

Big data and
Finance

大数据金融

李勇 许荣 编著



中国工信出版集团



电子工业出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
<http://www.phei.com.cn>

大数据金融丛书

PANCOAL 盘古智库

Big data and
Finance

大数据金融

李勇 许荣 编著

电子工业出版社

Publishing House of Electronics Industry

北京·BEIJING

内 容 简 介

大数据金融是利用大数据技术突破、革新并发展传统金融理论、金融技术和金融模式的一种全球性趋势。这一趋势既是现有技术进步的必然结果，又是未来金融发展的强劲动力。广度上，大数据金融重塑了银行业、保险业、证券投资业等金融行业的核心领域。深度上，大数据金融不仅推动了金融实务的持续创新，更催生了金融模式的深刻变革。

本书构建出一个系统性的分析框架，从多个角度对全球大数据金融的实践进行了全面的归纳与研究。沿着本书的分析框架，读者可以清楚地了解大数据金融在银行、保险、量化投资、资产管理、金融监管和国家金融安全这六大领域的机遇和挑战。同时，本书详细分析了国内外大数据金融应用的大量真实案例。这些案例生动鲜活、时效性强，既便于读者加深对理论内容的理解，又使得本书具备了一定的现实指导作用。

在保证专业性的前提下，本书兼顾了表述的通俗化。因此，本书既为金融业界、学界的深入研究提供了基本资料，又适合普通读者入门之用。

未经许可，不得以任何方式复制或抄袭本书之部分或全部内容。

版权所有，侵权必究。

图书在版编目（CIP）数据

大数据金融 / 李勇，许荣编著. —北京：电子工业出版社，2016.1

（大数据金融丛书）

ISBN 978-7-121-27626-2

I. ①大… II. ①李… ②许… III. ①金融—数据管理—研究 IV. ①F830.41

中国版本图书馆 CIP 数据核字（2015）第 281491 号

策划编辑：李 冰

责任编辑：李 冰

特约编辑：田学清 赵海军

印 刷：北京季蜂印刷有限公司

装 订：北京季蜂印刷有限公司

出版发行：电子工业出版社

北京市海淀区万寿路 173 信箱

邮编：100036

开 本：720×1000 1/16 印张：11.5 字数：184 千字

版 次：2016 年 1 月第 1 版

印 次：2016 年 1 月第 1 次印刷

定 价：58.00 元

凡所购买电子工业出版社图书有缺损问题，请向购买书店调换。若书店售缺，请与本社发行部联系，联系及邮购电话：（010）88254888。

质量投诉请发邮件至 zltz@phei.com.cn，盗版侵权举报请发邮件至 dbqq@phei.com.cn。

服务热线：（010）88258888。

“大数据是互联网金融的基石。基于大数据的互联网金融既是金融的新业态，也是金融体系功能提升和结构变革的推动者”

中国人民大学金融与证券研究所所长

吴晓求教授

除了你的才华，其他一切都不重要！

近年来，互联网和人工智能技术的飞速发展，推动传统金融大踏步前进，尤其是量化投资、互联网金融、移动计算等领域，用一日千里来形容亦不为过。2015年年初，李克强总理在政府工作报告中提出制定“互联网+”行动计划，推动移动互联网、云计算、大数据等与各行业的融合发展。2015年9月，国务院又印发了《促进大数据发展行动纲要》，提出“推动产业创新发展，培育数据应用新业态，积极推动大数据与其他行业的融合，大力培育互联网金融、数据服务、数据处理分析等新业态”。可见，大数据金融将会成为未来十年最闪亮的领域。2012年年初，中国量化投资学会联合中国工信出版集团电子工业出版社，共同策划出版了“量化投资与对冲基金丛书”，深受业内好评。在此基础上，我们再次重磅出击，整合业内顶尖人才，推出“大数据金融丛书”，引领时代前沿，助力行业发展。

