

法律声明

□ 本课件包括：演示文稿，示例，代码，题库，视频和声音等，小象学院拥有完全知识产权的权利；只限于善意学习者在本课程使用，不得在课程范围外向任何第三方散播。任何其他人或机构不得盗版、复制、仿造其中的创意，我们将保留一切通过法律手段追究违反者的权利。

□ 课程详情请咨询

■ 微信公众号：小象学院

■ 新浪微博：小象AI学院



互联网金融业申请评分卡的介绍

目录

信用违约风险的基本概念

申请评分卡的重要性和特性

贷款申请环节的数据介绍和描述

非平衡样本问题的定义和解决方法

信用违约风险的基本概念

□ 什么是信用违约风险

交易对手未能履行约定契约中的义务而造成经济损失的风险，即受信人不能履行还本付息的责任而使授信人的预期收益与实际收益发生偏离的可能性，它是金融风险的主要类型。

□ 组成部分

PD	违约概率
LGD	违约条件下的损失率
EAD	违约风险下的敞口暴露
RWA	风险权重资产
EL	期望损失

信用违约风险的基本概念

□ 信用违约的主体

个人违约

个人向金融机构
借贷后，没有在
规定的期限之内
还款的行为

公司违约

公司向金融机构
借贷后，没有在
规定的期限之内
还款的行为，或
者公司在发行债
券后，没有履行
或者延期履行利
息或本金的支付
义务

主权违约

一国政府无法按
时对其向外担保
借来的债务还本
付息的情况

信用违约风险的基本概念

□ 个贷中常用的违约定义

- M3 & M3+ 逾期
- 债务重组
- 个人破产
- 银行主动关户或注销
- 其他相关违法行为

□ M0, M1, M2的定义

- M0: 最后缴款日的第二天到下一个账单日
- M1: M0时段的延续, 即在未还款的第二个账单日到第二次账单的最后缴款日之间
- M2: M1的延续, 即在未还款的第三个账单日到第三次账单的最后缴款日之间

。 。 。

信用违约风险的基本概念

□ 案例

- 趣店上市后，其CEO罗敏宣称
“凡是过期不还的，我们这里就是坏账，我们的坏账，一律不会催促他们来还钱。电话都不会给他们打。你不还钱，就算了，当作福利送你了。”
- 宜信曾被曝8亿坏账难追回
- 陆金所CEO计葵生曾多次在公开场合揭露行业的坏账率，2015年11月，他说大概在15%-20%之间，2016年1月，这个数字降至13%-17%。而陆金所自身的坏账率大概是5%-7%。

目录

信用违约风险的基本概念

申请评分卡的重要特性

贷款申请环节的数据介绍和描述

非平衡样本问题的定义和解决方法

申请评分卡的重要性和特性

□ 什么是评分卡

信贷场景中的评分卡

- 以分数的形式来衡量风险几率的一种手段
- 是对未来一段时间内违约/逾期/失联概率的预测
- 有一个明确的(正)区间
- 通常分数越高越安全
- 数据驱动
- 反欺诈评分卡、申请评分卡、行为评分卡、催收评分卡

