

改进的支持向量机的个人信用评估模型

刘晓丽*, 李程, 王天珊



华北理工大学冀唐学院, 河北唐山 063210

摘要: 信用消费已经成为越来越多人的选择, 信用消费的膨胀在带来便利的同时, 随之而来的失信问题给信托机构带来不可估量的损失, 如故意欠款, 恶意透支消费等, 给信托机构在经营运行中造成巨大损失。一方面, 我国目前未能建立完善的个人信用记录, 客户信用信息难以共享, 获得相关数据的途径有限, 个人征信系统不够完善, 因而建立完善、自动化的个人信用评估体系, 科学识别个人信用风险, 从而实现银行信贷收益的最大化, 是十分必要的。在银行风险管理中, 面临的主要问题是如何衡量和规避银行的财务风险, 而个人信用评估是其中难度最大也是最重要的一部分。本文利用 XGBOOST 算法对特征进行重要度排序, 并根据模型准确度与特征数量之间的关系选取特征数量与具体的特征, 然后将数据中的 80% 作为训练集带入传统的支持向量机模型中, 并用剩余的数据进行测试, 准确率过低, 因此, 在个人信用风险评估方面, 对传统的支持向量机引入二次曲面, 建立无核二次曲面支持向量机模型, 准确率提高了 9.8 个百分点, 对银行引用风险控制具有指导意义。

关键词: 个人信贷评估; 支持向量机; 二次曲面; XGBOOST

DOI: [10.57237/j.cst.2022.01.004](https://doi.org/10.57237/j.cst.2022.01.004)

Improved Support Vector Machine Personal Credit Rating Model

Xiaoli Liu*, Cheng Li, Tianshan Wang

Ji Tang College of North China University of Science and Technology, Tangshan 063210, China

Abstract: Credit consumption has become the choice of more and more people. While the expansion of credit consumption brings convenience, the consequent problem of discredit brings incalculable loss to trust institutions, such as intentional arrears, malicious overdraft consumption, etc. to the trust in the operation of the huge losses caused. On the one hand, our country is unable to establish perfect personal credit record, customer credit information is difficult to share, access to relevant data is limited, personal credit system is not perfect, therefore, it is necessary to establish a perfect and automatic personal credit evaluation system, identify the personal credit risk scientifically, and realize the maximization of bank credit income. In the bank risk management, the main problem is how to measure and avoid the bank's financial risk, and personal credit assessment is the most difficult and important part. In this paper, we use the XGBOOST algorithm to sort the importance of features, and select the number of features and the specific features according to the relationship between the model accuracy and the number of features, then 80% of the data was taken into the traditional Support vector machine model as a training set and tested with the remaining data with low accuracy, so that, in terms of individual credit risk assessment, this paper introduces the quadric surface into the traditional Support vector machine and builds the kernel-free quadric surface Support vector machine model, which improves the accuracy

基金项目: 河北省人力资源和社会保障研究项目 (课题编号: JRS-2022-2025).

*通信作者: 刘晓丽, 1373325919@163.com

收稿日期: 2022-10-05; 接受日期: 2022-12-05; 在线出版日期: 2022-12-15

<http://www.computscitech.com>

by 9.8 percentage points and has a guiding significance for the bank reference risk control.

Keywords: Personal Credit Assessment; Support Vector Machine; Quadric XGBOOST

1 引言

信用消费已经成为越来越多人的选择，信用消费的膨胀在带来便利的同时，随之而来的失信问题给信托机构带来不可估量的损失，如故意欠款，恶意透支消费等，给信托机构在经营运行中造成巨大损失。一方面，我国目前未能建立完善的个人信用记录，客户信用信息难以共享，获得相关数据的途径有限，个人征信系统不够完善，因而建立完善、自动化的个人信用评估体系，科学识别个人信用风险，从而实现银行信贷收益的最大化，是十分必要的。

