

作者是资深风控算法专家，
在头部互联网公司和综合性国际化资产管理集团工作多年，经验丰富

中国科学院院士推荐



蔡主希 著

INTELLIGENT RISK CONTROL AND ANTI-FRAUD
Systems, Algorithms and Practices

智能风控 与反欺诈

体系、算法与实践

系统讲解智能风控和反欺诈在信贷领域应用，
从业务和技术双维度讲解传统风控体系与智能风控算法和模型的使用



机械工业出版社
China Machine Press

智能风控与反欺诈：体系、算法与实践

蔡主希 著

ISBN: 978-7-111-67625-6

本书纸版由机械工业出版社于2021年出版，电子版由华章分社（北京华章图文信息有限公司，北京奥维博世图书发行有限公司）全球范围内制作与发行。

版权所有，侵权必究

客服热线：+ 86-10-68995265

客服信箱：service@bbbvip.com

官方网址：www.hzmedia.com.cn

新浪微博 @华章数媒

微信公众号 华章电子书（微信号：hzebook）

目录

推荐序

前言

第1章 互联网金融与风险管理

1.1 互联网金融的发展和现状

1.2 风险管理类型划分

1.2.1 欺诈风险

1.2.2 信用风险

1.3 风险管理的重要性

1.3.1 风险评估

1.3.2 差异化定价

1.3.3 整体利润最优

1.4 本章小结

第2章 传统风险管理体系

2.1 人工审核

2.1.1 纸质材料评估

2.1.2 电话回访

2.1.3 线下走访尽调

2.2 专家模型

2.2.1 业务规则库

2.2.2 专家调查权重法

2.2.3 熵权法

2.3 评分卡模型

2.3.1 目标定义

2.3.2 样本选取

2.3.3 变量分箱

2.3.4 变量筛选

- 2.3.5 模型建立
- 2.3.6 模型评估
- 2.3.7 模型应用
- 2.4 传统方法的问题和挑战
- 2.5 本章小结
- 第3章 智能风控模型体系
 - 3.1 大数据平台
 - 3.1.1 原始数据清洗
 - 3.1.2 数据仓库管理
 - 3.1.3 数据标签应用
 - 3.2 决策引擎
 - 3.2.1 规则配置
 - 3.2.2 模型部署
 - 3.2.3 冠军挑战者
 - 3.2.4 版本和权限管理
 - 3.3 智能反欺诈模型
 - 3.3.1 无监督学习
 - 3.3.2 图计算
 - 3.4 智能信用风险模型
 - 3.4.1 专家模型
 - 3.4.2 逻辑回归
 - 3.4.3 决策树
 - 3.4.4 集成树
 - 3.4.5 深度神经网络
 - 3.4.6 循环神经网络
 - 3.5 智能模型带来的提升
 - 3.5.1 数据广度和深度
 - 3.5.2 模型快速迭代和主动学习

- 3.5.3 线上自动决策
- 3.6 统计学与机器学习
- 3.7 本章小结
- 第4章 风控大数据体系
 - 4.1 数据源类型
 - 4.1.1 征信报告
 - 4.1.2 消费能力
 - 4.1.3 资产状况
 - 4.1.4 基本信息
 - 4.1.5 黑名单
 - 4.1.6 多头借贷
 - 4.1.7 运营商
 - 4.1.8 地理位置
 - 4.1.9 设备属性
 - 4.1.10 操作行为
 - 4.2 特征工程方法
 - 4.2.1 统计量
 - 4.2.2 离散化
 - 4.2.3 时间周期趋势
 - 4.2.4 交叉项
 - 4.2.5 隐性特征
 - 4.2.6 用户画像
 - 4.3 数据测试与应用
 - 4.3.1 联合建模机制
 - 4.3.2 数据质量评估
 - 4.3.3 线上应用
 - 4.4 数据安全合规
 - 4.5 本章小结

第5章 智能风控中的常用算法

5.1 有监督学习

5.1.1 逻辑回归

5.1.2 决策树

5.1.3 随机森林

5.1.4 梯度提升决策树

5.2 无监督学习

5.2.1 聚类

5.2.2 孤立森林

5.3 深度学习

5.3.1 深度神经网络

5.3.2 循环神经网络

5.3.3 词嵌入

5.3.4 自编码器

5.3.5 迁移学习

5.4 图计算

5.4.1 社区发现

5.4.2 标签传播

5.4.3 图嵌入

5.5 强化学习

5.6 本章小结

第6章 智能模型训练流程

6.1 数据清洗

6.1.1 缺失值处理

6.1.2 异常值处理

6.1.3 重复值处理

6.1.4 一致性检验

6.1.5 有效性检验

6.2 特征工程和特征筛选

6.2.1 探索性数据分析

6.2.2 稳定性

6.2.3 重要性

6.2.4 相关性

6.2.5 解释性

6.3 模型训练

6.4 模型部署

6.5 监控预警

6.6 本章小结

第7章 反欺诈案例

7.1 案例背景

7.2 原始数据介绍

7.3 探索性数据分析

7.3.1 交易笔数

7.3.2 交易时间

7.3.3 交易类型

7.3.4 交易IP地址

7.4 特征工程

7.4.1 特征加工

7.4.2 特征筛选

7.4.3 特征分组

7.5 模型训练

7.6 模型评估

7.7 案例优化

7.8 本章小结

第8章 个人信贷风控案例

8.1 案例背景

- 8.2 原始数据介绍
- 8.3 特征工程
- 8.4 探索性数据分析
- 8.5 模型训练
 - 8.5.1 逻辑回归
 - 8.5.2 XGBoost
 - 8.5.3 Wide&Deep
- 8.6 模型评估
- 8.7 模型应用
- 8.8 案例优化
- 8.9 本章小结
- 第9章 企业信贷风控案例
 - 9.1 银行POS贷
 - 9.1.1 案例背景
 - 9.1.2 原始数据介绍
 - 9.1.3 特征工程
 - 9.1.4 模型训练
 - 9.1.5 模型应用
 - 9.2 汽车金融CP评级
 - 9.2.1 案例背景
 - 9.2.2 原始数据&特征工程
 - 9.2.3 模型训练
 - 9.2.4 模型评估
 - 9.3 案例优化
 - 9.4 本章小结
- 第10章 智能风控能力对外输出
 - 10.1 对外输出的意义
 - 10.1.1 内部能力“走出去”

