

一本书带你领略小数据之美、小数据之道、小数据之魅、小数据之巅

小数据 之美

精准捕捉未来的商业小趋势

陈辉 / 著

⊕ 精准定义自我与用户的需求，寻找小数据的高价值

⊕ 发现大数据忽视的盲点与机遇，用小数据发现新的商机与趋势

⊕ 谁控制了小数据谁就是赢家

中信出版集团

小数据之美

精准捕捉未来的商业小趋势

陈辉 / 著

美女甜媵小

数据小趋势的未来发展趋势

陈辉著

图书在版编目(CIP)数据

小数据之美:精准捕捉未来的商业小趋势/陈辉著

—北京:中信出版社,2019.4

ISBN 978-7-5086-9282-1

I. ①小… II. ①陈… III. ①数据处理 IV.

①TP274

中国版本图书馆CIP数据核字(2018)第167255号

小数据之美——精准捕捉未来的商业小趋势

著者:陈辉

出版发行:中信出版集团股份有限公司

(北京市朝阳区惠新东街甲4号富盛大厦2座 邮编 100029)

承印者:北京通州皇家印刷厂

开本:787mm×1092mm 1/16 印张:20 字数:200千字

版次:2019年4月第1版 印次:2019年4月第1次印刷

广告经营许可证:京朝工商广字第8087号

书号:ISBN 978-7-5086-9282-1

定价:59.00元

版权所有·侵权必究

如有印刷、装订问题,本公司负责调换。

服务热线:400-600-8099

投稿邮箱:author@citicpub.com

序言

在数据的江湖里，既有波澜壮阔的大数据（Big Data），也有微波涟漪的小数据（Small Data），二者相辅相成，才能相映生辉。目前大数据流行，人们就言必称大数据，这不是做学问的态度，不要碰到大量的数据，就给它戴上一顶“大数据”的帽子。大数据体现出规律，小数据蕴含着智慧，它们都闪烁着理想之光。

古人云：“圣人见微知著，睹始知终。”道家的一部重要著作《淮南子·说山训》中说：“以小明大，见一叶落而知岁之将暮，睹瓶中之水而知天下之寒。”意思是说，看见一片落叶，就知道秋天来临；看到瓶中水结冰，就知道天气的寒冷程度，这是对见微知著的形象比喻。

吴甘沙先生用《一代宗师》的台词来比拟大、小数据的区分，倒也甚是恰当。他说，小数据“见微”，作个人刻画，可用《一代宗师》中“见自己”形容之；而大数据“知著”，反映自然和群体的特征和趋势，可用《一代宗师》中的“见天地、见众生”比喻之。

大就是大数据，就是全量数据；小就是小数据，就是个体数据。所以，对于数据科学，我们必须在把情况搞清楚的基础上懂得哪些是大、哪些是小，知道怎样处理大小辩证关系，才能在具体数据应用中做到抓大放小、以大兼小，以小带大、小中见大。在研究小数据时，要以大兼小、以小见大，必须考虑目标的正确性、可操作性和决策的科学性、可行性。在研究大数据时，要抓大放小、以小带大，既要考虑整体共性，又要注重个体特征。这样，在数据应用中，大能与小数据量化的自我保持高度一致，小能与大数据预见的未来保持一致，既不能见小不见大，也不能见大不见小。对于数据科学，从数据中来，到数据中去，既要见大，也要见小，以小带大、小中见大，才能真正用好数据。

当认知科学领域发生“天翻地覆”的变化时，我们的未来又会是怎样？认知革命，特别是“真相时代”的到来，“预测”将被“预见”取代，那么，“预见未来”将不再是遥远的星辰。如果认知科学的本质是计算科学，那么，“大数据”和“小数据”争夺所谓“大小”的“江湖地位”意义何在？但我们需要觉悟的是：此“数据”非彼“数据”。面向未来，“大数据”和“小数据”将开启一个“全新故事”，一个“预见未来”的故事。

