

BIG DATA

致胜大数据时代的 50种思维方法

如何用大数据思维改变现状，预知未来，成功获利

杨倩◎编著

简单高效的实践方法论，改变亿万人生轨迹的案例总结与透析。
迄今为止最直给的大数据圣经，新思维致胜新时代有这一本就够了！

红旗出版社



BIG DATA

致胜大数据时代的 50种思维方法

如何用大数据思维改变现状，预知未来，成功获利

杨倩◎编著

红旗出版社

图书在版编目 (CIP) 数据

致胜大数据时代的 50 种思维方法 / 杨倩编著 .

-- 北京 : 红旗出版社 , 2014.11

ISBN 978-7-5051-3310-5

I . ①致… II . ①杨… III . ①思维方法—研究 IV . ① B80

中国版本图书馆 CIP 数据核字 (2014) 第 273838 号

书 名: 致胜大数据时代的 50 种思维方法

编 著 者: 杨 倩

出 品 人: 高海浩

责任校对: 张浩敏

总 监 制: 徐永新

封面设计: 博雅工坊·肖杰

责任编辑: 周 强

版式设计: 博雅工坊·肖杰 程海林

出版发行: 红旗出版社

地 址: 北京市沙滩北街 2 号

邮 编: 100727

编 辑 部: 010-82061212

E-mail: hongqi1608@126.com

发 行 部: 010-64024637

欢迎品牌图书项目合作

项目电话: 010-84026619

印 刷: 北京九天志诚印刷有限公司

开 本: 710 毫米 × 1000 毫米

1/16

字 数: 150 千字

印 张: 14.75

版 次: 2015 年 2 月北京第 1 版

2015 年 2 月北京第 1 次印刷

ISBN 978-7-5051-3310-5

定 价: 36.00 元

版权所有 翻印必究 印装有误 负责调换

前言

什么是大数据

最早提出“大数据”时代来临的是全球知名的咨询公司麦肯锡。麦肯锡公司称：“数据已经渗透到每一个行业和业务领域，成为重要的组成部分之一。人们对于海量数据的挖掘和运用，预示着新一波生产率增长和消费者盈余浪潮的到来。”

2008年9月，《自然》杂志推出了封面专栏——“大数据”，内容讲述了数据在生物、物理、工程、数学及社会经济等多方面学科所占据的位置和角色的重要性。

如今，“大数据”这个词汇俨然成了工商界和金融界的新宠。在哈佛大学担任社会学教授的加里·金说：“这是一场革命，庞大的数据资源使得各个领域开始了量化进程。无论学术界、商界还是政府，所有领域都将开始这种进程。”

说到这里，还是有人对“大数据”不明所以：到底什么是大数据？

“大数据”是从英语“Big Data”一词翻译而来的。“大数据”这一概念在近几年逐渐被人们所熟知，并为全球各大企业所重视。简单来说，“大数据”是一种巨量资料库，可以在合理时间内达到撷取、管理、处理并整理为帮助公司、企业经营和决策的资讯信息。

随着云时代的悄然到来，“大数据”渐渐得到了越来越多的企业关注。后来，业界将“大数据”概括成四个V，即大量化（Volume）、多样化（Variety）、快速化（Velocity）和价值化（Value）。

大数据不仅是量大而已

“大数据”的量到底有多大？根据2012年3月的一份调查结果显示：在

短短的一天之内，互联网产生的资料内容可以刻满 1.68 亿张 DVD；发出的社区帖子高达 200 万个，相当于《时代》杂志 770 年的文字数量；发出的邮件高达 2940 亿封，相当于美国两年的纸质信件数量；卖出的手机为 37.8 万台，高于全世界每天出生的婴儿数量 37.1 万……