- 10.1.2 外部资源“引进来”
- 10.2 头部玩家介绍
 - 10.2.1 互联网公司
 - 10.2.2 银行科技子公司
 - 10.2.3 第三方技术提供商
- 10.3 合作模式及案例
 - 10.3.1 SaaS+本地化模式
 - 10.3.2 对外输出案例
- 10.4 金融科技创新与监管
- 10.5 本章小结

推荐序

金融科技是以大数据、人工智能、区块链等为代表的新一代信息技术最重要的应用场景之一，而智能风控与反欺诈又是金融科技的核心。经过最近几年的发展，金融科技在概念上已经被广泛接受，但在具体实践中仍然困难重重。其中最根本的原因是金融科技是一项系统工程。技术是必要的基础，但业务、体制机制、商业模式等也是不可或缺的。

蔡主希的这本书最大的优点就是它从实际案例出发，系统地讨论了智能风控和反欺诈在业务、技术、体制机制、推广模式等各个方面的主要问题和解决方案。它不是一本高高在上的技术书，而是一本全方位地指导我们如何让金融科技落地的基础教材。

在这本书的背后，我看到的是“成熟”二字。我还清晰地记得蔡主希刚刚走出校门来到北京大数据研究院工作时的情景。短短几年，看到他从一个只有书本知识的学生成长为一个金融科技领域的专家，并能够写出这样的书来，我感到非常欣慰。还记得为了落实某银行的一个项目，他扎根当地两个月，夜以继日地工作，阅读了所有能够看到的（手写）资料，系统研究了相关的业务。这样的经历使他对银行的借贷和风控有了一个整体的了解。这也是他能够写出这本书的基础。

大数据和人工智能的落地是一个非常繁杂、艰难的过程，需要很多像蔡主希一样的年轻人参与，也需要像《智能风控与反欺诈：体系、算法与实践》一样的书的指导。

中国科学院院士，普林斯顿大学教授，北京大数据研究院院长

鄂维南

前言

为何写作本书

互联网金融是中国金融行业数字化升级的产物。2020年监管部门出台了一系列举措，标志着这个行业从成长期正式步入成熟期。互联网金融机构的快速发展离不开依托于大数据和机器学习的风控及反欺诈能力，这些风险管理能力帮助机构在过去十年通过信贷业务赚取了利润，也是未来十年各机构开展金融科技业务的核心竞争力。

笔者有幸于毕业后进入互联网金融行业，在北京大数据研究院和头部互联网公司有过丰富的项目经验，与传统金融机构的业务人员也有过较多接触。在实际工作及与同行交流的过程中，笔者发现这个行业内还是存在一些信息不对称的问题：传统金融机构（包括银行、消费金融公司、信托等）的业务人员仍然习惯于接受传统评分卡那一套风控理念，对于“黑箱”的机器学习模型存在较多质疑；而互联网公司的建模人员往往过于注重算法的性能优化，却忽略了其与线上风控策略和信贷场景的结合。笔者也调研了市面上与风控相关的图书，多数图书专注于评分卡或者智能风控的单一方面，缺乏将两者有机结合的资料，这使得刚入行的新人很难对整个行业有一个完整的认知。与此同时，业务人员和技术人员对风控的理解也存在鸿沟。

笔者撰写本书，一方面是想降低整个信贷风控行业的门槛，帮助更多对此感兴趣的学生快速了解并加入这个行业；另一方面是想帮助各大公司的业务人员和技术人员拉齐彼此对于风控的认知，提高合作效率，促进整个金融行业的科技化转型。当然，笔者并不鼓励读者去互联网平台借贷，希望读者可以提高风险意识，谨慎操作，远离不法贷款平台。

本书读者对象

本书主要面向信贷风控行业的广大业务人员、策略分析师、数据分析师、算法工程师以及对互联网金融和智能风控、反欺诈感兴趣的读

者。

如何阅读本书

本书集合了目前互联网金融行业风控方面的业务和技术内容，辅以项目案例和代码实现，力求帮助读者通过一本书了解智能风控和反欺诈技术的全貌。

第1~3章涵盖风控业务的基础知识，主要介绍了什么是信用风险和欺诈风险，传统风险管理体系中搭建评分卡的思路，以及智能风控时代下大数据平台、决策引擎和智能模型的技术框架。这一部分普及了风控中的一些常见术语，可帮助技术人员更深入地了解风控决策的整个过程。对传统金融机构风险管理体系感兴趣的读者，可以进一步阅读业界大咖陈建老师、乔杨老师的书。

第4~6章介绍智能风控模型中常见的数据源和算法，其中数学原理和公式较多，适合想了解更多模型知识的业务人员以及想从事建模工作的读者。由于篇幅有限，笔者只选取了与风控和反欺诈场景相关的算法，很多公式的推导和求解过程并没有详细展开。想了解更多机器学习和深度学习算法的读者，可以进一步学习周志华老师、李航老师以及Ian Goodfellow等人的著作。

