

文章编号:1673-9469(2020)04-0099-06

DOI: 10.3969/j.issn.1673-9469.2020.04.016

基于异类类内超平面的模糊支持向量机及其应用

陈继强,余志鹏,张峰,张丽娜

(河北工程大学数理科学与工程学院,河北邯郸 056038)

摘要: 针对如何基于不平衡信贷数据对借贷人的信用进行合理准确地评估问题,基于个人信用的统计数据,提出了一种新的个人贷款信用风险的评估方法。首先,构建了个人贷款信用风险评估指标体系,结合 IV 模型,对各特征进行重要性分析。其次,结合模糊数学理论,设计了一种基于异类类内超平面的隶属函数。最后,结合支持向量机,构建了一种新的个人贷款信用风险评估方法——基于异类类内超平面的模糊支持向量机。结果表明:个人贷款信用风险评估的可用额度比值、逾期 30~59 天的次数、逾期 90 天及以上次数以及逾期 60~89 天次数这 4 个指标的 IV 值均大于 0.3,重要性较强,表明对贷款人的信用评估影响较大;所设计的隶属函数能对不同样本赋予不同权重,可充分体现不同样本的重要性;基于异类类内超平面的模糊支持向量机在一定程度上可以提高贷款人的信用风险评估精度,证明了该方法的有效性和可行性。

关键词: 分类;不平衡数据;支持向量机;隶属函数;个人信用评估

中图分类号:TP301.6

文献标识码:A

Fuzzy Support Vector Machine Based on Heterogeneous Class Internal Hyperplane and Its Application

CHEN Jiqiang, YU Zhipeng, ZHANG Feng, ZHANG Lina

(School of Mathematics and Physics Science and Engineering, Hebei University of Engineering, Handan, Hebei 056038, China)

Abstract: Aiming at the problem of how to reasonably and accurately evaluate the credit of the borrower based on the unbalanced credit data, a new method for evaluating the personal credit risk was proposed. Firstly, the personal credit risk assessment index system was constructed, and in order to analyze the importance of each characteristic of the data, IV values of the credit characteristics of the data were calculated with IV model. Secondly, combining with fuzzy mathematics theory, a membership function based on the heterogeneous class hyperplane was designed. At last, combining the traditional support vector machine (SVM), a fuzzy support vector machine (FSVM) based on the heterogeneous class hyperplane was constructed. The results show that IV values of four indexes of personal loan credit risk assessment, namely the available amount ratio, the number of overdue 30 to 59 days, the number of overdue 90 days or more, and the number of overdue 60 to 89 days are all more than 0.3, which are of great importance, indicating that they have a great impact on the credit assessment of the lender. The constructed membership function can provide different weights for different samples, and can reflect the importance of different samples. The constructed FSVM based on the heterogeneous class hyperplane can effectively improve the accuracy of estimating credit risk for personal loans, which shows the feasibility and the effectiveness of the proposed method.

Key words: classification; unbalanced data; support vector machine; membership function; personal credit evaluation

收稿日期:2020-08-05

基金项目:国家自然科学基金资助项目(62006068);新能源电力系统国家重点实验室开放课题资助项目(LAPS19012);河北省高等学校科学技术研究资助项目(ZD2020185, QN2020188)

作者简介:陈继强(1982-),男,河北保定人,博士,副教授,从事统计学习、数据挖掘、模糊数学理论及其应用领域研究。

个人信用评估是指银行等金融机构通过对影响借贷人还款的各种主客观条件环境的综合考察,运用严谨的科学方法评估借贷人的借贷信用及其还款能力,从而对借贷人是否具有履行偿还银行等金融机构的借贷能力进行评估,本质上是一个不平衡数据集的二分类问题。然而,我国的信征体系发展较晚,个人信用评估模型的研究尚不完善,亟需进行深入研究。