截止到 2012 年，数据量已经从 TB 级别跃升到 PB、EB 乃至 ZB 级别。（ $1024\text{GB}=1\text{TB}$ ， $1024\text{TB}=1\text{PB}$ ， $1024\text{PB}=1\text{EB}$ ， $1024\text{EB}=1\text{ZB}$ ）。国际数据公司（IDC）经过详细的调查研究，得出一个结论：2008 年，全球产生的数据量为 0.49ZB；2009 年的数据量为 0.8ZB；2010 年的数据量为 1.2ZB；2011 年的数据量更是持续增长，竟高达 1.82ZB。这个数据量，相当于全世界的每个人产生 200GB 以上的数据。可见，大数据的信息量有多大。

IBM 公司称，截止到 2013 年 10 月，全世界所获得的数据中，有 90% 都是过去两年内产生的。预计到 2020 年时，全世界所产生的数据规模将达到今天的 44 倍。

当然了，“大数据”不仅是量大而已，它还具有多样化、快速化、价值化等优势。

多样化：数据的类型繁多。这种特质也让数据被分为两部分——结构化数据和非结构化数据。相对于以往那些以文本为主的结构化数据，非结构化数据越来越多，包括日志、图片、音频、视频、地理位置信息等。

快速化：处理的速度快。这是大数据区别于传统数据挖掘的最明显的特征。根据 IDC 的一份名为“数字宇宙”的报告，预计到 2020 年，全世界的数据使用量将高达 35.2ZB。在如此浩瀚的数据面前，处理数据的效率快慢决定了企业生命的长短。

价值化：价值密度低。价值密度的高低与数据总量的大小成反比。

我们来举个例子：一部时长为一小时的视频，在持续不间断的监控中，有用的数据仅仅有那么一两秒而已。因此，如何通过强大的计算方式迅速地完成数据的价值“精纯度”，已成为目前“大数据”背景下需要解决的难题。

大数据——潜藏的金库

随着技术的革新，我们已经踏进大数据时代，而数据背后潜藏着巨大的商业机会，值得我们去挖掘。

根据技术研究机构 IDC 的研究结果可知，近年来，大量的新数据无孔不入，它们以每年 50% 的速度在增长。或者说，它们每两年就要翻一番，完全超出人们的预料。

事实上，我们生活的方方面面，都会因大数据的存在而发生变化。如消费习惯、兴趣爱好、人际关系，以及整个互联网的走向与潮流等，都将成为 IT 行业所关注的重点。当然了，这一切的获取和分析都与大数据息息相关。

我们不能说数据的圈子越来越大，而是全新的圈子越来越多。比如，全世界有数不清的数字传感器依附在汽车、工业设备、电表和板条箱上，它们能准确地掌握方位、温度、湿度、运动、振动，以及大气中的化学变化。

从一方面来说，大众媒体基础上的大数据挖掘和分析，将衍生出令人意想不到的应用；从另一方面来说，基于数据分析的营销和咨询服务也正在崛起。这些专注于数据挖掘和数据服务的公司，将成为 IT 行业乃至互联网服务业中的新兴力量。

以往，只有像谷歌、微软这样的全球化公司能做关于大数据的深挖和分析。但现在，大数据偏向平民化，让越来越多的 IT 公司有机会进入这个领域。也因此，大数据领域有了不同的数据分析和服 务，促使人们不断地创新商业模式。比如在一分钟内，用户就会在 Facebook（脸谱网）上发布近 70 万条信息；在一分钟内，用户会在 Flickr（雅虎旗下图片分享网站）上传 3125 张照片；在一分钟内，用户就会在 YouTube（世界上最大的视频网站）上点击 200 万次观赏……

铁一般的事实告诉互联网从业人员，这些庞大数字意味着一种全新的致富手段。可以说，它的价值不可估量。

虽然在目前来说，大数据在中国还处于初级阶段，但是它的商业价值已

经告诉人们——凡是掌握大数据的公司，就相当于站在“金库的门口”。基于数据交易产生的经济效益和创新商业模式的诞生，能帮助企业进行内部数据挖掘，以便更准确地找到潜在客户，从而降低营销成本，提高企业的销售利润。

