

郭昕 孟晔◎编著

THE POWER OF BIG DATA

# 大数据的力量

息 改 变 生 活      数 据 创 造 价 值

大数据的核心并不是拥有数据，而是拿数据去做了什么

| 宽带资本董事长 田溯宁 | 当当网联合总裁 李国庆 |

倾情推荐



机械工业出版社  
CHINA MACHINE PRESS

郭昕 孟晔◎编著

THE POWER OF BIG DATA

# 大数据的力量

信 息 改 变 生 活      数 据 创 造 价 值



机械工业出版社  
CHINA MACHINE PRESS

联系当前大数据的发展现状,对照微软、IBM、GE 等世界 500 强企业  
对大数据的应用,从应用大数据的行业、大数据人才争夺战、大数据应  
用关键技术、中国大数据的发展前景等角度,立体解读了大数据对我们工  
作、生活和学习方式的改进,以及大数据在企业转型过程中的应用。

《大数据的力量》专注于用案例来分析企业在大数据应用环节对数据  
的理解、收集、应用以及如何实现数据价值最大化,使读者能够亲身体  
验领军者如何把握机遇,同时又深入细节,捕捉领军者应用大数据的具  
体方式、方法。

## 图书在版编目 (CIP) 数据

大数据的力量/郭昕,孟晔编著. —北京:机械工业出版社,2013.6  
ISBN 978-7-111-43137-4

I. ①大… II. ①郭… ②孟… III. ①数据处理—研究 IV. ①TP274

中国版本图书馆 CIP 数据核字 (2013) 第 145857 号

机械工业出版社 (北京市百万庄大街 22 号 邮政编码 100037)

策划编辑:谢小梅 责任编辑:谢小梅 杨 勋

责任印制:杨 曦

北京双青印刷厂印刷

2013 年 7 月第 1 版·第 1 次印刷

170mm×242mm·13.25 印张·2 插页·135 千字

标准书号:ISBN 978-7-111-43137-4

定价:39.00 元

凡购本书,如有缺页、倒页、脱页,由本社发行部调换

电话服务

网络服务

社服务中心:(010) 88361066

教材网:<http://www.cmpedu.com>

销 售 一 部:(010) 68326294

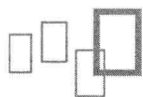
机工官网:<http://www.cmpbook.com>

销 售 二 部:(010) 88379649

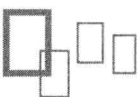
机工官博:<http://weibo.com/cmp1952>

读者购书热线:(010) 88379203

封面无防伪标均为盗版



## 推荐序一



大数据已成为一个流行词，有关大数据的书已出了几本，但是如何能进一步认识并应用它，仍需有一段路要走。我也担心过度泛滥使用这一词汇，盲目赶时髦，对产业发展反而不利。

近来在硅谷我看到一些企业已对大数据进行了有限的产业实践，我们将这些企业称之为有“反馈经济”的企业。它的特征是能把移动设备获知的各种数据实时传输到“云”中去，并通过大数据池进行比较、分析、计算，最终反馈到手机终端或其他设备上。这个闭环的形成对个人的行为乃至整个经济与社会有着丰富的意义。

例如，对于人的日常生活来讲，我们把一个人一生相关的生理和心理数据全部记录下来，大约需要 1000T 的数据量。这些数据通过“可佩戴计算”，忠实地记录着我们的衣、食、住、行及状态，并通过无所不在的移动网络，将大量的数据传输到“云”中去。“可佩戴计算”设备可以记录我们的生活，并对我们的习惯进行纠正。

就像互联网创造了搜索、电子商务、竞价排名等一系列商业模式一样，大数据会孕育更多新的公司类型，这也是大数据最有投资潜力的原

因。“数据库”就在那里，分析平台也已经成熟，现阶段是要用解决问题的视角，寻找数据分析师和懂得商业操作的人才，把数据分析产品化。

互联网时代的今天呼唤数据应用和服务。正如郭昕与孟晔编著的这本《大数据的力量》所言，我们正在感受大数据带来的迎面冲击，技术和需求的双重推动会让越来越多的政府机构、公司企业和个人意识到数据是巨大的经济资产，像货币或黄金一样，它将带来全新的创业方向、商业模式和投资机会。

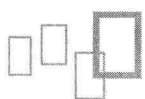
《大数据的力量》一书对大数据在国民经济各领域的应用，并对新的技术潮流及产业趋势进行了深刻剖析，为关注大数据的产业政策制定者、企业家及业界的学者提供了一本知识普及读物。

下一个阶段，产业界的机会和挑战是能否把大数据作为一个可以持续发展的商业模式；政策制定者可否很快制定出适合“数据时代”的数据所有权，数据公开与隐私的法规；教育和学术界能否尽快培养出大数据专家和学者。