专家评分模型^[1]是最早被应用于个人信用评估的模型,但是由于其评估结果易受主观因素干扰等问题,逐渐被舍弃。后来,Carter& Catlett^[2]将 Fisher 判别分析运用到个人信用评估领域,Orgler^[3]利用线性回归分析方法建立了个人信用评估模型,都取得了一定的成果。然而,这种基于统计学的方法难以处理信贷数据集中的噪声数据。后来,随着机器学习理论的发展,一些学者将决策树与支持向量机等机器学习算法应用于信贷评估^[4-12],均取得较好的效果。但是,这类方法在处理不平衡信贷数据集时的分类能力是有限的。随着科学技术水平的发展,一些学者致力于集成模型在个人信用评估领域中的应用研究。例如,丁岚骆等^[13]构建了 Stacking 集成策略的借贷人违约风险评估模型。夏国斌^[14]将基于 Bagging 集成法的集成支持向量机模型运用于个人信用评估中,饶希^[15]将基于 Boosting 集成法的逻辑回归与支持向量的集成模型运用于个人信用评估中。郭孝敬^[16]运用了基于逻辑回归与决策树的集成模型,王思懿^[17]将随机森林与逻辑回归的集成模型运用到个人信用评估中,张碧月^[18]实验结果表明 RF-APSOLSSVM 的预测精度比 RF 模型和 APSOLSSVM 模型的精度高。王黎等^[19]的实验结果表明 GBDT 模型明显优于支持向量机和逻辑回归的效果。赵天傲等^[20]将 XGBoost 算法应用在个人信用风评估中,与决策树、支持向量机等模型进行对比分析,实证结果表明 XGBoost 模型比单一模型的预测精度提升效果比较明显。虽然集成模型分类效果相对于其单一模型分类效果较好,但是它们同单一模型一样也没有考虑不平衡信贷数据集中噪声数据对于分类结果的影响。

综上所述,当前关于个人信用评估问题的研究,现有方法没有很好地处理不平衡信贷数据集中噪声对模型分类精度的影响。因此,为了降低噪声数据对模型分类精度的影响,本文考虑了不同样本在分类问题中的不同作用,构建了一种基

于异类类内超平面的模糊支持向量机,为信贷评估问题提供了一种新方法。

1 支持向量机简介

支持向量机是 20 世纪 90 年代中期由 Cortes 和 Vapnik 提出的一种有监督的学习方法。其基本思想是通过最大化间隔寻找最优分类超平面,从而对数据进行分类,自提出以来被广泛应用^[21-23]。

为不失一般性,这里以二分类问题为例。假设给定的训练集为

$$T = \{(x_1, y_1), \dots, (x_n, y_n)\} \quad (1)$$

其中 $x_i \in R^m, y_i \in \{-1, +1\}$ 为类标签, $i = 1, 2, \dots, n$ 。对于非线性分类问题,通过映射 $\varphi(x)$,将训练数据集从原空间映射到高维特征空间,使得映射后的数据集在特征空间中线性可分。因此,支持向量机的学习问题可用如下二次规划问题来描述:

$$\begin{aligned} \min_{\omega, b} \quad & \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^n \xi_i \\ \text{s. t} \quad & y_i(\omega \cdot \varphi(x_i) + b) \geq 1 - \xi_i \\ & \xi_i \geq 0, i = 1, 2, \dots, n \end{aligned} \quad (2)$$

其中 $\|\cdot\|$ 表示向量的模, $C > 0$ 为惩罚参数, ξ_i 为松弛变量。原问题(2)的求解可通过构造拉格朗日函数,转化为如下对偶问题来求解:

$$\begin{aligned} \max_{\alpha} \quad & -\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j K(x_i, x_j) + \sum_{i=1}^n \alpha_i \\ \text{s. t} \quad & \sum_{i=1}^n \alpha_i y_i = 0 \\ & 0 \leq \alpha_i \leq C, i = 1, 2, \dots, n \end{aligned} \quad (3)$$