第7~9章讲解笔者参与过的风控和反欺诈的实战项目，希望帮助读者通过实际案例更好地将风控理念和建模技术融会贯通，缺乏项目经验的读者可以重点关注这部分内容。如果还想了解更多风控算法的Python代码实现，可以关注梅子行老师的书。

第10章是行业内金融科技的头部玩家解析和案例介绍，可帮助读者初步了解智能风控和反欺诈在未来十年的应用方向。金融科技目前在国内仍然处于探索阶段，市面上还没有系统介绍这方面知识的图书，感兴趣的读者可以关注相关头部公司或者权威媒体的宣传账号，紧跟行业动向。

资源与勘误

笔者在写作过程中参考了市面上已有的许多图书、论文和权威媒体文章，力争书中的每一处内容都是有据可循的。但由于笔者水平有限，书中难免会存在一些错误和疏漏，欢迎各位读者和专家批评指正。如果有任何关于本书的意见或建议，欢迎发送邮件到yfc@hzbook.com。

致谢

首先感谢机械工业出版社华章公司的杨福川老师和朱巍老师，以及其他为本书出版做出贡献的工作人员。没有各位的耐心解答和专业指导，我一定没有办法顺利完成这本书的写作工作。

其次感谢北京大数据研究院、京东数科、滴滴金融（按照入职时间顺序）以及其他合作公司的领导和同事。本书中的很多风控理念和建模思想，都源自与业内各位专家的交流，感谢同人们为中国金融行业数字化升级做出的贡献。

最后要感谢我新婚的妻子陪伴我度过无数个码字的午夜和周末，默默地支持我完成整本书的创作。在繁忙的工作之余还要利用休息时间写书是一件很困难的事情，有她的鼓励我才得以坚持下来。

第1章 互联网金融与风险管理

互联网金融是传统金融业务与新兴互联网技术结合的一个交叉领域，例如互联网公司开展的金融业务，或者金融机构的线上化服务，都属于互联网金融的范畴。与传统金融行业的线下模式相比，互联网金融以网络和通信技术为载体，受众群体更广，辐射场景更多，服务效率更高，符合“普惠金融”的核心理念。然而在互联网金融业务的开展过程中，由于机构人员与客户并没有实际接触，因此金融风险无处不在。在此背景下，依托于大数据和机器学习的智能风控与反欺诈技术应运而生，为互联网金融行业的健康发展锦上添花。第1章将先介绍互联网金融行业及其场景下风险管理的背景内容，更多智能风控与反欺诈的细节和案例将在后面的章节中展开。

1.1 互联网金融的发展和现状

广义的互联网金融（简称“互金”）包括互联网贷款、第三方支付、众筹、数字货币、网上银行等多种商业模式。其中互联网贷款无疑是起步最早且体量最大的一个场景，截止到2018年末，我国互联网贷款余额超万亿，覆盖了上百家金融机构。可以说国内互联网金融的发展史也是互联网贷款的兴衰史。

国内互联网贷款起源于2007年，当时的背景是中国经济发展达到了一个新高度，消费金融、小微金融需求空前旺盛；同年11月阿里巴巴在港交所上市，标志着国内互联网行业已经初具规模。在此基础上，2007年6月阿里巴巴与建行推出了“e贷通”产品，被视为互联网贷款的雏形；2007年7月国内第一家P2P平台拍拍贷在上海注册成立，标志着P2P行业的开端。

2008年~2014年是互联网贷款的发展阶段。首先是政策上的扶持，小贷牌照被下放到地方的金融机构，省级金融办拥有最终的批复权，大大降低了机构的准入门槛。其间，头部的互联网公司如阿里、腾讯、京东等纷纷成立小贷公司，杀入互联网金融行业。其次，P2P业务蓬勃发展，不少P2P公司考虑到合规的需求也申请了小贷牌照。这两方面因素促使国内小贷公司数量在那几年快速增长。

2015年~2018年，随着“助贷”模式的兴起，互联网贷款行业进入井喷阶段。2015年两款标志性的贷款产品“蚂蚁花呗”和“微粒贷”正式上线，连接了银行的资金和互联网公司的流量，也点燃了整个市场的热情。在这个阶段，国内小贷公司累计超过8000家，互联网贷款余额进入万亿规模。

随着2016年《网络借贷信息中介机构业务活动管理暂行办法》（即我们常说的“P2P管理办法”）、2019年《数据安全管理办法》、2020年《商业银行互联网贷款管理办法》等规定相继出台，国内的互联网贷款行业逐步告别高杠杆、高风险的野蛮扩张时期，转而进入风险可控的健康发展阶段。在强调合规的现状下，银保监会对于各家互联网金融机构的风险管理能力提出了更高的要求，大数据、机器学习等金融科技技术必然会在互联网贷款发展的下一个阶段扮演更加重要的角色。

1.2 风险管理类型划分

互联网金融行业中的风险主要可以分为两类：欺诈风险和信用风险。

这两类风险产生的原因和背后的规律存在一定差异，因此需要风控人员采用不同的手段去甄别和预防。然而，这两类风险都会造成贷款的逾期和不良，轻则对于金融机构造成经济上的损失，重则引发严重的品牌和舆论危机，金融机构将陷于难以挽回的境地。

欺诈风险是指借款人带着欺诈的目的来申请贷款，资金被成功获取后再难收回。这是信贷场景中危害较大的一类风险，占比较低，但是必须严格防范。另外，欺诈风险存在专业性、团体性、变化性等特质，互联网金融行业的快速发展催生了一大批“羊毛党”和“黑中介”，他们通常拥有专业的欺诈技术，并且以团伙形式出现，不断攻击着各个平台的风控漏洞，长期考验着风控人员博弈的能力。