目前，关于个人信用风险的研究主要集中在评估模型的优化上，其评估模型的研究经历了从基于传统统计理论的模型到基于机器学习模型的过程。

基于传统统计理论的模型，主要以统计学领域的统计推断方法为主进行模型的构建。1968 年，Altman 提出了 Z-score 模型，该模型首次将统计学领域的方法应用于信用评估领域，突破了传统人工识别违约客户的主观性[1]。Steenackers 等提出了 Logistic 评估模型，解决了线性回归模型目标函数值不在 [0,1] 之间的问题[2]。方匡南等将信用卡信贷数据作为我国个人信用评估的实证数据，得到 Lasso-Logistic 模型优于全变量 Logistic 模型和逐步回归 Logistic 模型的结论[3]。

近年来，机器学习模型因其优秀的性能逐渐被应用到信用风险管理领域。Henley 等基于惰性学习理论设计了 K 邻近 (KNN) 模型，将其应用于评价个人信用状况，对比统计模型其拟合效果得到了较大提升[4]。Carter 等将决策树 (DT) 模型引入信用卡申请者的信用风险评估中，得到的模型拟合效果较优[5]。刘弘将神经网络用于贷款违约识别，得到

神经网络性能显著优于判别分析和决策树的结论[6]。庞素琳等对德国某银行的个人信贷数据建立了基于 C5.0 算法的银行个人信用评级模型[7]。

本文在个人信用风险评估方面，对传统的支持向量机引入二次曲面，建立无核二次曲面支持向量机模型，对银行引用风险控制具有指导意义

2 模型的建立及应用

2.1 个人信用数据集的预处理

由于数值型数据会受到量纲的影响，若未采用正确的表述方式会对模型性能造成负面影响。为使模型的训练效率得以提高，进行测试前需对数据进行标准化处理。

本文对 German Credit Data 数据集做如下处理

- (1) 缺失值。本文所选用的 German Credit Data 无信息缺失，也无需去除冗杂和噪声，即无需进行缺失值填充。
- (2) 指标赋值。本文样本中包含 20 个属性，有 7 个数值型属性变量，13 个离散型属性变量，离散型数据无法成为个人评估模型的输入，我们需要把所有的离散型字符属性转化为数值数据。
- (3) 标准化。我们对所有的数据集进行标准化处理，将 13 个离散型变量用数值 {1, 2, 3, ...} 表示，然后对每个属性的数值进行标准化处理，将数据映射到关于 0 上下波动的区间（大于 0 说明高于平均水平，小于 0 说明低于平均水平），本文选用 Z-Score 标准化方法进行处理，所用公式如下：

x^* 表示标准化后的数值

$$x^* = \frac{x - \bar{x}}{\sigma} \quad (1)$$

\bar{x} 为原始数据的均值：

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (2)$$

σ 为原始数据的标准差：

$$\sigma = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (3)$$

标准化后数据具体可见表格 1

表 1 个人信用标准化后的数据表

现有活期存款状态	-1.2539	-0.4588	1.13149	-1.2539	-1.2539	1.13149	1.13149	-0.4588
贷款期限/月	-1.2359	2.24707	-0.7383	1.74951	0.25682	1.25195	0.25682	1.25195
信用记录	1.34334	-0.5032	1.34334	-0.5032	0.42008	-0.5032	-0.5032	-0.5032
贷款用途	0.06267	0.06267	1.15579	-0.3017	-1.0305	1.15579	-0.3017	-0.6661
信贷金额	-0.7448	0.94934	-0.4164	1.63343	0.56638	2.04898	-0.1546	1.30255
储蓄账户/债券	1.83225	-0.6994	-0.6994	-0.6994	-0.6994	1.83225	0.56645	-0.6994
工作年限	1.33741	-0.3178	0.5098	0.5098	-0.3178	-0.3178	1.33741	-0.3178
债务占可支配收入的百分比	0.91802	-0.8698	-0.8698	-0.8698	0.02413	-0.8698	0.02413	-0.8698
性别及婚姻状况	0.4491	-0.9632	0.4491	0.4491	0.4491	0.4491	0.4491	0.4491
其他债务人/担保人	-0.3035	-0.3035	-0.3035	3.88314	-0.3035	-0.3035	-0.3035	-0.3035
现址居住年限	1.04646	-0.7656	0.14043	1.04646	1.04646	1.04646	1.04646	-0.7656
财产状况	-1.2931	-1.2931	-1.2931	-0.3409	1.5635	1.5635	-0.3409	0.61131