本书特点

李勇教授和许荣教授的这本《大数据金融》是“大数据金融丛书”中入门性质的书籍，通过浅显简单的案例，全面地介绍了大数据金融的方方面面。

作为为数不多的大数据金融的入门书籍，本书构建了一个系统性的分析框架，全面涵盖了国内外大数据金融发展的主要领域和最新成果。全书划分了银行、保险、量化投资、资产管理、金融监管、中国金融安全6个细分领域，对大数据金融在每个细分领域的影响进行了详尽分析，为读者提供了全景式的介绍。

第 1 章阐述了大数据对传统金融行业的影响与冲击，包括金融模式、金融业态、社会变革等各个方面。大数据绝不仅仅是一个技术性的辅助手段，而是从根本上对金融产品产生了颠覆性的影响。

第 2 章通过支付宝、余额宝等案例说明了大数据对传统银行的影响。银行的主要业务包括：储蓄业务、贷款业务、支付业务。支付宝的便捷性使得越来越多的用户愿意将支付宝作为前端的支付工具，目前已经从线上蔓延到线下，银行反而成为背后的管道。余额宝则大大推进了中国利率市场化的进程，提高了银行的资金成本。P2P 借贷则以更加灵活的方式让供需双方直接联系，降低了中间成本。

对于保险领域，大数据从思维模式、商业模式、经营能力和产品服务等方面全面提升了保险的品质。例如，可以利用大数据分析进行保险产品的精确定价，利用互联网进行保险产品的销售和客户服务。保险产品相对复杂，利用互联网方式可以精准地了解客户的需求，开发出适应性的保险产品，从而满足客户的需求。

至于量化投资，更是大数据金融的典型应用，从选股、择时、对冲套利、资产配置等各方面，投资的依据从以前主要靠分析师的主观判断，到现在更多依靠数据、模型和 IT 系统，这是历史的一个必然趋势。由于量化投资的各种优势，其在国外已经成为主要的投资方式之一，在国内也正在茁壮发展中。

资产管理行业毫无疑问是未来中国具有巨大潜力的领域，其中客户开发、风险管理、资产配置等，都离不开对大数据的分析。第 5 章以阿里金融的“跨界”业务、国金的“触网”业务为案例，为读者展示了未来的这幅画卷。

当然，随着新业务的发展，金融行业的监管和安全也面临着新的课题。例如，在金融监管中，如何保护个人隐私，如何控制数据风险，以及个别业务人员

的鲁莽行为等。现在的系统响应速度很快，个人的误操作可能会带来灾难性的影响，所以风险控制与监管尤其重要。既不能因为监管耽误了市场创新，也不能因为过度创新带来市场的不可控。

而在金融安全领域，大数据也有着非常重要的作用，包括利用大数据模型的分析来寻找异常交易行为、打击经济犯罪、反洗钱等。总而言之，大数据在金融行业中的应用才刚刚开始，未来发展空间异常广阔。李勇教授和许荣教授的这本《大数据金融》入门书籍，第一次全面地阐述了大数据金融的重要方面，值得读者仔细研读。

美好前景

中国经济经过几十年的高速发展，各行各业基本上已经定型，能够让年轻人成长的空间越来越小。未来十年，大数据金融领域是少有的几个有着百倍、甚至千倍成长空间的行业。在传统的、以人为主的分析逐步被数据和模型代替的过程中，从事数据处理、模型分析、交易实现、资产配置的核心人才（我们称之为“宽客”），将有广阔的舞台可以充分展示自己的才华。在这个领域中，将不再关心你的背景和资历，无论学历高低，无论有无经验，只要你勤奋、努力，脚踏实地地研究数据、研究模型、研究市场，实现财务自由并非遥不可及的梦想。对于“宽客”来说，除了你的才华，其他一切都不重要！

