

“十二五”

国家重点图书出版规划项目

sas WILEY



新信息时代商业经济与管理译丛

大数据应用

商业案例实践

too BIG to IGNORE
THE BUSINESS CASE FOR BIG DATA

卡内基梅隆大学 康奈尔大学

【美】Phil Simon◎著

漆晨曦 张淑芳◎译

邓煜熙◎审校



人民邮电出版社
POSTS & TELECOM PRESS

“十二五”

国家重点图书出版规划项目

sas

WILEY



新信息时代商业经济与管理译丛

F713.51

17

大数据应用

商业案例实践

too BIG to IGNORE

THE BUSINESS CASE FOR BIG DATA

卡内基梅隆大学 康奈尔大学

【美】Phil Simon◎著

漆晨曦 张淑芳◎译

邓煜熙◎审校

人民邮电出版社

北京

图书在版编目(CIP)数据

大数据应用：商业案例实践 / (美) 西蒙
(Simon, P.) 著；漆晨曦，张淑芳译. — 北京：人民邮
电出版社，2014. 11

(新信息时代商业经济与管理译丛)

ISBN 978-7-115-36526-2

I. ①大… II. ①西… ②漆… ③张… III. ①商业信
息学 IV. ①F713.51

中国版本图书馆CIP数据核字(2014)第164274号

版权声明

Phil Simon.

Too Big to Ignore: The Business Case for Big Data.

Copyright © 2013 by John Wiley & Sons Ltd.

All rights reserved. This translation published under license.

Authorized translation from the English language edition published by Wiley Publishing, Inc..

本书中文简体字版由 John Wiley & Sons Ltd 公司授权人民邮电出版社出版，专有版权属于人民邮电出版社。

-
- ◆ 著 [美]Phil Simon
译 漆晨曦 张淑芳
审 校 邓煜熙
责任编辑 刘 洋
责任印制 杨林杰
- ◆ 人民邮电出版社出版发行 北京市丰台区成寿寺路 11 号
邮编 100164 电子邮件 315@ptpress.com.cn
网址 <http://www.ptpress.com.cn>
北京隆昌伟业印刷有限公司印刷
- ◆ 开本：700×1000 1/16
印张：15.25
字数：223 千字
印数：1—4 000 册
- 2014 年 11 月第 1 版
2014 年 11 月北京第 1 次印刷
- 著作权合同登记号 图字：01-2013-7455 号
-

定价：49.00 元

读者服务热线：(010)81055488 印装质量热线：(010)81055316

反盗版热线：(010)81055315

内容提要

大数据现已成为越来越重要的话题。很多商业案例已经表明，从正涌向我们的形式多样、快速变化的大量信息中可获取巨大价值。本书收集了令人信服的大数据应用案例。全书首先为理解大数据提供了必需的背景知识，然后对大数据的主要特征和相关具体技术进行探讨。要让大数据真正鲜活起来，新的工具不可或缺，本书介绍一些最普及的工具，并深入讲解已经成功实施这些工具的来自不同行业的一些企业。随后本书对于如何开始运用大数据提供了一些建议。伴随大数据，也存在着一些危险，我们将会了解到大数据的另一面。本书最后会提供一些关于大数据将走向哪里，以及在即将到来的时间里，为何它会落入寻常人家而不再成为奢侈品的一些预测。

本书读者群广泛，许多人群都可从本书获益，例如想要对大数据引起的大震动有所理解的 CEO、CIO 和管理者；想要将大数据的巨变与现有和潜在客户分享的咨询公司以及软件厂商的人员；想要指导学生如何进入大数据世界的高校信息技术和经济管理等专业的教授及学者；其他对大数据及大数据能干什么等感兴趣的人员。

作者简介

Phil Simon 是一个受欢迎的演讲者，著有 5 本书，包括备受赞誉的《平台时代》(The Age of the Platform)。当没有写作和演讲时，他向公司提供如何优化他们的技术应用的咨询服务。NBC、CNBC、《纽约时报》、《彭博商业周刊》、《赫芬顿邮报》、《环球邮报》、《Fast Company》，以及其他众多的主流媒体都对他的贡献做过专题报道。他取得卡内基梅隆大学和康奈尔大学的学位。你可以通过推特@philsimon 找到他，他的主页是 www.philsimon.com，居住在内华达州的亨德森。

当我们现在能够看到社会交互的动态及其从发展到结束的全过程，而不仅仅局限于计算推导诸如市场指标平均值之类的现状，说实话，这确实让我感到大为惊诧。为了能够看到市场变化的细微情况，发现政治改革的始发迹象，为了预测甚至能够控制它们，这绝对是一种普罗米修斯之火的大事件。大数据可用于好的或坏的方面，但无论哪种方式，它都给我们带来了一个有趣的时代，我们将彻底改写人类社会存在的意义。