欺诈风险多发生在贷款的申请和支用阶段，由于手段样式繁多，目前业界对欺诈风险并没有明确的类型划分。笔者结合自身业务经验，将欺诈风险从如下三个角度进行分类：

- 从欺诈意愿上分为第一方和第三方；
- 从欺诈主体上分为账户级和交易级；
- 从欺诈组织上分为个人和团伙。

信用风险则是信贷场景中另一类较为常见的风险，与欺诈风险相比，信用风险出现频率更高，是贷款逾期和不良的最主要因素。个体的信用通常可以从还款意愿和还款能力这两个方面来衡量，一个信用良好的借款人一定是同时具有较强还款意愿和较高还款能力的人。深究信用风险的产生，主要还是个体的信用状况与平台方授予的额度、利率不匹配所导致的，并且由于个体信用会随着贷款的不同阶段而发生变化，因此对于信用风险的管理必定是贯穿准入、贷前、贷中、贷后整个贷款的生命周期。

欺诈风险和信用风险管理环节的组合顺序如图1-1所示，在风险可控的前提下，尽量提升各个环节之间的转化率。

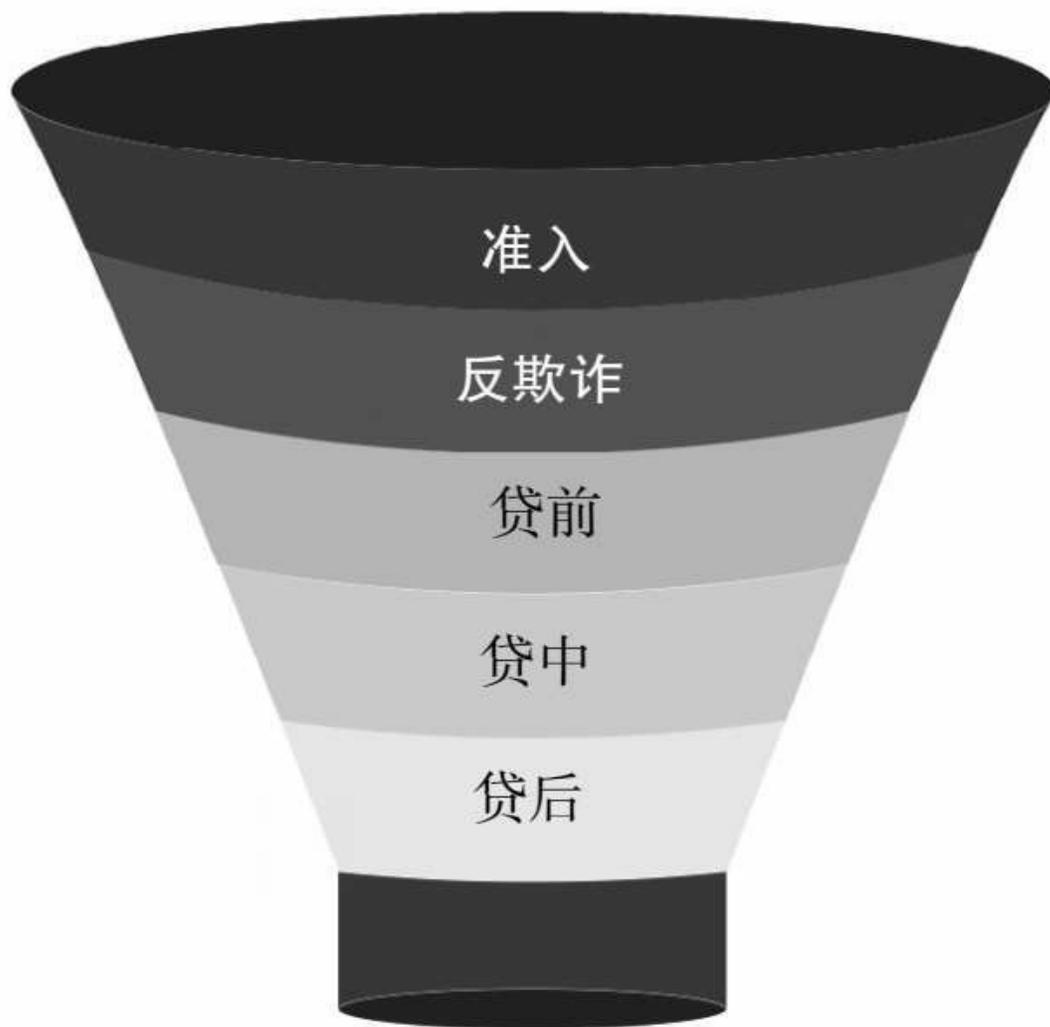


图1-1 欺诈风险和信用风险管理流程

1.2.1 欺诈风险

1. 第一方和第三方

所谓第一方欺诈，是指借款人主动发起的欺诈行为；相对的，第三方欺诈是指借款人在身份被冒用或者账户被盗用的情况下，被动发生的欺诈行为。第三方欺诈的排除是客户进入申请或者支用流程后需要进行的第一步工作，因为只有确认了本人操作，后续其他的欺诈风险和信用风险识别才有意义。对于第三方欺诈的识别，目前各大金融机构普遍采用四要素验证（姓名、身份证、手机号、银行卡号）和活体识别这两个技术手段，有效遏制了非本人操作行为的发生。但是，对于中介申请或者电信诈骗，由于借款人是在欺诈分子诱导下进行的本人操作，单纯利用四要素验证或者活体识别并不能很好地进行判断，因而还需要加入其他反欺诈策略，进一步提高召回率。