问题(3)中 $K(x_i, x_j) = (\varphi(x_i), \varphi(x_j))$ 为核函数, α_i 为拉格朗日乘子。

通过求解对偶问题(3)可得分类函数为

$$g(x) = \sum_{x_i \in SV} \alpha_i^* y_i K(x_i, x) + b^*$$

进一步可得决策函数为

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

这样,对于新的样本 x' ,其预测类别为 $f(x')$ 。

2 基于异类类内超平面距离的隶属函数

隶属函数是模糊数学中的一个重要内容,可用于为不同样本设定不同的权重^[24-27]。本文考虑到在支持向量机中对确定最优分类超平面起决定性作用的只有支持向量,而支持向量的位置一般

又位于距离异类点较近的区域。为此将模糊理论引入支持向量机中,设计了基于异类类内超平面距离的隶属函数。该方法的思想是根据每个向量到异类类内超平面距离的不同,将样本输入 x_i 到异类类内平面距离的函数作为隶属函数,对训练集中的输入 x_i 赋予相应的权重(隶属度)来提高支持向量在训练样本中的作用。

如图1所示,对于训练样本集 $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$, 假设正类样本(在图中用菱形表示)数目为 n_1 , 负类样本(在图中用圆形表示)数目为 n_2 , $n_1 + n_2 = n$ 。在线性不可分时,通过映射 $\varphi(x)$, 将数据集从样本空间映射到特征空间,数据集变为 $\{(\varphi(x_1), y_1), (\varphi(x_2), y_2), \dots, (\varphi(x_n), y_n)\}$ 。

由此可得正类样本的类中心为 $x^+ = \frac{1}{n_1} \sum_{i=1}^{n_1} \varphi(x_i)$, 负类样本的类中心为 $x^- = \frac{1}{n_2} \sum_{i=1}^{n_2} \varphi(x_j)$, 则以向量 $\bar{x} = x^+ - x^-$ 为法向量,且经过类中心 x^+ 的正类类内超平面为 $l_1: \bar{x} \cdot (\varphi(x) - x^+) = 0$, 其中 \cdot 表示向量的点积。同理可得以向量 $\bar{x} = x^+ - x^-$ 为法向量且经过类中心 x^- 的负类类内超平面为 $l_2: \bar{x} \cdot (\varphi(x) - x^-) = 0$ 。

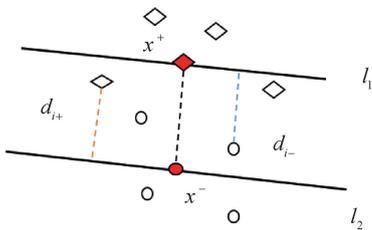


图1 基于异类类内超平面的距离示意图

Fig. 1 Diagram for the distance of heterogeneous hyperplane

因此,正类样本点 x_i 到负类类内超平面 l_2 的距离可定义为

$$d_{i+} = \frac{|\bar{x} \cdot (\varphi(x_i) - x^-)|}{\|\bar{x}\|} \quad (4)$$

同理,负类样本点 x_i 到正类类内超平面 l_1 的距离可定义为

$$d_{i-} = \frac{|\bar{x} \cdot (\varphi(x_i) - x^+)|}{\|\bar{x}\|} \quad (5)$$

令 $r^+ = \max_{y_i=+1} \{d_{i+}\}$, $1 \leq i \leq n_1$ 和 $r^- = \max_{y_i=-1} \{d_{i-}\}$, $1 \leq i \leq n_2$, 分别表示正类样本点到负类类内超平面的最大距离、负类样本点到正类类内超平面的最大距离。由于 d_{i+} 与 r^+ 的表达式中均含有相同的