丁鹏 博士

中国量化投资学会 理事长

《量化投资——策略与技术》作者

“大数据金融丛书”主编

2015.9 上海

前言

2013 年以来，大数据浪潮迅速席卷全球。短短数年间，大数据应用在政府管理、新闻传播、市场营销等各个领域蓬勃发展，取得了累累硕果。新的联系和规律被不断发掘，而传统的思维和模式则被纷纷颠覆：谷歌搜索比医学家更早地预报流感趋势，社交网站比朋友更熟知我们的性格特征，数据分析师比客户经理更了解市场的消费风向……这些都得益于大数据技术，一切信息都可能被深入利用，并创造出巨大的社会财富。为了发挥大数据对社会经济的推动作用，近年来，美国、日本等国家相继实施了大数据发展规划，开启了国家间的“大数据竞赛”。在国内，贵州、北京、安徽等地也先后将大数据作为重点支持的战略新兴产业，积极打造“数据高地”。2015 年 10 月，中共十八届五中全会明确将国家大数据战略纳入我国“十三五”规划。这些变革都标志着我国已经步入了崭新的“大数据时代”。

作为现代经济的核心，敏锐的金融行业正在积极拥抱大数据技术。大数据金融应运而生，为金融行业带来了广泛而深远的变革。首先，大数据金融成为金融创新的重要源泉。从互联网货币基金的异军突起，到量化投资的日渐风行，再到证券公司与科技公司的强强联合，革命性的产品服务和经营模式层出不穷，形成了百舸争流、竞相开拓的新局面。其次，大数据也给金融行业带来了大挑战。难以察觉的隐私泄露、猝不及防的鲁莽行为、无处不在的数据风险……诸多因素不仅扩大了金融行业的系统性风险，也带来了新的信息风险，为大数据金融的良

性发展埋下了隐患。最后，机遇和挑战的冲突也对金融行业的监管层提出了更高的要求，带着监管内容、监管方式和监管体系的不断升级。

在大数据金融的业务实践高歌猛进之时，大数据金融的理论研究却发展滞后。火热的市场，缺少冷静的审视与思考。先行的实践，急需理论的跟进与夯实。大数据金融为何兴起，又将如何发展？分散在金融行业各领域、各公司的大数据实践是否可以凝练出普适性的模式和理论？类型众多的大数据金融风险是否存在系统性的治理框架？立足于本土的大数据金融实践，中国金融行业如何实现持续的业务创新和模式升级？在大数据金融的发轫之初，这些问题的研究不仅具有前瞻性的理论价值，更能够为大数据金融的实务发展提供指导性的意见。但放眼国内，尚未有相关著作问世，实为憾事。为此，我们撰写了这本《大数据金融》，希望能在这一领域作出绵薄的贡献，起到抛砖引玉的作用。

在结构设计和内容写法上，本书具有如下三个特点：

(1) 内容系统全面。本书构建了一个系统性的分析框架，全面涵盖了国内外大数据金融发展的主要领域和最新成果。在这一框架中，本书划分了银行、保险、量化投资、资产管理、金融监管、国家金融安全 6 个细分领域，对大数据金融在每个细分领域的影响进行了详尽分析，从而为读者提供了大数据金融全景式的介绍。

(2) 研究扎实深入。本书并非是对现有实践经验的简单搜罗与陈述。相反，在归纳实践经验的基础上，本书侧重对大数据金融的发展态势和发展前景进行深层次、理论性的探讨和研究。在每一个细分领域，本书都总结、提炼了大数据金融的发展模式，发掘了现有问题，并结合中国实际提出了可能的解决方案。

(3) 案例丰富生动。在理论研究之余, 本文也为读者呈现了国内外大数据金融应用的大量真实案例, 并进行了细致入微的剖析。这些鲜活生动的案例缩短了读者与市场的距离, 有助于读者更加准确、深入地理解本书的各部分内容。

在本书的撰写过程中, 作者参考了国内外大量的文献和新闻资讯, 在此向这些资料的作者表示衷心的感谢。此外, 本书的写作也得到了研究生钱智俊、李刚等的大力支持和帮助, 在此一并表示感谢。由于大数据金融的发展日新月异, 且作者水平有限, 本书难免存在不足之处, 恳请读者批评指正。