百分点信息科技的联合创始人苏萌曾说过：“未来，数据可能成为最大的交易商品。但数据量大并不能算是大数据，大数据的特征是数据量大、数据种类多、非标准化数据的价值最大化。因此，大数据的价值是通过数据共享、交叉复用后获取的最大的数据价值。”在他看来，未来，大数据将会如基础设施一样，有数据提供方、管理者、监管者，数据的交叉复用将大数据变成一大产业。

据一项统计结果显示：截止到2012年10月，大数据所形成的市场规模在51亿美元左右。到了2017年，此数据预计会上涨到530亿美元。

由此，可见“大数据”的价值所在。

改变思维模式，用数据说话

近几年，“大数据”变成了香饽饽，成为各大企业、公司、媒体甚至是学者都津津乐道的东西。他们有着不同的见解和理论，唯一相同的观点就是——大数据时代对人类有着至关重要的影响，甚至即将成为改变未来社会的重要力量。

然而，大数据究竟改变了什么呢？在人们的脑中，这个问题仍然是一个乱糟糟的毛线团，想要找到毛线的一头，却又不知道从哪里入手。

在“第五届中国云计算大会”的第二天，中国电子学会云计算专家委员会候任主任委员、中国科学院院士怀进鹏发表了一篇演讲，题为《大数据及大数据的科学与技术问题》。在演讲中，他表示：“大数据的发展可能会改变经济和社会生活，可能会改变科学研究的途径，甚而改变人类的思维方式。”

我们可以来举一个简单的例子：

苹果公司创始人史蒂夫·乔布斯在罹患癌症后，不惜花费巨额抽取了全身的 DNA 信息。这样一来，医生就可以根据他全身的 DNA 信息来判断癌细胞的变化，从而对症下药。然而，对于普通的患者来说，是拿不出巨额钱款来做这种抽取检测的。换句话说，医生只能从普通患者身上提取很小的一部分 DNA 信息。两者的治疗效果有多大差异，可想而知。

对此，或许你还有疑问，大数据真的能获得最全面的信息，能够找到信息的源头和结果吗？

我们再来举个简单的例子：

如果你上淘宝，登录支付宝账户，点开电子对账单，你是不是能够看到自己一年的消费曲线图？是不是能够看到每个月的支出和收入？是不是能够看到自己的钱花到哪里去了？是不是比拿笔记账清晰和准确得多？根据网络购物的数据中，你还会发现：哪个城市的男人比女人购买的东西还要多；哪个城市的人用支付宝缴纳水电费的频率最高；什么星座的男性或女性在某年的消费额最高；在一年里，在父母或亲友身上花费了多少；节假日时，什么东西最畅销……

这样庞大的数据分析，在过去单一的小数据时代根本无法做到，甚至想都想不到。大数据专家维克托·迈尔-舍恩伯格在《大数据时代》中是这样解释的：云计算在获取海量数据的同时，也带来了数据的混杂性，这会给传统的数据分析带来一些困扰。在以往，我们习惯于由数据得出具体结果。而在大数据时代，我们应当关注的是数据之间的相关关系，而不是数据之间的因果关系。

数据之间的相关关系可以帮助我们捕捉现在的线索和预测未来。如 A 和 B 的情况经常一起发生，那么只要注意到 B 情况发生，就能预测 A 的情况是怎么样的。这种“A 和 B”的关系在零售行业和 IT 行业中已被广泛运用。7-11 便利店通过分析零售终端的数据，得出了这样的一个相关关系——温

度低于 15 摄氏度，暖宝宝的销售量便增加 5%。于是，只要温度低于这一度数，7-11 便利店内的暖宝宝就会上架；豆瓣电台会推荐一些你可能会喜欢的音乐；当你在当当网买某本书后，系统就会提醒你——购买这本书的人中，有 30% 也购买了另外一本书……

