

文章编号: 2095-2163(2024)02-0097-03

中图分类号: TP399

文献标志码: A

# 基于横向联邦学习的小微企业信贷决策模型研究

朱 军

(中国人民银行 贵州省分行, 贵阳 550001)

**摘要:** 本文对已有影响信贷决策的因素进行研究,综合得出影响银行信贷决策的指标,在使用PCA对指标数据进行简化的基础上,运用横向联邦学习算法,实现不同金融机构间信贷数据共享,得出精准信贷决策模型和小微企业贷款额度,这对解决小微企业融资难、融资慢以及银行担忧信贷风险而对小微企业惜贷的问题有重大意义。

**关键词:** PCA; 横向联邦学习; 数据共享

## Research on credit decision model of small and micro enterprises based on horizontal federal learning

ZHU Jun

(Guizhou Provincial Branch of the People's Bank of China, Guiyang 550001, China)

**Abstract:** This paper studies the factors influencing credit decisions and synthesizes indicators that affect bank lending decisions. Upon simplifying the indicator data using PCA (Principal Component Analysis), it employs a horizontal federated learning algorithm to facilitate the sharing of credit data among different financial institutions. This approach results in an accurate credit decision model and loan quotas for small and micro enterprises. This is significantly meaningful in addressing the difficulties small and micro enterprises face in financing, such as slow funding processes and banks' reluctance to lend due to concerns over credit risks.

**Key words:** PCA; horizontal federated learning; data sharing

## 0 引言

小微企业是小型企业、微型企业、家庭作坊式企业的统称,根据企业从业人员、营业收入、资产总额等指标以及行业特点进行分类<sup>[1]</sup>。小微企业数量占全国各类市场主体比例超过95%,是中国经济和社会发展中不可或缺的重要力量,在增加就业岗位、提高居民收入、保持社会和谐稳定等方面发挥着举足轻重的作用<sup>[2]</sup>。小微企业与成熟完善的大型企业相比,由于规模小、缺乏抵押物、信誉度低、财务报告不完善以及抵抗风险能力差等特点,出于信贷风险考虑,商业银行通常会拒绝向小微企业发放贷款。本文通过横向联邦学习隐私计算技术,实现不同金融机构间小微企业相关的财务指标、非财务指标以及企业信息共享,解决小微企业与银行之间的信息不对称问题,得出有效的小微企业信用评估方法,以此提升金融机构服务小微企业的意愿、能力和可持续性,助力稳市场主体、稳就业创业、稳经济增长。

目前应用最多的信贷管理模型主要是评分模型,包括Probit选择模型、多元判别模型、logistics等回归分类模型,这些模型通过分析信贷相关指标数据得到企业违约概率<sup>[1]</sup>。在众多模型中,logistics回归模型效果最好,对样本数据要求不严格,样本数据可以不服从正态分布,样本数据类型可以为多种类型,包括连续变量和离散变量,概率结果在 $[0, 1]$ 区间,结果直观便于理解。

## 1 小微企业信贷风险评价指标

综合对比财务指标、非财务指标、企业实力评价指标、企业信誉指标以及企业信息等指标。因为企业供应稳定性和销售稳定性指标能够体现企业的进货和销售情况,企业进货规模和企业销售规模在财务指标中已有体现,所以本文删除企业实力评价指标中的销售次数、有效进货次数、企业进货规模、企业销售规模指标,最终选取的指标为:财务指标(现金比率、速动比率、流动比率、贷款总额占销售收入

作者简介: 朱 军(1991-),男,硕士,中级工程师,主要研究方向:网络安全。Email: zhujun\_wenhan@163.com

收稿日期: 2023-02-09

哈尔滨工业大学主办 ◆ 系统开发与应用

比率、资产负债率、主营业务利润率、销售净利率、净资产收益率、总资产收益率、主营收入增长率、净资产增长率、存货周转率、应收账款周转次数、总资产周转率);非财务指标(员工数量、企业专利数量、企业贷款余额);企业实力评价指标(企业供应稳定性、企业销售稳定性、企业对上游企业的影响力、企业对下游企业的影响力);企业信誉指标(信用评级、违约情况);企业信息指标(最大股东持股比例、经营场所权属、企业经营年限)。

