

隐藏在 “大数据”

背后的巨大财富

[韩] 宋吉永 著
安胜煜 译

富士康总裁郭台铭倾情作序

清华大学出版社



隐藏在
“大数据”

背后的巨大财富

[韩] 宋吉永 著
安胜煜 译

清华大学出版社
北京

北京市版权局著作权合同登记号 图字：01-2013-7717

여기에 당신의 욕망이 보인다

Copyright © 2013 by Gilyoung Song, All Rights Reserved. Original Korean edition published by SAM&PARKERS, Co., Ltd. Simplified Chinese translation copyright © 2014 by Tsinghua University Press. Simplified Chinese language edition rights arranged with SAM&PARKERS, Co., Ltd. through Beijing GW Culture Communications Co., Ltd. 本著作物中文简体字版权由北京水木双清文化传播有限责任公司代理。

本书封面贴有清华大学出版社防伪标签，无标签者不得销售。

版权所有，侵权必究。侵权举报电话：010-62782989 13701121933

图书在版编目 (CIP) 数据

隐藏在“大数据”背后的巨大财富/ (韩) 宋吉永著; 安胜煜译. —北京: 清华大学出版社, 2015

ISBN 978-7-302-38904-0

I. ①隐… II. ①宋… ②安… III. ①商业信息学 IV. ①F713.51

中国版本图书馆 CIP 数据核字 (2015) 第 005706 号

责任编辑: 朱敏悦

封面设计: 汉风唐韵

责任校对: 王荣静

责任印制: 李红英

出版发行: 清华大学出版社

网 址: <http://www.tup.com.cn>, <http://www.wqbook.com>

地 址: 北京清华大学学研大厦 A 座 邮 编: 100084

社总机: 010-62770175

邮 购: 010-62786544

投稿与读者服务: 010-62776969, c-service@tup.tsinghua.edu.cn

质 量 反 馈: 010-62772015, zhiliang@tup.tsinghua.edu.cn

印 刷 者: 三河市君旺印务有限公司

装 订 者: 三河市新茂装订有限公司

经 销: 全国新华书店

开 本: 148mm×210mm 印 张: 6.75 字 数: 150 千字

版 次: 2015 年 9 月第 1 版

印 次: 2015 年 9 月第 1 次印刷

印 数: 1~3000

定 价: 29.00 元

世界何其小

我一直期待着在台北的一家书店再获惊喜，因为我在寻觅一本关于“大数据”的参考书。店员向我热情推荐此书，说它当下广受欢迎，绝对洛阳纸贵。我通读了此书，发现其海量的信息中，隐藏了对产业趋势的分析与洞见，真是字字珠玑！

故而我希望，我所有的高阶主管们也都能亲其笺页，俾获心得。近两个月来，只要一谈到大数据的话题，我就迫不及待地从我公文包中取出这本书，秀给大家看看我目前正在读它，并提醒大家也要读读此书，来补补脑。

最近我到韩国出差，遇到了之前一次旅行中结识的宋博士，他向我介绍他的研究。我很受感染，邀请宋先生共进晚餐。他赠我一本也是有关大数据的参考书。我很吃惊，宋先生送我的书居然与我公文包中的书是一样的！我告诉他，此书我两个月前已买，然后从公文包掏书给他看。他好奇我有此书，我好奇他送我此书。真是英雄所见略同啊！

古人说无巧不成书，我要说无巧不买书！我便问他，“为什么会向我特别推荐这本书？”他说，“为什么不呢？我是作者嘛！”我



知道作者是一位宋先生，以为他是一名中国人，我从来没想到彼“宋先生”就是此“宋博士”啊。在异国他乡，目睹迥然风物，手捧醒世佳作，更遇作者指点天下趋势，确为一件幸事，亦直教人感叹这个星球，真是非常非常的小啊！是为序。

郭台铭 谨序

二〇一四年十一月十八日

序言 解读人心

“（4·27）在补缺选举中制造舆论的走向当中，社交媒体起到了十分重大的作用。让每一个人都成为‘新闻工作者’，把媒体尚未报道的或减少或冲淡的事实完全公开，并且进行传播。”

——曹国，韩国首尔大学教授，权威推特用户

“在这次选举中谁会当选呢？让我们一起来看看民心的走向吧。”