这些结论或预测，都是基于大数据分析而来的。当然了，大数据也为我们带来了另一种生活方式，那就是还有更多的事物都可以数据化。通过对关键词的分析和搜索，我们可以看到购物的习惯被数据化、人际关系被数据化、社会热点和考试重点的走向也被数据化。这些数据可以导出商业潜能，更能导出社会走向。

阿里巴巴创始人马云敏锐地捕捉到大数据的巨大潜能。在 2012 年，他提出大数据战略，通过资源共享与数据互通创造商业价值。在每年一度的“双十一”销售热潮中，阿里巴巴以云计算为基础的数据服务，对数以亿万计的消费者需求信息进行详细地捕捉，并帮助电商随时调整销售决策。

的确，大数据的出现，不仅改变了人们的思维方式，还让更多的企业和社会决策有足够的力量和依据——以数据说话。

致胜大数据时代的
50种
思维方法

前言 1

第一章 大数据是一场革命 001

- 1 定律思维 \ 002
- 2 集合思维 \ 008
- 3 创新思维 \ 011
- 4 转型思维 \ 013
- 5 战略思维 \ 017

第二章 应对传统模式的大数据思维 021

- 6 扩散思维 \ 022
- 7 颠覆思维 \ 024
- 8 量变思维 \ 027
- 9 优劣思维 \ 029
- 10 冲击思维 \ 031

第三章 新数据时代的新思维 035

- 11 幂律思维 \ 036
- 12 效率思维 \ 038
- 13 邮件思维 \ 041
- 14 定位思维 \ 043
- 15 逆向思维 \ 047

第四章 大数据能量的积累与爆发 051

- 16 关联思维 \ 052
- 17 双刃思维 \ 054
- 18 价值思维 \ 057

第五章 大数据引导的商业变革 061

- 19 知识思维 \ 062
- 20 变革思维 \ 064
- 21 生产思维 \ 068
- 22 自动化思维 \ 071
- 23 导向思维 \ 074

第六章 大数据的营销策略 079

- 24 测量思维 \ 080
- 25 谈判思维 \ 084
- 26 预算思维 \ 087
- 27 选定思维 \ 093

第七章 教育、教学正在大数据化 097

- 28 院校思维 \ 098
- 29 教学思维 \ 101
- 30 教育思维 \ 105

第八章 以大数据进行高效管理 107

- 31 安全思维 \ 108
- 32 公平思维 \ 110
- 33 真实思维 \ 113

第九章 大数据的全方位力量 117

- 34 渗透思维 \ 118
- 35 隐私思维 \ 123
- 36 监督思维 \ 128
- 37 管理思维 \ 132
- 38 节流思维 \ 140

第十章 更加透明的大数据时代 143

- 39 俘虏思维 \ 144
- 40 调控思维 \ 150
- 41 顺风车思维 \ 153
- 42 公众思维 \ 161
- 43 平衡思维 \ 166

第十一章 在竞争的环境中运用大数据 171

44 竞争思维 \ 172

45 优势思维 \ 175

46 阶段思维 \ 183

第十二章 大数据民主——不仅是说说而已 189

47 民主思维 \ 190

48 透明化思维 \ 195

49 大众思维 \ 199

50 保值思维 \ 202

结语 未来正在发生 207



第

致胜大数据时代的

50种 思维方法

一

章

大数据是一场革命

1

定律思维

大数据中潜伏着很多潜在的规律，只有找到这些规律，大数据才有价值。建设新数据时代和平台的必要手段，就是通过积累数据，预测提升服务和管理水平来实现。

此前，在大数据中，有两个较为突出的定律：一秒定律或秒级定律和摩尔定律。

什么叫一秒定律或秒级定律呢？指的是对处理速度有要求，一般要在秒级时间给出准确的分析结果。如果时间过长，就会失去原有的“一秒定律或秒级定律”的价值。也正是这个速度要求，才区分出大数据挖掘技术和传统的数据挖掘技术的不同。

