

平滑削边绝对偏离惩罚截断 Hinge 损失支持向量机的财务危机预报

刘遵雄, 黄志强*, 刘江伟, 陈 英

(华东交通大学 信息工程学院, 南昌 330013)

(* 通信作者电子邮箱 huangylqf@gmail.com)

摘 要: 针对传统支持向量机(SVM)分类存在对离群点敏感、支持向量(SV)个数多和分类面参数非稀疏的问题, 提出了平滑削边绝对偏离(SCAD)惩罚截断 Hinge 损失 SVM(SCAD-TSVM)算法, 并将其用于构建财务预警模型, 同时就该模型的求解设计了一个迭代更新算法。结合沪深股市 A 股制造业上市公司的财务数据进行实证分析, 同时对比 L1 范数惩罚 SVM、SCAD 惩罚 SVM 和截断 Hinge 损失 SVM(TSVM)构建的 $T-2$ 和 $T-3$ 模型, 结果发现 SCAD-TSVM 构建的 $T-2$ 和 $T-3$ 模型都具有最好的稀疏性和最高的预报精度, 而且其不同训练样本数上的平均预测准确率都要比 L1 范数 SVM(L1-SVM)、SCAD-SVM 和 TSVM 算法的高。

关键词: 支持向量机; 平滑削边绝对偏离惩罚; 截断 Hinge 损失支持向量机; 财务预警; L1 范数惩罚

中图分类号: TP181 **文献标志码:** A

Financial failure prediction using truncated Hinge loss support vector machine with smoothly clipped absolute deviation penalty

LIU Zunxiong, HUANG Zhiqiang*, LIU Jiangwei, CHEN Ying

(School of Information Engineering, East China Jiaotong University, Nanchang Jiangxi 330013, China)

Abstract: Aiming at the problems that the traditional Support Vector Machine (SVM) classifier is sensitive to outliers and has the large number of Support Vectors (SV) and the parameter of its separating hyperplane is not sparse, the Truncated hinge loss SVM with Smoothly Clipped Absolute Deviation (SCAD) penalty (SCAD-TSVM) was put forward and was used for constructing the financial early-warning model. At the same time, an iterative updating algorithm was proposed to solve the SCAD-TSVM model. Experiments were implemented on the financial data of A-share manufacturing listed companies of the Shanghai and Shenzhen stock markets. Compared to the $T-2$ and $T-3$ models constructed by SVM with L1 norm penalty (L1-SVM), SVM with SCAD penalty (SCAD-SVM) and Truncated hinge loss SVM (TSVM), the $T-2$ and $T-3$ model constructed by the SCAD-TSVM had the best sparseness and the highest accuracy of prediction, and its average accuracies of prediction with different number of training samples were higher than those of the L1-SVM, SCAD-SVM and TSVM algorithms.

Key words: Support Vector Machine (SVM); Smoothly Clipped Absolute Deviation (SCAD) penalty; Truncated Hinge loss SVM (TSVM); financial early-warning; L1 norm penalty

0 引言

为了对上市公司未来的财务状况进行有效预测, 许多财务危机建模方法被相继提出, 其通常可分为传统统计类方法和人工智能方法。传统统计类方法主要包括单变量分析 (Univariate Analysis)^[1]、多元判别分析 (Multiple Discriminant Analysis, MDA)^[2]、逻辑回归 (Logistic Regression, LR)^[3] 等, 这些方法在财务预警中应用广泛且较为成功, 但仍旧存在一些不足之处, 如: 单变量分析不能全面地反映公司营运状况, MDA 要求假定指标变量满足正态分布或者协方差矩阵相等, LR 存在多重共线性以及分界点的确定缺乏足够理论依据等问题。随着对建模方法的不断研究, 人工智能方法被引入且广泛应用于危机建模, 其中就包括人工神经网络 (Artificial Neural Network, ANN)^[4]、支持向量机 (Support Vector

