清华大学数据科学研究院
清华大学数据产业联合会
联合力荐

大数据

BIG DATA IN EVERYTHING

应用启示录

陈海滢　郭佳肃　主编

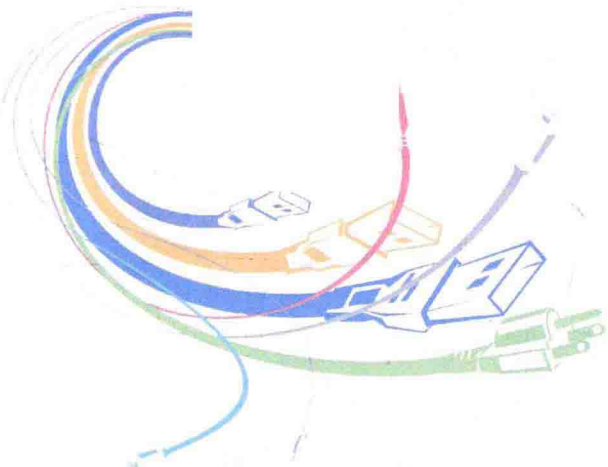
◎ 本书汇集了如此多的案例，而且从理论上做了一些探索。
◎ 案例覆盖领域之广可以使各方面的人士获得启示，理论探索可以推动人们的思考，值得一读。
◎ 此书属于开卷有益之类。

——联办财经研究院院长、国家税务总局前副局长 许善达

机械工业出版社
CHINA MACHINE PRESS

大数据应用启示录

主　编　陈海滢　郭佳肃
参　编　张　爽　陈利人　阮郑福　刘岩岩　马鸿飞
　　沈　洋　王小康　苏　波　任士勇　邵菁芃



机械工业出版社

本书在对大数据的概念特征及发展现状进行了梳理的基础上，列举了近20个不同行业中大数据应用的经典案例，以点带面，展现了各行业实战中体现出的大数据新思维，并以充足的实例为基础，提炼出大数据应用思维的新常态，给出了大数据应用的切实建议。为读者，特别是向位于政府或企业决策岗位的读者全面展现了大数据的应用行业图景。

本书适合企业管理者、信息化部门和营销部门人员、投资行业、金融行业的从业者、数据分析师以及其他有意了解大数据应用的各界读者阅读，同时对于电子、计算机、经济社会类专业的大学生、研究生及高校教师来说，本书也是良好的阅读材料。

图书在版编目（CIP）数据

大数据应用启示录/陈海滢，郭佳肃主编. —北京：
机械工业出版社，2016.9

ISBN 978-7-111-54534-7

Ⅰ.①大… Ⅱ.①陈… ②郭… Ⅲ.①信息学－研究 Ⅳ.①G201

中国版本图书馆CIP数据核字（2016）第190192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：吕 潇 责任编辑：吕 潇

责任校对：周文龙 樊钟英 封面设计：张 静

责任印制：李 洋

北京新华印刷有限公司印刷

2017年1月第1版第1次印刷

184mm×240mm · 21.5印张 · 426千字

0001—4000册

标准书号：ISBN 978-7-111-54534-7

定价：69.00元

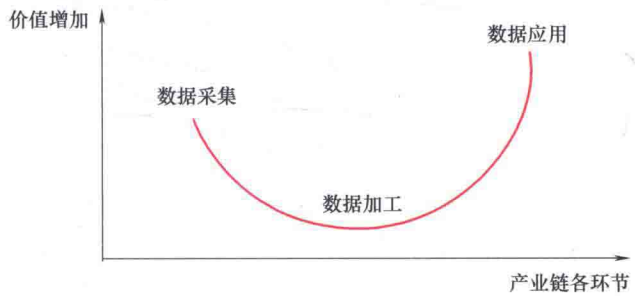
凡购本书，如有缺页、倒页、脱页，由本社发行部调换

电话服务	网络服务
服务咨询热线：010-88361066	机 工 官 网：www.cmpbook.com
读者购书热线：010-68326294	机 工 官 博：weibo.com/cmp1952
010-88379203	金 书 网：www.golden-book.com
封面无防伪标均为盗版	教育服务网：www.cmpedu.com

推荐序一

随着大数据在我国被确定成为国家战略，“大数据”成了时下的热词。很多场合，很多领域，很多人的口中言必称“大数据”的现象已屡见不鲜。但即便是在关心大数据的人群中，也存在着这样或那样的迷茫：听了不少大数据神奇的故事，可总无法与自己的工作联系在一起？在着力搞好大数据基础建设（IDC、网络带宽、WIFI 全城覆盖等）之后，一些地方政府和企业也有同样迷茫：下一步大数据产业走向何方？

