

VOLUME/VARIETY/VELOCITY/VALUE

BIG DATA **活用大数据**

打造成功的商业和生活方式

[日] 高桥范光◎著

刘阳◎译



引导大数据活用成功并获取成果的
八个要点

 机械工业出版社
CHINA MACHINE PRESS

VOLUME/VARIETY/VELOCITY/VALUE

BIG DATA **活用大数据**

打造成功的商业和生活方式

[日] 高桥范光◎著

刘阳◎译



大数据并不是一时的流行，能否顺利活用大数据已成为商业成功与否的判断标准。即使如此，当前除了一部分先进企业外，大部分企业仍处于摸索如何活用大数据的阶段。本书以活用大数据成功和失败的各种实例为基础，详细介绍了在商业领域活用大数据的方法和具体实施步骤，以及活用大数据给生活方式带来的变化。

本书适合企业管理人员、中小企业主和自主创业者阅读，也可供从事大数据应用相关工作的人员参考。

DOGUTOSHITENO BIG DATA by Norimitstu Takahashi

Copyright © 2015 Norimitstu Takahashi

All rights reserved.

Original Japanese edition published by Nippon Jitsugyo Publishing Co., Ltd.

Simplified Chinese translation copyright © 2019 by China Machine Press

This Simplified Chinese edition published by arrangement with Nippon Jitsugyo

Publishing Co., Ltd., Tokyo, through HonnoKizuna, Inc., Tokyo, and Shinwon Agency Co.

Beijing Representative Office, Beijing

本书由日本实业出版社授权机械工业出版社在中国境内（不包括香港、澳门特别行政区及台湾地区）出版与发行。未经许可之出口，视为违反著作权法，将受法律之制裁。

北京市版权局著作权合同登记 图字：01-2017-0496 号。

图书在版编目（CIP）数据

活用大数据：打造成功的商业和生活方式 /（日）高桥范光著；
刘阳译. —北京：机械工业出版社，2019. 10

ISBN 978-7-111-63607-6

I. ①活… II. ①高… ②刘… III. ①数据处理
IV. ①TP274

中国版本图书馆 CIP 数据核字（2019）第 195240 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：王华庆

责任编辑：王华庆

责任校对：孙丽萍

封面设计：张 静

责任印制：孙 炜

保定市 中画美凯印刷有限公司印刷

2020 年 1 月第 1 版第 1 次印刷

148mm × 210mm · 6 印张 · 109 千字

标准书号：ISBN 978-7-111-63607-6

定价：29.80 元

电话服务

客服电话：010-88361066

010-88379833

010-68326294

封底无防伪标均为盗版

网络服务

机 工 官 网：www.cmpbook.com

机 工 官 博：weibo.com/cmp1952

金 书 网：www.golden-book.com

机工教育服务网：www.cmpedu.com

活用大数据

前 言

P r e f a c e

自“大数据”一词从2013年被人们热议以来，已有很多有关大数据的图书出版。

此前，大数据神奇得简直像法力无边的魔法一样受到万众瞩目，每天与大数据相关的新闻不绝于耳，人们对大数据的讨论和认知日趋理性。然而参与讨论的人少，隔岸观望的人多，人们对大数据的认知程度并不高，并且积极参与的企业也占少数。

为什么受到如此关注的大数据很难走进企业呢？

“数据分析太难了！”“不能保证一定成功，所以不费劲了。”社会上也许会有这样的声音。

的确，现在涉及大数据概念的事例以及各种统计分析方法和实践的书籍有很多，但是明确表明“数据分析很简单”“只要开始做就肯定会成功”的却很少。

事实上，大数据并不是法力无边的魔法，但它确实是一种非常有效的工具。掌握其使用方法需要下一番功夫，但熟练掌握后一定可以取得成果。

随着科技的进步，面向初学者以及专业人士的大数据种

类越来越多。

我此前一直从事大数据相关工作，有大数据使用培训经验。本书在这些经验的基础上，整理了活用大数据的相关知识以及在使用大数据的过程中常见的误区。

为了让更多的人对大数据感兴趣并参与其中，本书以大数据的活用为中心，没有复杂的数学公式介绍。

那么接下来让我们一起走进大数据的世界吧！

著者

目 录

前言

绪论 大数据从热潮到活用 / 001

- 0-1 利用大数据可以做什么? / 002
- 0-2 大数据带来的未来一片光明! ? / 004
- 0-3 不是做不做, 而是如何做 / 008

第1章 大数据的特点 / 011

- 1-1 为什么大数据如此受关注? / 012
- 1-2 大数据的三个“V” / 018
- 1-3 作为工具的大数据: 第四个“V” / 021
- 1-4 大数据所处的社会环境 / 026
- 1-5 熟练掌握工具的“数据科学家” / 029

第2章 大数据活用要有目标 / 033

- 2-1 大数据活用无法顺利进行的三个类型 / 034
- 2-2 没有目标的大数据工程不会顺利进行 / 037
- 2-3 从《点球成金》学习大数据活用要点 / 038
- 2-4 商业活动中大数据活用的四种模式 / 041
- 2-5 利用大数据开展“新商业” / 048