申请评分卡的重要性和特性

□ 申请评分卡的概念

用在申请环节，以申请者在申请当日及过去的信息为基础，
预测未来放款后的逾期或者违约概率

□ 为什么要开发申请评分卡

- 风险控制：借贷生命周期的第一个关口
- 营销：优质客户识别
- 资本管理：可作为PD模型的一个因子

申请评分卡的重要性和特性

□ 评分卡的特性

➤ 稳定性

当总体逾期/违约概率不变时，分数的分布也应不变

➤ 区分性

违约人群与正常人群的分数的分数应当有显著差异

➤ 预测能力

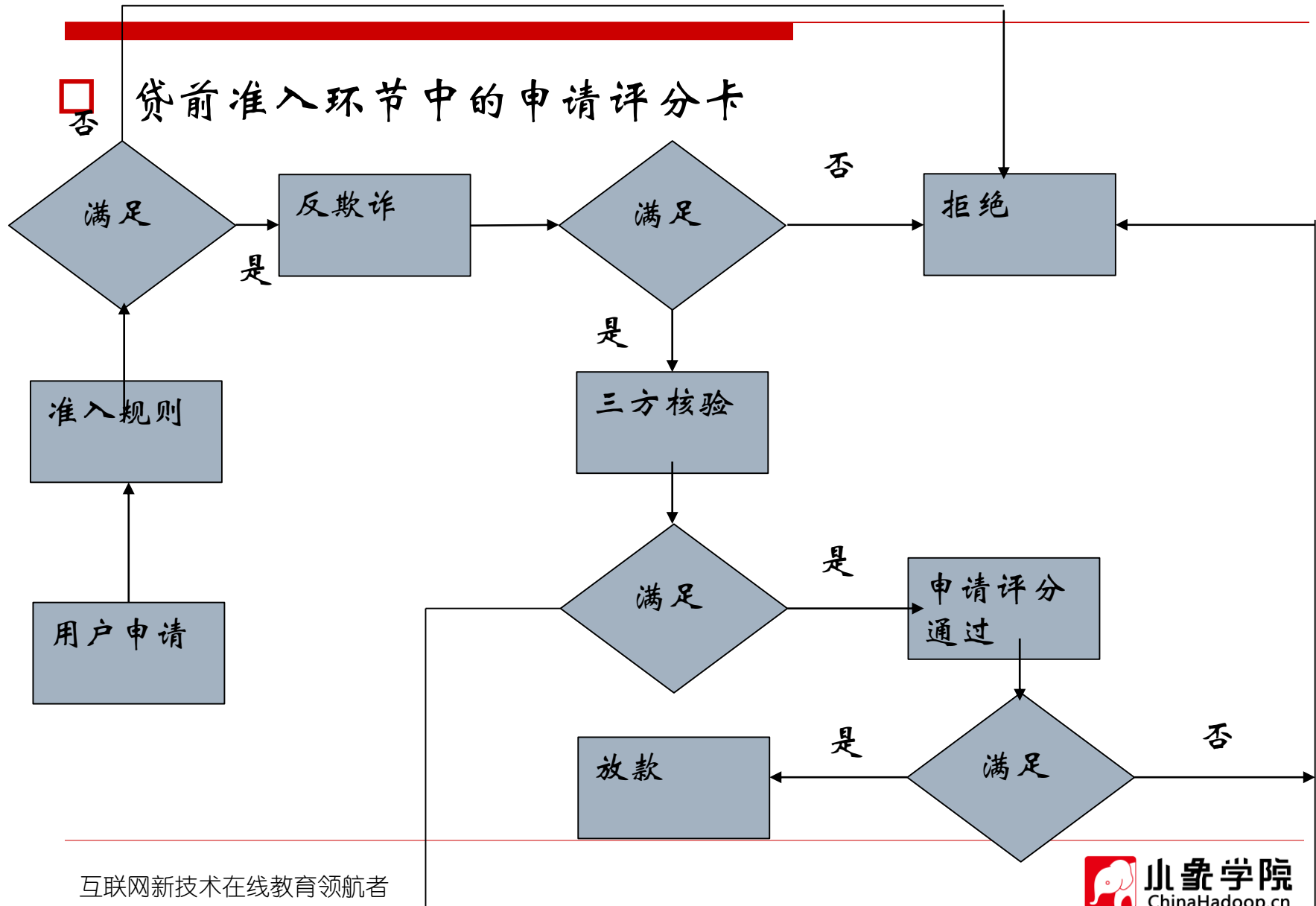
低分人群的违约率更高

➤ 和逾期概率等价

评分可以精准地反映违约/逾期概率，反之亦然

申请评分卡的重要性和特性

□ 贷前准入环节中的申请评分卡



申请评分卡的重要性和特性

□ 评分卡模型开发步骤

I. 立项：场景、对象、目的

II. 数据准备与预处理：账户、客群、内部／外部数据

III. 模型构建：逻辑回归 VS 机器学习，单一模型 VS 集成模型

IV. 模型评估：区分度、预测性、平稳性

V. 验证/审计：是否有计算错误、逻辑错误、业务错误

VI. 模型部署：从开发环境到生产环境、容量、并发度

VII. 模型监控：性能是否减弱，是否需要优化甚至重开发

申请评分卡的重要性和特性

□ 评分卡开发的常用模型

• 逻辑回归

优点: 简单, 稳定, 可解释, 技术成熟, 易于监测和部署

缺点: 准确度不高

• 决策树

优点: 对数据质量要求低, 易解释

缺点: 准确度不高

• 其他元模型

• 组合模型

优点: 准确度高, 不易过拟合

缺点: 不易解释; 部署困难; 计算量大

目录

信用违约风险的基本概念

申请评分卡的重要特性

贷款申请环节的数据介绍和描述

非平衡样本问题的定义和解决方法

贷款申请环节的数据介绍和描述

□ 申请评分卡常用的特征

个人信息

学历 性别 收入

负债信息

在本金融机构或者其他金融机构负债情况

消费能力

商品购买纪录，出境游，奢侈品消费

历史信用记录

历史逾期行为

新兴数据

人际社交 网络足迹 出行 个人财务

贷款申请环节的数据介绍和描述

□ 数据与描述

变量

loan_amnt
term
int_rate
emp_length
home_ownership
annual_inc