大数据的生态圈或产业链大致可分为三种业务：数据采集、数据加工和数据应用。大数据公司可以是单一业务，也可以是三种业务的任意组合。借用台湾宏碁集团创办人施振荣先生 1992 年提出的“微笑曲线”，大数据的三种业务也符合“微笑曲线”的特征。



数据加工（特指数据的传输、存储和计算）部分当然也有很大的技术挑战，但其符合摩尔定律，性能不断提高，成本不断降低，这个环节的附加值相对较低；数据采集（特指传感器部署，可穿戴设备的普及，网络行为数据的采集等）部分的产业附加值相对高一些；右“嘴角”更加上扬说明最有价值的是数据应用（特指数据挖掘、BA、BI、AI 和机器学习等）部分。而且有了应用，可以让采集更有目的，加工更有意义。

由于我国 IT 基础相对薄弱，一些大数据基础设施建设工作最早进入人们的视线也是自然而然的，从各地争相建设大数据云计算中心即可窥见一斑。只是完成了这些基础设施的建设，仅仅是解决了“方便传”“能够存”“可以算”的问题，是产业链中附加值相对低的部

分，仍然没有解决“传什么”“存什么”“算什么”的核心问题。即便是全城 Wi-Fi 覆盖，摄像头密布，大量数据在实时产生，完成了数据的采集工作，也还是要回答：这些数据怎么用，有什么价值？这是我们迫切需要解决的问题。大数据产业的核心价值在于应用，大数据产业的出路也在于应用。

大家都为大数据产业发展寻找出路时，《大数据应用启示录》一书应运而生。通观本书内容，作者从厘清大数据的概念开始，为读者介绍了大数据发展的缘由、推动力、现状和前景。在此基础上，本书横跨各个行业，囊括了近百个国内外具有代表性的大数据行业应用，将“大数据”这个概念充分实体化了。这就犹如让大数据自己来发声，逐一“证明”它的特性和它所具备的巨大潜能，因此反而具有更强的真实性和说服力。

在国内，目前大数据应用的先行者多数是互联网公司和 IT 基础较好的传统企业，因为他们有数据，意识又相对超前。特别是一些直接面对消费者的行业，个性化服务，精准营销的应用已经开始，书中的国内案例多集中在这些领域。当然有些例子并非符合“大咖”们定义的“大数据”，但它们简单实用容易理解，并且有很好的可复制性，不失为数据应用的优秀案例。内容丰富的外国案例更展示了大数据的应用前景。多数国外的案例都是可以直接为我所用的，即便是一些不适合中国国情的案例，也可以为我们提供借鉴参考。书中的不少案例都是从想解决一个实际生活或工作中的小问题开始的，运用了大数据的思想方法，挖掘自己采集的数据或其他数据源，甚至是来自公开数据源的数据，得到了解决问题的新思路新手段。我们推动大数据的应用，确实不必非要从宏大的问题开始，着眼于身边的小问题，用数据把它解决了，就是最好的起步。只有让人们从数据的价值挖掘中尝到甜头，才会有越来越多的人参与其中，才会有越来越多的数据源参与其中。

人类在经历了科学发现的三个范式——实验科学、理论科学、计算科学——之后，在大数据时代迎来了科学发现的第四范式：数据科学。今天我们可以用来描述这个世界的数据维度越来越多，相信在浩瀚的数据间呈现出来的某些现象与关系可能超出人类现在的想象。这无疑为今后的科学探索和科学发现提供了新的方向。很多大数据的价值就是在对数据探究中，不经意发现的。本书中，除了在第十一章重点介绍了基础科学领域的大数据应用外，贯穿全书的多个案例也都是科学发现第四范式的佐证。

大数据的经典案例中，“啤酒与尿布”的例子常常会被提及。本书几个扩展阅读的精彩之处在于不迷信这些“经典中的经典”，提出了自己的观察。不论是对“啤酒与尿布”案例的思考，还是对“人体情绪热量图”的质疑，都是在帮助读者辨识大数据营销中的“善意谎言”。其实，虽然“啤酒与尿布”和“人体情绪热量图”的案例在是否符合科学规律，是否可以被反复应用、反复验证方面不禁推敲，极有以讹传讹的可能，但在提示和引领人们从

