

评分卡模型（B卡）的开发



讲师：安迪生

- 行为评分卡在风控领域中起着违约监控和额度管理的作用。此外，在巴塞尔模型里，内部评级法模型（Internal Rating Based Model）也非常依赖于行为评分卡。

目录

- ◆行为评分卡的基本概念
- ◆行为评分卡的特征构造
- ◆行为评分卡模型的开发

行为评分卡的基本概念

• 行为评分卡的基本概念

基本定义

- 根据贷款人放贷后的表现行为，预测未来违约风险概率的模型。

使用场景

- 和申请评分卡不同，行为评分卡用在贷款发放之后、到期之前的时间段，即“贷中”环节。

使用目的

- 监控贷款人在贷款结束之前的逾期 / 违约风险
- 在此基础上构建内部评级法中的PD模型

行为评分卡的基本概念

- 适用的信贷产品

还款周期长的信贷产品

- 长周期的产品有房贷、车贷、某些信用现金贷

循环授信类的信贷产品

- 信用卡、某些信用贷

注：

对于还款周期短的产品，一般无法也无需构建行为评分模型。可以用申请评分预测还款表现

对于较新的账户（例如开户时长在半年内），行为评分模型也不适用对其进行监控

行为评分卡的基本概念

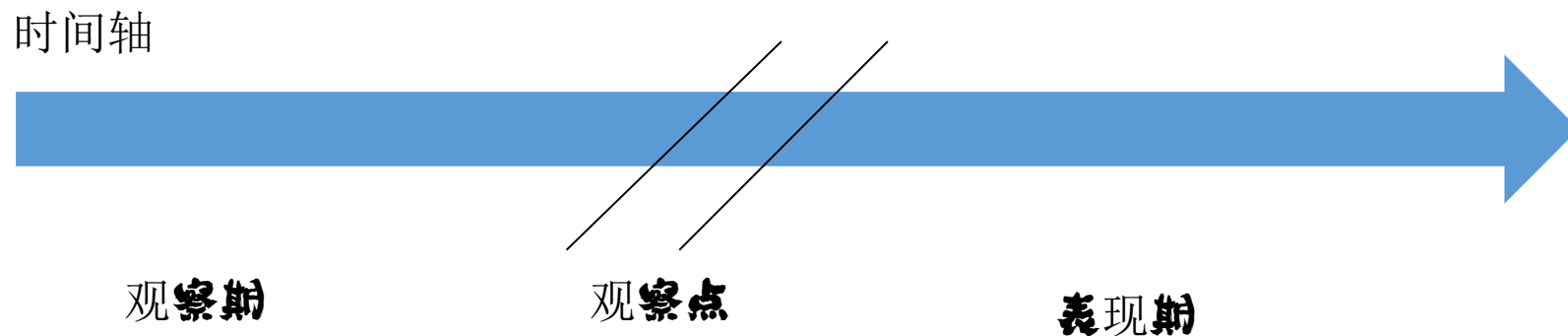
- 观察期与表现期

行为评分卡预测的是条件概率：

$$P(\frac{\text{未来违约}}{\text{当前未违约}})$$

构建和使用行为评分卡的两个时间段

- 观察期：当前及过去一段时间，用于搜集变量、特征的时间窗口，通常1年或半年以内
- 表现期：未来一段时间



行为评分卡的基本概念

- 观察期与表现期(续)

MOB

- Month on Book(MoB) = 观察点 - 贷款发放日

表现期的设定

不宜太短，否则

- 失去预测的意义
- 时间长度无法保证
- 概率难以预测

观察期的设定

- 不宜太长，否则MoB过长，大量客户无法进入模型
- 不宜太短，否则构建的变量，有效性不够

行为评分卡的特征构造

- 时间切片

定义：两个时刻间的跨度

例：观察日期之前30天内信用卡帐户的总消费额

基于时间切片的衍生

- 观察日期之前180天内，平均每月(30天)的逾期次数

常用的时间切片

- (1、2个)月，(1、2个)季度，半年，1年

时间切片的选择

- 不能太长：保证大多数样本都能覆盖到
- 不能太短：丢失信息

行为评分卡的特征构造

- 还款率类型特征

与还款行为有关的变量。还款行为由用户的还款能力与还款意愿决定。还款能力强、还款意愿高的客户，发生违约的可能性较小。通常情况下还与上月末欠款余额有关。因此在定义还款行为时，需要将还款额转换成还款率：