Machine, SVM)^[5]、组合预测^[6]等。张红梅等^[6]构建了一种组合预测模型, 同时实证发现该模型要优于 LR 和反向传播 (Back-Propagation, BP) 神经网络模型; Min 等^[7]将 SVM 用于财务危机预测, 并得出了 SVM 方法要优于 MDA、LR 和 BP 神经网络的结论; 王宝富等^[8]进行了危机预测方法的比较研究, 结果发现 SVM 和神经网络的误判率要比 LR 和判别分析的低, 并且 SVM 的总误判率最低。

自 20 世纪 90 年代被提出, SVM 的研究受到广泛关注和拓展, 研究发现: 1) SVM 训练对样本的离群点敏感, 将产生过拟合问题; 2) 处理一些问题 (如较难分类或者指标变量较多的问题) 时, 模型的支持向量个数较多, 将导致模型构建需要更长的计算时间; 3) 传统 SVM 从样本选择角度构建稀疏模型, 但是其无法实现变量选择, 模型的经济解释性不够强。因此, 许多学者提出了各种改进的 SVM。Zhang 等^[9]将平滑削

收稿日期: 2013-07-25; **修回日期:** 2013-09-26。 **基金项目:** 国家自然科学基金资助项目 (61065003, 71361009); 教育部人文社会科学规划基金资助项目 (13YJC630192); 华东交通大学 2013 年度研究生创新专项资金资助项目 (YC2013-S172)。

作者简介: 刘遵雄 (1967-), 男, 江西瑞昌人, 教授, 博士, 主要研究方向: 机器学习、数据挖掘; 黄志强 (1989-), 男, 江西抚州人, 硕士研究生, 主要研究方向: 机器学习、优化算法; 刘江伟 (1983-), 男, 河南许昌人, 硕士研究生, 主要研究方向: 数据挖掘; 陈英 (1991-), 男, 江西南昌人, 硕士研究生, 主要研究方向: 聚类算法。

绝对偏离 (Smoothly Clipped Absolute Deviation, SCAD) 惩罚^[10]引入到 SVM 中,提出了 SCAD-SVM 算法并用于同时实现基因选择和癌症分类,结果发现其在分类精度等方面要优于传统 SVM、L1 范数 SVM (L1-SVM) 等算法。Wu 等^[11]根据截断 Hinge 损失的思想提出了稳健截断 Hinge 损失支持向量机 (Robust Truncated Hinge Loss SVM, RSVM), 实验表明 RSVM 要优于传统 SVM。受这两者的启发,为构建稀疏且稳健的预警模型,本文提出了 SCAD 惩罚截断 Hinge 损失 SVM (Truncated Hinge Loss SVM with SCAD Penalty, SCAD-TSVM) 并将其用于财务预警实证研究。由于 SCAD-TSVM 的优化问题涉及非凸损失和非凸惩罚,相比 SVM 和 RSVM 更难求解,故本文借鉴文献[9]的思路设计了一个迭代更新算法。为了检验 SCAD-TSVM 模型的预警效果,利用沪深股市 A 股制造业上市公司 2010 年至 2012 年首次因财务状况异常而被实施特别处理 (Special Treatment, ST) 的 39 家公司和配对的 39 家正常公司的财务指标数据进行实验,同时对比 L1-SVM 模型、SCAD-SVM 模型和截断 Hinge 损失 SVM (Truncated Hinge Loss SVM, TSVM) 模型的预警效果进行分析。

1 模型构建

财务危机预警可归结为模式识别中的二元分类问题:一类是因财务状况异常而被实施 ST 的公司 (ST 公司, 标记为 -1), 另一类是财务状况正常的公司 (正常公司, 标记为 +1)。给定 ST 公司和正常公司共 n 家且每家公司具有 d 个财务指标变量, 如此构成训练样本集 $\{(x_i, y_i)\}_{i=1}^n \in \mathbf{R}^d \times \{-1, +1\}$, 其中 x_i 表示第 i 家样本公司的财务指标数据以及相应的 y_i 表示第 i 家样本公司的标记值。根据损失 + 惩罚的准则, 正则化机器学习算法的模型框架如下:

$$\min l(y, f(x)) + p(\lambda, w) \quad (1)$$

其中前者为损失函数, 后者为惩罚函数, $y \in \mathbf{R}^n, x \in \mathbf{R}^{n \times d}$, $\lambda > 0$ 是正则化参数且 $w \in \mathbf{R}^d$ 为 $f(x)$ 模型系数。