只有通过商业、政府、学术几方面通力合作，我们才能再认识大数据，掌握新时代给我们的机遇，实现跨越式发展。

田溯宁

宽带资本董事长



## 推荐序二



大约两年前，我曾为帕科·昂德希尔的《顾客为什么购买》一书做序。在那本书中，作者向读者展示了实地调研的独特方法和惊人发现，呈现了消费者在购物时的行为特征，为零售业者提供了理解消费者的有效途径。

两年过去了，“移动互联网”、“云计算”、“大数据”技术接踵而来，新技术不再只是专家学者手中的玩物，也深深地影响着商业实践。我仍然坚信，商家如果不能紧随经济、技术、社会的发展而不断变化，就可能遭受惨败。

大数据技术和应用的不断成熟，为商家特别是电子商务企业，站在昂德希尔等学者肩上，更进一步地了解消费者，揣摩其心理、满足其需求提供了更有力的支撑。

当当网作为中国电子商务领域的领军企业，经过多年经营，积累了丰富的数据资源，并且一直在采取数据分析的方法指导企业决策和运营，可以说大数据在当当网是进行时。

电子商务经济是名副其实的“注意力经济”。没有客流的导入，转

化率就是空谈。我们基于数据分析和搜索引擎优化，获得了海量访问；通过对页面的优化，提升了转化率。

满足消费者个性化需求，体现在以数据为基础的计算上。根据消费者留下的点点痕迹，我们可以为其推荐最符合其需求的商品，不管是在购物过程中，还是通过营销邮件取得联系。我们也可以了解他从其他网站转来的原因，弥补他的缺憾。“大数据”营销已经跨越了传统的分组、分类、抽样的 CRM，直接利用海量数据，一对一营销成为了现实。

数据分析可以解决传统商业运营的弊端。长期以来，商业经常面临的难题是供需信息交流不畅，产品积压此起彼伏。当当网经营了上百万种商品，积压和断货现象却很少，就在于我们可以建立模型，通过对数据的处理，准确预测销售走势。

当当网是国内图书销售的龙头企业，也是数字化阅读的积极探索者。无论是电子书终端还是数字版图书，都在不断推陈出新。读者的注释、点评、分享等都在提供着纸质版图书无法提供的海量信息。基于此，当当网将有可能为内容的创造提供依据，引领国内数字出版产业的走向。

我们注重数据的积累。当当网从交易、运营、多种渠道中积累数据资源。在移动商务深入普及的今天，当当网的移动端交易获得了极快的增长，移动端服务水平也不断提高。甚至我们的内部运营，包括物流、仓储、支付等环节也都与移动通信、物联网应用等紧密集成。数据资源本身也可以产生新数据产品，如当当网可以在实时广告平台上大展拳

脚，将细分客户资源提供给不同的广告客户。当当网数据的积累和应用，已经形成了良好循环。

当然，关注“大数据”的人们（包括各类决策者在内）仍然对这一新生事物了解不多，或者浮于概念表面，或者对技术词汇的堆积心生厌倦。而郭昕、孟晔编著的这本《大数据的力量》，以丰富的案例、通俗的解读，为解决这一困惑提供了帮助，是一本很好的“大数据”知识普及本。

田溯宁先生曾向我描绘：将企业长时间积累的大数据集中存储并冶炼，也许可以挖出很多宝石来。对此我非常赞同。当然大数据在企业中的广泛运用也面临着许多挑战：如能源方面，大数据应用催生存储和计算的变革，大量电能的消耗难以避免；如技术创新方面，大数据的真正普及有赖于存储和计算成本的大幅下降，而这需要技术上的不断突破；如隐私方面，大数据收集和一对一营销需要用户许可，否则对隐私的肆意侵犯将令人不安，许可营销时代才是基业长青的。

我真心希望各界有识之士，与当当网一道充分利用“大数据”带来的历史机遇，在时代进步的浪潮中留下积极进取的痕迹。

李国庆

当当网创始人、联合总裁，中国书刊发行业协会副会长



# 前言

也许您刚刚访问了淘宝网，仍然回味着刚刚拍到的宝贝；也许你不经意打开新浪页面，突然发现它好像能猜透您的心思，不断推送您喜欢的商品。是不是有了些许当“上帝”的感觉？当然也不全是好消息，近日前美国特工爱德华·斯诺登披露了“棱镜”计划，让刚玩熟了智能手机和平板电脑的人们倒吸一口凉气，对于自己的隐私平添几分顾虑。眼下对大数据充满了追捧、探索、困惑与质疑，人们对大数据千差万别的态度恰恰反映了对这一事物的普遍关心。将大数据的理念与实践清楚地介绍给各位读者，是我们编著本书的愿望。