相较于第三方欺诈，第一方欺诈隐蔽性更强，并且手段灵活多变，提高了风控人员的工作难度，目前业界常用的手段是针对首支首逾和“羊毛党”这两类特定客群进行识别。

首支首逾是指借款人在首次支用后的首个还款日逾期的行为，如果逾期时间较长，则会被机构认定为欺诈客群。首支首逾率通常是各个机构考核反欺诈人员的重要指标，搭建针对性的模型和策略方案，是反欺诈人员的工作重点之一。

还有一类常见的第一方欺诈是“羊毛党”，“羊毛党”分子通过非法手段获取个人信息，注册虚假账户，用来“薅取”互联网金融机构发放的现金券、礼品券等，让机构受损失。对于“羊毛党”，可以从设备指纹、地理位置、WiFi MAC地址、注册时间这几个方面来识别，找出批量操作的客群。

2. 账户级和交易级

区分了欺诈风险发生的意愿后，下一步需要确定欺诈的主体，进而制定不同的管控策略。例如首支首逾、“羊毛党”等第一方欺诈案件，在核实无误的情况下，可以将账户或者设备拉入黑名单，杜绝该账户或者

设备之后发生的任何申请和支用行为。利用首支首逾的历史样本，还可以搭建离线的欺诈评分模型，定期在全量账户上进行批量预测，对于分数较低的账户可以提前预警或者冻结。

而对于中介申请、电信诈骗这类第三方欺诈案件，以及“羊毛党”等通过单一账户属性难以识别的第一方欺诈案件，则需要实时采集客户每笔交易的地理位置、设备属性和操作行为等数据，在保证拦截率的情况下尽可能提高召回率。交易级的实时模型和策略对于机构系统的实时采集和计算能力是个比较大的考验，但是只有真正具备了这一能力，机构才能更好地应对层出不穷的欺诈手段。

3.个人和团伙

常见的欺诈类型从组织上又可以分为个人和团伙两种。传统的反欺诈策略和模型多针对个人案件，而团伙案件一直是反欺诈人员比较头疼的问题。为了解决这一痛点，关联图谱被逐渐应用到团伙反欺诈的场景中。关联图谱基于客户的身份证、手机号、银行卡号、设备指纹、地理位置、WiFi MAC地址等属性，构建客户与客户之间的关系，从个人欺诈出发，深挖背后的团伙组织。关联图谱的应用方式有很多，简单的可以直接制定规则，提取客户之间的聚集度，识别中介申请和“羊毛党”；复杂的可以利用社区发现和标签传播等算法，智能地划分团伙并且量化个人的潜在欺诈风险。

总的来说，目前在欺诈风险领域，有标签的数据量很少，导致机构对于欺诈案件的感知较为被动，人工调查多集中在事后，已经造成的损失难以追回。不过随着反欺诈技术的进步，例如样本增强、无监督学习等智能算法已经被投入实际业务场景中，使得机构对于欺诈分子的识别更加主动，客户的个人信息和财产也得到了更全面的保护。

1.2.2 信用风险

1.白名单准入

白名单是信用风险管理的第一道门槛，与整个平台贷款产品的设计和定位有紧密的联系。白名单设立的初衷是圈定目标客户，有了目标客群之后才能更好地进行精准营销，并且使得后续的风控流程利润最大化。同时在贷款产品上线初期，由于缺乏足够的数据积累，难以搭建完善的风控模型，因此白名单也是冷启动阶段较为常见的一种风控手段。白名单的制定通常会从政策要求、风控能力和客户画像3个方面考虑。

首先，政策要求是重中之重，监管部门的相关政策是每个金融从业人员应该守住的底线，在互联网金融机构的信贷业务中，年龄和定价是两根红线。2017年年底出台的《关于规范整治“现金贷”业务的通知》，明确要求金融机构不得为在校学生提供借贷撮合业务，因此目前银行和大型互联网金融机构大多将22岁作为年龄的准入门槛。而对于60岁以上的客户，由于存在较大的不确定性和社会道德压力，一般也会被准入规则排除在外。对于定价，监管条例中规定，现金贷产品的综合年化利率不得超过36%，因此对于前期测算综合成本有可能超过36%的客群，也不应当纳入白名单中。

其次，作为风险管理的一部分，风控能力自然是影响白名单制定的重要因素。与后面的几个信用风控环节相比，白名单部分侧重于硬规则的制定，触碰硬规则的客群由于存在重大风险隐患，也会被风控人员排除在外。例如，大多数平台都会将注册时长和活跃度这两个因素放到准入规则中，理由是如果借款人在平台的注册时间过短或者活跃度过低，一方面平台方无法判断借款人是否是为了骗贷而特意注册开户的，另一方面这类客户在平台侧的数据量太少，给风控模型的信用判断造成了极大困难，因而不被白名单所接受。

最后是客户画像，客户画像规则通常与贷款产品关注的人群有关。对于纯粹的现金贷产品，由于此类产品适用于大多数人群，在白名单中可以不考虑特定的客户画像指标。但是，对于具有特定场景和人群的贷款产品，例如教育分期、医美分期、滴滴司机贷等，抓住这些指向性客群背后的信用规律，并且归纳成规则放入白名单准入环节中，可以大大

降低风险，提升产品的运营效率。

另外，白名单的制定不宜过于严格，随着数据和风控能力的逐步提升，白名单的准入策略应该逐步放开，转而更信任之后的贷前识别和贷中管理，这样才能向着全流量风控运营的大方向进发。