1.1 L1-SVM 模型和 SCAD-SVM 模型

传统 SVM 由 Hinge 损失函数与 2 范数惩罚函数构成, 即对应的优化问题为:

$$\min_{w, b} \sum_{i=1}^n [1 - y_i(w^T x_i + b)]_+ + \lambda \|w\|_2^2 \quad (2)$$

其中: $w \in \mathbf{R}^d$ 和 $b \in \mathbf{R}$ 决定了 SVM 最优分类面, $[u]_+ = \max(u, 0)$ 。将式(2)中 2 范数惩罚函数分别换成 1 范数惩罚函数和 SCAD 惩罚函数, 可以得到两个新的预警模型, 即 L1-SVM 模型和 SCAD-SVM 模型。

对于 L1-SVM 模型, 其相应的最小优化问题为:

$$\min_{w, b} \sum_{i=1}^n [1 - y_i(w^T x_i + b)]_+ + \lambda \|w\|_1 \quad (3)$$

针对上述优化问题的求解, 已有许多较好的算法, 如: 分段线性解算法^[12]、分解算法^[13]等。而对于 SCAD-SVM 模型, 其相应的优化问题为:

$$\min_{w, b} \sum_{i=1}^n [1 - y_i(w^T x_i + b)]_+ + \sum_{j=1}^d P_\lambda(|w_j|) \quad (4)$$

其中:

$$P_\lambda(|w_j|) =$$

$$\begin{cases} \lambda |w_j|, & |w_j| \leq \lambda \\ -\frac{(|w_j|^2 - 2a\lambda |w_j| + \lambda^2)}{2(a-1)}, & \lambda < |w_j| \leq a\lambda, \text{ 且 } a > 2 \\ \frac{(a+1)\lambda^2}{2}, & |w_j| > a\lambda \end{cases}$$

由于 SCAD 惩罚函数是非凸的, 故求解 SCAD-SVM 模型比求解 SVM 模型或者 L1-SVM 模型都要困难。针对式(4)的求解, 已有迭代算法^[9]、差分凸 (Difference Convex, DC) 算法^[14]等。

1.2 SCAD-TSVM 模型

Wu 等^[11]将截断 Hinge 损失思想引入 SVM 中, 提出了可用于多元分类的 RSVM。对 RSVM 进行特例化, 可得出针对二元分类问题的 TSVM 模型, 其对应的正则化问题为:

$$\min_{w, b} \sum_{i=1}^n [1 - y_i(w^T x_i + b)]_+ - \sum_{i=1}^n [s - y_i(w^T x_i + b)]_+ + \lambda \|w\|_2^2 \quad (5)$$

其中 $s \leq 0$ 为截断参数且决定了截断位置。若 $s = -\infty$, 则 TSVM 模型就变成了传统 SVM 模型。对于 TSVM 模型的求解, Wu 等^[11]给出一般的求解算法, 即 DC 算法。若将式(5)中的 2 范数惩罚函数换成 SCAD 惩罚函数, 就得到本文所提出的 SCAD-TSVM 模型, 即相应的最小化问题为:

$$\min_{w, b} \sum_{i=1}^n [1 - y_i(w^T x_i + b)]_+ - \sum_{i=1}^n [s - y_i(w^T x_i + b)]_+ + \sum_{j=1}^d P_\lambda(|w_j|) \quad (6)$$