2011年9月7日，笔者公司开设了为10月26日韩国首尔市长再补缺选举中提供实时分析选举舆论的网站。关于无党派朴元淳候补和大国家党罗卿瑗候补之间的首尔市长补缺选举对决，在初期的误差范围内已经预测了罗卿瑗候补的优势，电话舆论调查也是相同的结果，从候补的认知度或是支持层的投票率等客观层面上来看，情势对罗卿瑗候补是有利的。但是，选举当日投票站调查却倾向于朴元淳候补，开票结果也证实是政治无名的无党派朴元淳当选了。

在这之前举行的4月27日地方补缺选举当中，人们在博客（Blog）、推特（Twitter）、脸谱（Facebook）等多种社交媒体（SNS）上发表了个人意见。笔者认为SNS可以对选举产生很大影响，故利用此数据为基础，进行分析、预测并指导结果。因而在首



尔市长补缺选举时甚至对数据流向进行了现场直播，并在推特上统计出提及各个候选的次数，情况发展以及各个候选的支持率之间存在着怎样的关联性。

回顾过去的3周时间，图1左侧是对两名候选的推特占有率的分析，整体来看，朴元淳候补的占有率高。其中，接受民主党的竞选法则的情况下朴候补接受大企业捐款的疑惑成为热点的时候以及被确定为在野党唯一候选人的时候其占有率特别高。相反，罗卿瑗候补在推特上被多次提到是什么时候呢？是10月6日的MBC《100分钟讨论》节目中罗候补的答辩人申志镐喝完炸弹酒出来的时候。这时候不仅是门户网站，而且SNS上提到罗候补的次数也瞬时上升。然而，这会是正面关注吗？我们都心知肚明，当时关于罗候补的推特中约90%都属于否定的舆论。

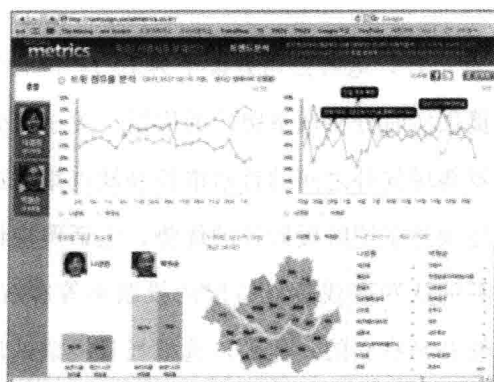


图1 补缺选举时推特走势

解读一条条推特内容，分辨出正负面是不可能的。在韩国一天生成的推特就有1 000万条信息，怎么能全部读完呢。但是，如果只考虑特定时期的热点问题的话，就可以揣测出当时的内容。同



时，20%左右的推特用户会公开本人现在所处的地理位置。并且，通过用户的名字或自身表现的信息来推断相关用户的性别可以达到90%以上的准确性。通过分析性别、各地区的关注度，可以大略推测出参选人的支持率。之后，罗候补在江南出入VVIP美容室的时候或是帮助残障儿童洗澡的新闻，都会在推特上引起争议。可以说，我们可以借由发文者的语气来推测其对参选人的想法。

这个故事的启示是什么呢？推特和电话舆论调查的性质完全不同。在推特上即使没有人问，人们也会自觉地把自己的想法说出来，而调查当事者收集这些对话内容。也就是说，可以看到朋友之间的隐私对话内容。如果转换成熟悉的表达方式，也就是说是可以看到“民心”的意思，是没有介入任何刻意、思想的原原本本的民心。

以2011年的两场补缺选举为基础，社交媒体在韩国开始正式引起社会的关注，包括笔者公司在内的韩国多种社会数据分析企业开始分析推特等SNS资料。新闻、广播等现有媒体描述朴元淳和罗卿媛候补的对决是势均力敌，相反社会数据分析企业从SNS等网络舆论初期开始，分析指出情势对朴元淳候补更为有利，最后结果也认证了数据分析的准确性。通过这个观察数据可以得知舆论的走向，而且也广而告知可以将其作为选举预测的一种模式来应用的事实。

最近在这里点燃了“大数据”的热潮。全世界在每一分、每一秒都有数以万计的标准和非标准的数据，用现存的分析工具是无法承受如此规模庞大的资料的。政府机关或企业认为这里埋藏着商业的契机，对此极为关注。从2011年末开始的6个月间，只是我参



与的大数据会议就超过了6个。

为什么人们对大数据如此狂热呢？因为是总统说的？还是因为隐藏着新的商机？在美国谷歌有人通过对大数据的分析来进行预防流感的活动。如果我们也利用大数据来分析的话，可能会对缓解失业难现况有所帮助呢。

