

大数据丛书 “十二五”国家重点图书出版规划项目

Broadview  
www.broadview.com.cn

上海研究院智库丛书

# 互联网金融 与 大数据分析

庞引明 张绍华 宋俊典 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

大数据丛书

项目

上海研究院智库丛书

# 互联网金融 与 大数据分析

庞引明 张绍华 宋俊典 著

电子工业出版社

Publishing House of Electronics Industry

北京·BEIJING

## 内 容 简 介

互联网金融作为传统金融行业与互联网精神相结合的新兴产物,近年来可谓风生水起,对传统金融业的影响已不容小觑。在大数据时代,如何从理论高度审视互联网金融与大数据分析的方方面面,是本书所关注的内容。

本书从金融业务、客户服务、金融监管、信息技术、大数据和金融本质 6 个维度深入剖析互联网金融的内涵,阐明了互联网金融与大数据的相互关系,描述了互联网金融环境下的大数据结构、元数据和大数据应用架构,分享了大数据分析模型在互联网金融风控中的应用,以及互联网金融与大数据发展展望。本书具有一定的可读性和可操作性,对互联网、金融及大数据领域相关人士具有较高的参考价值。

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

互联网金融与大数据分析 / 庞引明, 张绍华, 宋俊典著. —北京: 电子工业出版社, 2016.5  
(大数据丛书)

ISBN 978-7-121-28419-9

I. ①互… II. ①庞… ②张… ③宋… III. ①互联网—应用—金融—数据处理 IV. ①F830.41

中国版本图书馆 CIP 数据核字(2016)第 057352 号

策划编辑: 刘 皎

责任编辑: 郑柳洁

印 刷: 三河市华成印务有限公司

装 订: 三河市华成印务有限公司

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: 14.75 字数: 240 千字

版 次: 2016 年 5 月第 1 版

印 次: 2016 年 5 月第 1 次印刷

定 价: 69.00 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, 88258888。

质量投诉请发邮件至 [zltz@phei.com.cn](mailto:zltz@phei.com.cn), 盗版侵权举报请发邮件至 [dbqq@phei.com.cn](mailto:dbqq@phei.com.cn)。

本书咨询联系方式: (010) 88254395。

## 推荐序一

互联网金融作为一种新兴的金融业态，它不是互联网和金融业的简单结合，而是在电子商务快速发展的基础上，自然而然为适应新的需求而产生的新金融模式和新业务形态，它对当前国家经济转型和“互联网+”计划的实施具有重要的实践意义。

互联网金融与传统金融的区别不仅仅在于从事金融业务所采用的渠道媒介不同，更重要的在于互联网金融参与者深谙互联网“开放、平等、协作、分享”的精髓，通过互联网、移动互联网等工具，使得传统金融业务具备了透明度更强、参与度更高、协作性更好、中间成本更低、操作上更便捷等一系列特征。互联网金融能够对传统金融形成非常有益的补充，为国家经济的发展提供有力支持。

近期，互联网金融发展迅猛，一方面，各类互联网金融平台如雨后春笋般涌现，各类互联网金融产品琳琅满目。另一方面，不断出现互联网金融平台提现困难甚至倒闭的现状，真可谓冰火二重天。我们密切关注互联网金融的行业态势，看到了这一行业的巨大价值，也顺势成立了绿地互联网金融平台，但是我们觉得互联网金融野蛮生长的态势应该稳中求进，互联网金融行业的健康发展需要适当的金融监管，行业的持续发展需要一定的理论支持。

本文从专业的视角详细解读互联网金融和大数据分析的相关理论，并结合大数据技术的应用，全面系统地分析了互联网金融的风险控制模型，从一定程度上解决互联网金融面临的一些共性问题，为互联网金融行业的健康持续发展提供技术支持和理论指导。我相信，本书的出版对互联网金融行业的发展具有很强的指导意义。

绿地集团副总裁 绿地金融控股集团董事长、总裁  
耿靖

## 推荐序二

数据是“互联网+”的基础，只要有互联网的应用，就会产生海量的数据沉淀，形成大数据。过去，在行业与行业之间，政府各个部门之间，数据都是局限的、难以共享的。如果可以把涉及商业、农业、民政、医疗等不同领域的数据进行汇集、聚合、打通，形成一个共享、开放的大数据，将对经济发展起到很大推动作用。