2.2 客户信息特征提取

介于已给的数据训练集，客户信息特征有 20 个因素，会导致后期处理数据变得非常复杂，所以我们选用 GBDT 和 XGBOOST 两种分类算法对客户信息特征的重要性进行打分，并对特征重要度进行排序。

2.2.1 BDT 分类算法

损失函数表示为：

$$L(y, f(x)) = \log(1 + e^{-yf(x)}) \quad (4)$$

其中， $y \in \{-1, 1\}$ 。则此时的负梯度误差为：

$$r_i = -\left[\frac{\partial L(y_i, f(x_i))}{\partial f(x_i)} \right]_{f(x)=f_{t-1}(x)} = \frac{y_i}{1 + \exp(y_i f(x_i))} \quad (5)$$

对于生成的决策树，我们各个叶子节点的最佳残差拟合值为：

$$c_{ij} = \arg \min_c \sum_{x_i \in R_{ij}} \log(1 + \exp(-y_i(f_{t-1}(x_i) + c))) \quad (6)$$

由于上式比较难优化，我们一般使用近似值代替：

$$c_{ij} = \frac{\sum_{x_i \in R_{ij}} r_i}{\sum_{x_i \in R_{ij}} |r_i|(1 - |r_i|)} \quad (7)$$

2.2.2 XGBOOST 分类算法

对于给定数据集 $D \{(x_i, y_i)\} (|D| = n, x_i \in R^n, y_i \in R)$

数据集有 n 个样本与 m 个特征，CART 空间

$F = \{f(x) = W_{q(x)}\} (q: R^m \rightarrow T, w \in R_T)$ ，树结构叶子节点系数为 q ， T 为叶子的数目，而 f_k 对应独立的叶子权重 w 和树结构 q ，xgboost 通过对以下的目标函数最小化得到：

$$L(\theta) = \sum_j I(\hat{y}, y_i) + \sum_k \Omega(f_k) \quad (8)$$

其中

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda_w^2 \quad (9)$$

$L(\theta)$ 为损失函数， Ω 为模型复杂度的惩罚项， γ 为 L1 正则化的系数， λ 为 L2 的正则化系数。目标函数可化为：

$$L(\theta) = \sum_T^{j=1} \left[\left(\sum_{i \in I_j} g_i \right) W_j + \frac{1}{2} \left(\sum_{i \in I_j} h_i + \lambda \right) W_j^2 \right] + \gamma T \quad (10)$$

令：

$$G_i = \sum_{i \in I_j} g_i, H_i = \sum_{i \in I_j} h_i \quad (11)$$

则目标函数变换为：

$$L(\theta) = \sum_T^{j=1} \left[G_j W_j + \frac{1}{2} (H_j + \lambda) W_j^2 \right] + \gamma T \quad (12)$$

此时目标函数变换为 W 的二次函数，目标函数取最小值时的最优权重为：

$$w_j^* = -\frac{G_j}{H_j + \lambda} \quad (13)$$

则目标函数成为：

$$L(\theta) = -\frac{1}{2} \sum_{j=1}^J \frac{G_j^2}{H_j + \lambda} + \gamma T \quad (14)$$

分类过程中同时对特征重要度进行了计算。每建立一个新树需重新查找分支点，即特征变量，计算这些变量的重要度，以此来决定树的分支点，用贪心算法搜索所有的特征，选择增益值最大的特征作为树的分裂点。增益计算公式为：

$$G_{\text{gain}} = \frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{(G_L + G_R)^2}{H_L + H_R + \lambda} - \gamma \quad (15)$$