是为序。

中央财经大学保险学院院长 李晓林

2018年3月于北京

“忽如一夜春风来，千树万树梨花开。”似乎在一夜之间，大数据就红遍了南北半球，在神州大地更是一时风头无二。与此同时，大数据也被神化得无处不在，无所不包，无所不能。这里面有认识上的原因，也有跟风的成分。我们以为，越是在热得发烫的时候，越是需要有人在旁边吹吹冷风。在这里谈谈小数据的重要性、预测态、未来式，并非要否定大数据的价值。相反，只有我们充分认识了小数据的特点和应用，才能更好地利用大数据，才能通过小数据和大数据的互补，更好地展现数据之美、数据之道、数据之魅、数据之巅。

数据本无大小，但应用场合、处理方式的不同却分出大小。数据表示的是过去，但表达的是未来，所以应用数据不仅需要全量数据，也需要样本数据；不仅要了解相关性，更要明白因果关系；不仅要预见未来，更要量化自我。这就迫使我们从更广泛的角度理解小数据，梳理小数据与大数据的分野，从而将相关思路投射、印证于小数据，考察其核心特

点和应用特质。除传统统计学以外，人工智能、复杂系统等技术的发展，小数据也能学习，小数据也能复杂化，小数据远比我们想象的要强大。这本书主要通过一些小案例来理解小数据之美，如中医的“望闻问切”、保险的“大数法则”、军事的“战略与战术”、信用的“能力与意愿”、生态的“小环境与大环境”、企业的“账面与实际”、教育的“应试教育与素质教育”等，进而揭示数据的局面变化、逻辑更新、未来演进。

传统思想要求我们“知己知彼，百战不殆”“运筹帷幄，决胜千里”“抽丝剥茧，明察秋毫”，互联网思维要求我们“精简”“取舍”“极致”“专注”，偶尔也思考“黑天鹅”“灰犀牛”，但大多数时间都沉陷于“大数据”和“大数据时代”。我们迫切需要一点新的思路、新的视角，我们需要更系统、更强大的小数据，我们需要更全面地理解小数据。重新审视我们的时代。小数据，尽管仍笼罩在迷雾中，但已经开始在我们脑海中浮现出其整体的轮廓，这就是我们选择写《小数据之美》的初衷，也是我们选择写《小数据之美》的本义。

本书由陈辉编著，中央财经大学硕士研究生李明子、李冰，招商银行易斯琪，央财国际研究院特约研究员庞博参与了本书主要章节的资料整理，由陈辉统一修改定稿。

本书在编写过程中，参考了国际国内的相关著作、论文、报告和案例，中央财经大学中国精算研究院的多位专家学者提出了许多有益的修改意见，在此一并表示感谢。由于时间紧迫和编者水平有限，书中疏漏、错误之处在所难免，敬请读者批评指正。