分母 $\|\bar{x}\| = \|x^+ - x^-\|$, 为了简便计算可以把该分母去掉,于是式(4)中的 d_{i+} 可仅考虑 $|\bar{x} \cdot (\varphi(x_i) - x^-)|$, r^+ 取 $\max_{y_i=+1} \{|\bar{x} \cdot \varphi(x_i) - x^-\|$; 同理式(5)中的 d_{i-} 可仅考虑 $|\bar{x} \cdot (\varphi(x_i) - x^+)|$, r^- 取 $\max_{y_i=-1} \{|\bar{x} \cdot \varphi(x_i) - x^+\|$ 。因此,当 $y_i = +1$ 时, d_{i+} 的分子及 r^+ 可分别表示为

$$\begin{aligned} & |\bar{x} \cdot (\varphi(x_i) - x^-)| \\ &= |(x^+ - x^-) \cdot (\varphi(x_i) - x^-)| \\ &= \left| \left(\frac{1}{n_1} \sum_{k=1}^{n_1} \varphi(x_k) - \frac{1}{n_2} \sum_{m=1}^{n_2} \varphi(x_m) \right)^T \left(\varphi(x_i) - \frac{1}{n_2} \sum_{j=1}^{n_2} \varphi(x_j) \right) \right| \\ &= \left| \frac{1}{n_1} \sum_{k=1}^{n_1} K(x_i, x_k) + \frac{1}{n_2} \sum_{m=1}^{n_2} \sum_{j=1}^{n_2} K(x_m, x_j) - \frac{1}{n_2} \sum_{m=1}^{n_2} K(x_i, x_m) - \frac{1}{n_1 \cdot n_2} \sum_{k=1}^{n_1} \sum_{j=1}^{n_2} K(x_k, x_j) \right|, \\ r^+ &= \max_{y_i=+1} \left\{ \left| \frac{1}{n_1} \sum_{k=1}^{n_1} K(x_i, x_k) + \frac{1}{n_2} \sum_{m=1}^{n_2} \sum_{j=1}^{n_2} K(x_m, x_j) - \frac{1}{n_2} \sum_{m=1}^{n_2} K(x_i, x_m) - \frac{1}{n_1 \cdot n_2} \sum_{k=1}^{n_1} \sum_{j=1}^{n_2} K(x_k, x_j) \right| \right\} \end{aligned}$$

同理,当 $y_i = -1$ 时, d_{i-} 的分子及 r^- 可分别表示为

$$\begin{aligned} & \left| \frac{1}{n_1} \sum_{k=1}^{n_1} K(x_i, x_k) - \frac{1}{n_1} \sum_{k=1}^{n_1} \sum_{h=1}^{n_1} K(x_h, x_k) - \frac{1}{n_2} \sum_{m=1}^{n_2} K(x_i, x_m) + \frac{1}{n_1 \cdot n_2} \sum_{h=1}^{n_1} \sum_{m=1}^{n_2} K(x_h, x_m) \right|, \\ r^- &= \max_{y_i=-1} \left\{ \left| \frac{1}{n_1} \sum_{k=1}^{n_1} K(x_i, x_k) - \frac{1}{n_1} \sum_{k=1}^{n_1} \sum_{h=1}^{n_1} K(x_h, x_k) - \frac{1}{n_2} \sum_{m=1}^{n_2} K(x_i, x_m) + \frac{1}{n_1 \cdot n_2} \sum_{h=1}^{n_1} \sum_{m=1}^{n_2} K(x_h, x_m) \right| \right\} \end{aligned}$$

这样,为了表示第 i 个样本隶属于对应类标签的程度,可设计如下基于样本点到异类类内超平面距离的隶属函数

$$u_i = \begin{cases} 1 - \frac{d_{i+}}{r^+ + \sigma}, y_i = +1 \\ 1 - \frac{d_{i-}}{r^- + \sigma}, y_i = -1 \end{cases} \quad (6)$$

式(6)中 σ 是一个给定的很小的正数,目的是为了 保证隶属函数 μ_i 的取值大于0。

3 基于异类类内超平面的模糊支持向量机

传统的支持向量机是把所有的训练样本看作

同等重要的,这使得获得的分类超平面对噪声数据或者非支持向量样本较为敏感,从而导致获得的最优分类超平面存在偏差,进而影响分类器的分类精度。为了克服上述问题,本小节将上节设计的隶属函数赋予每个向量不同的权重,并结合传统支持向量机,构建基于异类类内超平面的模糊支持向量机。