金融行业一直是大数据的重要产生者和消费者。金融系统中涉及的账户、交易、价格、风控、投资分析等都可以成为数据的来源。同时，金融行业也高度依赖数据，依赖数据进行风险测量，依赖数据进行产品定价。因此，可以认为金融业是数据驱动的行业。

在互联网金融环境中，大数据作为一种重要的核心资产，对它的分析和使用，将会极大地帮助金融机构更好地了解客户，与客户建立更良好的客户关系管理，同时可以根据数据分析进行精准营销。大数据分析日益成为互联网金融发展的利器，没有大数据分析，互联网金融领域的发展缺少智力和理性基础；离开互联网金融领域的土壤，大数据分析难以体现其应用价值和发挥其用武之地。互联网金融的发展，产生了数量大、结构复杂、格式多样、时间和空间多维度的大数据流，同时也对数据分析的实时性、准确性、定制化等方面提出更高要求。

本文深入分析了互联网金融领域的大数据结构、构建了面向互联网金融的大数据应用架构、建立互联网金融的风控模型和征信框架，是国内少有的系统分析大数据与互联网金融的理论书籍。相信本书的出版可以为互联网金融行业的发展提供重要的技术支持，具有非常重要的参考意义。

阿里巴巴副总裁 阿里研究院院长

高红冰

### 推荐序三

作为最近几年在媒体上出现频率最高的热门词汇，大众对“互联网金融”和“大数据”的讨论，一直都超乎寻常的热烈。尽管互联网金融只是金融业在互联网时代商业模式转型的一种过渡性现象，但不可否认的是，它正从业务渠道、成本结构、运营效率及赢利模式等诸多方面改变金融行业。当然，这一切的充分实现，都需要以一个完善、多元化征信体系的建立为前提。这可以包含央行征信信息、社交、电商及其他网络行为等大数据，全面真实、客观可靠、开放分享的数据体系，以及在此基础上展开的大数据挖掘与分析，必将大幅降低信息不对称，从而有效缓解国内个人和中小企业获取金融服务既困难又昂贵的问题，进而提升中国金融运行的效率。因此，对于互联网金融与大数据分析的研究，尤其是从实务操作角度进行的有益探讨，其社会需求的急迫性和实践指导意义的重要性就愈发凸显。

当下，市面上存在一些分别以互联网金融和大数据为主题展开分析的书籍，但或许受制于新兴及跨学科交叉领域研究的难度，从大数据分析的角度探讨互联网金融运行的国内著作目前当属凤毛麟角。作为中国互联网金融和大数据领域资深的行业专家，作者不仅具备扎实的理论功底，而且拥有丰富的实战经验，因此，本书融合互联网金融与大数据分析所涉及的多学科知识，不仅写出了理论的高度，同时也写出了实践的深度。从某种意义上说，本书也是作者长期践行跨学科研究，推动协同创新的产物。

本书兼顾理论沉淀与实践创新，将跨学科融合的思维贯穿全程，以独特的视角观察和研究互联网金融与大数据分析的相互关系，并从互联网金融环



境下的大数据应用及大数据分析支持下的互联网金融两大角度出发，重点研究了互联网金融领域的大数据结构、具体应用及架构，同时建立了互联网金融的风控模型，也就大数据分析处理相关的应用展开探讨，针对如何解决互联网金融实践过程中的一系列问题，提出了独特见解。

本书记录了作者在长期一线工作实践过程中的理解和思考，对互联网金融和大数据分析领域进行了有意义的探讨和总结。相信它能给相关领域的学者、从业人员、有兴趣了解或有志进入该领域的人士提供一些新的思路，引发一些新的思考，进而推动实践，共同促进中国互联网金融的发展。

林采宜

2015年11月

## 推荐序四

伴随着信息技术的飞速发展，人类社会已经进入大数据时代。数据已经被人们公认为企业和社会发展的关键要素，企业的重要资产，数据洞察力已经成为企业的核心竞争力之一。金融行业作为大数据的重要产生者和消费者，收集和加工数据成为互联网金融行业重要的基础工作。

在竞争日益加剧，环境日益复杂的经济形势下，通过对大数据的收集和整理，利用数据分析和数据挖掘的方法，对所授信用个人或企业进行充分有效的评估，建立完善的征信和风险评估体系是互联网金融企业必备的基本功。能否利用大数据构建一个完善的风险识别和评估系统，是互联网金融企业生死存亡的关键。