$$\text{本月还款率} = \frac{\text{本月总还款额}}{\text{上月末总欠款额}}$$

常用的还款率类型特征

过去6个月时间内，最大（小）的还款率

$$Max_Payment_L6M = \max\left\{\frac{payment_amount_i}{outstanding_{i-1}}, i = 1, 2, \dots, 6\right\}$$

过去6个月内，平均还款率

$$Avg_Payment_L6M = \frac{\sum_{i=1}^6 payment_amount_i}{\sum_{i=0}^5 outstanding_{i-1}}$$

行为评分卡的特征构造

• 额度使用率类型特征

关于授信额度使用情况的特征。使用额度较多的账户，未来的还款压力较大，相对容易发生违约。同时使用额度也受到授信总额的影响，需要将使用额度转换成使用率：

$$\text{额度使用率} = \frac{\text{本月使用额度}}{\text{授信总额度}}$$

注：

分母是授信总额度而非当前可以使用的余额

常用的额度使用率类型特征

过去6个月内平均额度使用率

$$Avg_Urate_L6M = \frac{1}{6} \sum_{i=1}^6 \frac{Spending_i}{Limit}$$

过去6个月内额度使用率升高的月份数

$$Increase_Urate_L6M = \#\left\{ \frac{Spending_{i+1}}{Limit} > \frac{Spending_i}{Limit}, i = 1, 2, \dots, 6 \right\}$$

行为评分卡的特征构造

- 逾期类型特征

关于M1，M2&M3状态的特征。当前较高的逾期天数容易导致较大的违约概率。注意，如果将M1，M2或M3作为特征，则违约定义是M3+。此处M1是指从最后还款日的后一天开始到第30天，M2是指从最后还款日的第31天起到第60天，以此类推。

常用的逾期类型特征

- 过去6个月 的最大逾期状态
- 过去6个月M1,M2,M3的次数

行为评分卡的特征构造

- 消费类型特征

反应用户消费行为的特征。例如在信用卡客户中，可以建立：

- “境外消费” 类型特征
- “提现” 类型特征
- “线上” 或 “线下” 类型特征
- 消费类目型特征

等等

行为评分卡的特征构造

- 其他类型的特征

除了上述介绍的常用的类型外，机构还可以结合其他内部数据或者三方数据构建特征。例如，可以计算观察期内在其他平台的申请次数，或者通过运营商数据构建通话详情类的变量

特别地，如果是抵押型产品如房贷，还可将抵押品估值类的变量放入模型中。例如，在按揭贷款的行为评分模型中，CLTV (current loan to value) 是一类重要的变量：

