



主 编：盖瑞瑞  
副主编：王高峰

# 数据化风控

## 信用评分建模教程

单良 乔杨◎著



中国工信出版集团



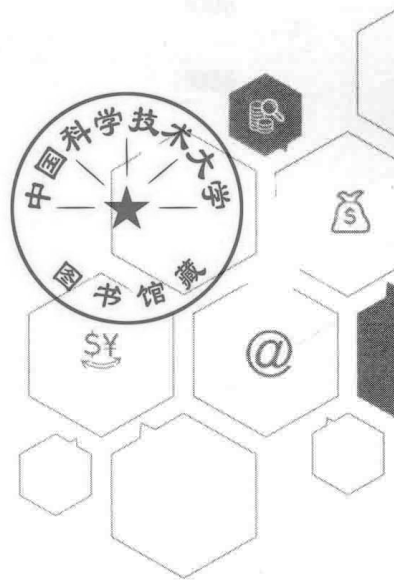
电子工业出版社  
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY  
<http://www.phei.com.cn>

主 编：盖瑞瑞 副主编：王高峰

# 数据化风控

## 信用评级建模教程

单良 乔杨◎著



电子工业出版社  
Publishing House of Electronics Industry  
北京·BEIJING

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

数据化风控：信用评分建模教程 / 单良，乔杨著. —北京：电子工业出版社，2018.9  
ISBN 978-7-121-34629-3

I. ①数… II. ①单… ②乔… III. ①信用评级—系统建模—教材 IV. ①F830.5  
中国版本图书馆 CIP 数据核字（2018）第 137808 号

出版统筹：刘声峰

策划编辑：黄菲

责任编辑：黄菲      文字编辑：王欣怡      特约编辑：徐学锋 刘广钦

印刷：三河市鑫金马印装有限公司

装订：三河市鑫金马印装有限公司

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：720×1 000 1/16 印张：15 字数：213 千字

版次：2018 年 9 月第 1 版

印次：2018 年 9 月第 3 次印刷

定价：65.00 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 [zlts@phei.com.cn](mailto:zlts@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：1024004410（QQ）。

## 单良

本科毕业于美国纽约哥伦比亚大学，复旦大学、台湾大学EMBA，曾任职于香港维信理财有限公司、台北富邦银行、中国信托商业银行、澳商澳盛银行及台新银行等机构；兼任台湾金融研训院特约讲师、VISA中国区兼职顾问。

具备台湾银行业消费金融风险管理与大陆小贷、P2P风控管理完整资历，长期关注两岸消费金融产业风控管理的发展与创新。曾发表前瞻性评论，并为台湾金融研训院、中国P2P网贷实务研修班授课。

著作有《信用评等模型关键12堂课》《互联网金融时代消费信贷评分建模与应用》。

## 乔杨

ZRobot CEO。曾担任知名互联网金融公司联合创始人兼首席风险官，美国发现金融芝加哥总部风险策略及模型业务高级经理，发现金融上海大数据风控中心风控策略及大数据建模业务负责人。曾参与美国通用电气公司财务管理领导力项目（FMP），研究商品期货和货币的对冲策略。拥有美国爱荷华大学经济学及MBA双硕士学位，芝加哥大学计算机科学硕士学位，SAS认证师，Teradata认证SQL专家，微软认证系统工程师（MCSE）等。

## 本书编委会

主 编：盖瑞瑞

副 主 编：王高峰

成 员：张艳华 王秋瑞 刘 军 周 巍 徐夏楠

邵长春 张昊天 孙 坚 李文贤

联合出品：ZRobot 微衡科技 翼勋金服 蜡笔分期 逸能资产



ZRobot是一家专注于数据技术与应用的科技公司，由中国极具创新力的互联网金融企业京东金融和美国著名大数据公司ZestFinance合资建立。依托京东金融的本土金融实践，以及来自美国硅谷领先的模型技术，利用领先的建模能力完成数据价值的优化配置，提供中国本土有效的领先数据技术解决方案，包括反欺诈、信用评估、ABS资产定价和精准营销等全方位数据技术，主要服务于金融相关领域，如银行、消费金融、汽车金融等。



上海微衡信息科技有限公司是一家以结构化金融服务为核心，以数字化资产技术为驱动力，专注于服务小额消费信贷领域的创新型金融科技企业。旨在帮助小微资产更经济、更合规、更透明地连接金融机构资金，助力金融机构更高效、更精准、更安全地配置小微资产。