### 1.1 数据来源及共享

除非财务指标的企业贷款余额通过隐私计算求取外,其余数据可以向银行机构填写问卷调查得到。企业在各机构的贷款余额可使用隐私计算加密技术对企业各机构的贷款余额加密汇总得到。

如何对小微企业信用状况进行量化分析,对规避小微企业信用风险,降低小微企业贷款不良率极其重要。当前,多数金融机构主要通过白名单机制和风险模型对小微企业的信用状况进行判断,白名单机制带有主观性,而单个金融机构能应用到贷款风险模型中的数据往往非常有限。

为得到更加精准的小微企业信用贷款风控模型,多家金融机构可以合作一起搭建基于联邦学习的数据共享联邦学习风控模型。比如,金融机构 A 拥有小微企业信用标签  $Y$  和特征数据  $X_1$ ,金融机构 B 只拥有小微企业的特征数据  $X_2$ ,此时金融机构 A 和 B 可以通过横向联邦学习,先使用隐私求交 (PSI) 技术找到共同的小微企业(例如纳税人识别号),然后利用各自的数据一起训练模型,并将中间的结果进行加密传输,进而在无需共享各金融机构明文数据的同时达到提高各金融机构小微企业风控模型准确度的目的。

### 1.2 数据归一化处理

由于各指标数值相差较大,若直接计算,得到的实用模型不准确。如企业贷款余额总数可能达到千万级别,而经营场所权属最大值为 5,违约情况最大值为 1。为消除各指标数值大小带来的差异,需要对各指标进行归一化处理。本文指标均为正向指标,归一化公式如式(1)所示:

$$X_j = \frac{x_j - \min x_j}{\max x_j - \min x_j} \quad (1)$$

其中,  $x_j$  表示某企业指标  $j$  的数值;  $\max x_j$  表示某企业指标  $j$  的最大数值;  $\min x_j$  表示某企业指标  $j$  的最小值。

经过归一化处理后,各指标值区间为  $[0, 1]$ ,消

除了归一化处置前较大数值指标对较小指标的影响,有利于提高信贷模型结果的准确度。

### 1.3 基于 PCA 信贷指标降维

26 个原始指标进行归一化处理后,消除了指标不同量纲的影响。为进一步对 26 个指标进行处理,本文采用 PCA (Principal Component Analysis) 主成分分析方法,通过正交变换,将 26 个可能存在相关关系的指标变为若干个线性无关变量表示的数据。

使用 PCA 主成分分析方法,将原来 26 个高维度空间指标数据映射到低维空间,在降低信贷模型计算复杂度的同时保留了原始指标中的大部分信息。

## 2 横向联邦学习 logistics 信贷模型

### 2.1 实现方式

使用 PCA 主成分分析方法对原始指标降维后,对各金融机构的简化指标数据进行模型训练过程如图 1 所示。

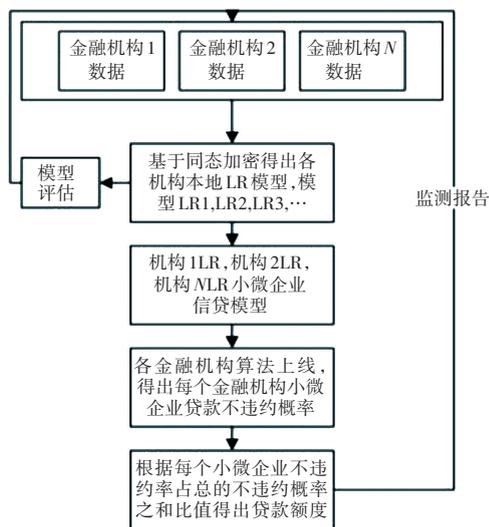


图 1 基于横向联邦学习的小微企业信贷决策模型训练过程

Fig. 1 Flow chart of credit decision model for small and micro enterprises based on horizontal federal learning

如图 1 所示,通过共享不同金融机构间小微企业信贷决策数据,得出小微企业精准信贷模型,最后根据每个小微企业的贷款不违约概率,得到某个金融机构对某个小微企业的信贷额度。

假设银行机构 A 拥有小微企业的特征为  $x_A$ , 银行机构 B 拥有小微企业的特征  $x_B$ , 逻辑回归值范围为  $0 \sim 1$ , 将银行 A、B 对客户违约的概率分别映射为  $0 \sim 1$  之间 10 等份的某一个值,如计算得到某用户信用卡不违约概率为 0.63, 小数点后第二位进行四舍五入,该值变为 0.6。