——Sandy Pentland，美国麻省理工学院教授

知识是有用的。

——卡通片《动物屋》(Animal House)中Faber学院的格言

专家赞誉

“今天，大数据影响着每一个人，并将在可预见的未来继续影响着。在《大数据应用：商业案例实践》一书中，Phil Simon 使该话题变得易于理解且深入人心。这本重要的书为人们展示了如何运用大数据为组织效力。”

——William McKnight, McKnight 咨询公司董事长

“Simon 拥有一种将商业案例与复杂技术原理结合起来的不可思议的能力，最重要的是，他还能清晰解释这一切是如何结合在一起的。在本书中，Simon 揭开了大数据的神秘面纱。Simon 的观点有助于我们理解这种不断发展和流行的手段今天在如何影响我们的企业。”

——Dalton Cervo, 《主数据管理实务：实现真正的客户主数据管理》
(Master Data Management in Practice—Achieving True Customer MDM)

一书合著者, Data Gap 咨询公司董事长

“从 Twitter 所形成的照片流信息到 RFID 的交互信息,大数据领域正在迅速扩大,这为了解当前及放眼未来提供了前所未有的机遇。挖掘大数据潜力的同时避免其缺陷,并不需要魔法,需要的是一张地图。在《大数据应用：商业案例实践》一书中,Phil Simon 为企业进入下一个数据前沿,提供了一个全面的、富有洞察力的和令人愉悦的指南。”

——Chris Berdik, 《驾驭心灵：期望所带来的惊人力量》
(Mind over Mind: The Surprising Power of Expectations) 作者

“商业领袖正沉浸在大量的数据中,而大数据的浪潮才刚刚开始。在《大数据应用：商业案例实践》一书中, Simon 深入研究了大数据世界,并探讨了采集、组织、分析和可视化企业可获得的大量信息的商业案例。该书为企业提供了将数据转化成智能,再将智能转化成行动所需的优势。”

——Paul Roetzer, PR 20/20 公司的创建者和首席执行官,《营销机构蓝图》
(The Marketing Agency Blueprint) 作者

“Phil Simon 的《大数据应用：商业案例实践》一书,清晰地表明大数据日益增长的作用和价值。他的行文风格生动,案例研究清晰直观,将消除高层管理者对于当今经济环境中大数据如何驱动成功可能产生的任何疑问。”

——Adrian C. Ott, 《24 小时客户》(24-hour Customer) 获奖作者

致谢

向 Wiley 团队的 Tim Burgard, Shelly Sessoms, Karen Gill, Johnna VanHoose Dinse, Chris Gage, Stacey Rivera 致敬，是他们让本书能够如此快速付诸出版，对此，你们大家提供了“大”帮助。

感激智慧的小甜饼们，包括 Charlie Lougheed, Jim McKeown, Jason Crusan, Jag Duggal, Jim Kelly, Clinton Bonner, William McKnight, Scott Kahler 和 Seth Grimes，感谢他们无私付出的时间和专业知识，与这些家伙们交谈，充满学识乐趣。感谢 Hope Nicora, Andy Havens, Adrian Ott, Brad Feld, Chris Berdik, Terri Griffith, Jim Harris, Dalton Cervo, Jill Dyché, Todd Hamilton, Tony Fisher, Ellen French, Dick 与 Bonnie Denby, Kristen Eckstein, Bob Charette, Andrew Botwin, Thor 与 Keri Sandell, Clair Byrd, Jay 与 Heather Etchings, Karlana Kuder, Luke “Heisenberg” Fletcher, Michael, Penelope, 以及 Chloe DeAngelo, Shawn Graham, Chad Roberts, Sarah Terry, Jeff Lee, Mark Cenicola, Brenda Blakely, Colin Hickey, Bruce Webster, Alan Berkson, Michael West, John Spatola, Marc Paoella, Angela Bowman, 还有 Brian 与 Heather Morgan 及他们可爱的 3 个孩子。

接下来是一直伴随我身边的你们：卡内基梅隆的老友们——Scott Berkun, David Sandberg, Michael Viola, Joe Mirza 和 Chris McGee。