中央财经大学中国精算研究院 陈辉博士

2018年3月于北京

目录

第一章 小数据时代的到来

一、小数据之蜕变 / 003

二、小数据之崛起 / 006

三、小数据大潜能 / 014

第二章 大数据时代的小数据

一、数据变化趋势 / 023

二、数据本无大小 / 025

三、大小相得益彰 / 030

四、打通数据孤岛 / 035

第三章 智能时代的小数据

一、智能无处不在 / 043

二、智能产品到嵌入程序 / 047

三、小数据与人工智能 / 051

四、小数据与智能产品 / 052

五、小数据与虚拟现实 / 054

六、小数据与增强现实 / 054

第四章 小数据蕴含的大智慧

一、小数据的逻辑 / 059

二、什么时候想起小数据? / 072

三、贝叶斯真的能推理吗? / 075

四、回归分析是力大无穷的! / 076

五、让小数据逼近真相 / 079

六、与小概率平起平坐 / 081

第五章 小数据赋能人工智能

一、小数据决定人工智能的未来 / 085

二、谁控制了小数据谁才是赢家 / 090

三、用小数据也能机器学习 / 099

四、会学习的小数据更强大 / 102

第六章 未来科技依赖小数据

一、小数据驱动复杂系统的优化 / 113

二、小数据推动自适应系统的进化 / 121

三、小数据助力模糊系统的应用 / 130

四、小数据创新小数据系统的发展 / 140

第七章 小数据的再认识

一、小数据的现象本质 / 153

二、小数据的原因结果 / 158

第八章 预见自我

一、智能化人生 / 167

二、智能化健康 / 174

三、智能化学习 / 178

四、智能化医疗 / 181

第九章 预见世界

一、认知的革命 / 187

二、经营的本质 / 192

三、信用的价值 / 198

四、军事的科学 / 202

第十章 预见未来

一、小数据与黑天鹅 / 207

二、小数据与灰犀牛 / 216

三、小数据与独角兽 / 223

四、小数据与长尾分布 / 231

五、小数据与随机漫步 / 234

第十一章 大数据时代的小数据革命

一、大数据时代隐忧 / 241

二、让遗忘回归常态 / 244

三、让精简成为王道 / 247

四、从细节预见未来 / 249

五、从量化自我开始 / 252

六、迎接小数据时代 / 259

第十二章 小数据思想成就大未来

- 一、厘清小数据的关系 / 265
- 二、形成小数据思维 / 278
- 三、读懂小数据逻辑 / 287
- 四、掌握小数据的智慧 / 294
- 五、成为小数据科学家 / 301

参考文献 / 307

- 一、小数据思维 / 307
- 二、小数据思维 / 307
- 三、小数据思维 / 307
- 四、小数据思维 / 307
- 五、小数据思维 / 307

第十三章 小数据赋能人工智能

- 一、小数据赋能人工智能 / 311
- 二、小数据赋能人工智能 / 311
- 三、小数据赋能人工智能 / 311
- 四、小数据赋能人工智能 / 311

第十四章 未来科技催动小数据

- 一、小数据赋能人工智能 / 311
- 二、小数据赋能人工智能 / 311
- 三、小数据赋能人工智能 / 311
- 四、小数据赋能人工智能 / 311

第十五章 小数据的再认识

- 一、小数据的再认识 / 311
- 二、小数据的再认识 / 311

第一章

小数据时代的到来

一、小数据之蜕变

(一) 小数据的成长之路

在过去的几个世纪里,以抽样数据为特征的小数据研究使自然科学、社会科学和人文科学的发展取得了跨越式的进展。小数据的成长过程可以分为两大阶段:数据的产生与科学数据的形成。在此基础上,大数据悄然诞生。

数据起源于古人对事物的“多”或“少”的认识。从中国古代的结绳记事、苏美尔人的串珠计数,到古埃及的十进制象形文数字、古巴比伦的六十进制数字,再到中国商朝的甲骨文数字、唐代的汉字数字,最终到后来的罗马数字、阿拉伯数字,逐步实现了数的抽象性和可计算性。而量赋予了事物的测度和比较标准的语言,它是能体现事物内在本质特征以及事物之间差异程度的一种载体,包括量的规模、关系、变化、界限与规律。从数到数据的发展过程,客观反映了人类认识事物本质属性的必然过程。如果说抽象的数的分析产生了数学,那么有计量单位的数据分析则产生了统计学。相应地,统计学中所使用的就是传统意义上的小数据。

随着对数据需求的增加,以及收集与运用数据经验的不断积累,人们在科学研究的过程中逐渐形成了能够科学使用的小数据——科学数据。科学数据是人们用以探求现象变化规律或验证已有理论假设的数据,数据收集的方式主要是观察、测量和实验。科学数据的特点是科学设计、可重复获得、相对精确和具有共享性,它不仅提高了人类认识事物的精确性,而且逐步形成的数学化思维与方程表

达式使人类解决了不同物理量之间的数值关系表达问题，从而为人类开展相关事物之间的定量研究提供了途径。自然哲学家开普勒通过对第谷天文观察数据的使用，推导出了行星运动三大定律；伽利略通过对地球表面物体运动数据的测量，发现了自由落体运动规律；牛顿利用大量的天文观察数据和实验测量数据，创立了牛顿力学体系。

当然，在自然科学对数据进行科学研究的同时，社会科学也同样对数据进行了科学范式的研究，并发现了如平均人、恩格尔系数、基尼系数、经济周期等定律。

（二）小数据 2.0 的诞生

传统意义上的小数据因其数据体量的窄小、抽样采集方式的传统而得名，其实质是通过目前主流统计工具在合理时间内采集、存储、处理的数据集。经典的数理统计和数据挖掘知识，可以较好地解决这类问题。而大数据时代下的小数据，是一类新兴的数据，它是指需要新的应用方式才能体现出具有高价值的、个体的、高效率的、个性化的信息资产。