假设逻辑回归的参数为  $\theta$ ,  $\theta$  为一个列向量, 维度为信贷指标个数, 则在两方银行机构共同参与的逻辑回归模型中, 令  $h(x) = \theta^T x = \theta_A^T x_A + \theta_B^T x_B$ , 则包含两个银行机构小微企业信用数据的逻辑回归的输出模型, 式(2):

$$p(x) = \frac{1}{1 + e^{-h(x)}} \quad (2)$$

根据极大似然估计原理, 即使得函数  $-\prod_i p(y_i x_i) = \prod_i \frac{1}{1 + e^{-h(y_i x_i)}}$  值最小, 因为对函数求对数不影响其单调性, 故将该函数转化为求对数的最小值, 转化后即最小化函数  $\sum_i \log(1 + e^{-h(y_i x_i)})$ 。对该最小化目标函数进行二阶泰勒展开, 得到二阶泰勒展开式为  $\sum_i \log(1 + e^{-h(y_i x_i)}) = \sum_i \log 2 - \frac{1}{2} y_i \theta^T x_i + \frac{1}{8} (\theta^T x_i)^2$ , 令  $u_i = \theta^T x_i$ ,  $x_i = \theta_A^T x_{Ai} + \theta_B^T x_{Bi}$ , 则得到式(3):

$$L = \sum_i \log 2 - \frac{1}{2} y_i u_i + \frac{1}{8} (u_i)^2 \quad (3)$$

根据式(3), 分别求银行金融机构 A 和银行金融机构 B 的梯度, 式(4)和式(5):

$$\frac{\partial L}{\partial \theta_A} = \sum_i -\frac{1}{2} y_i x_{Ai} + \frac{1}{4} u_i x_{Ai} \quad (4)$$

$$\frac{\partial L}{\partial \theta_B} = \sum_i -\frac{1}{2} y_i x_{Bi} + \frac{1}{4} u_i x_{Bi} \quad (5)$$

参数的更新基于梯度、学习率以及初始参数值, 最后得到每个变量的权重值, 即生成并更新了本地模型参数。

## 2.2 小微企业信贷额度确定

传统的小微企业信用风险评估方法, 如评级方法、专家系统等属于定性分析, 这些方法简单易操作, 但是具有主观性强、小微企业信用评估误差大等缺点。究其原因, 主要是小微企业信用评分的指标体系不全面, 各指标权重没有统一标准, 进而导致各

金融机构对小微企业的信用量化不够准确。本文将小微企业贷款不违约率作为贷款额度的量化标准, 解决金融机构人工主观上对小微企业信贷额度不易度量的问题。

将基于横向联邦学习的逻辑回归模型应用于银行机构信贷决策中: 假设银行金融机构某年小微企业的贷款总额为  $A$ , 由模型得出每个小微企业能发放贷款且不违约的概率为  $p_k$ , 则该银行金融机构可以授予小微企业  $k$  的贷款额度  $y_k$  由式(6)得出:

$$y_k = \frac{p_k}{\sum_{k=1}^n p_k} \times A \quad (6)$$

其中,  $n$  为银行金融机构每年准备给与贷款的小微企业个数。

定量得出每个小微企业贷款不违约的概率以及每年的授信额度, 可以在科学信贷管理的同时减少信贷过程中的人工审批, 解决小微企业融资难、融资慢问题。

## 3 结束语

本文对已有小微企业信贷决策模型进行研究, 综合形成与小微企业相关的财务指标、非财务指标以及信誉度指标等 26 项指标, 并使用 PCA 主成分对指标进行简化, 得出最能代表原始指标信息的成分; 在简化的指标基础上, 基于横向联邦学习技术, 将不同金融机构之间小微企业的数据进行共享, 提高模型的准确率; 最后, 根据某金融机构小微企业贷款的不违约率, 按照占该金融机构总的不违约率的比重, 得出某金融机构可以授予小微企业的贷款额度。

## 参考文献

- [1] 王军锋. 基于 Logit 模型的威海地区商业银行小微企业信贷风险评估研究[D]. 山东: 山东大学, 2019.
- [2] 符建斌. 关于商业银行小微企业信贷风险评估模型的实证研究[J]. 黑龙江科技信息, 2015(30): 294.