数据中发现相关性，在发现叠加新维度后产生新现象等方面，这些案例已经完成了它们的使命。《大数据应用启示录》中的扩展阅读部分对“经典案例”的辩疑告诉我们，大数据时代的确让我们和过去相比对这个世界多了解了很多，但是我们未知的世界依旧很大。对“经典案例”的辩疑还告诉我们，一个新相关性的发现如果不能被验证，不能被应用，它的意义会大打折扣。

在本书的结尾，作者提出了“大数据应用新思维”。篇幅不长，但却是这些实例中精炼得出的思维方式，正是本书最具精华之处。今天，大数据技术固然面临很多挑战，但观念的束缚和机制的制约仍然是中国大数据产业的相对缓慢发展的主要因素。也就是说，有相当一部分人不是“不会做”，而是“不想做”。此时，观念的更新与机制的突破就尤为重要。

作为国内第一部以大量应用事实为主体的大数据书籍，本书独有的实用性和趣味性也令人耳目一新。不仅适于普通读者拓展知识领域，也适合大数据技术的研究与实践者用于作为参考。

大数据可以应用的领域十分广阔，尽管本书内容已经十分丰富详实了，也仍然有很多领域尚未涉及。没有政府治理（反恐、打拐、整治套牌车也许可以算子领域的应用）方面的案例是本书的一个缺憾。希望在没有涵盖的那些领域中从事工作的读者能够触类旁通，从本书中获得启发，运用大数据的思想方法和工具手段于自己的领域中，创造出自己领域的成功应用案例。

从 1980 年著名未来学家 Alvin Toffler 在《第三次浪潮》提出“大数据——第三次浪潮的华彩乐章”之后，对大数据的定义和大数据的特征描述一直有众多的版本。用几个“V”能够完整定义大数据，抑或“全量数据替代抽样数据”、“只关心相关性不关心因果性”等是不是涵盖了大数据的特征，都不那么重要。重要的是：数据（大数据，还有不合乎大数据定义的那些数据）能不能被用起来，能不能对经济发展、社会进步以及人们生活产生越来越大的价值。

大数据的应用应该遵循如下“三部曲”：

一、不拘泥于大数据的概念，用好现有的数据；

二、有意识采集更多、价值密度更高的数据，以利再用；

三、放眼留意其他维度的数据，数据互通，使数据应用的价值进一步释放。

应用，应用，还是应用，重要的事情说三遍。我们关心、重视和推动大数据应用之日，必将是美丽的大数据之花绽放之时。

韩亦舜
清华大学数据科学研究院
执行副院长
2016 年底

推荐序二

“大数据”已经成为近年来备受关注的热词，越来越多的人逐渐意识到，大数据将是新一轮产业革命的新动力、新引擎。当各行各业中潜藏的数据陆续觉醒，行业的先行者能够借助大数据分析推动行业的发展时，新一轮产业革命的高潮便来临了。当今，已经有一批生活服务领域的先行者抓住了“信息时代”中计算机和互联网技术的巨大潜力，与计算机专业人才一同开发了包括电商网站、餐馆点评、网络约车等一系列成功的产品，成为生活服务领域新一轮产业革命的带头人。

作为生活服务领域重要的组成部分，餐饮行业大数据的应用目前正如火如荼展开。无论是前期的选址、装潢，中期的运营、备货，还是后期的口碑营销、客户聚拢，大数据的应用在每一个环节中都展现出它的巨大潜力。然而餐饮行业大数据的应用面临着巨大的挑战，获取餐饮行业的大数据具有几个难点：首先，是这个行业企业数量庞大但组织化程度很低，搜集数据十分困难；其次，是行业的非结构化数据占比较大，消费者消费行为留下的痕迹复杂多样，其表现出的评价和反馈包罗万象；第三，是线下的数据随意性较强，标准不一，采集、处理和分析更加困难。

面对这些困难，餐饮界人士需要借鉴各行业的先进经验，与信息技术领域专家一起共同努力，加快从数据向决策转变的进程，让大数据尽快成为行业发展与变革的新一代引擎。来自屏芯科技的郭佳肃先生，潜心创作的《大数据应用启示录》这本书，为我们借鉴其它行业的经验提供了大量案例。这本数十万字的作品，内容横跨多个领域，从商业领域专业的个性化消费行为大数据归纳，到金融领域牵动人心的股市行情预测；从制造行业的助力安全生产，到人工智能万众瞩目的围棋对战；从交通拥堵的大数据分析，到空气污染的大数据监控；从公共安全的大数据反恐，到体育比赛的大数据“制胜”……充分体现了“大数据”概念的应用范围以及外延的广泛性，对有志于大数据应用的人们获取跨界启示提供了难得的资料来源。