2-6 “营业额增长”是大数据活用的第一步 /052

2-7 大数据有助于品质提升和降低风险 /062

第3章 大数据活用的推进方法 /067

3-1 理解大数据活用的两个企划 /068

3-2 大数据活用计划企划的步骤 /071

3-3 大数据稳定化企划的步骤 /081

3-4 小行动引导大成果 /085

第4章 引导大数据活用成功的八个要点 /087

4-1 大数据是工具，不是目的 /088

深层次讲解 大数据、IoT 和工业 4.0 /091

4-2 意愿会引导数据产生启示 /101

4-3 思维定式易产生误解 /105

4-4 组合多个数据以正确把握实情 /108

4-5 仅进行一次数据验证无法得到结果 /111

访谈1 数据假设验证的摸索尝试理所应当 /115

4-6 大数据之神藏于细节 /124

4-7 不要光分析，行动才会产生结果 /126

访谈2 分析结果（output）与商业

成果（outcome）是两回事 /128

活用大数据

打造成功的商业和生活方式

4-8 听取一线声音，转动一线，持续使用 /137

访谈 3 将一线想获取的数据可视化 /141



第5章 活用大数据，我们的生活将如何变化 /149

5-1 大数据活用的未来是什么样的？ /150

5-2 大数据活用在和谐社会开花 /156

访谈 4 用于表现手法的大数据 /163

后序 /178

参考文献 /180

绪 论

大数据从热潮到活用

近年“大数据”一词频繁出现在报纸及电视新闻中。它被称为法力无边的魔法，活用实例也逐渐变多。本部分主要介绍大数据给我们带来的改变以及目前我们该思考的东西。

◆ 0-1

利用大数据可以做什么？

◆ 根据大数据，我们可事先预测“××将会发生”

如果我们事先知道“××将会发生”，那我们的日常生活将会发生什么变化呢？

例如天气，如果能够准确掌握今后 24 小时内全国的天气变化，我们的日常生活将会变得非常方便。

例如，“14:00 之后 40 分钟内会有阵雨，因不是上下班时段，可不必带伞”“明天 18:00 会有 5 厘米降雪，需准备长靴和防寒用具”“明天会有大雷阵雨，需防护农作物”等。据此，在工作和生活中，我们可事先采取预防措施，天灾带来的灾害也可以减少。

如果事先知道接下来 24 小时内铁路、公路、飞机、船等交通设施的状况，又将会怎样呢？

“那条高速从早上会持续 5 小时拥堵，到达目的地会是

15:00。”“明天山手线搭乘人数会有××万人，高峰是8:00~10:00和18:00~20:00，在此期间将调整运行间隔，从涉谷站到东京站将花28分钟。”

我们如果可事先知道上述信息，就可调整出勤时间和外出时间，另外换乘也会很顺畅。你如果是销售人员，就不会错过重要贸易谈判。你如果在物流企业上班，就会合理安排配送日程。

不仅商业如此，个人生活也同样受到影响。比如，在饮食方面，可以根据自己的身体状况选择最适合的店铺和菜单，以预防肥胖和避免不好的生活习惯。

还有，可根据体温、血压和脉搏数等充满活力的数据提供合适的睡眠和运动建议，我们的生活将会更健康。

在打折贱卖时大量购买，但最终压箱底的衣服也可能会减少。如果能根据手头上衣服的形状、颜色、品牌、种类等进行最合适的搭配，我们就不会大量购买没用的衣服。

另外，路过商场附近时，如果手机可接收合适的打折或打折券信息，我们就可以在控制支出之余还能享受满意度较高的消费生活。

提高购买效率，在日常生活的食材购买中将发挥更大的作用。

手机软件会配合超市特价信息提供便宜又美味的菜谱，这样就不会浪费冰箱里的食材了。另外，配合标明健康状况的活力指数，可推荐合适的家庭菜谱进而预防不良的生活习惯。

然而，现实生活中“通勤时间段内下大暴雨，电车延迟，上班迟到”“没想到遇到交通拥堵，延误了开会”等时有发生。

如果仅仅发生那还好，而对于物流、批发零售等受天气影响较大的企业，预估稍有偏差就会产生较大损失。

另外，对于从事农业、水产、林业等第一产业的人来说，如果能准确知道天气变化，营业额和成果将会天差地别。

此处列举的“事先知道××将会发生”的一系列事件，将它们变为可能的就是大数据。

所谓大数据活用就是指活用世界上存在的各种数据，并将它转化为现实的技术。进一步讲，活用数据的意识以及现象都包含在内。

0-2

大数据带来的未来一片光明！？

◆ 因经常被定位而“心情不爽”

大数据使我们的生活变得更加便利，这一点毋庸置疑。但是，因为获取用户数据而经常被定位导致的心情不爽，以及对“被数据统治”的厌恶感也确实存在。

前面说的“心情不爽”的例子之一就是被称为“追踪型

（定位型）广告”的标签广告。

所谓追踪型（定位型）广告是指通过将用户曾浏览过的商品店铺的标签广告显示在网络浏览器上，再根据网络浏览器的网上信息模块（将用户信息暂存在计算机中或参照相关机能而得出的信息）识别浏览用户的广告。