我心目中的英雄 Rush 乐队 (Geddy, Alex 和 Neil)、Dream Theater 乐队

(Jordan, John, Mike 和 James)、Marillion 乐队 (h, Steve, Ian, Mark 和 Pete), 还有 Porcupine Tree 乐队 (Steven, Colin, Gavin, John 和 Richard), 是他们的音乐激励了我这么多年的创作激情。请继续, 一直继续吧! Vince Gillian, Aaron Paul, Bryan Cranston, Dean Norris, Anna Gunn, Betsy Brandy, RJ Mitte 以及《绝命毒师》团队中其他的演员们, 是他们激励我要做出大事情。

接下来是我的父母双亲, 没有你们也就没有我。

中文版序

移动互联网时代，社交网络成为推动移动互联网迅猛发展的生力军。互联网花了 30 年时间才达到 7.5 亿用户；至 2012 年，成立于 2004 年的 Facebook 只花了 8 年时间便达到与之不相上下的用户数。

社交网络的核心价值，在于人和人的社交关系，马克·扎克伯格说：“人们分享得越多，他们就能够通过自己信赖的人，获得更多有关产品和服务的信息。他们能够更加轻松地找到最佳产品，并提高生活品质 and 效率。在这一过程中，企业获得的益处是：他们能够制造更好的产品，即以人为本的个性化产品。与传统商品相比，那些基于社交关系、社交图谱、社交圈推广的产品更富有吸引力。”可见，社交网络为人们开拓了新的信息分享和交流空间，也为企业创造了利用社交关系更开阔、更深入、更高效开展客户销售、服务和营销的机会。对于企业来说，谁更早抓住机会研究了解自身客户的社交网络关系，谁就更具核心市场竞争力。

根据博雅公关 Burson-Marsteller 和互联网监测分析公司 Visible 联合发布的“2012 年度《财富》100 强公司社交媒体使用报告”显示，2010 至 2012 年，100 家公司平均拥有 Twitter 账号分别为 4.2 个、5.8 个和 10.1 个，Facebook 账号分别为 2.1 个、4.2 个和 10.4 个，YouTube 账号分别为 1.6 个、2.7 个和 8.1 个。而根据 Allstate 公司 2012 年 6 月发布的美国公众使用社交网络的一份调查，最近一个月，2/3 的成人使用社交媒体，64% 的社交媒体用户说他们喜欢企业利用社交媒体作为客户服务的渠道，而

59%的用户说企业利用社交媒体让他们感到企业触手可及且值得信赖。可见，确实如制定企业社交媒体实践“黄金标准”，著述《营销和公共关系的新规则》一书的营销专家大卫·米尔曼·斯科特（David Meerman Scott）所言：“我们正在经历一场人们沟通方式的变革。我认为这是自印刷机发明以来人们沟通方式最显著的革命……社交媒体已经在革命性地改变商业沟通。”

我们知道，这是移动互联网时代，这是社交网络时代。而同时，人们的数字化生存，让有关人们生活甚至工作的行为信息都数字化，而这些以单个个体为对象的形形色色、包罗万象、细致入微、支撑个体兴趣需求和喜好的数字化信息构成大数据。所以，这个时代更是一个大数据时代。到今天，世界上所有印刷材料数据量是 200PB，全人类说过所有对话的数据量大约是 5EB，每天我们产生的数据大约是 2.5PB，这就意味着当今世界全部数据的 90%都由近两年产生。

如果我们有相应的 IT 技术、分析手段驾驭大数据，大数据是金矿；如果没有相应技术和手段，大数据则将成为淹没我们的海洋。谈论大数据在整个社会确实已成为一种时髦，但是根据麦肯锡在 2012 年 4 月的调查，仅有 1/5 的受访者所在公司已经在业务单元或职能部门完全部署大数据和分析，以获得客户洞察；仅有 13%的受访者表示，公司全面使用数据获得洞见。可见，大数据要从谈论和研究到技术和应用实现，路途或许还很漫长。所以，如何客观审慎地对待已有的大数据优势，提前思考并规划、架构、完善、部署数据从采集、清洗、存储、分析、应用以及管理监控的全企业层面的 BI（商业智能）平台，并培养贯穿企业运营管理流程的 BA（商业分析）体系，用数据说话，实现全企业层面的精确管理和精确营销、销售、服务，也就是大数据时代我们最终能够成为时代弄潮儿抑或被潮水淹没者的“To be or not to be”的关键问题。

中国电信股份有限公司广州研究院市场运营研究部的专业人员，长年从事企业运营管理及市场研究的实践和方法总结，研究时间最长的已达 16 年，并在行业竞争、商业模式创新、精确营销、品牌、舆情、口碑营销、数据分析及挖掘、数据仓库/BI 架构及规范等细分领域长年支撑企业运