在互联网金融大环境下，通过对数据的收集、规整、分析和挖掘，以数据信息为基础建立客户经理工作管理平台，推动客户的精准营销，建立端到端的营销管理流程，帮助金融机构更好地服务客户和开拓市场。在与客户建立初步合作意向后，通过大数据分析建立客户风险评级模型，充分评估客户的信用风险水平，帮助金融机构准确定位客户的风险水平，提高抗风险识别和风险管理能力。

本书全面深入地分析了互联网金融的特点，剖析了大数据与互联网金融的关系，概述了大数据在互联网金融行业中的应用，构建了互联网金融领域内的大数据分析框架和征信评估模型，对互联网金融行业的发展路径进行了深入的探索。本书为大数据和互联网金融行业发展提供了重要的理论支持和

技术实践，是互联网金融领域内不可多得的好书。另外，本书对于大数据相关的技术和发展前景也进行了较为深入的分析，可以作为这方面知识的参考书籍。

吉贝克信息技术（北京）有限公司董事长，国家“千人计划”入选人  
中科院大学金融科技研究中心主任，XBRL 中国执行委员会副主席  
刘世平

## 推荐序五

互联网金融与大数据是当前最热的两个名词，互联网金融代表金融业务的最新发展方向，而大数据是信息技术发展到一定阶段形成的一种新的技术手段。正是大数据技术在数据获取、处理和分析技术上突破性的进展，为互联网金融的创新发展提供了技术前提。目前，对大数据技术论述的论文和书籍已经比较多了，但是在互联网金融的背景下，对大数据技术进行论述的书籍还不多。由于大数据是一个和应用背景高度结合的技术，因此本书的撰写对大数据技术和互联网金融的发展具有重要的意义。

本书的作者是三位活跃在互联网金融和大数据领域第一线的技术和管理人员，他们具有深厚的计算机和大数据技术基础，同时他们对互联网金融业务有着创造性地理解，所以他们的组合为本书的成功撰写提供了良好的基础。

书的内容从互联网金融出发，对大数据处理技术和互联网金融及其关联模式进行了系统介绍。特别是作者对互联网金融的六个维度进行了论述，很好地说明了互联网金融的内涵和基础，也明确了大数据在互联网金融中的作用。进而针对互联网金融的业务需要从数据采集、预处理、元数据管理、存储管理、分析等角度对大数据技术进行系统的论述，特别对互联网金融的风险控制及特殊的大数据处理方式进行了深入地探讨。

本书是少有的将互联网金融与大数据技术进行系统地融合的著作，相信本书的出版将对互联网金融的从业人员和大数据应用的技术人员有很大的指导意义。

复旦计算机学院副院长  
汪卫教授

## 前言

本书的作者都是工作在互联网金融和大数据领域第一线的人员，具有一定的理论与实践经验。作者们希望能以此书为契机，记录目前工作实践的理解和思考，给业内同行、专家学者、在校学生等带来一些新的思路和实践途径，进而能够共同在互联网金融和大数据分析领域进行探索、尝试和归纳总结，在理论和实践两个方向上，走出一条清晰而又合乎逻辑的业务与技术结合之路。

本书在内容上，从独特的视角审视互联网金融和大数据分析。其一，从一定理论高度，结合信息技术的应用，宏观分析互联网金融的风险控制模型和征信框架，并结合大数据共性技术的研究，进行理论上的思考和实践上的创新，解决部分互联网金融面临的共性问题；其二，本书拟从互联网金融环境下的大数据应用，以及大数据分析支持下的互联网金融两个角度，抽象出大数据分析的共性特征、模型等，从而为诸如互联网金融风控相关问题提供解决之道。

本书的写作思路是，在互联网金融和大数据分析等知识的基础上，结合相关的理论和应用分析，将工作一线的实践，以及实践过程中的理论思考也写入本书，供读者细细品读和思考。

本书作为电子工业出版社规划出版的“十二五”国家重点图书出版项目“大数据丛书”中的一本，它针对的读者对象主要是：经济金融专业、管理科学和工程专业、信息技术类专业的大学高年级学生和研究生，互联网金融