其中 $P_\lambda(|w_j|)$ 同式(4)中的一样。

SCAD-TSVM 模型中的 SCAD 惩罚项能同时满足无偏性、稀疏性和连续性这三个条件, 即 SCAD 惩罚项具有 Oracle 性质。而 L1-SVM 模型中的 L1 惩罚项不满足无偏性, 还有传统 SVM 模型中的 L2 惩罚项不满足稀疏性, 故它们都不具有 Oracle 性质。可见, SCAD-TSVM 模型要优于 L1-SVM 模型和传统 SVM 模型。不仅如此, 令 $t = y_i(w^T x_i + b)$, 有 Hinge 损失 $[1 - t]_+$, 当 $t < 1$ (满足该条件, 即为离群点) 且 t 不断减小时, 该损失值将不断增大。所以离群点将严重影响 SVM 的分类结果, 使 SVM 的分类效果变差。但如果采用截断 Hinge 损失 $[1 - t]_+ - [s - t]_+ (s \leq 0)$, 当 $t < 1$ 且 t 不断减小时, 损失值将先不断增大, 然后当满足 $t = s$ 以后, 损失值将不会继续增大且一直保持在 $1 - s$ 。这样便可很好地消除离群点对于分类的影响。可见, 不管是从惩罚函数还是从损失函数来讲, SCAD-TSVM 模型都明显优于传统 SVM 模型。从上述分析中, 不难看出 SCAD-TSVM 算法比其他算法更适合于模型需要变量选择功能、样本特征变量较多、样本中存在离群点等问题。我国上市公司的指标变量相对较多, 因此在构建财务预警模型时, 就希望模型能够选择出一些对预警有较大作用的指标。同时, 部分公司还存在会计舞弊等修改真实数据的可能性, 这样将给危机预报研究带来困难。所以, SCAD-TSVM 算法还是比较适合于财务危机预警研究。针对 SCAD-TSVM 模型的求解, 本文借鉴文献[9]的思路设计了一个迭代更新算法。

2 求解算法

SCAD-TSVM 模型包含非凸损失和非凸惩罚,所以一般的二次规划或者线性规划方法难以求解式(6)的优化问题。本文参照文献[9]的策略,设计了一个迭代更新算法。对于 SCAD-TSVM 优化问题的求解,该算法只需计算一些线性等式即可。

对于每一家样本公司($y_i^2 = 1$),式(6)的前两项可通过下面两个等式进行替换:

$$\begin{cases} [1 - y_i(\mathbf{w}^T \mathbf{x}_i + b)]_+ = \frac{1 - y_i(\mathbf{w}^T \mathbf{x}_i + b)}{2} + \frac{|y_i - (\mathbf{w}^T \mathbf{x}_i + b)|}{2} \\ [s - y_i(\mathbf{w}^T \mathbf{x}_i + b)]_+ = \frac{s - y_i(\mathbf{w}^T \mathbf{x}_i + b)}{2} + \frac{|sy_i - (\mathbf{w}^T \mathbf{x}_i + b)|}{2} \end{cases} \quad (7)$$

假定初始值 (b_0, \mathbf{w}_0) , 对式(7)中的两个等式的后两项进行局部二次近似(Local Quadratic Approximation, LQA), 可以得出:

$$\begin{cases} |y_i - (\mathbf{w}^T \mathbf{x}_i + b)| \approx \frac{1}{2} \frac{[y_i - (\mathbf{w}^T \mathbf{x}_i + b)]^2}{|y_i - (\mathbf{w}_0^T \mathbf{x}_i + b_0)|} + \frac{1}{2} |y_i - (\mathbf{w}_0^T \mathbf{x}_i + b_0)| \\ |sy_i - (\mathbf{w}^T \mathbf{x}_i + b)| \approx \frac{1}{2} \frac{[sy_i - (\mathbf{w}^T \mathbf{x}_i + b)]^2}{|sy_i - (\mathbf{w}_0^T \mathbf{x}_i + b_0)|} + \frac{1}{2} |sy_i - (\mathbf{w}_0^T \mathbf{x}_i + b_0)| \end{cases} \quad (8)$$

同样利用一个二次函数对 SCAD 惩罚函数进行近似,有:

$$P_\lambda(|w_j|) \approx P_\lambda(|w_0|) + \frac{P'_\lambda(|w_0|)}{2|w_0|}(w_j^2 - w_0^2) \quad (9)$$