那什么叫摩尔定律呢？指的是简单地评估出半导体技术进展的经验法则，其重要的意义是对于长期来说的，IC 制程技术是以一直线的方式向前推进，使得 IC 产品能持续降低成本，增加功能和提升性能。

1998 年，台湾积体电路制造公司董事长张忠谋曾说过：摩尔定律在过去 30 年是非常有效的，在未来 10 ~ 15 年也依然适用。但很快，就有新的研究结果推翻了他的言论。研究发现，摩尔定律的时代将会结束。由于研究和实验室的成本需求非常高昂，而有财力投资在创建和维护芯片工厂的企业少之又少。再加上，制程越来越接近半导体的物理极限，将很难再缩小化。

大数据时代正在聚集改变的能量，其定律也在发生着一定的变化。社科院世界经济与政治研究所副所长何帆在一次讲座中，曾说过这样的话：

大数据时代，人们更要重视统计学。比如说，随着大数据时代的来临，

人们开始重视大数据，要重视统计学。可当数据变得足够强大后，人们突然发现，社会上的一切现象都是有一定的统计规律的。它无法像物理学可以准确地描述出前后的因果关系，而只是一个统计的规律。关于这点，有人就玩笑似的说过：只要统计学学好了，再去学别的都战无不胜，因为社会上的一切现象都有一个统计规律。

与此同时，有人就觉得疑问：为什么要强调统计学呢？那是因为人们在认知能力中，统计思维算是最差劲的。要知道，人的大脑中有一些功能比较优良，甚至超过人们自身的想象，比如人们的语言能力。著名的语言学家乔姆斯基就曾经说过：“语言不是你学来的，而是你天生就会的。要是从一出生，开始学语言的话，那是根本学不会的。事实上，一个人在出生的时候，大脑中就已经预装了一套操作系统，那就是语言的操作系统。因此可以说，语言是人们天生就会的。再比如，人们察言观色的能力，也是天生就会的，但有一些是人们不会或不愿意学的。”

诺贝尔经济学的美国心理学家丹尼尔·卡尼曼写过一本书，书名是《思考，快与慢》。在这本书中，有这样的言论，大致意思是说：人有很多思维都是靠直觉的快思维，这是人们经过数百年、千年慢慢演化而来的，最终被留下和被记忆的直接感受，就是所谓的第六感觉。举个例子：当一个人在深夜行走时，会敏锐地察觉到周边的变化。一旦感受到危险或不安的情绪时，就会立即逃跑，甚至大喊大叫。而与此同时，人的大脑之中还有另外一套操作系统，是用来做逻辑推理以及进行统计分析的，只是这个系统不怎么完善。于是，人们天生就缺乏逻辑推理能力和统计思维能力。

所以，在大数据飞速发展的今天，人们应该锻炼自己的逻辑推理能力和统计思维能力！

为什么大数据变成了一个最热门的词汇？能够让大数据变成一个热门词汇，主要的原因有两个。

第一个原因是，由于IT革命后，人们有了处理数据的多方面能力，有对计算机数据的处理能力、对计算机的存储能力以及对计算机的计算的能力，等等。再加上，人类储存信息量的增长速度要比世界经济增长的速度快四倍

(这仅仅是在金融危机爆发之前的世界经济增长的速度)。而计算机数据处理能力的增长速度,要比世界经济增长的速度快九倍。

第二个原因是,社会上的一切现象以及企业的发展,能够被数据化的东西越来越多。在最早时,仅仅是数字可以被数据化,于是就有了阿拉伯的计数,后来又出现了二进位,再后来人们发现文字也可以处理成数据,于是又发现图像也可以处理成数据。比如,有人要去旅行,但是不知道要去的地方的具体位置和周边的信息,那就可以利用搜索引擎搜索;当人们在与微信中的朋友聊天,用微博分享一天的见闻……就已经被数据化了。因此,这就是为什么现在要谈大数据时代,那是因为大数据能够处理和分析的东西太多了,多到人们无可预计。