2. 贷前识别

贷前识别是整个信用风险管理中最重要的一环，一套良好的贷前识别体系能够规避70%以上的潜在风险。同时，贷前识别环节具有客户触达人数多、数据维度丰富、信用风险存在规律性等先天优势，是大数据风控模型应用最为成熟的一个模块。如本节开头所述，信用主要包含还款意愿和还款能力这两个方面，大多数信用风险的出现都是由于平台授予的贷款额度与客户这两方面情况不匹配所造成的。因此基于客户申请授信环节准确的信用评估，并且给出最合理的额度和利率，是风控从业者永恒的话题。

对于还款意愿的衡量，行业内通常的做法是搭建申请评分卡（Application Card）。申请评分卡是在有一定存量客户和风险表现的基础上，将客户申请时的多维指标和未来一段时间内是否发生逾期建立关系，从而实现对于新增客户预测风险的目的。申请评分卡出现时间较早，并且已经经历了从专家模型、到评分卡模型，再到大数据模型的三个阶段，贷前识别能力也得到了长足的提升。国内的大型互联网金融机构在前几年线上贷款业务蓬勃发展阶段已经积累了大量的客户样本和逾期表现数据，同时国内相对开放的第三方数据环境，也为这些样本带来了丰富的指标维度，这都是训练出良好大数据模型的必要前提。在业界，目前M1+的模型KS基本可以做到0.35，M3+的模型KS甚至可以达到0.45以上，已经远高于人工审核的准确率。

在贷前识别环节，只是评估客户的还款意愿显然是不够的，风控人员还需要将还款能力纳入评估体系内。这就要用到价值模型（Value Model）。相较于申请评分卡有是否逾期这一明确的目标定义，客户价值的高低在不同平台的定义各不相同，有些平台考虑客户的收入，有些平台则会考虑客户带来的利润，因而价值模型的目标可以根据平台的实际业务需求来制定。模型变量方面，价值模型通常更关注金融属性方面的指标，传统金融机构一般从收入、资产、负债三个角度来选取指标，而对于互联网金融机构来说，则会基于自身数据的特色，加入客户的商品消费习惯和金融场所出行习惯等，一定程度上丰富了价值模型的维

度。

搭建兼具准确性、区分度和稳定性的申请评分卡和价值模型，是智能时代的风控从业者必须面临的挑战和课题，在这两个模型基础上，风控人员才能判断借款人的申请是否可以通过，并且给出相匹配的额度和利率。由于贷前识别是整个信用风险管理中最核心的部分，本书之后的章节会大多围绕贷前环节展开，关于申请评分卡的搭建和应用，也会在后面相关章节中具体阐述。

3. 贷中管理

借款人一旦在贷前环节通过了授信申请，就进入了贷中管理环节，这个环节主要是监控和调整，对于贷前识别的结果进行查漏补缺。目前市面上的互联网金融贷款产品额度授信方式主要分为一次性额度和循环额度两类，不同的授信方式带来的贷中管理策略也会有很大的不同。

首先解释下一次性额度和循环额度的区别。一次性额度，就是在客户的贷款申请通过后，机构主动地、一次性地将全部额度打到客户的银行账户上，解决客户短期内的燃眉之急，常见于银行的房贷、经营贷等大额商业贷款。而循环额度是指客户在贷款申请通过后获得预授信的额度，在需要的时候由客户主动发起支用，并且在客户偿还账单后额度可以实时恢复，例如银行的信用卡业务就是一款非常典型的循环额度产品。对于一次性额度的产品，由于全部金额已经在第一时间打到客户的账户上，贷中管理主要集中在风险的监控和预警上，一旦客户出现连续几期的逾期，就需要进入贷后催收的环节。而对于循环额度的产品，由于初始只是预授信，机构完全可以借助更多贷中行为的数据，通过支用拦截和额度管理等贷中手段，压降潜在的风险。同时对于借款人来说，循环额度类的贷款即需即用，随借随还，给客户提供了更好的产品体验，客户完全可以根据自己目前的资金现状选择支用和还款的时间节点，减少不必要的账期带来的贷款利息，符合国家普惠金融的大方向，也是目前市面上大多数消费贷款的产品模式。

对于循环额度类的产品，搭建行为评分卡（Behavior Card）是一种常见的技术手段。与申请评分卡一样，行为评分卡的预测目标也是客户在未来一段时间内是否会发生逾期。不过模型指标在申请评分卡的基础上，又多了贷中行为这一大类数据，通过加入客户支用、还款、逾期等数据，行为评分卡可以更好地刻画出客户近期的还款能力和还款意愿，帮助机构做出相应的支用和额度策略的调整。

4. 贷后催收

目前市面上头部的互联网金融机构，由于贷款业务起步较早，时至今日已经积累了较大规模的贷款余额，并且在经济周期处于低谷的大环境下，整个行业逾期率走高，形成了较多的贷后资产。针对这些已经逾期的账户，平台侧必须采取贷后催收的手段，才能尽可能地收回账款，在控制整体逾期率和不良率的同时，最大限度减少经济上的损失。

相比于贷前识别和贷中管理，贷后催收更重运营，如何在可控的成本范围内，根据逾期案件的严重程度，为不同账户分发不同的催收方式和催收团队，是贷后管理环节的风控人员需要考虑的问题。同时，在互联网金融野蛮生长的前几年，暴力催收案件屡见不鲜，对社会和整个行业造成了极大的危害，引发了监管部门的高度关注。从2018年年底开始，监管部门联合各地公安机关，查处了许多涉黑的催收公司，暴力催收的行为也被整个行业严厉杜绝。在整个行业合规的大环境下，精细化和智能化的贷后运营成为了提高催回率最有效的方式。