令式(6)的目标函数为 $L(b, \mathbf{w})$, 将式(8)和式(9)都代入该目标函数中,然后化简并去掉其中不包含 (b, \mathbf{w}) 的项,最终得到同式(6)等价的优化问题:

$$\begin{aligned} \min \hat{L}(b, \mathbf{w}) = & \frac{1}{4} \sum_{i=1}^n \frac{(\mathbf{w}^T \mathbf{x}_i + b)^2}{|y_i - (\mathbf{w}_0^T \mathbf{x}_i + b_0)|} - \\ & \frac{1}{2} \sum_{i=1}^n \frac{y_i(\mathbf{w}^T \mathbf{x}_i + b)}{|y_i - (\mathbf{w}_0^T \mathbf{x}_i + b_0)|} - \\ & \frac{1}{4} \sum_{i=1}^n \frac{(\mathbf{w}^T \mathbf{x}_i + b)^2}{|sy_i - (\mathbf{w}_0^T \mathbf{x}_i + b_0)|} + \\ & \frac{1}{2} \sum_{i=1}^n \frac{sy_i(\mathbf{w}^T \mathbf{x}_i + b)}{|sy_i - (\mathbf{w}_0^T \mathbf{x}_i + b_0)|} + \sum_{j=1}^d \frac{P'_\lambda(|w_0|)}{2|w_0|} w_j^2 \end{aligned} \quad (10)$$

在样本公司的财务数据 \mathbf{x} 的第一列之前增加一列全 1 向量得到 $\mathbf{X} \in \mathbf{R}^{n \times (d+1)}$, 同时令 $\varepsilon_i = y_i - (\mathbf{w}_0^T \mathbf{x}_i + b_0)$, $\eta_i = sy_i - (\mathbf{w}_0^T \mathbf{x}_i + b_0)$, $\mathbf{D}_1 = \text{diag}\left[0, \frac{P'_\lambda(|w_0|)}{2|w_0|}, \dots, \frac{P'_\lambda(|w_0|)}{2|w_0|}\right]$, $\mathbf{D}_2 = \text{diag}[1/|\varepsilon_1|, 1/|\varepsilon_2|, \dots, 1/|\varepsilon_n|]$ 和 $\mathbf{D}_3 = \text{diag}[1/|\eta_1|, 1/|\eta_2|, \dots, 1/|\eta_n|]$, 可将式(10)用向量形式表示为:

$$\min \hat{L}(b, \mathbf{w}) = \frac{1}{4} \begin{pmatrix} b \\ \mathbf{w} \end{pmatrix}^T \mathbf{A} \begin{pmatrix} b \\ \mathbf{w} \end{pmatrix} - \mathbf{B} \begin{pmatrix} b \\ \mathbf{w} \end{pmatrix} \quad (11)$$

其中: $\mathbf{A} = 4\mathbf{D}_1 + \mathbf{X}^T \mathbf{D}_2 \mathbf{X} - \mathbf{X}^T \mathbf{D}_3 \mathbf{X}$, $\mathbf{B} = \frac{1}{2} \mathbf{y}^T (\mathbf{D}_2 - s\mathbf{D}_3) \mathbf{X}$ 。式

(11)是关于 $\begin{pmatrix} b \\ \mathbf{w} \end{pmatrix}$ 的二次形式,所以求解式(11)就等于求解下面的线性等式:

$$\begin{pmatrix} b \\ \mathbf{w} \end{pmatrix}^T \cdot \mathbf{A} = 2\mathbf{B} \quad (12)$$

根据上述分析,可以得到 SCAD-TSVM 模型的求解算法。

给定 收敛条件 $\mu > 0$, 正则化参数 $\lambda > 0$, 截断参数 $s \leq 0$, SCAD 惩罚参数 $a > 2$, 初始化 $k = 1$ 和循环差异初值 $dv = 1000$ 。

1) 利用传统 SVM 计算出初始值 $\begin{pmatrix} b^{(1)} \\ \mathbf{w}^{(1)} \end{pmatrix}$;

while $dv > \mu$

2) 赋初始值 $\begin{pmatrix} b_0 \\ \mathbf{w}_0 \end{pmatrix} = \begin{pmatrix} b^{(k)} \\ \mathbf{w}^{(k)} \end{pmatrix}$, 然后根据式(12)计算出