$$\text{CLTV} = \frac{\text{未还清贷款}}{\text{房屋实时估值}}$$

在分期付款中，DSR (Debt Service Ratio) 也是一类常用的变量

$$\text{DSR} = \frac{\text{每期还款本息}}{\text{当期收入}}$$

行为评分卡模型的开发

- 特征构造

观察期设为12个月。在观察期内，构造了时间窗口为1、3、6、12个月的时间切片

每种时间切片内包含的行为变量类型：

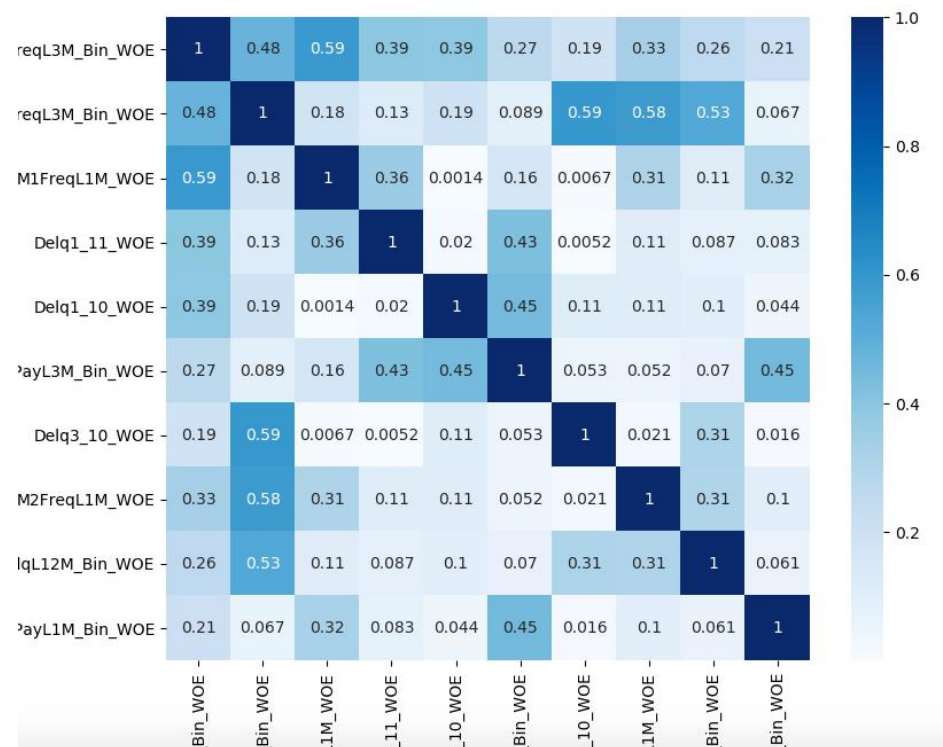
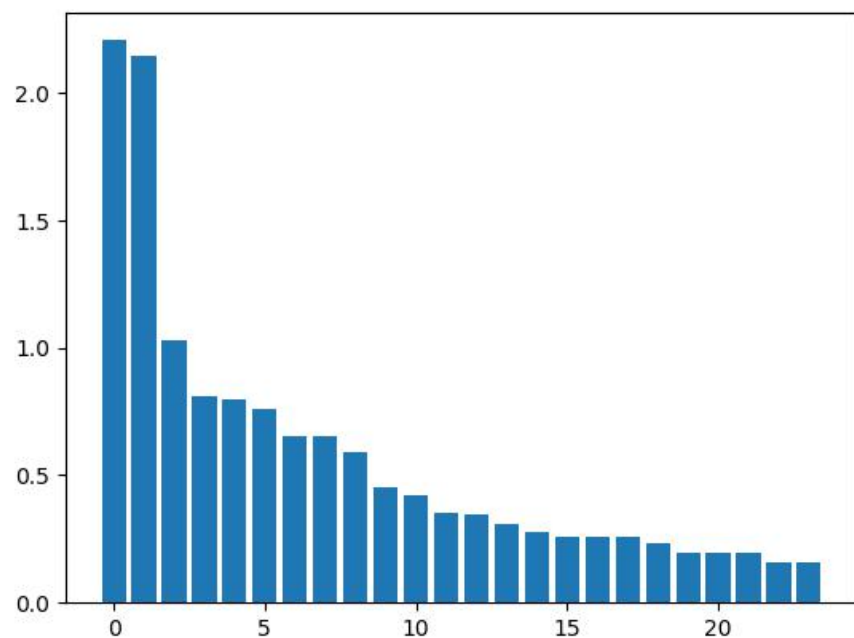
- 逾期类：最大逾期状态，M0/M1/M2的次数
- 额度使用类：最大月额度使用率，平均月额度使用率，月额度使用率增加的月份
- 还款类：最大月还款率，最小月还款率，平均月还款率