verification_status
purpose
zip_code
addr_state
delinq_2yrs
inq_last_6mths
mths_since_last_delinq
mths_since_last_record
open_acc
pub_rec
total_acc
pub_rec_bankruptcies
loan_status

解释

申请额度
产品期限
利率
工作年限
是否有自有住宅
年收入
收入核验状态
贷款目的
联系地址邮政编码
联系地址所属州
申贷日期前2年逾期次数
申请日前6个月咨询次数
上次逾期距今月份数
上次登记公众纪录距今的月份数
征信局中记录的信用产品数
公众不良记录数
正在使用的信用产品数
公众破产记录数
贷款状态

目录

信用违约风险的基本概念

申请评分卡的重要性和特性

贷款申请环节的数据介绍和描述

非平衡样本问题的定义和解决方法

非平衡样本问题的定义和解决方法

□ 非平衡样本的定义

在分类问题中，每种类别的出现概率未必均衡

例：

信用风险：正常用户远多于逾期/违约用户

流失风险：留存客户多于流失客户

□ 非平衡样本的隐患

降低对少类样本的灵敏性

非平衡样本问题的定义和解决方法

□ 非平衡样本的解决方案

过采样

- 优点: 简单, 对数据质量要求不高
- 缺点: 过拟合

欠采样

- 优点: 简单, 对数据质量要求不高
- 缺点: 丢失重要信息

SMOTE(合成少数过采样技术)

- 优点: 不易过拟合, 保留信息
- 缺点: 不能对有缺失值和类别变量做处理

非平衡样本问题的定义和解决方法

□ SMOTE算法

I. 采样最邻近算法，计算出每个少数类样本的K个近邻

II. 从K个近邻中随机挑选N个样本进行随机线性插值

III. 构造新的少数类样本

$$New = x_i + rand(0,1) \times (y_j - x_i), j = 1, 2, \dots, N$$

其中 x_i 为少类中的一个观测点， y_j 为k个邻近中随机抽取的样本

IV. 将新样本与原数据合成，产生新的训练集

疑问

□ 问题答疑：<http://www.xxwenda.com/>

■ 可邀请老师或者其他人回答问题

联系我们

小象学院：互联网新技术在线教育领航者

- 微信公众号：小象学院
- 新浪微博：小象AI学院