在引入隶属函数 μ_i 后,训练样本集 $\{\varphi(x_i), y_i\}$ 变为 $\{\varphi(x_i), y_i, \mu_i\}$ 其中 $0 \leq \mu_i \leq 1, i = 1, 2, \dots, n$, 它表示第 i 个样本隶属于对应类标签的程度。这样,可建立如下二次规划问题:

$$\begin{aligned} \min_{\omega, b, \xi} \quad & \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^n \mu_i \xi_i \\ \text{s. t} \quad & y_i(\omega \cdot \varphi(x_i) + b) \geq 1 - \xi_i \\ & \xi_i \geq 0, i = 1, 2, \dots, n \end{aligned} \quad (7)$$

其中 $C > 0$ 为惩罚参数, μ_i 为式(6)给出的隶属函数, ξ_i 为松弛变量。对于原问题(7),通过构造拉格朗日函数,可转化为如下对偶问题:

$$\begin{aligned} \max_{\alpha} \quad & -\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j K(x_i, x_j) + \sum_{i=1}^n \alpha_i \\ \text{s. t} \quad & \sum_{i=1}^n \alpha_i y_i = 0 \\ & 0 \leq \alpha_i \leq \mu_i C, i = 1, 2, \dots, n \end{aligned} \quad (8)$$

其中 $K(x_i, x_j) = (\varphi(x_i), \varphi(x_j))$ 为核函数, α_i 为拉格朗日乘子。结合传统支持向量机,可设计求解上述优化问题的求解算法,该算法就称为基于异类类内超平面的模糊支持向量机。

下面将通过 UCI 数据集和信用评估问题的实验验证所设计算法的有效性。

4 实验

UCI 数据库是加州大学欧文分校(University of CaliforniaIrvine)开发的用于机器学习的数据库,UCI 数据集是众多学者常用的标准测试数据集。为了验证上述构建的基于异类类内超平面的模糊支持向量机的有效性,本节将首先利用 UCI 数据集选择部分数据进行实验,然后在个人信用评估问题中验证该方法的有效性。

4.1 模型评估指标

考虑到数据集的不平衡性,选取 F_1 值。 F_1 是非平衡数据集分类问题中有效的评价准则之一^[28],它是基于混淆矩阵的一个评估指标,详见式(9)。

在分类任务中,混淆矩阵是一个 $N \times N$ 的矩阵,其中 N 是被预测的类别数,在本文中由于是一个二

分类任务,所以混淆矩阵是一个 2×2 的矩阵。表 1 是一个二分类任务中的混淆矩阵实例表。

表 1 基于二分类的混淆矩阵

Tab. 1 Confusion matrix based on binary

	预测正类	预测负类
真实正类	真正类(TP)	假负类(FN)
真实负类	假正类(FP)	真负类(TN)

F_1 的计算表达式为^[25]

$$F_1 = \frac{2TP}{2TP + FP + FN} \quad (9)$$

4.2 基于 UCI 数据集的实验

本小节选取 UCI 数据库中 4 个不同的数据集进行实验,各数据集的样本数量、属性数量、正负类样本数量详见表 2。

表 2 数据集描述

Tab. 2 Description of datasets

	Diabetes	Banknote	Bupa	Haberman
样本数量	768	1 372	345	306
属性数量	8	4	6	3
正样本数量	268	615	200	225
负样本数量	500	757	145	81

为了验证本文提出的基于异类超平面的模糊支持向量机(DHFSVM)的有效性,将它与传统的支持向量机(SVM)、决策树(DTC)、高斯朴素贝叶斯模型(GNB)三种模型分别在 4 种 UCI 数据集上进行实验,实验结果如图 2 所示。

从图 2 中可以看出,DHFSVM 算法在 Bupa、Diabetes、Haberman 等 3 个数据集上的 F_1 值为最高,在 Banknote 数据集上为第二高。这说明,本文所构建的 DHFSVM 算法考虑了不平衡数据集中不同样本在分类过程中的不同作用,有效地提高了不平衡数据分类的准确性。