增益值越大的分裂点表明特征的重要性越高，总体来看，对输出的所有树中计算每个特征的重要度均值，按重要度排序即可知道每个特征在分类中的重要性。

2.3 基于支持向量机的客户信用评价模型

2.3.1 支持向量机原理

支持向量机 (Support Vector Machine, SVM) 是一种线性分类器，它是一种二分类模型。其基本原理是将样本数据集通过变换得到高维特征空间，并在特征空间把样本数据集进行分类。支持向量机最初是为了解决二分类问题而提出的。它的目的是寻找一个最优超平面，使该平面在保证分类结果准确率较高的情况下，同时将超平面不同侧面样本间的间隔尽量拉大。支持向量机有着较好的泛化能力以及自主学习能力，可以在统计样本数据集较少的情况下，获得很好的统计规律。

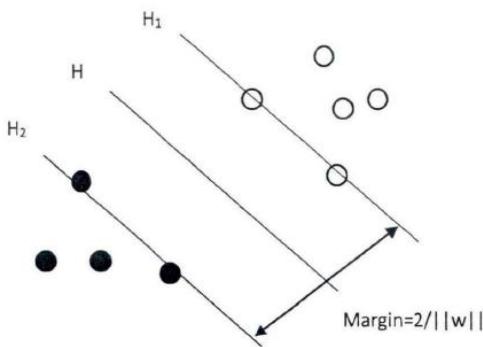


图 1 支持向量机原理

上图中的实心点和空心点表示两类样本， H 为分类超平面， H_1 、 H_2 是在所有的类里面距离 H 最近的样本的超平面， H_1 、 H_2 不但互相平行而且平行于 H ，二

者之间的距离叫做分类间距，最优分类超平面就是能将两类正确分开时的最大间隔的超平面。

SVM 算法可以简单描述如下：

对于给训练样本集 $\mathbf{T} = \{(x_1, y_1), (x_2, y_2) \dots (x_m, y_m)\}$

其中 $x_i = [x_i^1, x_i^2, \dots, x_i^n]^T$

$x_i \in R^n, y_i = \{-1, +1\}, i = 1, 2, 3, 4, \dots, m$

学习分类基本思路就是根据训练集 T 在样本空间中生成一个划分超平面，从而把不同类别的样本用超平面分开。此划分超平面应当对于训练样本的局部扰动“容忍”性最好。即超平面的鲁棒性更好，对未知样本的泛化能力更强。

支持向量机目的就是为了寻找到一个分类超平面 $W^T x + b = 0$ ，将正负类样本正确地划分开，同时要保证间隔最大化。

(1) 硬间隔支持向量机

假设超平面 (W, b) 能够将训练样本正确分类，即对于 $(x_i, y_i) \in T$ ，当 $y_i = 1$ ，则有 $W^T x + b > 0$ ；当 $y_i = -1$ ，则有 $W^T x + b < 0$ 。

令

$$\begin{cases} W^T x_i + b \geq +1, & y_i = +1 \\ W^T x_i + b \leq -1, & y_i = -1 \end{cases}$$

使得正类点都在 $W^T x + b = 1$ 外侧，负类点都在 $W^T x + b = -1$ 外侧。

如图所示，两条虚线间的距离代表支持向量机分开的两个类别之间的距离。我们的目标是在能把点集分开的情况下最大化这个距离。

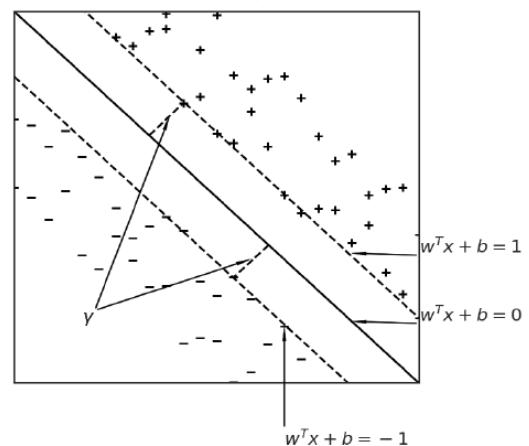


图 2 硬间隔支持向量机

两个不同类别的支持向量之间的距离为:

$$\gamma = \frac{2}{\|w\|}$$

硬间隔支持向量机最终需要寻找一个分类超平面 $\mathbf{W}^T x + b = 0$ 。其对应的分类决策函数为:

$$f(x) = \text{sign}(w^T x + b)$$

(2) 软间隔支持向量机

在现实生活中，很多数据点集是不可能被线性分割开，在这种情况下，我们允许 $w^T x + b = 1$ 和 $w^T x + b = -1$ 这两条线划分虚线变成一个软限制条件，部分点可以跨越所在分类的界限，如图 3 所示。