目 录

| | |
|--------------------------------|----|
| 第 1 章 绪论 | 1 |
| 1.1 迈入大数据时代..... | 1 |
| 1.1.1 大数据是什么..... | 2 |
| 1.1.2 大数据的产生：计算机与互联网技术的发展..... | 5 |
| 1.1.3 大数据与社会变革..... | 8 |
| 1.1.4 大数据与思维变革..... | 9 |
| 1.2 大数据金融..... | 12 |
| 1.2.1 大数据金融的理论基础..... | 13 |
| 1.2.2 大数据金融模式..... | 16 |
| 1.2.3 大数据对金融业态的影响..... | 19 |
| 1.3 大数据，大挑战..... | 19 |
| 第 2 章 大数据与银行 | 22 |
| 2.1 大数据金融与商业银行的特征对比..... | 23 |
| 2.1.1 业务来源..... | 23 |
| 2.1.2 经营风格..... | 24 |
| 2.1.3 成本..... | 24 |
| 2.1.4 门槛..... | 25 |

| | | |
|--------------|----------------------------------|-----------|
| 2.1.5 | 大数据的使用 | 25 |
| 2.2 | 目前几种大数据金融模式对商业银行业务的影响 | 26 |
| 2.2.1 | 以支付宝为代表的互联网第三方支付对银行 的影响 | 26 |
| 2.2.2 | 以余额宝为代表的互联网货币基金对银行 的影响 | 29 |
| 2.2.3 | P2P 借贷及阿里金融 | 32 |
| 2.3 | 大数据金融与商业银行的竞争趋势 | 35 |
| 2.3.1 | 大数据金融尚未形成较为成熟的商业模式 | 36 |
| 2.3.2 | 大数据金融还将承受外部的压力 | 37 |
| 2.4 | 总结 | 41 |
| 第 3 章 | 大数据与保险 | 43 |
| 3.1 | 大数据对传统保险理论的影响 | 44 |
| 3.1.1 | 大数据对保险定价机制的影响 | 44 |
| 3.1.2 | 大数据对于保险去中心化的影响 | 46 |
| 3.1.3 | 大数据对于可保风险的影响 | 47 |
| 3.2 | 大数据对保险公司经营的影响 | 48 |
| 3.2.1 | 大数据对保险业思维模式的影响 | 49 |
| 3.2.2 | 大数据对保险业商业模式的影响 | 49 |
| 3.2.3 | 大数据对保险公司经营能力的影响 | 51 |
| 3.2.4 | 大数据对保险公司产品服务的影响 | 53 |

| | | |
|------------|------------------|-----------|
| 3.2.5 | 大数据对保险业业务链条的具体影响 | 54 |
| 3.3 | 大数据与保险的案例分析和总结 | 58 |
| 3.3.1 | 中国保信 | 58 |
| 3.3.2 | 众安在线 | 59 |
| 3.3.3 | 去哪儿网获得保险经纪业务牌照 | 62 |
| 3.3.4 | 案例总结 | 63 |
| 第4章 | 大数据与量化投资 | 66 |
| 4.1 | 量化投资概述 | 66 |
| 4.1.1 | 量化投资的发展历程 | 66 |
| 4.1.2 | 量化投资的主要策略 | 70 |
| 4.1.3 | 量化投资的主要工具 | 74 |
| 4.2 | 大数据在量化投资中的应用 | 75 |
| 4.2.1 | 高频交易数据的应用 | 75 |
| 4.2.2 | 非结构化数据的应用 | 78 |
| 4.2.3 | 大数据应用的挑战性与局限性 | 83 |
| 第5章 | 大数据与资产管理 | 86 |
| 5.1 | 大数据资管时代来临 | 86 |
| 5.2 | 金融机构资产管理变革 | 90 |
| 5.2.1 | 大数据与风险管理 | 90 |
| 5.2.2 | 大数据与客户开发 | 93 |