行为评分卡模型的开发

• 特征挑选

要求

- $IV > 0.1$
- WOE编码后，两两线性相关性低于0.6
- WOE编码后，共线性VIF < 10



行为评分卡模型的开发

- 第一次逻辑回归

```
...  
              coef      p-value  
M1FreqL3M_Bin_WOE -0.830779 0.000000e+00  
M2FreqL3M_Bin_WOE -0.348182 1.091914e-04  
M1FreqL1M_WOE      -0.099797 1.509626e-03  
Delq1_11_WOE       -0.219532 2.382291e-09  
Delq1_10_WOE       -0.260345 2.095687e-09  
minPayL3M_Bin_WOE  0.260876 1.689237e-06  
Delq3_10_WOE       -0.285445 9.591112e-03  
M2FreqL1M_WOE      0.283429 9.376783e-03  
maxDelqL12M_Bin_WOE -0.059895 2.667239e-01  
avgPayL1M_Bin_WOE  -0.474006 9.747042e-18  
intercept          -1.778571 0.000000e+00
```

发现变量minPayL3M_Bin_WOE和M2FreqL1M_WOE的系数为正；maxDelqL12M_Bin_WOE系数不显著。要单独检验这三个变量。

```
sm.Logit(y, trainData['minPayL3M_Bin_WOE']).fit().params # -0.825956  
sm.Logit(y, trainData['M2FreqL1M_WOE']).fit().params    # -29.060739  
sm.Logit(y, trainData['maxDelqL12M_Bin_WOE']).fit().pvalues # 0.00  
...
```

三个变量分别做单变量回归时，系数的符号和p值都是符合要求的。说明依然有多重共线性的存在。

行为评分卡模型的开发

- 第二次逻辑回归

我们将变量按照IV值进行降序排列，并按照以下步骤将合适的变量挑选后放入模型中

1. 将第1个变量放入模型中

2. For i in 2:N

- 将第i个变量放入模型中

- 如果模型中所有的变量的系数都为负且显著，则将第i个变量保留在模型中

- 否则剔除

3. 结束

行为评分卡模型的开发

- 第二次逻辑回归的结果

	Coef.	Std.Err.	z	P> z	[0.025	0.975]
M1FreqL3M_Bin_WOE	-0.8372	0.0207	-40.3954	0.0000	-0.8778	-0.7966
M2FreqL3M_Bin_WOE	-0.2194	0.0492	-4.4635	0.0000	-0.3158	-0.1231
M1FreqL1M_WOE	-0.0902	0.0310	-2.9124	0.0036	-0.1509	-0.0295
Delq1_11_WOE	-0.1556	0.0337	-4.6240	0.0000	-0.2216	-0.0897
Delq1_10_WOE	-0.1666	0.0391	-4.2645	0.0000	-0.2431	-0.0900
Delq3_10_WOE	-0.4313	0.0766	-5.6326	0.0000	-0.5813	-0.2812
avgPayL1M_Bin_WOE	-0.3650	0.0498	-7.3360	0.0000	-0.4625	-0.2675
intercept	-1.7851	0.0206	-86.5733	0.0000	-1.8255	-1.7446