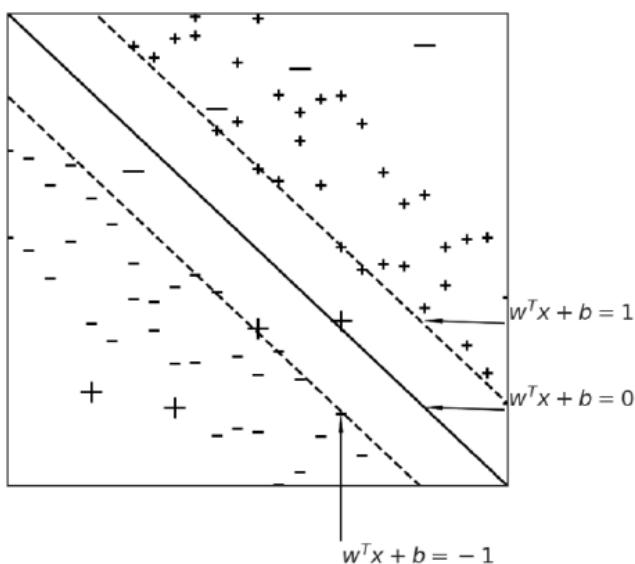


图 3 软间隔支持向量机

同时我们对于打破这种界限的行为给予一定的惩罚，引入了惩罚因子和松弛变量，惩罚大小与惩罚因子 C 和点集打破界限的距离有关，从而我们可以提出软间隔支持向量机模型:

$$\begin{aligned} & \min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^m \varepsilon_i \\ \text{s.t. } & y_i (w^T x_i + b) \geq 1 - \varepsilon_i \\ & \varepsilon_i \geq 0 \quad i = 1, \dots, m \end{aligned}$$

(3) 核函数支持向量机

在现实任务中大多数原始样本空间并不存在一个能正确划分两类训练样本的超平面。因为软间隔支持向量机本质上仍然还是一个线性的支持向量机其处理

近似线性分类问题的效果比较有效。因此，当训练点集不能被一个超平面分开，但能被一个非线性曲面分开时，软支持向量机的分类效果就不令人满意了。为了间接克服这个弱点，Cortes 与 Vapnik (1995) 通过一个非线性的投影函数 $\phi(x_i): \mathbb{R}^m \rightarrow \mathbb{R}^l$ ，将 m 维上的每个训练点 x_i 投射到 l 维 ($m \leq l$) 空间里的相应点 $\phi(x_i)$ 。将训练集数据从低维空间映射到高维空间。

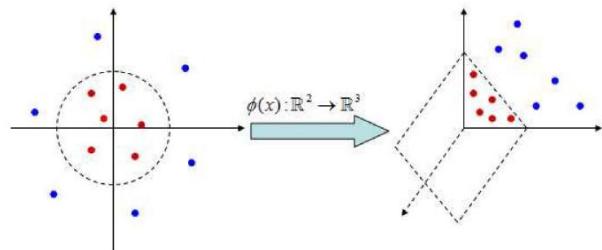


图 4 核函数支持向量机

2.3.2 无核二次曲面支持向量机

对于无核二次曲面支持向量机而言，我们希望能找到一个二次曲面 $g(x) = \frac{1}{2} x^T W x + b^T x + c$ ，并把 $g(x) = 0$ 作为分类曲面，而该分类曲面能完美的把正负类样本分割开来。与传统支持向量机最大化支持向量的最小间隔不同，二次曲面支持向量机的目标是使得所有正类样本点在 $g(x) = 1$ 外侧，而所有负类样本点在 $g(x) = -1$ 外侧，与此同时最大化所有样本点到分类曲面的相对几何距离。而对于任意一个新样本点 x，其对应的分类决策函数可以如下表示:

$$f(x) = \text{sign}(g(x))$$

所有样本点均被二次曲面正确分类应表达为:

$$y_i \left(\frac{1}{2} x^T W x + b^T x + c \right) \geq 1, \quad i = 1, 2, \dots, m$$