为了实现精细化的分案，主要的方法是利用催收评分卡（Collection Card）衡量客户逾期的严重程度，根据严重程度分发给不同力度的催收方式和团队。催收评分卡基于已经发生逾期的样本群体建立，目标是预测这些样本是否会在未来一段时间内发生更严重的逾期。相比于贷前和贷中的特征，催收评分卡又多了与催收相关的贷后特征，比如进入催收队列的持续时长、催收阶段成功还款的金额等，这些特征都提升了模型对于贷后人群的区分能力。根据业务的需求，在样本量满足条件的情况下，还可以根据客户所处的贷后阶段，进一步将通用模型拆分为早期催收评分卡、中期催收评分卡和后期催收评分卡。

贷后环节也是当前AI技术探索较多的环节，机构希望通过更智能的算法和模型，降低单纯人力的运营成本。目前像失联修复、语音自动外呼、催收机器人等技术都已经有了正式落地的案例，随着这些AI系统的进一步成熟，整个催收行业也会更加合规和可控。

1.3 风险管理的重要性

互联网金融行业发展至今，早期头部机构的流量红利已经消失殆尽，取而代之的是互联网时代下半场的精细化运营。在这一阶段，风险管理就显得尤为重要。在传统认知中，风控人员只是一味“踩刹车”，大家普遍认为控制坏账率就是风控人员的全部工作。其实防范坏账发生只是风控人员工作的一部分，风控的本质应该是在坏账率满足条件的情况下使得整体利润最优，风控人员更像是一个控制“刹车”和“油门”的节拍器，风险管理应该是从风险评估、差异化定价到最终实现整体利润最优的三级火箭模式。

1.3.1 风险评估

风险评估是风控人员的本职工作，也是整个风险管理流程中的“第一级火箭”，是实现整体利润最优的基础和约束条件。风险是所有互联网金融业务的红线，一旦触碰红线就不用谈利润了。对于传统银行来说，贷款可以按照风险程度划分为正常、关注、次级、可疑和损失五类，其中后三类合称为不良贷款，根据银保监会要求，不良贷款余额占总贷款余额的比重不得超过5%。对于互联网金融行业，监管部门暂时还没有明确的不良贷款率警戒线，但是控制贷款风险仍然是各家机构的基本准则。

对于借款人风险的评估，主要还是从综合欺诈和信用两个方面入手。对于存在欺诈风险和高信用风险的客户，平台一般不授予其贷款额度；对于信用风险较低的客户，风控人员会利用申请评分卡量化其风险水平，将不同的人匹配到不同的区间内，实现风险分层的目的，指导下一步的差异化定价。风险分层如表1-1所示，其中申请评分卡700分以上的为低风险客群，640分以下的为高风险客群，640~700分的为中风险客群。

表1-1 风险分层映射关系

分数区间	区间逾期率 (%)	累计逾期率 (%)	风险评估
(820,830]	0.0	0.00	低风险客群
(810,820]	0.0	0.00	
(800,810]	0.0	0.00	
(790,800]	0.0	0.00	
(780,790]	0.0	0.00	
(770,780]	0.3	0.19	
(760,770]	0.7	0.50	
(750,760]	0.5	0.49	
(740,750]	0.9	0.70	
(730,740]	1.4	1.01	
(720,730]	2.2	1.48	
(710,720]	2.5	1.84	
(700,710]	2.8	2.17	
(690,700]	3.5	2.55	中风险客群
(680,690]	5.5	3.20	
(670,680]	6.9	3.89	
(660,670]	10.0	4.67	
(650,660]	13.3	5.45	
(640,650]	17.3	6.29	

(续)

分数区间	区间逾期率 (%)	累计逾期率 (%)	风险评估
(630,640]	23.7	7.20	高风险客群
(620,630]	28.1	7.79	
(610,620]	39.5	8.10	
(600,610]	48.7	8.27	
(590,600]	52.2	8.32	
(580,590]	69.0	8.34	
(570,580]	100.0	8.34	
(560,570]	100.0	8.34	

1.3.2 差异化定价

在风险评估的基础之上，下一步要做的事情就是给不同的客群分配不同的贷款额度和利率，也就是风险管理流程中的“第二级火箭”——差异化定价。对于机构来说，定价模型直接决定了产品最终的利润。如果机构对于借款人都给予相同的额度和利率，势必会造成好客户额度过低、坏客户额度过高的情况，又由于坏客户的逾期概率一定高于好客户，则会导致在逾期人数相同的情况下，该机构会损失更多的利润。同时，由于好客户一定是市面上所有互联网金融机构的目標客群，如果一个机构给出的额度和利率吸引力不足，这些好客户必定会流向其他平台，这对于平台来说是一种更大的损失。考虑到以上两个问题，差异化定价应当是所有风控人员必备的能力。

简单的差异化定价，可以直接利用客户风险分层的结果，对于风险较小的客群给予高额度和低利率，对于风险较高的客群则给予低额度和高利率。而对于技术能力较强的机构，如果在申请评分卡以外还搭建了一个效果较好的价值模型，则可以通过二维矩阵的方式，综合考虑客户的还款能力和还款意愿，给出更为合理的额度。在定价模型中，利率通常直接与申请评分卡所预测的逾期风险挂钩，最高利率不得超过监管所设置的上限；额度则需要风控人员和财务人员一同制定，在财务人员测算的综合成本的基础上，风控人员考虑逾期损失，根据经验和计算给出能够盈利的最合理的额度。结合风险模型和价值模型的额度矩阵如图1-2所示，整个矩阵从左上角到右下角额度递增。