营管理实践，不仅对企业运营有深刻理解和独到见解，且基于企业运营管理实践完成大量方法创新和应用研究，出版多本论著和发表数十篇专业论文，为各细分专业领域积累了众多的方法、经验和模型。

近几年，随着移动互联网—社交网络—大数据的迅猛发展，也因为企业转型的需要，市场运营研究部在邓煜熙部长的带领下，研究人员围绕如下两大问题进行深入探索：（1）企业如何建立自己的社交媒体策略并进行社交网络分析；（2）为实现精确管理、精确营销、销售和服务，企业如何架构 BI 平台和 BA 体系。部门集中有关资源有计划、有步骤、层层推进地开展研究，完成相关科研项目并撰写论文若干。

接下来，研究团队将围绕客户关系管理、客户体验管理搭建体系，以支撑企业生产运营管理流程各环节的运作，预计对企业大数据体系架构和分析、应用等方面进行深入研究。

最后，借用狄更斯的话：“这是最好的时代，也是最坏的时代；这是智慧的年代，也是愚蠢的年代；这是信仰的时期，也是怀疑的时期；这是光明的季节，也是黑暗的季节；这是希望的春天，也是失望的冬天；大伙儿面前应有尽有，大伙儿面前一无所有”。现在就让咱们大伙儿一起，掌握商业智能和商业分析两大工具，驾驭社交媒体，洞察社交网络，弄潮大数据。

中国电信股份有限公司广州研究院院长 蔡康

蔡康

2014 年 8 月于广州

前言

因数据不足导致的错误远远好过那些根本不用数据的错误。

——Charles Babbage（英国数学家）

2011 年 10 月 26 日的清晨大概 7:30，我驾车行驶在内华达州拉斯维加斯大道上。哦，不，我不是去掷骰子玩赌博也不是去看 Celine Dion。（她虽然是个天才，但她确实不是我的菜。）此行是出于一个更专业的原因。从 2011 年年中什么时候开始，我听到越来越多关于“大数据”的事情。在那年 10 月的清晨，我被邀请参加 IBM 的需求方信息（IOD）研讨会，这是学习大数据这个新事物的很好的机会，因为在计算机前学习的东西很有限。

除了对技术之类的知识孜孜不倦的渴求之外，老实说，我去 IOD 还有一大堆其他的原因。首先，很显然，拉斯维加斯大道离我家只有 15 分钟车程。其次，价格合理，我可以以免费的方式参会。因为自从我为一些用户众多的网站撰写文章之后，一些人就当我是媒体人。（这是 10 年前我始料不及的，当然我也不会对此评价吹毛求疵。）再次，这是个很好的宣传机会，我的第 4 本书《平台时代》（The Age of the Platform）正好刚出。我对图书出版业非常了解，所以知道如果想要一本书引领风潮并且销售良好，作者必须出门去做宣传。这就是当时所有正当理由，但还有一个额外的原因，那就是我能有幸亲自去到研讨会聆听两位主要演讲者的演讲，一位是 Michael Lewis（我最喜欢的作家之一），还有

一位是 Billy Beane。

就 Lewis 而言，他这次来 IOD 并非跟我一样来宣传他的新书。相反，他是去那里讲他那本 2003 年的书《点球成金：赢得不公平游戏的艺术》（Moneyball: The Art of Winning an Unfair Game）。有赖于最近一部由 Brad Pitt 担纲主演的同名电影，这本书竟然重新尽享商业成功之美妙光环。我这些年并没有阅读过《点球成金》，但是我仍记得当年快速读完的情景。Lewis 的写作风格如果不深入探究谈不上迷人。（他甚至在《大空头》（The Big Short）中体现出对次级抵押贷款和合成抵押债务的兴趣。）

我一直是位统计极客，《点球成金》一下触及了我的神经。它讲述的是奥克兰运动家、棒球队总经理 Beane 面对预算挑战的故事。虽然他的团队财务窘迫，但他们较其他三流球队赢得了更多比赛——甚至比肩实力和财力均十分雄厚的纽约扬基队。显然，问题是，他们如何做到的？Beane 颠覆传统，无视棒球长期形成的常规，过程中也因此常遭嘲笑。但是，Beane 是将其管理风格扎根于一个称为“Sabermetrics”（赛伯计量学：棒球记录统计分析）的鲜为人知的统计方法之上。他签约那些他认为被其他球队低估的自由球员，也即他努力从市场的低效率中挖掘被埋没的价值。