分类决策函数可表示为:

$$f(x) = \text{sign}(x^T W^* x + b^{*T} x + c^*)$$

2.4 模型准确率

2.4.1 混淆矩阵

我们按照模型输出结果和实际情况的组合可以把

数据划分为以下四类: 真正例, 假正例, 真负例, 假负例。即为混淆矩阵, 如表 2。

表 2 混淆矩阵

真实情况	预测结果	
	正例	负例
正例	真正例 (TP)	假正例 (FP)
负例	假负例 (FN)	真负例 (TN)

2.4.2 模型准确率检验

- (1) 查准率: 查准率代表了预测值为正中真正为正的比例:

$$precision = \frac{TP}{TP + FP}$$

- (2) 查全率: 查全率代表了真实为正中预测为正的比例:

$$recall = \frac{TP}{TP + FN}$$

3 实证分析

3.1 数据分析与处理

介于已有的数据, 客户信用特征因素过多, 会造成后期信息处理复杂程度增加, 会降低模型的效率和可用性。所以本文首先进行了特征提取。分别选取了 GBDT 与 XGBOOST 两个分类算法对特征重要性进行打分, 原理主要是通过测试多个阈值, 来求出特征重要性得分。介于 XGBOOST 对代价函数进行二阶泰勒展开, 收敛速度大大提升, 最终本文选择了 XGBOOST 进行特征提取, 结果如图 5 和图 6 所示。

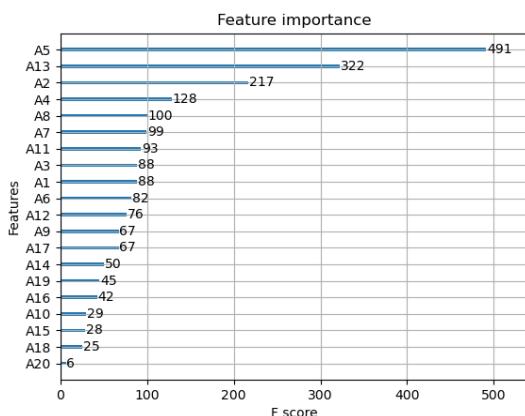


图 5 特征提取结果图



图 6 特征提取结果准确率图

为了在模型简便程度和准确率之间找到一个平衡, 本文对特征选取个数对整体预测准确性的影响进行了分析。从表中可以看出从 20 个指标降到 6 个指标, 虽然模型准确率从 82.00% 降到了 78.00%, 但模型却大大地简便了。10 个指标准确率虽比 6 个的高, 但两者准确率相差不大, 6 个指标有效地增加了模型的效率。

本文对表 1 进行了比对最终决定选择前六个指标。

上图 6 可以看到, 前六个指标按重要性得分从大到小依次排列分别为 A5、A13、A2、A4、A8、A7 与之相对应的分别为信贷金额、客户年龄、贷款期限、贷款用途、债务占可支配收入的百分比、工作年限。如表 3

表 3 模型的准确率

选取特征值的个数	模型准确率
1	69.50%
2	72.00%
3	76.00%
4	77.50%
5	76.50%
6	78.00%
7	78.00%
8	74.50%
9	76.00%
10	79.00%
11	77.50%
12	75.50%
13	78.50%
14	78.50%
15	78.50%
16	78.50%
17	79.00%
18	78.50%
19	81.50%
20	82.00%

3.2 支持向量机的训练

为了能进一步得出个人信用风险评估模型, 利用了 Matlab 里的工具箱 libsvm, 根据客户属性矩阵和标签建立起分类和预测模型。取 800 个客户 (560 个优质客户和 240 个不良客户) 的数据作为训练集, 对模型进行训练。