上海翼动互联网金融信息服务有限公司（简称“翼动”）成立于2015年4月，是一家互联网金融综合服务平台，总部位于上海。专注于网络信贷撮合服务和消费金融等，致力于用自身强大的平台建设、卓越的风控技术、专业的大数据驱动能力为广大用户提供专业、简单、优质、便捷的普惠金融服务。2016年，凭借良好的行业口碑及领先的金融科技技术，翼动成为中国互联网金融协会会员、上海互联网金融行业协会理事单位、上海金融信息行业协会副会长单位。



蜡笔分期成立于2015年1月，以移动互联网、大数据科技、生态链发展模块为基础，在行业领域建立起了两大独特优势：“移动+教育场景+生态链”的发展模式、大数据驱动的信用风控与系统技术服务。为用户提供简单、高效、优质、便捷的普惠移动教育分期服务，致力于为有梦想的用户学业发展、技能发展提供合理及正当的助力，包括助学教育、培训和课题科研等教育及消费信用服务。



江苏逸能资产管理有限公司专注于为互联网金融机构提供催收坐席外包服务和不良资产系统解决方案，为资产管理行业提供技术外包、呼叫中心标准化作业解决方案，为金融机构提供辅助贷款决策服务。

## 反侵权盗版声明

电子工业出版社依法对本作品享有专有出版权。任何未经权利人书面许可，复制、销售或通过信息网络传播本作品的行为；歪曲、篡改、剽窃本作品的行为，均违反《中华人民共和国著作权法》，其行为人应承担相应的民事责任和行政责任，构成犯罪的，将被依法追究刑事责任。

为了维护市场秩序，保护权利人的合法权益，我社将依法查处和打击侵权盗版的单位和个人。欢迎社会各界人士积极举报侵权盗版行为，本社将奖励举报有功人员，并保证举报人的信息不被泄露。

举报电话：(010) 88254396；(010) 88258888

传 真：(010) 88254397

E-mail: [dbqq@phei.com.cn](mailto:dbqq@phei.com.cn)

通信地址：北京市万寿路 173 信箱

电子工业出版社总编办公室

邮 编：100036

# 目 录

## 第一章 信用评分基础认识与应用 /001

---

第一节 信用评分卡简介 /003

第二节 评分卡建立与验证 /008

第三节 评分应用 /026

## 第二章 信用评分模型规格与设计 /031

---

第一节 数据收集、质量检验 /031

第二节 应排除的数据样本 /033

第三节 样本期间、好坏客户定义 /034

第四节 范例 /039

## 第三章 分组 (Segmentation) 目的与分析选择 /041

---

第一节 分组目的 /041

第二节 分组分析 /043

第三节 范例 /046

---

## 第四章 细致分析与自变量分析 /049

第一节 细致分类 (Fine Classing) /051

第二节 范例 /052

第三节 单因子分析 (Single Factor Analysis) /057

第四节 粗略分类 (Coarse Classing) /064

第五节 范例 /065

---

## 第五章 模型建立方法讨论 /071

第一节 线性回归 (Linear Regression) /073

第二节 逻辑回归 (Logistic Regression) /077

第三节 两阶段式建立方法 /082

第四节 初始模型讨论 /084

第五节 范例 /085

---

## 第六章 拒绝推论 (Reject Inference) 的原因与方法 /089

第一节 拒绝推论的原因 /090

第二节 拒绝推论的方法 /092

---

## 第七章 最终模型选择与风险校准 (Calibration) /099

第一节 最终模型产出 /101

第二节 设定风险校准 (Risk Calibration) /105

第三节 模型验证 /109



<b>第八章 决策点 (Cut-off) 设定</b>	<b>/115</b>
第一节 决策点策略设定方式	/116
第二节 核准点应用方式	/118
第三节 范例	/119
<b>第九章 信用评分模型监控报告</b>	<b>/123</b>
第一节 前端监控报告	/126
第二节 后端监控报告	/135
<b>第十章 信用评分模型策略运用</b>	<b>/151</b>
第一节 业务策略制订方式	/152
第二节 业务策略应用方式	/154
第三节 范例	/158
<b>第十一章 信用评分模型案例 (消费产品分期)</b>	<b>/161</b>
第一节 数据样本	/162
第二节 样本好坏表现定义	/163
第三节 变量分析	/167
第四节 模型建立与验证	/170
<b>第十二章 信用评分模型案例 (现金贷)</b>	<b>/173</b>
第一节 数据样本	/174
第二节 样本好坏表现定义	/175