Beane 最得意的交易是一位非常便宜的球员却打出非常高的上垒率（OBP）*。概括来说，Beane 简单却不容置疑的逻辑可总结为：乐于上垒的球员更易上垒得分。长远来说，得分更高的球队比得分更低的手能赢得更多场的比赛。但 Beane 并没就此止步，他偏爱那些对于首个投球不挥棒的球员（当然，身价要合适），Beane 喜欢那些有耐心的、不断让对方投手沉醉于得分的击球手，当对方投手手倍感倦怠，再对每一个更好的投球机会予以反击。（再次强调，有成效的跑垒越多，取胜的概率越大。）

* OBP 代表击球手上垒率指标，（安打+四坏保送+触身球）/（打数+四坏保送+触身球+高飞牺牲打），上垒率代表一个球员能够上垒而不是出局的能力，是现代棒球最重要的数据之一。Beane 还将 OBP 跟 OPS 加在一起，评估球员的攻击能力，OPS 等于球手 OBP 加长打率（垒打数 / 打数）的总和。



图1 Michael Lewis 和 Billy Beane 与 Katty Kay 在 2011 年 IBM 需求方信息大会上¹ (来源: Todd Watson)

当时，基于比如那些被传统棒球循环赛视为异端的非正统的统计方法，对球员进行评估受到很多人的抵制，而这种抵制还不仅仅来自棒球界局外人，在 20 世纪 90 年代末至 21 世纪初，在 A's（奥克兰运动家）Beane 和他的最突出员工——Art Howe 队长——之间出现冲突。Howe 12 年间是 3 个球队的前内野手，因为 Beane 的（说得委婉些）所谓“非常规计划”而没能上场。正如 Lewis 在《点球成金》中所言，Howe 资格很老，他显然不乐意总经理告诉他一些新式的基于统计的有关这家或那家球队的棒球知识。

奥克兰内部冲突并没持续很久，所有运动中，总经理和队长必须站在同一立场，棒球也不例外。彻底解雇 Howe（以 150 万美元）队长职务之后，Beane 重获活力。他劝导纽约队对 Mets 放手，而非使用说服 Mets 使 Mets 主动加入自己球队的方式。这支球队很快签了新的队长，以 940 万美元的 4 年合同方式。而 Howe 这时已经获得球队 53% 的认同，因此他看起来更像是队长。Howe 态度强硬，他将会如何对待像 Mets 这么一个重金聘请的队长，可想而知。

Howe 对待 Mets 的方式比较卑鄙。在 Howe 的见招拆招中，球队只赢了 42% 的比赛。两个赛季后，Mets 意识到 Beane 早已认识的道理：Howe 与其强硬态度更适合理论而非实践。2004 年 9 月，Mets 与其队长分道扬镳。

Beane 可能是首位拥抱赛伯计量学的总经理，他也立刻赢得了球队的拥戴。他的成功理论不仅适用于棒球，也适用于棒球之外的世界。其中就有 Theo Epstein，目前是芝加哥俱乐部的棒球运营总裁，之前任波士顿红袜队（Red Sox）总经理职务时，他甚至聘用赛伯计量学教父 Bill James，而且确实很有作用。Epstein 为红袜队打破 86 年的僵局，赢了两场世界职业棒球大赛。休斯敦火箭队总经理 Daryl Morey 也将《点球成金》的理念引入 NBA，正如 2012 年 11 月《体育画报》（Sports Illustrated）一篇文章指出，MBA 毕业生对球员的获取和培养为他的同行带来了完全不同的方法论。²

这里还有个关于 Pulaski Academy（阿肯色州小石城的一所高中学校）足球总教练 Kevin Kelly 的经典案例。Kelly 并非一平常教练，他“2005 年阅读一篇有关第一局下来将主动权交与其他球队的统计学术研究报告之后，弃踢悬空球。”³Kelly 教练只是弃踢悬空球，而且一直如此。即使是在距离自方 10 码线 20 码处的第 4 次进攻。结果出乎意料的好，曾经一次在 Pulaski 得分后，Kelly 利用反向思维让其球员试着用非越位方式使球尽快被踢回。还有一场比赛，Kelly 的球队比对手领先 29 分，对手甚至没能有机会触碰足球。⁴结果呢？教练利用非常规思维模式，Bruins 球队赢得了多次州冠军赛。

那么为什么 Lewis 和 Beane 在 IOD 这类企业信息技术研讨会上会作为主要演讲者呢？因为正如《点球成金》所表现出的强有力的说服力一样，如今新型的数据资源正以非传统和创新性方式应用于各种不同领域并产生出令人侧目的结果——而越来越多的人、行业以及政企组织等都已经认识到这个情况。

本书要讲述的正是为什么大数据是了不起的大事。例如，马萨诸塞州波士顿市的居民主动通过智能手机报告路面坑洞及危险；先进保险公司对