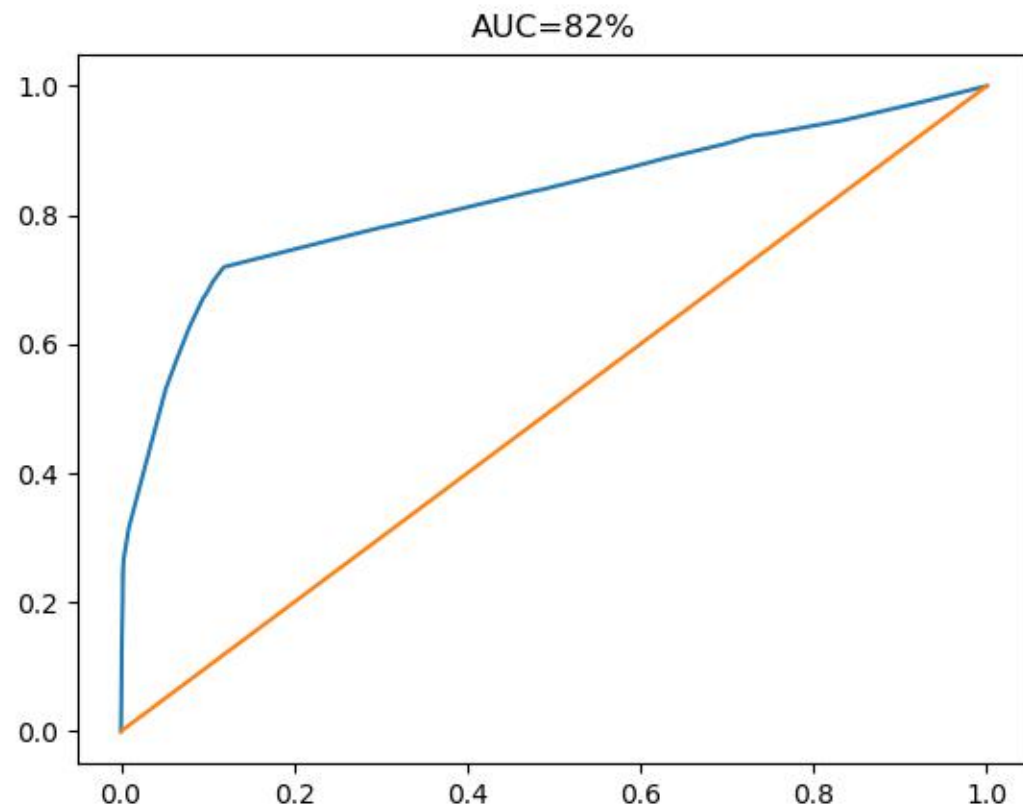
行为评分卡模型的开发

- 模型的性能

将概率按照尺度化的公式转换成分数：

$$score = Base\ Point + \frac{PDO}{\ln(2)}(-y)$$

在训练集上检验了模型的KS和AUC，分别达到60%和82%。



—— 秦路主讲 ——
七周成为数据分析师
七周为期，Get一条数据分析师职业黄金通道！



—— Python ——
数据分析与挖掘
集Python爬虫、数据采集、数据处理、数据分析与数据挖掘于一体，打造Python全栈工程师
主讲老师: 韦玮
VIP会员群+在线答疑+录播复习+1年反复观看

参团课程

案例为师, 实战为王
开启Python机器学习之路
科学规划全套课程体系，从入门到进阶，从理论到技巧，嵌入丰富课程案例讲解，逐步推进
讲师: 唐宇迪 深度学习领域多年一线实践研究专家

**独一无二的
数据仓库** 建模指南系列教程升级版
• 从企业视角进行数据规划以及数据仓库模型的搭建
• 高质量的数据库模型和技巧，以及丰富的例子
• 数据仓库架构理论和实践要领
资深讲师: BAO胖子 15年+BI从业经验
涉足电力、快消品、医药、信息服务行业的BI老兵

业务知识一站通
技术+业务，挣钱有门路！
—— 讲师: 陈文 ——



自己动手 丰衣足食
Python3网络爬虫实战案例
— 循序渐进，案例为王，诠释全面，思路制胜 —
讲师: 崔庆才 北航硕士，百万级热度爬文博主



讲师 丘祐玮
人人都爱数据科学家
Python数据科学精华实战课程



**数据分析
报告制作**
秘籍升级版
讲师: 陈丹奕 知乎大神，前百度资深数据分析师

**先机致胜
破冰AI**
—— 深度学习模型/框架与实战 ——
讲师: 唐宇迪 同济大学硕士
深度学习领域多年一线实践研究专家



BI、商业智能
数据挖掘 大数据
数据分析师
R语言 Python
机器学习
深度学习
人工智能
Hive Hadoop
Tableau
BIEE ETL
数据科学家
PowerBI