第三节	变量分析	/176
第四节	模型建立与验证	/178
<b>第十三章</b>	<b>催收框架</b>	<b>/183</b>
第一节	催收管理流程	/185
第二节	催收管理系统简介	/190
第三节	催收模型系统	/191
第四节	催收策略系统	/195
<b>第十四章</b>	<b>催收技巧及 KPI 标准</b>	<b>/213</b>

## 第一章

# 信用评分基础认识与应用

金融机构在高度市场竞争与业绩成长压力下，对风险管理的精准度与作业效率的要求越来越高。不管是申贷时还是审批后，每一位客户在不同的阶段都有不同的潜在风险，这些风险征兆可能存在于各种容易令人忽略的细节之中，一旦错失实时发现的先机，后续将造成难以弥补的损失。

早期银行的风险管理主要是以人工操作为主，客户的各种风险皆依赖于人员的经验判断。其中的问题是，风险因子众多且彼此交互极为复杂。一是有经验的人员需要长时间的培养；二是人毕竟不是机器，不但质量无法完全一致而且人工处理能量有限，难以大量作业。因此，必须借由计量方法在早期侦测客户潜在风险，以期减少人为主观判断的失误并减轻风险管理人员的负担。

关于客户个人风险的预测，目前使用最普遍的工具为信用评分卡，其应用的范围包括进件、贷后管理及催收等。不但可以筛选高风险客户，减少损失发生，也可以找出相对优质的客户群，发掘潜在机会。

从经营管理层面来看，信用评分的出现使科学化及自动化程度向前迈进了一大步，也大幅提升了银行竞争力。无疑，信用评分是近年来最受银行喜爱的风险管理工具之一，且使用范围日趋广泛。本章就信用评分的发展历史、种类、建立方法及运用加以介绍。

## 第一节 信用评分卡简介

信用评分的应用最早源自 20 世纪 30 年代，主要的概念是仿照有经验的风险分析专家设计信用判断条件，使授信质量一致。这种革命性的构想将风险量化技术应用在银行授信作业中，使信贷审批有了客观的依据，不再完全依赖于经验判断。在此阶段之前，判断条件是以资深授信人员及主管的经验制定，虽是集合所有专家的意见精华，但基本上还是经验给分，在选择风险因子及设定权重时即使反复摸索修改，也无法确定因子间的关系。到了 20 世纪 50 年代，回归分析等统计技术开始运用于信用评分，才将人类经验与数学实证进行了结合。

经过数十年的发展改良，再加上信息科技的进步——计算机现在可存储大量历史数据并可以轻易进行各种复杂运算，目前，信用评分已成为银行极为倚重的风险评估工具。

信用评分主要的功能可归纳为以下 3 项。

### 一、以科学方法将风险模式数据化

风险因子若未经量化处理，就是一种意象及概念，授信的判断若取决于“个人感觉”，固然弹性大，但风险判断的灰色地带必然无法缩小。信用评分以科学方法将风险量化，使得风险评估有所依据，从而弥补了上述缺憾。

信用评分建立于完整的历史数据之上，借由数据汇整、清理、分组及

探勘等技术，将大量数据转化成为有用的风险信息。信用评分模型建立后，可将风险数据化。此时，可清楚地呈现客户违约概率及风险排序，使风险管理单位得以确切掌握客户风险，并且制定更为精准的授信政策。

## 二、提供客观风险量尺，减少主观判断

授信作业最大的挑战在于如何客观完整地将客户各种质化特性加以综合评断，并做出最合适的决定。例如，经验告诉授信人员男性客户的风险较女性客户的高，但若再加上年龄、职业、地区、收入和负债等分析维度，又会出现什么状况呢？面对不同的因子组合，授信人员要如何决定案件的准驳、额度及利率？

传统授信为人所诟病之处就在于过度依赖于个人经验。不可否认，过去的经验的确有一定的价值，不过影响客户行为的风险因子本来就极为复杂且难以掌控。人为的经验可能是从多次错误之中学习而来，可能是由资深人员传承而来，也有可能是由主观的直觉认定而来，但是经验的培养及传承皆非易事，此外，不同人员养成过程及环境未必相同，对各种案件的经验及认知也会不同，从而产生看法上的分歧。事实上，即使是同一授信人员在不同时期对同一案件可能也会有不同的审批决定。