小数据 2.0 下的定义并不是指数据量小，而是围绕个体的全方位数据及其配套的收集、处理、分析和对外交互的综合系统。个体产生的数据，包括生活习惯、社交行为、财务状况等，全部被各种智能设备或传感器收集和利用并进行分析，并对外形成一个富有个人色彩的数据系统。小数据范畴下的个体，不仅仅包括自然人个体，还包括一切具有完整内部环境的个体，如家庭、企业、学校、医院、生态环境等。通俗来讲，以我们每个人为例，小数据能够观察个体行为如发送电子邮件、短信或推特，购物或锻炼，步行上班还是乘车上班，在家看电视剧还是看电影，或者在跟 Amazon Echo（亚马逊公司研制的一款智能音箱）对话等。这些痕迹反映了我们是谁、

在哪、和谁、做什么。通过数据整合，以可视化的方式让你能够更了解自己。

因此，小数据又被称为“量化的自我”，目的与大数据相同，提供个体决策的依据。虽然小数据迄今为止的应用还十分简单，但小数据若与智能时代的各项新技术相结合，它能提供的信息将远不止于此。

(三) 小数据的时代表征

(1) 具有鲜明的个体独特性

小数据是围绕个体用户所感知的数据集合，体现了较高的数据价值密度，具有鲜明的个体独特性。不同层次的个体用户如企业层次和个人层次，其数据广度和深度大相径庭；同一层次的个体用户，其行为特征同样具有一定的特异性，这导致涉及的小数据集也各不相同。即使对于同一个事件，不同的个体用户由于所处环境、学习背景、认知能力、思维方式不同，其产生的行为方式也是不同的。这也决定了小数据的个体独特性。而这一特征恰好为获取用户的个性化需求创造了条件。

(2) 具有复杂多样的数据特性

随着高性能传感器、可穿戴设备等智能产品和嵌入程序的广泛普及和运用，这些设备所产生的数据类型体现出广泛的半结构化和非结构化特征，并且随着时间的推移以及个体用户活动范围的扩大，其产出的数据量也在逐渐增多，从而进一步加剧了数据的复杂性。另外，在 2.0 时代，小数据已经不拘泥于传统观念中为了解决特定的问题才去收集、分析、处理指定的相关数据的内涵，而是来到了针对特定个体进行数据生产的阶段。从这个角度来讲，小数据与大数据兼具预测功能，只不过预测的视野和方式有所区别，这种“预

测”也体现了小数据复杂多样的数据特性。此外，个体在日常的生活、学习和工作中，其自身产生的数据内容也具有多样性，如视频数据、图片数据、文本数据等，同时这些数据集具有一定的主观性、离散性和随机性。

(3) 具有高度的实时动态性

具有高度的实时动态性是由于对个体的感知和监控是全天候、全方位的，因此小数据的获取和收集也是实时更新、动态存储的。对于同一个体用户而言，在不同的时间，其所处的情景状态是不一样的，从而使小数据集合具有动态性和不确定性。另外，个体用户所承担的任务和遇到的问题也是不断变化的，用户的需求行为随之变化，这也将促使小数据集合的实时变动。

(4) 具有明显的人机交互性

1.0时代，小数据无需强调个人的表达需求，而更关注样本对总体的代表性以及对所研究的科学问题的解释程度。2.0时代，小数据可以以两种方式走进我们的生活：第一种是人为采集小数据，这个层面上小数据是消费者与采集者分享的有关其独特偏好的微数据，比方说他们订阅哪些报纸、青睐哪家餐厅以及他们在旅行时喜欢去哪些地方，通过人工对数据进行理解和加工；第二种则是通过各类智能设备和传感设备采集个体的所有数据，这个层面上小数据带有很明显的人机交互性，需要机器系统以适当的方式回应不同风格的用户。

二、小数据之崛起

图像、语音、文本、情绪成为近两年最热门的概念，所谓需求驱动市场，市场驱动技术发展。在现在的人工智能时代，“快速”“准