20年前互联网大师尼葛罗庞蒂提出了数字化生存概念。20年后的今天，信息技术真正将我们带入了以云计算、移动互联、社会化和大数据为新特征的“信息社会”。信息设备不断推陈出新，从大型机到笔记本电脑，再到移动智能终端，最近又出现了可穿戴设备，如智能眼镜、智能腕带等。这些设备“忠实”、“不知疲倦”地记录着我们的一举一动、一言一行，并把这些数据发送至“云”中。

大数据技术的出现，让我们认识人类的思想和行为的能力得到增

强。《爆发》作者艾伯特·巴拉巴西认为，93%的人类行为都可以通过数据分析予以预测。美国科学家最新研究发现，当不同的数据流被整合到大型数据库后，如把使用手机的时间、地点与使用信用卡或电子支付系统的数据相匹配，就能获得几乎全部的个人行为信息。《自然》杂志上发布的一项研究显示，只要有4条移动通话的时间和地点数据，就能有95%的几率确定通话者的身份。

IBM公司估计：由于智能手机、平板电脑、社交媒体网站、电子邮件和其他形式数字通信的广泛使用，每天全球会产生250亿亿字节的新数据。而且在全球现存数据中，90%是在过去两年中新产生的。国际数据公司（International Data Corporation，IDC）预计，一直到2020年，世界数据的规模将每两年翻一番。

数据爆炸性增长，数据分析能力也在快速进步。六度分隔理论、小世界理论、复杂系统理论的研究，Hadoop等开源软件以及内存计算等的发展，让我们能以前所未有的洞察力去关注、去发现新现象。如孩子是如何习得语言，防疫部门如何及时预测疫情的爆发，商家如何精准营销，企业如何通过软件去定义自身的边界、产品和任务，医生们如何在病人还没有走进医院的大门就已经拿到各种化验报告……

我们离不开大数据，这是历史进化、技术进步的必然结果。大数据时代是人类智慧和汗水创造出来的历史新篇章。我们有责任和义务趋利避害，让大数据改善生活、助力工作，将梦境化为现实。

这是一本大数据入门读本，我们试图通过鲜活的案例和创新的探

索，把大数据的力量展示出来。书中的应用案例有趣，又充满智慧，有益于打开我们的思路。这些内容尽管覆盖了我们工作生活的多个方面，但也仅仅是大数据应用的冰山一角。漫长而艰辛的大数据探索之旅依然在我们前方。

世界末日没来，数字生存时代来了。

# 机工经管读者俱乐部反馈卡

完整填写本反馈卡将可以参加幸运抽奖

每月我们将会抽出 10 位幸运读者，免费赠送当月新书一本

加入俱乐部，将会收到我们定期发送的新书信息

获奖名单将公布在 <http://www.Golden-book.com> 及  
<http://www.cmpbook.com> 上

## 个人资料

姓名：\_\_\_\_\_ 性别：☐男 ☐女 年龄：\_\_\_\_\_  
E-mail：\_\_\_\_\_ 联系电话：\_\_\_\_\_  
传真：\_\_\_\_\_ 手机：\_\_\_\_\_  
就单位及部门：\_\_\_\_\_ 职务：\_\_\_\_\_  
通讯地址：\_\_\_\_\_ 邮政编码：\_\_\_\_\_

## 单位情况

单位类型：

- ☐国有企业      ☐私营企业      ☐政府机构      ☐股份制企业  
☐外资企业（含合资）      ☐集体所有制企业  
☐其他（请写出）\_\_\_\_\_

单位所属行业：

- ☐食品/饮料/酿酒      ☐批发/零售/餐饮      ☐旅游/娱乐/饭店  
☐政府机构      ☐制造业      ☐公用事业  
☐金融/证券/保险      ☐农业      ☐多元化企业  
☐信息/互联网服务      ☐房地产/建筑业      ☐咨询业  
☐电子/通讯/邮电      ☐其他（请写出）\_\_\_\_\_

单位规模：

- ☐500 人以下      ☐500—1000 人      ☐1000—2000 人      ☐2000 人以上

## 关于书籍

1. 您购买的图书书名: \_\_\_\_\_ ISBN: \_\_\_\_\_
2. 您是通过何种渠道了解到本书的?  
☐报刊杂志 ☐电视台电台 ☐书店 ☐别人推荐 ☐其他\_\_\_\_\_
3. 您对本书的评价  
内容 ☐好 ☐一般 ☐较差  
编排 ☐易于阅读 ☐一般 ☐不好阅读  
封面 ☐好 ☐一般 ☐较差
4. 您在何处购买的本书  
☐书店 ☐网络 ☐机场 ☐超市 ☐其他\_\_\_\_\_
5. 您所关注的图书领域是:  
☐投资理财 ☐人力资源 ☐销售/营销 ☐财务会计  
☐管理学与实务 ☐其他\_\_\_\_\_
6. 您愿意以何种方式获得我们相关图书的信息?  
☐电子邮件 ☐传真 ☐书目 ☐试读本
7. 如果您希望我们发送新书信息给您公司的负责人, 请注明所推荐人的:  
姓名\_\_\_\_\_ 职务\_\_\_\_\_ 电话\_\_\_\_\_  
地址\_\_\_\_\_ 邮件\_\_\_\_\_