信用评分提供了一个客观的风险量尺，消除了人为判断所造成的差异，使得风险评量有一致性的标准。

## 三、提高风险管理效率，节省人力成本

强大的处理效率也是信用评分的重要功能，以消费金融产品而言，其客户数量远远超过其他产品，为了迅速消化申请案件并有效管理往来客户，势必需要高度的自动化流程及系统协助，信用评分模型搭配评分计算引擎（Scoring Engine）可针对个别案件进行实时风险运算，也可以用批次方式处理大量数据，与传统人工审核方式相比，大大节省了处理时间。由于信用评分卡能事先将风险计算完成，所以中间灰色地带可大幅缩小。授信人员仅须将重心放在对这些灰色案件的处理上，不用为了配合不同的营销活

动而增加人力。

另外，若银行拥有信用评分卡，就可针对现有客户进行定期或不定期的大规模期中复审，尤其是信用卡，因其客户数量庞大、产品生命周期长且风险为动态性质，风险因子多样且复杂，若无信用评分的协助，很难切实掌握客户的风险变化。以信用卡循环利率为例，目前各家银行都采用定价模式，信用评分可协助银行划分客户风险等级，以设计适合各等级的合理定价。

尽管信用评分有许多优点，为风险管理带来了许多的便利，但事实上，它并非完美，在使用时仍需注意一些限制。举例来说，数据库是信用评分的基础建设，建立评分模型之前须有完整且质量良好的历史数据，否则模型预测效力将会大打折扣。因此，必须先行建立完善的数据库，信用评分才能顺利开展。

再则，信用评分模型是以历史数据为基础，其准确性建立在未来的信用表现与过去相同的假设前提之上。一旦遭遇重大经济变动，假设不再成立，模型预测能力就会受到影响。在现实世界中，不可能期望所有状况永远固定不变，因此必须每月定期检测模型的准确性与偏离程度。

另外，信用评分的发展是以风险评量为出发点，为求主题明确，聚焦于风险因子的探讨，收益面并非考虑重点。因此，风险极低者未必是获利贡献度最高的客户。

需要特别提醒的是，信用评分模型以“户数”为单位计算好坏比(Good/Bad Odds)，我们可借此得知各分数下可能倒账的户数比率，而一般评估风险所惯用的逾期比率则是以“金额”为计算单位，两者计算基础不同，户数倒账率与金额倒账率无法互相比较。

信用评分的应用领域广泛，为适应多样化的需求，也发展出不同种类的评分卡，下面就两种常用的分类方式进行介绍。

## （一）依发展母体区分

### 1. 通用型评分（Generic Score）

这类评分多为顾问公司依据手中收集的资料所开发的评分卡。由于建立模型所需数据收集不易，且其定位为多用途信用评分，因此，未针对特殊主题或条件做个别考虑。

此种评分卡的优点是不需要从头开始发展模型，可节省开发成本，并可快速部署于风险管理流程中，但其精准度比其他类型信用评分卡略逊一筹。

### 2. 征信机构评分（Bureau Score）

各国征信机构皆拥有庞大的客户信用数据，而由这些数据所建立起来的评分卡被称为 Bureau Score。征信机构除拥有完整的数据之外，还有极高的公信力。对于未开发专属信用评分卡的中小型金融机构而言，Bureau Score 具有很大的吸引力。由于它并非以特定金融机构的数据建立模型，因此也可视为通用型评分卡的一种。

### 3. 客制化评分（Customized Score）

客制化评分也是完全依照各家金融机构的客户资料及特殊需求量身打造的评分卡。由于目标明确且针对性高，若历史数据完整且质量良好，其预测效果就会高于通用型及征信机构两种评分。其缺点为整体建立时间较长，初期开发成本较高，后续也需要自行负责模型监控与调校的工作。

## （二）依使用时机区分

### 1. 申请评分（Application Score）

申请评分用于客户进件审核，信用评分模型所使用的变量除了申请资料数据之外，还包括联合征信中心的信用资料。在进件评分的协助下，授信人员可将重心放在介于准驳边缘的案件上。



## 2. 行为评分 (Behavior Score)

行为评分多用于预测信用卡客户的动态风险，评分模型的变量以客户的交易及缴款形态数据为主。由于动态行为的风险预测较为复杂，因此，行为评分模型所采用的变量通常比进件评分模型多。

## 3. 催收评分 (Collection Score)

催收评分多被使用于案件量较多的前段催收，许多人将其与上述两种信用评分卡合称为“ABC Cards (Application Scorecard, Behavior Scorecard & Collection Scorecard)”。其主要功能是预测客户还款概率，属于行为评分的延伸应用。