社科院世界经济与政治研究所副所长何帆说:“当你能够被数据化的东西越来越多,当你能够拿到的数据越来越多时,就跟原来不一样了。原来的统计学得有一个抽样,因为你不可能拿到整体,因为整体太多了,而且无法去计算。而现在,当存储能力无限扩大,处理数据的计算能力不断进步,致使现在我们所处理的往往不是一个样本数据,而是一个整体的数据。”

不仅如此,何帆还总结出了大数据的三个规律:第一个规律是知其然而不必知其所以然,外行打败内行;第二个规律是彻底的价格歧视,商家比你更了解你自己;第三个规律是打破专家的信息优势,病人给医生解惑。

关于第一个规律,他先举了一个葡萄酒的案例——如何品葡萄酒。

在以往,靠品酒方面的专家拿起葡萄酒时,会先闻一闻,准确说出酒的什么味道、富有什么样的香味。接着,品酒专家会看是不是挂杯。最后,他会准确地说出:葡萄酒的产地,大约是什么年份的。但是,当品酒师在品新酒的时候,由于葡萄酒真正的品质还没有形成,因此,他的鉴定是不那么准确的。此外,当一个品酒师的声誉越来越高的时候,由于要顾及自己声誉和名望,所以在大多情况下,他不敢做大胆的推测和判断。

在普林斯顿大学，有一位经济学家很喜欢收藏葡萄酒。有一天，他想试试自己能不能预测出某年某地的葡萄酒品质如何，于是，他就去查找大量的数据，经过分析和研究后得出一个秘诀——葡萄酒的品质与冬天的降雨量、收获季节的降雨量、生长期的平均气温、土壤的成分等因素有关。1989年，葡萄酒的新酒刚刚下来，他就大胆预测：今年的葡萄酒是世纪佳酿。在1990年，他又大胆地预测出：今年的葡萄酒比1989年的好。要知道，一般的品酒师都不敢如此判断，但他却如此大胆，因而着实为自己带来了一些非议。不过事实证明，他说的完全正确！

有句话叫：要知其然，还要知其所以然。但是在大数据时代，人们可以知其然，却不一定非要知其所以然。如果你去问普林斯顿大学的教授：为什么说这个酒好？这个酒到底有什么香味？酒回甘是什么？他未必会说得很清楚。但是他能够知其然，所以才能够大胆地做判断。这是为什么呢？这或许是人们以往的认知里，执意去要寻找一些线性的、双边的直接因果关系，而忽略了其他方面的东西。而人们忽略的方面，恰恰又是最需要的。事实告诉人们：万物之间的联系比人们想象中的要复杂得多，它可能是非线性的，也可能是多元化的。所以说，出问题的不是大数据，而是人们原来的认知模式。那么，在这个时候，人们怎么办呢？最佳的办法，就是退而求其次，要先去寻找相关关系，再去找是否有因果关系。

第二个规律，是彻底的价格歧视。商家比你更了解你自己，他也有着自己的见解。比如说，一个机构是专门做信用卡的刷卡记录的，当他们积累了大量的数据后，经过分析和处理，就会找到很多规律。再比如，一个人的离婚与否，和信用卡上的还款记录以及驾驶车辆出车祸的概率有一定关系。这还真是个奇怪的规律。在大数据时代，比较有名的规律就是：尿布和啤酒的销售量有一定的关系。啤酒和尿布怎么会联系在一起？市场调查人员经过一番调查后才发现：原来当有新生儿出世后，买尿布的这个任务就给新爸爸了。尽管新生的宝贝出世以后，爸爸亲手照顾孩子的机会并不多，但他也有一种自豪感。在去买尿布的时候，为了庆祝，他会顺手去买啤酒。