紧接着我们用剩余样本（140 个优质客户和 60 个不良客户）的数据信息作为测试集，通过上述模型进行了分类预测。最终准确率达到了 71.5%。

但是传统支持向量机对训练样本难以实施如果数据量很大，SVM 的训练时间就会比较长，并且传统支

持向量机对参数和核函数选择敏感，目前比较成熟的核函数及其参数的选择都是人为的，根据经验来选取的，带有一定的随意性与主观性。这也导致我们的模型分类不够客观，因而本文又对传统支持向量机进行了改进，引进了无核二次曲面支持向量机。

表 4 两种支持向量机的对比表

	均值	方差	中位数	25%分位数	75%分位数
传统支持向量机	71.5%	3.7%	67.6%	69.7%	73.4%
无核二次曲面支持向量机	81.3%	3.4%	81.9%	80.3%	84.2%

由表中可以看到无核二次曲面向量机的分类正确率比传统支持向量机高了 9.8%。因此本文的新模型泛化性更高，对于广泛的客户信用数据具有良好的适应性与可靠性。并且无核二次曲面支持向量机的方差也更小，说明本文的新模型的鲁棒性更好。

为评价模型判定“好客户”中的真正“好客户”比例，以及“坏客户”中真正的“坏客户”的比例，本文引进了两个性能度量，一类错误和二类错误。其中预测为“坏客户”实际为“好客户”的称为假正类，即一类错误。预测为“好客户”实际为“坏客户”的称为假负类，即二类错误。基于此，它们和查全率、查准率共同将传统支持向量机与无核二次曲面支持向量机两个模型进行了对比。

1) 一类错误和二类错误，如表 5 所示

表 5 两种支持向量机的两类错误对比表

		传统支持向量机	无核二次曲面支持向量机
一类错误	误判个数	25	8
	总数	140	140
	误判率	17.8%	5.7%
二类错误	误判个数	12	4
	总数	60	60
	误判率	20%	6.6%