我相信无论是各行各业的从业者还是大数据技术的研究者，或者没有任何大数据技术

基础的读者，都能够从本书极为丰富的内容中找到自己的兴趣点，增加对大数据应用价值的了解与认知，获得大数据应用中那仿佛“灵机一动”般的启示。希望本书作为推进行业发展的重要大数据经典案例诠释，发挥它应有的作用，不仅在大数据技术领域激发研究人员的兴趣和动力，更能够得到各行各业“掌舵人”的欣赏，特别是能对餐饮行业如何搜集大数据，分析大数据，应用大数据有所帮助，为大数据理念的实体化、产业化，为新一轮产业革命作出贡献。

中国烹饪协会会长　姜俊贤

2016. 8. 3

前言

当面对宇宙中难以计数的群星时，人们会感到难以用语言描述的震撼，以及对人类力量有限的认知。同样，庞大的尺度或是巨大的数量往往会令人们惊叹。“大数据（big data）”这一平实的词汇，或许也表明了人们在意图描述这一现象级概念时的无力感。

数百万年以来，人类首次面对如此巨大而驳杂的数据之海，以致于无法在经验里找到相应的词汇，能够准确描述其中蕴藏的复杂力量。可见，理解“什么是大数据”也并不像想象中那样容易。仅凭“顾名思义”，人们很难立刻理解或把握大数据的内涵，而把它作为一个独特的概念，与其他传统数据处理区分开来。

国内外研究者们在大量的大数据相关著述中，条分缕析、抽丝剥茧，正是为了将这一具有广泛外延的概念解释清楚，让人们能够在理解大数据概念的基础上，发现大数据的巨大潜能，从而善用大数据技术推进行业发展或是帮助决策。

本书编著者的目的也并不例外。事实上，理解和运用大数据，需要比信息时代更为深刻的思维转型，这一过程不仅需要专精的大数据技术人才作为支撑，更需要各个领域专家的推动和引领。尤其是在一些传统领域，只有跳出长期运转的固定模式，才能以崭新的大数据思维来达到更高效、更安全、更透明的行业目标。而这一切都建立在对大数据概念与潜能充分理解的基础上。

本书特殊之处在于，我们在追寻大数据“为什么”富有价值的因果逻辑之前，用横跨15个行业和领域的百余个真实案例，为大家展示了大数据和它的应用价值究竟“是什么”。

这一构思源于作者在行业工作中多年的积累。作为将物理世界的数据进行采集、传输和应用的实践者，我们深刻感受到大数据在揭示客观规律、推动行业进步方面的潜能。尽管在不同的领域，海量数据的产生方式、类型、含义以及积累的厚度存在不同的侧重点，人们希望利用海量数据达到的目标也千差万别。然而，在考察大量实践案例之后，我们不得不说，大数据的魅力或许就在于，你可以尽情发挥想象力来设定你的目标，无论它有多么不可思议，大数据或许都可以用更为不可思议的方式来为你实现它。

在本书中我们可以看到，大数据已经在各个行业中发挥重要的作用。在这个大数据时代里，奇思妙想比以往任何时代都更容易成为现实。我们衷心希望，无论你是哪个领域的从业者，或是关注该领域的研究者，甚至是并没有大数据基础而仅抱有兴趣的读者，都可以从本书中找到具有启发性的章节。

在翔实的案例之外，编著者还针对大数据的应用价值、工具和思维方法做出了总结和建议，作为读者进一步运用大数据的参考。总体而言，本书不仅适于大数据企业以及政府机构等从事大数据应用、规划和推广的人士阅读，对各行业的大数据应用和潜在应用有着直接的借鉴意义，也可供大数据相关专业的教师和学生作为参考。

感谢清华大学数据科学研究院和清华大数据产业联合会给予本书的大力帮助和支持。

同时也对参与本书编校的刘丽娜、李世鹏、梁晓东、何汝星、李丹等人做出的工作表示衷心的感谢！

欢迎对本书内容及大数据产业抱有兴趣的各界朋友与我们共同探讨。

陈海滢　郭佳肃

2016 年 5 月

于北京恒升科创科技有限公司

目 录

CONTENTS

中篇　数据之力——行业实战

第四章　电子商务行业

第五章　传统零售行业

第六章　互联网领域

第七章　金融行业

商业

工业

科技

第十七章 日常生活领域

民生

第十八章 文化与体育领域

下篇 数据之思——价值、新思维与建议

第十九章 大数据的应用价值

上 篇

数据之观——概念、特征与发展