例如，美国的经济专业周刊《福布斯》在2012年5月12日发表了10年前并不存在、现在却很有前途的10种职业，即应用程序开发者、数据挖掘师（信息收集、分析家）、社会媒介人、用户体验设计师、云计算开发者等。其中还有CLO，你听过这个名称吗？CLO就是“首席倾听官”（Chief Listening Officer），是随着社会媒介受欢迎而产生的新型职业。它通过诸如脸谱、推特、YouTube等社会媒介来倾听顾客的心声。数据挖掘师是在网络中铺天盖地的数据洪流当中寻找出有益信息的人，若能将此技术应用于制订治安计划或选举策略等，将能给学习应用统计学的人群提供一个绝佳工作岗位。

说实话，每当听到这种故事的时候都很郁闷。

不久前，这种内容还像是IT调查公司高德纳（Gartner）或是《华尔街日报》处理的主题。但现在属于非经济杂志的《纽约时报》上也开始登出有关大数据的报道。所谓的“主流”也开始重视大数据了。

2012年2月的《纽约时报》报道称大数据很重要，仅仅在美国就需要14万~19万名资深数据分析家。借此鼓励具备良好物理能力的人从现在开始认真准备加入这个职业当中。只要知道哪里有适合的人才，企业都愿意，拿着优厚的待遇条件来洽谈。



美国需要 14 万~19 万名具有专业深层分析经验的资深工作者以及至少 150 万名资料分析师。

说实在的，每当听到这种报道时总是让我感到很苦恼。我们以需要 150 万名可以分析和应用数据的人们为例。企业可以通过对现有职员进行再教育或寻找新的人才来确保该人才能够读取数据背后的线索。通常说企业的业务要领可以概括成阅历、胆量和热情三种。如果发挥阅历优势，激发胆量和热情的话，就能成为数据解析家吗？这很难说。

我苦恼的地方就在于此。数据本身再多也没有什么意义，重要的是会解读那些具有独特性的数据。我公司向企业销售的就是这个模式的分析。比如说，我是一名矿工，但是我挖掘的不是铀或铁矿石，而是你的心思和欲望。被埋葬在人们心中的矿产就是社会媒介。但是随着社会媒介的数据量变得日益庞大，我自然而然地踏入大数据的领域。（所以，在书中所谓的“社会媒介”看成“大数据”也不为过。）

在创业初期，笔者主要从事向企业提供数据分析的工作。当企业要求资料时，通过高尖技术非常快速地分析后交付给企业，特别是掌握企业特别敏感的顾客投诉、产品缺陷等热点问题或分析相比竞争对手本公司产品的优越性等资料提供给公司。之后，通过更进一步的深层分析，介绍“热销”的产品与品牌战略、产业关键字等相结合的策略。越看资料越想进一步了解，我的工作也就随着时间的推移更加深化了。

但问题就是“分析”。即使拥有数十亿件需要分析的复杂数据，如果没有洞察力也是枉然。最近几年听说野营产业突然很火，这是为什么？而且，在这种流行野营趋势的环境下，也不知该做什么样



的事业。

所以，现在销售的不是数据而是分析。如果有企业问：“我们的产品一直很畅销，今年初为什么突然滞销了呢？是什么原因？”经过数据调查之后我们会告诉他们：“你们的产品没有市场了。”对于这样的答复企业很满意。因为我们确保其未投入更多成本，企业对这样清晰简明的整理结果表示感谢。

过去由于没有查看数据的方法，所以只能相信感觉。特别是经营者的感觉对创业、扩张以及改组起到了至关重要的作用。雄心勃勃的新入职员提交计划书时，中层主管或经营者会说：“这个不行，10年前就已经做过了，不行。”直接将其否决了。今天，比起10年之前已经变成了另一个完全不同的世界，但他们还只是说不行。如果员工提出建议说“没有市场了”，就会被指责道：“你懂什么呀！”。但是，现在出现了可以合理化决策疑问的科学性资料。

为了找出这种洞察力，我查看了数十亿条数据。更准确地讲是看出数十亿个将数据发布出去的人们的“心思”。企业的核心力量归根结底不就是“怎样了解顾客吗”？通过数据分析就会出现多样的阶层、职业群等，从而可以解读连顾客自己都认识不到的要求和欲望，甚至比顾客更加了解顾客自己。如果依据这个理念制定战略的话，在巨大的变化潮流当中会先人半步等待顾客的光临。