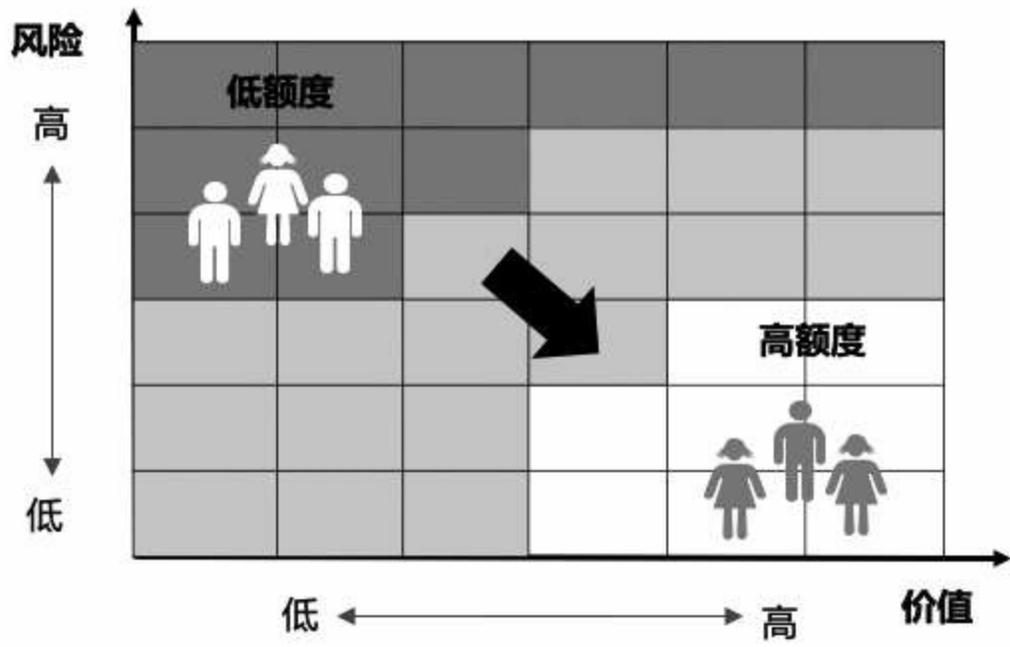


图1-2 额度矩阵

1.3.3 整体利润最优

上述的差异化定价是一种比较理想的情况，其中忽略了一些实际工作中存在的变量，这就要求专业的风控人员具备让整体利润最优的能力，也就是风险管理流程中的“第三级火箭”。在实际操作中，客户对于贷款额度通常比较敏感，额度过低没有吸引力会导致客户的支用意愿不强，减少机构的收入；额度过高超过借款人还款能力又会增大客户逾期的风险，给机构带来不必要的损失。因此，机构的整体利润会随着额度的增加先增大再减小，如何找到额度的最优解，是风控人员需要解决的实际问题。额度与利润的关系如图1-3所示。

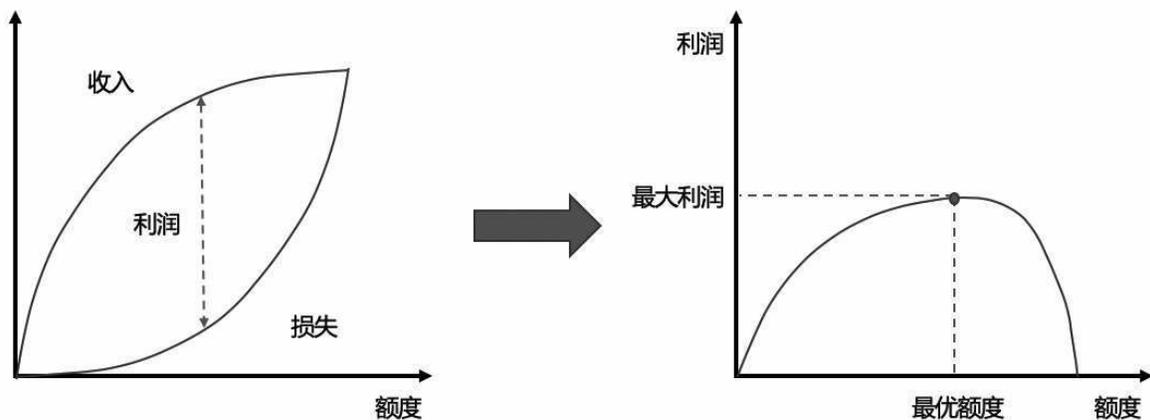


图1-3 额度与利润的关系

目前业内比较常见的方案有两种：息费敏感性测试和利用模型的在线学习。

息费敏感性测试，是指机构在线上配置多个实验组，不同的实验组给予不同的额度，利用线上的实际表现搜集客户关于额度的使用率、逾期表现等数据，再将这些实际数据纳入利润公式中，找出最优额度。不过由于风险表现的滞后性，通常需要3个月以上才能够观察到部分逾期表现，因而息费敏感性测试所需要的时间成本较高，适用于针对长期阶段的定价方案调整。

为了短期内优化已有的定价方案，在有一定放款数据的基础上，可

以尝试利用模型的在线学习方式。这种方法拟合额度关于使用率、逾期表现等变量的曲线，以模型的方式进行预测，给出初始的授信额度，然后在模型上线后，实时调整使用率和逾期表现的拟合曲线，使得线上额度的分配逐步接近最优方案。

在实际应用中，建议两种方案相结合使用，模型的在线学习可以解决短期内额度不合适的痛点，但要想彻底根治，还是需要通过息费敏感性测试长期积累数据来优化。