在很多超市里却随处可见纳米除污剂、纳米装饰品，甚至还有什么纳米食品。

可见，一个新概念的出现，即使本身正确，孕育着推动科技进步和社会发展的巨大潜能，其真正实现也绝非易事，更可能的是由于社会环境的制约，被否定，被歪曲，被庸俗化。

今天，在中国相当范围内，开始流行一个全新的概念——大数据。我得承认，这个概念的流行中我自己起了一点作用。在大数据开始被讨论差不多一年的时候，它开始沦入其他新概念类似的命运：越来越说不清楚了。个中原因也不难解释：首先，大数据的概念是个舶来品，在发达国家尤其是美国逐渐成形、讨论和实践。在潮涌般的媒体报道、论文分析和专著论述中，大数据这个概念如同瞎子摸象一样，被不同视角、不同利益和不同

水平的人描述出来，让人难以琢磨。一个被严格定义并被广泛接受的大数据概念还没有出现，它更像是一个宽泛的现象描述，各种各样的东西都被装在里面。这样的好处是百家争鸣，共识会逐渐形成；坏处是鱼龙混杂，存在走歪走偏的可能性。其次，大数据是个发展中的事物，人们对其理解、阐释也在思想的碰撞、利益的竞争和技术的发展中不断深化，在概念层次和理论框架上说不清楚是必然的，和历史上许多新理念的形成过程并无二致。再者，关心大数据的主力军是网络业和IT业人士，他们目前面临着沉重的竞争压力和创新突围的激烈竞争，难免情不自禁地把自己的战略、产品、技术和服务装进大数据这个筐里，图存发展。

在维基百科网站上，大数据开始是这样被定义的：“大数据通常包括这样一些数据集，其体量超出了业内常用软件工具的能力，无法在可以容忍的时间内获取、把握、管理和处理。”这个定义显然是描述性的、单向思维的、自相矛盾的：如果大数据仅仅等于数据体量大，那么大数据与其他数据有什么本质区别？这种区别仅仅在于软件处理能力上吗？近年在处理体量相对庞大的数据方面，最流行常用的软件程序叫Hadoop，那么能够被它处理的数据算不算大数据呢？说不算吧，Hadoop被很多人称为第一个大数据软件；说算吧，它面对的数据并没有超出它的能力。可见，这种大数据概念一定来自每日和数据纠缠在一起的软件工程师和数据工程师们，是一种具体的、狭隘的、操作性的定义，经不起时间和逻辑的检验。

2001年，在全球IT咨询服务巨头Gartner工作的分析师Doug Laney写了一篇研究报告，第一次提出了一个三维模型，用以分析数据增长所带来的挑战和机会。这个三维模型的第一维是高速增长的数据体量（Volume），第二维是高速进出的数据运动（Velocity），第三维是高度异质的数据种类（Variety）。由这三维形成的空间里充满的就是我们今天称之为大数据的东西。Gartner采用了这个3V模型，从此成了广为人知，也被IT业界普遍接

受的大数据定义。2012年,作为Gartner资深专家的Laney又在一篇新分析报告中更新了自己的大数据概念定义:“大数据是体量庞大、高速变动和/或种类繁多的信息资产,需要采用全新的处理形式以有助于提高人们在决策形成、视野拓展和过程优化中的能力。”

这是一个很不错的定义。首先,明确了大数据是一种以信息形态存在的资产,具有3V特性;其次,盘活这种资产需要全新的处理形式;再者,这种资产增值创利主要体现在决策、视野和过程优化三方面。这个定义比前面提过的定义有进步,至少把大数据从狭窄的数据处理领域扩展到了整个IT业,也就是信息技术业。而且,由于信息技术已经普及到各个产业,应用于社会生活的方方面面,所以,大数据也就应该被社会各领域的人们所关注。

但是,我对这样一个大数据概念的定义仍然有些不满足,有些困惑,有些疑虑。这个定义仍然是描述性的,有些含混不清、难以把握。例如,“体量庞大”是什么意思?体量为一个MB的数据等于1024 KB,一个GB数据等于1024 MB,一个TB数据等于1024 GB,一个PB数据等于1024 TB……数据体量大到什么程度就算大数据了?“高速变动”是什么意思?1 GB/s还是1 TB/s?或者是数据体量每年翻番?“种类繁多”是什么意思?1000种不同数据?100种数据来源?10种数据格式?“全新的处理方式”是什么意思?今天的全新方式也许明天就过时了,明天的全新方式也许后天就落后了,怎样的处理方式才能在本质上算是大数据处理方式呢?另外,除了“有助于提高人们在决策形成、视野开拓和过程优化中的能力”,大数据就不能再干点别的什么事了?在这三种用途之外就不存在大数据现象和大数据生存空间了吗?尽管有大量的论文、书籍试图对此详加说明,但好像至今没有谁能说得很清楚。

可见,这样的定义仍然是技术性的、应用性的,隐约可见IT咨询服务