最近，随着数据分析技术的进步，追踪定位已可跨越设备（计算机内部装置或周边机器），例如通过台式计算机上网络浏览器看的鞋广告可出现在平板计算机的网络浏览器上，这样的事情现已并不陌生。

虽然广告信息网中可识别浏览记录，但并不能锁定住址和姓名，不过对“被追踪”这一事情觉得心情不爽的人并不少。

进一步说，对于自己的真实行动踪迹被掌握，不知不觉数据被人出卖，应该会有很多人喊“No”吧。即使是为了个人不会被锁定而进行加工的安全数据，也可能产生“真的没有被锁定吗？”等疑问，一旦怀疑，后续问题就会源源不断。

即使大量数据中个人并不会被锁定，但是如果与自己相关的数据被卖到与自己无关的地方，对此感到不愉快的人也应该存在吧。

◆ 因被协调统治而感到厌恶

因被协调统治而感到厌恶的例子较为复杂。

例如，有这样一种软件，分析个人推特（Twitter）状态的

倾向，关注人所关注的网络新闻事件倾向，根据倾向向个人推送感兴趣的新闻。这的确很方便，可马上知道自己感兴趣领域的新闻，对很多方面都有帮助。

但是“想读这条新闻（曾经想读这条新闻）”，做出此决定的是计算机，还是人类本身呢？

甄选“适合自己的物或事”的基准是根据以往的生活和生活环境形成的。计算机推测的基准是“你想看的新闻是这个”“你适合的衣服是这件”。被计算机抢先提示，可能会产生“自己被计算机协调统治”的感觉。

这种“人类的意识行动被计算机、政府或别的东西控制”的状态，以乔治奥威尔的《1984年》为首，曾在各种科幻作品中被谈到，人类的尊严受到了被帕斯卡称为“会思考的芦苇”的挑战。

◆ 计算机将超过人类智能——“2045年问题”

如果再敷衍，警钟将会敲响。这就是被称为“2045年问题”的未来预测，即根据“计算机芯片的性能一年半时间翻倍”的摩尔定律导出“2045年计算机将拥有凌驾于人类智能之上的人工智能”的预测。

计算机芯片相当于人类的大脑。根据摩尔定律，它的性能以指数倍速度更新发展，总有一天计算机中将会产生具有优秀信息分析能力的人工智能。

单纯的芯片性能增加并不能产生人工智能，但是计算机芯片相当于人脑，分析庞大数据信息的人工智能技术、自发的“思考”性能迅速提高，以及“机器学习”技术发展很显著。

那么，会产生什么后果呢？人工智能集结所有智能创造出更优秀的智能，然后再开发新智能……如此循环，人类智能将赶不上计算机智能的发展速度。

超越人类智能的世界是何种世界，人类的想象已力所不能及。这种不可预测未来到来的瞬间被称为“技术特异点（奇点）”，这指的就是2045年。

2045年可能发生的现象如下：

- 计算机给人类造成巨大威胁。
- 计算机离人类而去。
- 大脑通上电极，在计算机描绘出的宇宙（世界）中生活。

也有可能可能会出现科幻电影中出现的情节。

◆ 计算机不仅仅是带来，同时也会夺去

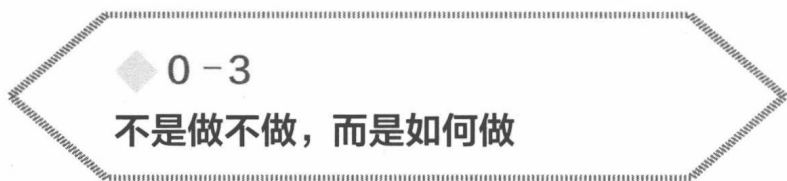
是否真的会像上述那样发展，现在谁都不能得到结论。可以肯定的一点是，许多被认为“只有人才能做的工作”在不断被计算机夺去。

例如客服咨询，通过可分模型和数据积累，计算机可根据过去的模型提出最合适的问题对策。

还有像商业企划案等“只有人才能做的工作”也有可能被计算机夺去。

现实中有制定市场策略方案的系统，将来计算机有可能提出“根据分析结果，对于20~34岁的女性来说，这样的活动最有效果”的提案。

大数据技术不断发展在给我们的生活带来便利的同时，也有可能将我们带入一个完全未知的世界。



◆ 如何跨越大数据的旋涡

像科幻作品所描绘的人类未来、“自己的数据被人利用”产生的不愉快，以及生活、商业、产业课题得以解决的便利性等，都与大数据有关。

是否应该加入大数据，选择权在每个个体及企业，但是现实中并不存在“不参加”这个选项。

究其原因，则是因为大数据并不是一家企业参与就可完结的，它是“不断被卷入”的过程。

将本企业保存的数据，社交媒体数据、评论网站的投稿、