| | |
|--|------------|
| 5.3 案例分析 | 95 |
| 5.3.1 阿里“跨界” | 96 |
| 5.3.2 国金“触网” | 103 |
| 5.3.3 总结 | 109 |
| 第 6 章 大数据与金融监管 | 110 |
| 6.1 如何监管金融企业对大数据的使用 | 110 |
| 6.1.1 保护金融大数据中的个人隐私：从“个人授 权制”到“企业负责制” | 110 |
| 6.1.2 规范金融大数据的应用：控制“数据风险” 和“鲁莽行为” | 119 |
| 6.1.3 明确金融大数据的产权：激活交易与创新， 打破封闭和垄断 | 124 |
| 6.2 如何利用大数据加强对金融市场的监管 | 130 |
| 6.2.1 技术问题和市场监管特殊性 | 130 |
| 6.2.2 美国证券市场的大数据监管实践 | 131 |
| 6.2.3 中国证券市场和银行系统的大数据监管实践 | 135 |
| 第 7 章 大数据与中国金融安全 | 138 |
| 7.1 大数据对中国金融安全的新挑战 | 138 |
| 7.1.1 大数据与中国金融信息安全 | 138 |
| 7.1.2 大数据与中国金融市场风险 | 141 |

| | |
|--------------------------------|-----|
| 7.2 各国对大数据与金融安全问题的处置经验 | 150 |
| 7.2.1 应对大数据带来的金融信息安全问题 | 150 |
| 7.2.2 应对大数据带来的金融市场风险 | 151 |
| 7.3 我国应对大数据与金融安全问题的原则及措施 | 158 |
| 7.3.1 应对挑战的“两个原则” | 158 |
| 7.3.2 综合处置的“三个方面” | 159 |
| 参考文献 | 162 |

第 1 章 绪 论

1.1 迈入大数据时代

清晨，当你独自走入小区花园散步，手机 GPS 已经将你的位置信息报告给了运营商；当你打开电脑开始工作，你的所有操作信息已经即时传输给了客户端；当你拿起手机，给朋友打个电话，或者是发一条微博，这些信息已经立刻被终端抓取存储；即使你只是在走廊徘徊了两步，你的身影也早已经被监控记录。

无声无息地，我们迈入了这个充满奇妙又令人畏惧的大数据时代。奇妙之处在于，即使足不出户我们就可以掌控天下大势，不费吹灰之力就可以得到来自世界的信息；而令人生畏的是，个人所有的言论、行为甚至是思考都赤裸裸地暴露在他人面前。我们可以随时在网站查看、购买心仪的商品，但有些时候或许仅仅多搜索了几个关键词（比如雷管、炸药、菜刀），就会突然地遭到“查水表”。在这个时代里，一切都被记录，一切都被量化，一切都被搜索和窥探。

从 330 年君士坦丁堡建立到 15 世纪的 1200 多年间，欧洲所有的手抄本书籍约 800 册；1439 年古登堡发明了印刷机，1453—1503 年的 50 年间，欧洲印刷的书籍就已经超过了这个数目；也就是说，50 年记录下的信息等价于过去 1200 年。而到今天，大约每 3 年，世界的信息存储量就翻一番。到 2013 年，世界上存储的数据超过了 1.2 ZB^①，其中非数字数据的占比不到 2%。把这些数据全部记载到书中，可以覆盖美国 52 次；如果存储在光盘上，高度可以达到地球和月球之间距离的 5 倍。曾经，亚历山大图书馆号称可以代表世界所有的知识；而在今天，每个地球人即时可获取的信息就超过了它的 320 倍^②。

毫无疑问，我们迈入了大数据时代！

1.1.1 大数据是什么

根据维基百科的定义，大数据（Big Data）又称为巨量数据或海量数据，是指涉及的数据量规模巨大到无法通过人工或者目前主流软件工具，在合理时间内达到截取、管理、处理并可以被人类解读的信息。网络上的每一次搜索、购物网站上的每一笔交易、资金的每一次流动，都被纳入到一个数据集合中，通过计算机的筛选、整理、分析，能够从这些数据中得到可信的规律和结论，用于帮助政府管理、企业决策及个人行为判断。

① 1024GB=1TB, 1024TB=1PB, 1024PB=1EB, 1024EB=1ZB。

② [英]维克托·迈尔-舍恩伯格，等. 大数据时代. 杭州：浙江人民出版社，2013.1.