4.3 个人信用评估中的应用

在个人信用评估问题中,信贷数据涉及到个人隐私等问题,众多信贷数据无法开放获取,因此本文中采用来自 Kaggle 数据科学竞赛平台上的信贷数据集,该数据集的名称为“Give Me Some Credit”。在该数据集中,包含 10 个解释变量 $x_i, i = 1, 2, \dots, 10$, 一个被解释变量 y 。具体如表 3 所示。

由于该数据集数量较大(将近 15 万条),本实验随机选取了其中的 1 000 条数据,其中违约数据为 200 条,未违约数据为 800 条。在图 3 中,给出

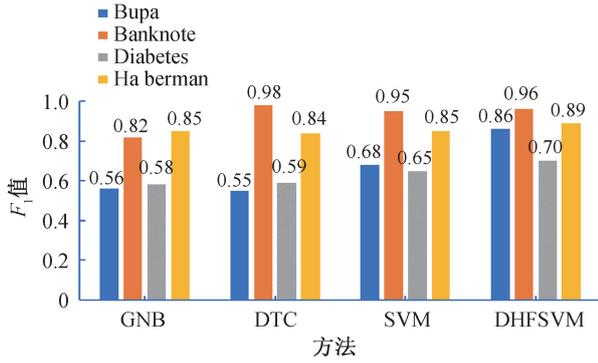


图 2 基于 GNB、SVM、DTC 与 DHFSVM 在数据集上的 F_1 值柱状图

Fig. 2 Histogram of F_1 value with GNB, SVM, DTC and DHFSVM

了 GNB、SVM、DTC 以及 DHFSVM 基于“Give Me Some Credit”数据集的分类混淆矩阵。

由式(9) F_1 值及如图 3 所示的混淆矩阵可以计算出各个模型对应的 F_1 值,结果如图 4 所示。

由图 4 所示的结果可以看出,在 GNB、SVM、DTC 和 DHFSVM 4 种算法中,本文所构建的 DHFSVM 算法表现最好,取得了最大的 F_1 值 0.87。这说明,在分类算法的构建时,充分考虑不平衡数据中不同样本点(包含噪声)所起的不同作

表 3 变量及其描述
Tab. 3 Variables and their description

变量	变量描述	数值类型
y	是否违约	文本
X_1	可用额度比值	数值
X_2	年龄	数值
X_3	逾期 30~59 d 的次数	数值
X_4	负债率	数值
X_5	月收入	数值
X_6	信用贷款数量	数值
X_7	逾期 90 d 的次数	数值
X_8	固定资产贷款量	数值
X_9	逾期 60~89 d 的次数	数值
X_{10}	家属数量	数值

用,可有效提高分类算法的精度,也说明了本文所构建的 DHFSVM 算法可以较好地应用于个人信用评估问题。

5 结论

本文结合模糊数学理论、经典支持向量机,构建了一种基于异类类内超平面的模糊支持向量机(DHFSVM)。与 GNB、DTC、SVM 3 种方法对比发现:(1) DHFSVM 算法在 Bupa、Diabetes、Habeman 等 3 个数据集上的 F_1 值最高,在 Banknote 数据集上第二高,表明所构建的 DHFSVM 算法可通过赋