感谢合作! 请确认我们的联系方式

联系人: 董琛

地址: 北京市西城区百万庄大街 22 号机械工业出版社经管分社

邮编: 100037

电话: 010-88379081

传真: 010-68311604

电子邮箱: [cmpdong@163.com](mailto:cmpdong@163.com)

登记表电子版下载请登录:

<http://www.golden-book.com/clubcard.asp> 或 <http://www.golden-book.com>

敬请惠赐名片, 谢谢!

# C 目录。 CONTENTS

推荐序一

推荐序二

前言

第1章 破解大数据兴起之谜 ..... 1

1.1 大数据真的是大热门么? ..... 1

1.2 雾里看花, 大数据到底是什么? ..... 10

1.3 何种力量推动了大数据的兴起? ..... 18

第2章 感受大数据迎面冲击 ..... 23

2.1 拓宽科学研究视野 ..... 23

2.2 变革教育与人力资源管理观念 ..... 28

2.3 提升医疗与健康服务水准 ..... 33

2.4 改善能源利用效率 ..... 39

2.5 保障城市生活质量 ..... 42

2.6 助力国家事务处理 ..... 50

2.7 推动工农业转型升级 ..... 54

|            |                         |            |
|------------|-------------------------|------------|
| 2.8        | 促进金融业务创新 .....          | 62         |
| 2.9        | 加速运动产业信息化 .....         | 65         |
| 2.10       | 塑造零售业运营新模式 .....        | 69         |
| 2.11       | 释放电子商务与互联网应用潜力 .....    | 72         |
| <b>第3章</b> | <b>争夺大数据战略资源 .....</b>  | <b>81</b>  |
| 3.1        | 战略储备——商业巨头拥“数”自重 .....  | 81         |
| 3.2        | 捷足先登——欧美列强抢滩布局 .....    | 88         |
| <b>第4章</b> | <b>跟踪大数据领军者动向 .....</b> | <b>93</b>  |
| 4.1        | SAS——高性能数据分析研发的重镇 ..... | 93         |
| 4.2        | IBM——智慧一切的王者 .....      | 106        |
| 4.3        | SAP——企业资源管理的巨擘 .....    | 121        |
| 4.4        | 天云趋势——大数据开发的新锐力量 .....  | 133        |
| 4.5        | 维富友——淘宝双十一狂欢的背后英雄 ..... | 138        |
| <b>第5章</b> | <b>引爆大数据人才战争 .....</b>  | <b>147</b> |
| 5.1        | 大数据人才征集令 .....          | 147        |
| 5.2        | 呼唤创新的教育培训变革 .....       | 153        |
| 5.3        | 开发数据宝藏的创业精神 .....       | 158        |
| <b>第6章</b> | <b>揭秘大数据关键技术 .....</b>  | <b>167</b> |
| 6.1        | 智慧之道——数据挖掘技术 .....      | 167        |
| 6.2        | 高效之道——分布式计算技术 .....     | 169        |
| 6.3        | 实时之道——内存计算 .....        | 171        |

|                                         |            |
|-----------------------------------------|------------|
| 6.4 投向更广阔的技术视域——云计算、物联网、<br>移动计算 .....  | 173        |
| <b>第7章 描绘大数据发展前景 .....</b>              | <b>177</b> |
| 7.1 大数据时代的企业管理变革 .....                  | 177        |
| 7.2 大数据丰富了统计方法并对商业模式及经济效益<br>产生影响 ..... | 185        |
| 7.3 国内外大数据产业的主要动向及挑战 .....              | 188        |
| 7.4 我国发展大数据产业的对策建议 .....                | 191        |



## 第 1 章

---

# 破解大数据兴起之谜

### 1.1 大数据真的是大热门么？

可以给您一个肯定的回答，没错，“大数据”确实是近两年来最热门的 IT 词汇之一。从新闻媒体长篇累牍的报道（记者有了好的写作题材），到学术文章的数量剧增（专家学者也在跟进研究），再到图书市场的追捧（新的获利来源），都说明了这一点。国内与大数据相关的会议及论坛活动中，参与人员也不再局限于 IT 专业人士，包括政府官员、企业家甚至文化学者在内的社会各界人士均有参与。这个话题真的是太吸引人了。

对“大数据”以势如破竹之势改变世界的“魔力”，人们更多是从新闻报道中获知的。让我们也从新闻热点中感受一下全社会对“大数据”的关注与热情吧！

#### 地震预测呼唤大数据

突如其来的地震灾难，造成了严重的人员伤亡和财产损失，让民众