在过去10年间，我是一直按照人们留下的想法和感觉的原本形式分析社会媒介的。并且在本书中，我对通过分析社会媒介领会的大数据分析方向和应用方向进行了说明（在书中提到的“我”不仅是本书作者，还指在这期间费心分析社会媒介的笔者公司的同事们。如果没有这些人出色的头脑和洞察力，也将无法完成这项分析



社会媒介的繁杂的工作)。从发现产品的缺陷或评价从消极的需求中出发的社会媒介分析,如今已经进化到产品策划、解读趋势,甚至理解人的程度。在这个过程中我切身感受到,重要的不是产品开发者的直观想法,而是顾客的想法。要想在市场上生存,重要的不是权威,而是和大众的意见一致。此外,如果你能够适应大环境的变化来分析大数据的话,手上的这份资料可以将大众的声音传递到更高的层面上来。

回忆过去那段被数以万计的资料包围着的时间,让我相信我所做的是持续不断地挖掘有意义的信息结果。但面对从哪儿挖掘的质问,如果说是“从文本中”的话,无疑是一个非常愚蠢的答案。文本只是手段但绝对无法成为目的。从各个地方收集来的多样数据堆成山,但我的关注点只是集中在其中“人”的真正想法上。相较于理性思考,这种情绪反应反而更重要。归根结底我的工作就是挖掘思想,这也成了我们公司的社训(口号):挖掘思想(Mining Minds)。

本书是来解读这个疑问的,同时也建议企业和社会能应用大数据方式的提案。为了帮助各位读者更好地理解及感兴趣地阅读此书,书中尽可能多地介绍了我亲自参与的分析案例。希望本书能成为让更多人了解现状并解读未来的工具,让人们意识到能够了解社会媒介是多么有趣的一件事。不要停留在只读一遍的泛泛浏览上,如果能够思考社会大数据在我们的生活和商业活动中应该怎样应用并应用得更好的话,就再好不过了。

宋吉永

目 录

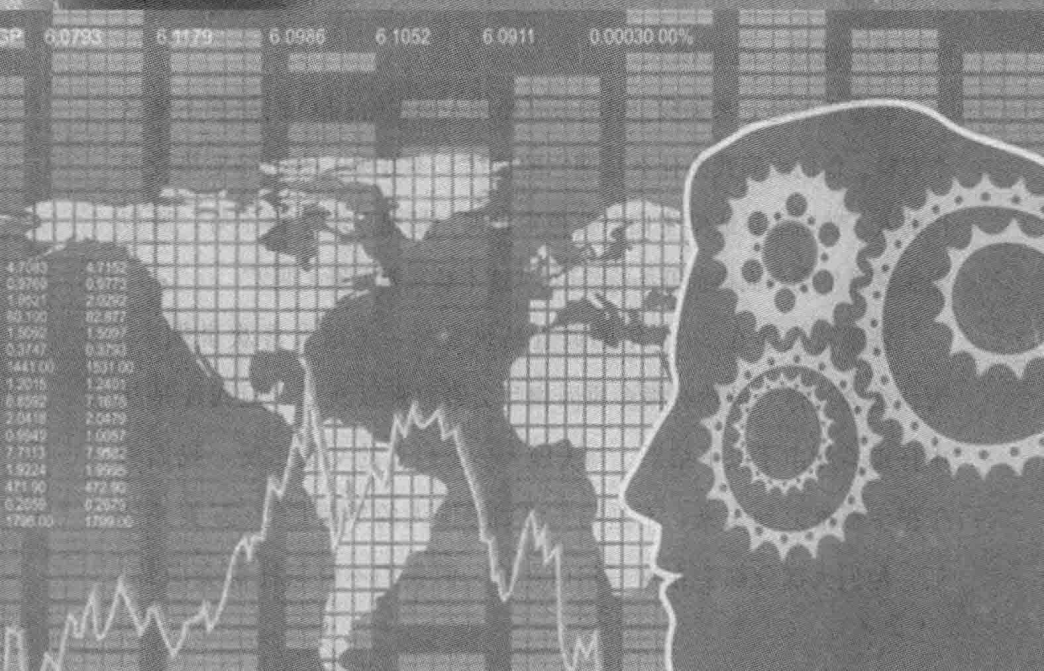
1. 现在 & 未来——日常积累，成就未来	1
1.1 周末销售忧郁症药，卖得特别好?	2
1.2 未来存在于现在当中	6
1.3 都是贤人的时代即将到来	18
1.4 “问题”优先	33
1.5 提出疑问	44
给 IT 专家的建议 机会一定会再次到来	59
2. 理解“为什么”——解读隐藏在现象背后的理由	63
2.1 信息的不对称正在瓦解	64
2.2 从“反映”到“形成”	70
2.3 购买欲望源自哪里?	75
2.4 指出市场的“空位”	82
2.5 寻找解决问题的信息	93
2.6 发现新市场	106
2.7 开发领先半步的产品	110
2.8 不看产品，而看消费者内心	120