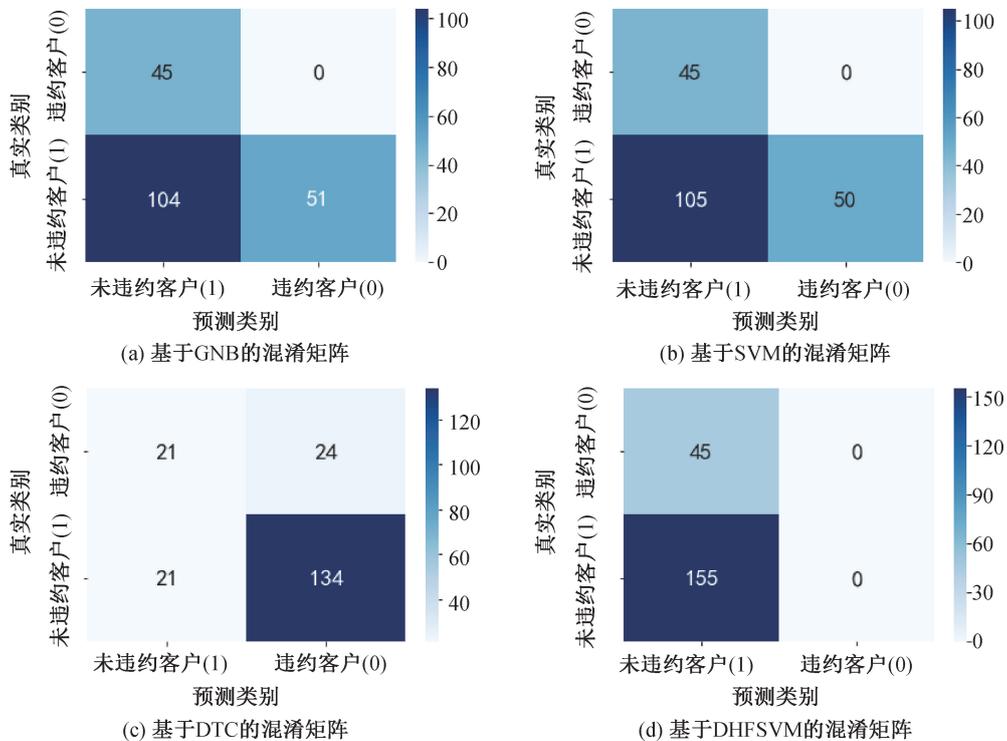


图 3 基于 GNB、SVM、DTC 和 DHFSVM 的混淆矩阵

Fig. 3 Confusion matrices with GNB, SVM, DTC and DHFSVM

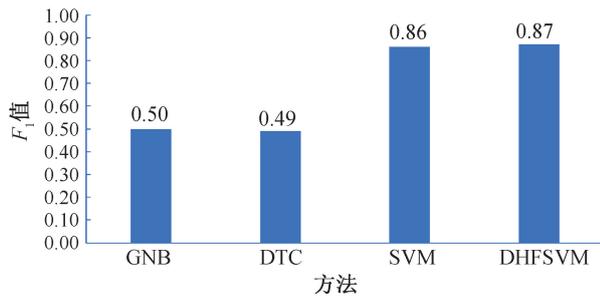


图4 基于GNB、SVM、DTC与DHFSVM
在信贷数据集上的 F_1 值柱状图

Fig. 4 Histogram of F_1 value on the credit data with GNB,
SVM, DTC and DHFSVM

予不同样本不同的权重来降低噪声对风险评估带来的影响;(2)在Kaggle数据科学竞赛平台上提供的真实信贷数据集进行的实验发现,DHFSVM算法表现最好,取得了最大的 F_1 值。表明DHFSVM算法可有效提高不平衡信贷数据集的分类准确性,可为个人贷款业务中个人信用风险评估问题提供借鉴。

参考文献:

- [1] 严鸿和,陈玉祥,许昭明,等. 专家评分机理与最优综合评价模型[J]. 系统工程理论与实践,1989(2):19-23.
- [2] CARTER C, CATLETT J. Assessing Credit Card Application using Machine Learning[J]. IEEE Expert Magazine, 1987, 2(3): 71-79.
- [3] ORGLER Y. A Credit Scoring Model for Commercial Loans[J]. Journal of Money, 1970, 2(4):435-445.
- [4] CHATTERJEE S, BARCUN S. A Nonparametric Approach to Credit Screening[J]. Journal of the American Statistical Association,1970,65(329):150-154.
- [5] 李旭升. 贝叶斯网络分类模型研究及其在信用评估中的应用[D]. 成都:西南交通大学,2006.
- [6] 金香. 基于逻辑回归的个人信用评分模型的研究[D]. 长春:东北财经大学,2016.
- [7] LAN J, YAN G. Reasearch on Individual Credit Evaluation Model Based on Iterative Self-adaptive Stochastic Decision Tree[J]. Economic Mathematics, 2018, 35(3):83-90.
- [8] 张玥,赵凯,黄全生. 组合核支持向量机个人信用评估模型[J]. 重庆大学工商学报:自然科学版,2019(6):37-43.
- [9] 黄巍,张靓,唐友. 基于SVM算法的个人信用评估方法的完善[J]. 黑龙江八一农垦大学学报,2016(2):105-110+114.
- [10] WANG X, LI Y. Research on Individual Credit Risk Evaluation Based on Support Vector Machine[J]. Journal of West China Normal University, 2017, 38(2): 195-198.
- [11] BAESENS B, VAN GESTEL T, VIAENE S, et al. Bench Marking State-of-the-art Classification Algorithm for Credit Scoring[J]. Journal of the Operation Research Society, 2003, 6(54): 627-635.
- [12] 贺方超. 基于 Logistic 回归和 SVM 的客户流失问题的研究[D]. 大连:大连理工大学,2018.
- [13] 丁岚骆,品亮. 基于 Stacking 集成策略的 P2P 网贷违约风险预警研究[J]. 投资研究,2017(4):41-54.
- [14] 夏国斌. 基于 Bagging 算法的集成支持向量机在个人信用评级中的应用[D]. 济南:山东大学,2018.
- [15] 饶希. 基于 LASSO-SVM 和 Logistic 的组合模型在个人信用评估中的应用[D]. 南昌:江西财经大学,2019.
- [16] 郭孝敬. 基于机器学习组合模型的个人信用评估[D]. 南昌:江西财经大学,2019.
- [17] 王思懿. 基于几种常见模型的 P2P 网贷借款人信用风险控制对比研究[D]. 武汉:华中师范大学,2019.
- [18] 张碧月. 基于 RF 和 APSOLSSVM 的两阶段信用评估研究[D]. 成都:西南财经大学,2016.
- [19] 王黎,廖闻剑. 基于 GBDT 的个人信用评估方法[J]. 电子设计工程,2017(15):68-72.
- [20] 赵天傲,郑山红,李万龙,等. 基于 XGBoost 的信用风险分析的研究[J]. 软件工程,2018(6):29-32.
- [21] 赵建差,刘俊娥,石祥锋. 基于 GA-LSSVM 深基坑墙体侧斜滚动预测模型研究[J]. 河北工程大学学报:自然科学版,2018(2):49-52+57.
- [22] 杨扬,哈明虎. 以竞赛带动教学模式下 BIM 教学质量影响因素研究[J]. 河北工程大学学报:自然科学版,2018(3):106-109.
- [23] 戴薇,石崇,张金龙. 浅埋地下洞室围岩岩体力学参数反分析研究[J]. 河北工程大学学报:自然科学版,2019(1):59-63.
- [24] 丁欣,谢祥俊,赵春兰,等. 基于动态隶属度的模糊时间序列在我国居民消费水平预测上的应用[J]. 模糊系统与数学,2019(1):164-174.
- [25] 邱望仁,刘晓东. 模糊时间序列模型研究综述[J]. 模糊系统与数学,2014(3):173-181.
- [26] 刘贤锋,刘枚莲,肖吉军,等. 网络消费者品牌选择的模糊 LOGIT 模型研究[J]. 模糊系统与数学,2014(4):171-179.
- [27] 汤春华,曹仁保,刘茂丹. 模糊环境下零售商竞争的供应链协调研究[J]. 模糊系统与数学,2018(2):159-169.
- [28] JOSHI M. On Evaluating Performance of Classifiers for Rare Classes[C]//Proceedings of the 2nd IEEE International Conference on Data Mining, Maebishi, 2002: 641-644.

(责任编辑 王利君)