

通过数据挖掘手段构建信贷预期预警模型流程及案例分析

唐汉

August 9, 2019

汇报内容:

- 信贷预期预警模型构建的一般思路
- 国内外信贷预期预警模型构建案例分析

1. 信贷预期预警流程图

流程图将以 CRISP-DM (Cross Industry Standard Process of Data Mining) 为基础, 结合信贷风险预警专业特征进行构建。

(a) 业务理解 & 数据理解 (Business Understanding & Data Understanding):

目标特征 - 客户违约状况

自变量根据场景不同有所变化

- i. 客户属性数据
- ii. 交易行为数据
- iii. 还款行为数据
- iv. 透支行为数据

Rivai 构建的用于信贷预警分析的 6C 分析法 [1]:

- i. 个人情况 (Character)
- ii. 贷方资产 (Capital)
- iii. 偿还能力 (Capacity)
- iv. 担保 (Collateral)
- v. 经济状况 (Condition of Economy)
- vi. 其他限制 (Constraint)

(b) 数据准备 (Item Preparation):

数据准备通常包含如下步骤:

- i. 数据初步探索: 观察变量各个值分布
- ii. 缺失值处理:
通常舍弃缺失值超过 70% 的变量。对于其他缺失值和包含缺失值的变量根据情况采取不同的处理办法 (0, mean, median, mode, etc.)
- iii. 异常者检测和处理:
通过观察数据分布 (z-score, percentile) 检验异常值, 变更异常值使其符合数据分布特征。
- iv. 以及根据情况对变量采取的其他处理方式 (Binning Variable to Categorical variable, etc.)
- v. 特征工程 (Feature Engineering):
通常使用数据原始特征建模并不能得到准确率高的模型。通常需要变更, 增减特征来改善模型。信贷预警分析模型构建过程中常用的特征工程手段有:
 - A. 构造多项式特征
 - B. 通过相关性检验筛选特征
 - C. 依据信贷风险管理专业知识筛选特征
 - D. 其他根据具体情况所采取的特征工程手段

(c) 建模 (Modeling):

信贷预警分析常用逻辑回归模型作为基本模型 (Base model)。实际运用中使用决策树类 (Decision tree) 的算法建立模型, 原因在于决策树模型对于包含结构的数据 (Structured data) 有较好的预测能力, 并且决策树模型的拥有很好的自我解释能力, 便于刻画客户预警级别画像。其中 Xgboost gradient descent, light boosting gradient descent 都展现出优秀的优化效果。



Figure 1: CRISP-DM graph

- (d) 模型评估 (Evaluation):
常用的模型评估方式同样适用于信贷预警分析 (Confusion matrix, AUC, Classification Accuracy, F1-score, etc.)。其中漏报率 ($FN/(FN + TN)$) 是比误报率 ($FP/(FP + TP)$) 跟重要的模型衡量手段。原因在于漏报会逾期客户对业务损失更大。
- (e) 模型部署 (Deploying):
CRISP-DM 的最后一步, 部署模型后同时检测模型运行状况, 注意预测准确性的变化。

2. 信贷预期预警模型构建案例分析

- (a) 农村小额信贷预警模型构建 [2]:
印尼的 Mandala 构建的用于预测农村银行中小额贷款违约状况的模型, 利用 C5.0 decision tree 将银行的不良贷款率 (Non Performing Loans(NPL)) 从 11.95% 降为了 3.0%。并且根据模型, 担保额度是衡量信贷违约状况最重要的依据。
- (b) CSDN 多个作者通过 Xgboost, Random Forest, and logistic regression 构建了准确率较高 (AUC from 0.9 to 0.95) 的信贷预警模型 [3][4][5][6][7][8]。

References

- [1] Rivai, Veithzal dan Permata, Andria V. 2007, *Credit Management Handbook: Theory, Concept, Procedures and Application - Principle Guide for Students Banks and Customers*. PT RajaGrafindo Persada, Jakarta, Indonesia.
- [2] Mandala, G., N., Nawangpaiupi, C., B., Prakitikto, F., R. (2012). Assessing Credit Risk: an Application of Data Mining in a Rural Bank. *Procedia Economics and Finance*, 4:406-412.
- [3] <https://blog.csdn.net/u014281392/article/details/81121122>
- [4] <https://blog.csdn.net/u014281392/article/details/81137799>
- [5] <https://blog.csdn.net/u014281392/article/details/81152310>
- [6] <https://blog.csdn.net/u014281392/article/details/81177794>
- [7] <https://blog.csdn.net/luoganttc/article/details/77435064>
- [8] <https://blog.csdn.net/kMD8d5R/article/details/84453307>