从误判的结果来看，无核二次曲面支持向量机的误判率明显低于传统支持向量机。为了更加直观地显示两个模型的分类能力，本文引用了 ROC 曲线来体现。如图 7 所示

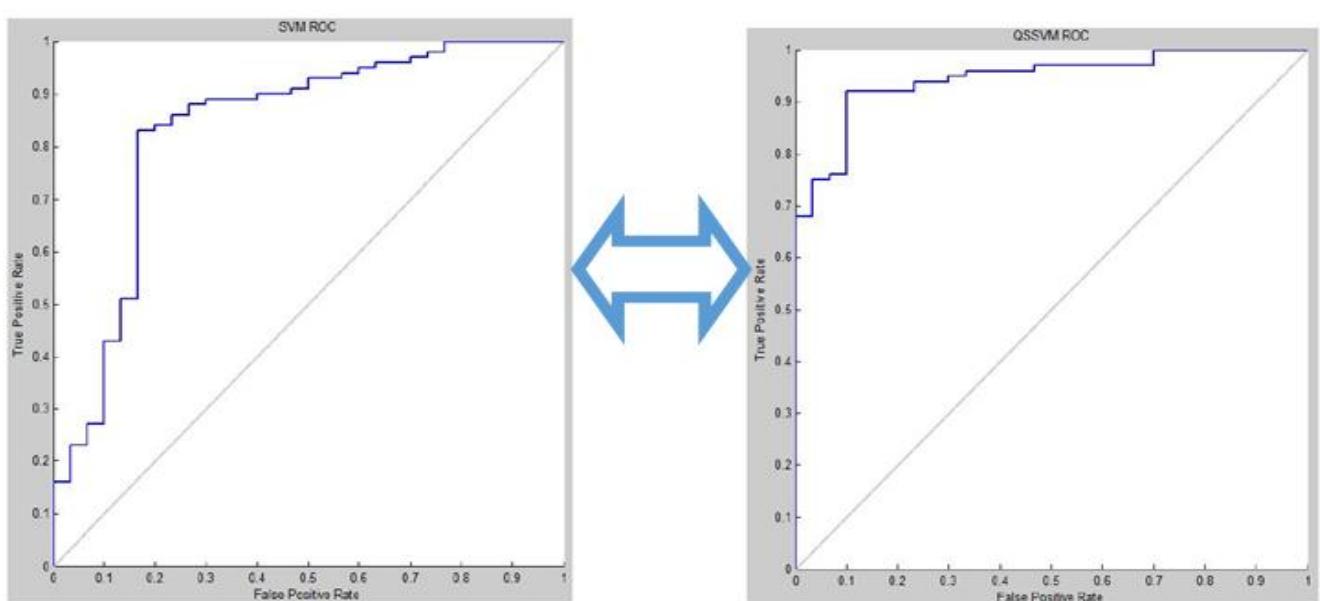


图 7 ROC 曲线图

图中可看出，无核二次曲面支持向量机模型的 ROC 曲线相比起传统向量机的 ROC 曲线 AUC 值更靠近左上方，故无核二次曲面支持向量机有更好的分类能力。

4 结论

本文构建了评价信用模型，引入了 XGBOOST 算法对特征进行重要度排序，在个人信用风险评估方面，对传统的支持向量机引入二次曲面，建立无核二次曲面支持向量机模型，准确率提高了 9.8 个百分点，对银行引用风险控制具有指导意义。

参考文献

- [1] Altman E. I. Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy [J]. Journal of Finance, 1968 (4): 589-609.
- [2] Steenackers A., Goovaerts M. J. A Credit Scoring Model for Personal Loans [J]. Insurance: Mathematics and Economics, 1989 (1): 31-34.
- [3] 方匡南, 章贵军, 张惠颖. 基于 Lasso-logistic 模型信用风险预警方法 [J]. 数量经济技术经济研究, 2014 (2): 125-136.
- [4] Henley W. E., Hand D. J. A k-Nearest-Neighbour Classifier for Assessing Consumer Credit Risk [J]. Journal of the Royal Statistical Society, 1996 (1): 77-95.
- [5] Carter C., Catlett J. Assessing Credit Card Applications Using Machine Learning [J]. IEEE Expert, 1987 (3): 71-79.
- [6] 刘弘. 基于神经网络的贷款违约识别研究 [J]. 统计研究, 2008 (7): 61-65.
- [7] 庞素琳, 巩吉璋. C5.0 分类算法及在银行个人信用评级中的应用 [J]. 系统工程理论与实践, 2009 (12): 94-104.
- [8] Koutanaci F. N., Sajedi H., Khanbabaei M. A Hybrid Data Mining Model of Feature Selection Algorithms and Ensemble Learning Classifiers for Credit Scoring [J]. Journal of Retailing and Consumer Services, 2015 (27): 11-23.
- [9] 莫贊, 张灿凤, 魏伟, 等. 基于 Bagging 集成的个人信用风险评估方法研究 [J]. 系统工程, 2019 (1): 143-151.
- [10] 白鹏飞, 安琪, Nicolaas Frans de ROOIJ, 等. 基于多模型融合的互联网信贷个人信用评估方法 [J]. 华南师范大学学报 (自然科学版), 2017 (6): 119-123.
- [11] 钟金宏, 邵晶晶, 李兴国. 基于组合分类策略的个人信用风险评估研究 [J]. 合肥工业大学学报 (自然科学版), 2020 (7): 996-1002.
- [12] 查志成, 梁雪春. 基于特征选择算法的个人信用预测模型 [J]. 计算机工程与技术, 2022, 43 (06): 1678-1685.
- [13] 陈舒期, 梁雪春, 改进的 SSVM 集成算法在信用风险评估中的应用 [J]. 计算机工程与设计, 2019, 040 (10): 2822-2826.
- [14] Niu K, et al., Resampling Ensemble Model Based on Data Distribution for Imbalanced Credit Risk Evaluation in P2P Lending [J]. Information Sciences. 2020, 536.
- [15] 夏金莲. 我国个人征信体系建设中信用信息采集制度的构建 [J]. 成都大学学报 (社会科学版), 2020 (03): 36-44.