2.9 打破偏见,问题迎刃而解	130
给市场营销者的建议 不看市场,读人心	133
3. 寻找机会——从动态变化中寻找可能性	137
3.1 跳出市场就能看到机会	138
3.2 解读情绪的变化	150
3.3 不再局限于既有消费者	163
3.4 绘制情绪图	170
3.5 从现在了解未来	181
给梦想成为大数据分析家的学生们 练习提问	187
送给为是否引进大数据而苦恼的 CEO 们 以客观的资料分析作为决策的依据	190
结语 大象绝对不会忘记	194
参考文献	198

1. 现在 & 未来

——日常积累，成就未来





1.1 周末销售忧郁症药，卖得特别好？

你可曾注意过你！哪天情绪最好？一般来说不是星期五吗？

以上只是推测没有可信度，因此，我们进行了一次分析。调查结果显示，从星期一到星期五，情绪真的在逐渐变好。相对来说，周末时情绪较低沉，而星期一情绪则跌落谷底。很多人都说星期日晚上的搞笑综艺节目结束之后，一想到第二天要上班，情绪就开始变得忧郁。疲于公司生活的人们，在星期六晚上的综艺节目结束的时候就开始进入忧郁模式。

有意思的是，参加我主讲的最高经理人课程的 CEO 们却看不懂这个资料。他们反问我为什么星期六的情绪会不好呢？他们都是热爱工作的人。普通职员或学生们只要想到星期一情绪就会变得不好，所以无法享受星期六而提前开始抱怨说：“这不是又要上班了吗？”

佛蒙特大学在 2011 年的调查结果显示，美国人喜欢星期六。为什么会这样？虽然有很多原因相互交错而难以断言，但我还是从中仔细推测出了以下结果。与美国人享受当下的文化相反，我们善于把幸福推迟到未来，不是吗？假如美国人的年薪是 10 万美元的话，他们大部分都会在这一年内花光，而我们不管怎样都会节约、



积攒后再花在教育 and 养老上。只有对未来无所担忧才能安心地享受现在，但是我们需要担心的事情太多了。比如，养老问题、明年是否能加薪、如果孩子的成绩没有下滑就阿弥陀佛了——根本就没有安心的日子可言。这种不安因素导致人们只要到了星期六就不知不觉地开始担心星期一了。一天的工作也是如此，下班之前的情绪是最好的，而睡觉前最坏，因为第二天醒来又要出去工作。

如此说来，韩国人的幸福时刻不像在现在。因为对未来心存不安，所以现在不够幸福。在书店的心理学区域摆放的书大多不是描述人们心理的书，而是治疗类的书，是吧？

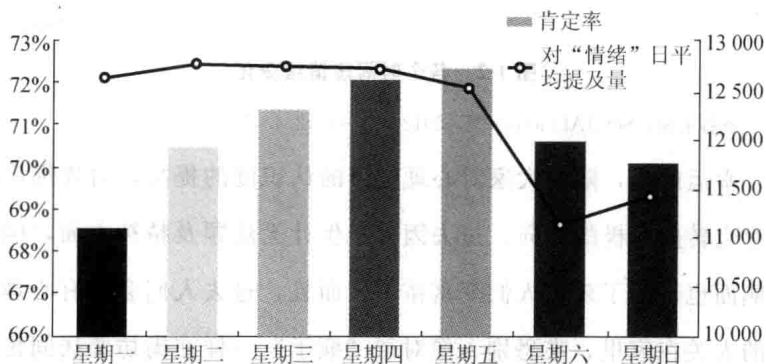


图 1-1 每周各天的情绪变化

资料来源：SOCIALmetrics™，2011.5.1~2011.10.31

不管怎样，我们得到了有意思的数据，我便开始考虑在什么地方能加以应用。这时，出现了能将这些数据适用于治疗忧郁症的机会。

当今，忧郁症已成为严重的社会问题，现在的你也感到忧郁吗？我们遇到高兴的事情时，情绪就会很好；遇到伤心的事情时，情绪就会变坏。在新闻中听到名人自杀的消息时，整个社会也会感到忧郁。结交阳光的朋友有益的原因就在于此，如果朋友忧郁、消