行业的骨干人员，金融单位的大数据分析和信息技术人员，互联网公司金融板块和大数据分析从业者，以及对互联网金融领域有兴趣、有志进入这一领域的各类人士。

互联网金融与大数据的分析相得益彰。大数据分析日益成为互联网金融发展的利器，没有大数据分析，互联网金融领域的发展将缺少智力和理性基础；离开互联网金融领域的土壤，大数据分析则难以体现其应用价值和发挥其用武之地。互联网金融的发展，产生了数量大、结构复杂、格式多样、时间和空间多维度的大数据流，同时也对数据分析的实时性、准确性、定制化等方面提出了更高要求。厘清互联网金融领域的大数据结构，构建面向互联网金融的大数据应用架构，建立互联网金融的风控模型和征信框架，并开展相关的应用，是本书的主要工作。

在互联网金融快速发展和大数据技术日新月异的时代，尽管作者全部工作在一线，也难免会有所疏漏；更重要的是，该领域内人才济济，各类精英正不断踌躇满志地进入这一蓝海领域。所以，如果本书在理论方法、论述手段、基本结论方面有什么错误和局限，首先希望能够得到各界巨擘的指导和斧正；其次，希望本书能够起到抛砖引玉的作用，引起读者的讨论和思考、推动实践，共同推动该领域的工作向前进。

本书共八章。第1~3章、第6章主要由庞引明博士撰写，第4章、第5章、第7章和第8章主要由张绍华博士、宋俊典博士撰写。在本书撰写过程中，戴炳荣、杨琳、高洪美、赵增奎、陈美等人参与了具体章节的编写，在此对各位的努力和贡献表示衷心的感谢！写书是一件辛苦的事情，单纯从收益上来看，这与作者所从事的行业收益不成比例，但是书籍是行业发展的基础工程，总得有人做，我们愿意为之贡献我们的精力、智慧和时间。

希望本书能够有益于您——我们的读者。

庞引明 张绍华 宋俊典  
2015年夏于上海

第 1 章

观察互联网金融的  
六个维度

- 1.1 金融业务：传播手段和颠覆效果 / 2
- 1.2 金融客户：金融普惠化和金融民主化 / 11
- 1.3 金融监管：风险控制的复杂化 / 13
- 1.4 信息技术：互联网基因导致的  
粒度混合化 / 25
- 1.5 大数据：数据碎片化和信息立体化 / 36
- 1.6 金融的本质：回归本源化带来的  
融资新模式 / 44

第 2 章

互联网金融与大数据  
的相互关系

- 2.1 人类数据科学的发展历程及大数据  
处理的历史 / 47
- 2.2 互联网金融产生大数据和大数据  
处理需求 / 56

- 2.3 大数据分析支撑互联网金融在  
多方面快速发展 / 58
- 2.4 互联网金融时代大数据  
处理面临的挑战 / 59

### 第 3 章

## 互联网金融领域复杂 的大数据结构

- 3.1 大数据领域主流数据结构 / 62
- 3.2 结构化、半结构化和非结构化数据 / 66
- 3.3 流数据 / 76
- 3.4 时空数据 / 80
- 3.5 传感数据和全息数据处理 / 84

### 第 4 章

## 元数据及其在互联网 金融领域的应用

- 4.1 元数据的基本概念 / 87
- 4.2 XML 的主要概念和基本结构 / 100
- 4.3 基于 XML 的元数据模型  
描述和应用 / 108
- 4.4 互联网金融领域的元数据描述分析 / 123

### 第 5 章

## 面向互联网金融的 大数据应用架构

- 5.1 大数据应用概述 / 132
- 5.2 互联网金融大数据应用研究 / 134



- 5.3 互联网金融服务模式及平台研究 / 139
- 5.4 基于大数据的互联网金融业务创新 / 144

## 第 6 章

### 互联网金融的 风控模型

- 6.1 互联网金融领域的风险分析 / 150
- 6.2 基于图的互联网金融风控模型 / 161
- 6.3 基于集合论的互联网金融风控模型 / 171

## 第 7 章

### 面向互联网金融的 大数据分析处理

- 7.1 大数据采集与存储 / 177
- 7.2 大数据分析与挖掘 / 188
- 7.3 SaaS 与互联网金融大数据 / 202
- 7.4 物联网金融大数据 / 206

## 第 8 章

### 互联网金融及大数据 发展展望

- 8.1 互联网金融的发展趋势 / 209
- 8.2 大数据的发展趋势及展望 / 211

参考文献 / 216