$\begin{pmatrix} b^{(k+1)} \\ \mathbf{w}^{(k+1)} \end{pmatrix}$, 接着更新 $dv = \left\| \begin{pmatrix} b^{(k+1)} \\ \mathbf{w}^{(k+1)} \end{pmatrix} - \begin{pmatrix} b^{(k)} \\ \mathbf{w}^{(k)} \end{pmatrix} \right\|_2$;

3) 更新 $k = k + 1$;

end

最终,获得预警模型的指标系数向量 $\mathbf{w}^{(k)}$ 和偏移特征量 $b^{(k)}$ 。

3 实证分析

3.1 样本与指标选取

本文以沪深股市 A 股制造业上市公司作为研究对象,同时按照上市时间不少于 5 年、财务指标数据完整、因财务状况异常而被实施 ST 等要求选取 ST 公司,最终从 2010 年至 2012 年间选出了 39 家 ST 公司,其中 2010 年 20 家、2011 年 6 家和 2012 年 13 家。对于配对的正常公司,按照上市时间不少于 5 年、规模相差不超过 10%、相同时间段、从未被实施 ST 等要求以 1:1 的比例选取出 39 家正常公司,其中各年的具体数目同 ST 公司的一样。从 RESSET 金融研究数据库 (<http://www2.resset.cn/product/>) 取得这些公司的财务指标数据作为预警研究的实验数据。

公司财务指标变量的选取对预警模型的构建起到至关重要的作用,好的指标有利于建立合理有效的预警模型,反之则不然。根据 271 篇国内有关财务预警的文献,统计出每个财务指标出现的总频次,然后选出总频次高的一些财务指标,最终选择了反映公司每股情况、盈利能力、偿债能力、发展能力、营运能力和现金流量的 29 个财务指标变量(见表 1)。

依据我国上市公司的年报披露制度可知,利用公司 $T-1$ 年的财务数据来预测公司在 T 年是否会被实施 ST 没有很大的实际应用价值,所以本文采用样本公司 $T-2$ 年和 $T-3$ 年的财务指标数据(如 2010 年的样本公司的 2008 年和 2007 年的财务数据)来分别建立危机预警模型,简记为 $T-2$ 模型和 $T-3$ 模型。

表 1 实证分析所采用的财务指标

| 类型 | 指标名称 |
|--------|---|
| 每股情况指标 | 每股收益、每股净资产、每股经营活动现金流量 |
| 盈利能力指标 | 净资产收益率、资产留存收益率、资产报酬率、资产收益率、销售净利率、销售毛利率、成本费用利润率 |
| 偿债能力指标 | 流动比率、速动比率、产权比率、资产负债率、股东权益比率、营运资本总资产比、经营性现金流量债务总额比、经营性现金流量流动负债比、利息保障倍数 |
| 发展能力指标 | 净利润增长率、净资产增长率、总资产增长率 |
| 营运能力指标 | 存货周转率、应收账款周转率、流动资产周转率、固定资产周转率、总资产周转率 |
| 现金流量指标 | 销售现金比率、总资产现金回收率 |

3.2 实验结果及分析

将 L1-SVM 算法、SCAD-SVM 算法、TSVM 算法和 SCAD-TSVM 算法结合选定公司的财务数据进行实证对比研究。本文进行了三类对比实验：第一，将前两年公司的数据作为训练样本构建上述 4 种算法的模型，然后对比模型的稀疏性、样本指标的相关性以及剩余样本数据在它们上的预警效果；第二，比较 4 种算法在各种训练样本数目下进行 10 次随机实验的平均分类效果；第三，对比分析 4 种算法在分别重复 10 次、20 次、30 次和 40 次实验的预报精度的均值和标准差。L1-SVM 算法由 Zheng 等^[15]编写的一个综合稀疏学习包来实现；SCAD-SVM 算法由 Zhang 等^[9]提供的 Matlab 代码来实现；TSVM 算法由本文的 SCAD-TSVM 算法的实现在替换 $D_1 = \text{diag}[0, \lambda, \dots, \lambda]$ 的情况下具体实现。至于，传统 SVM 将由 Lin 等^[16]编写的 LIBSVM 工具箱来实现，还有实验所涉及到的 SVM 都采用线性核^[5]。

在进行实验之前，先要确定各算法的参数取值。SCAD-TSVM 算法涉及到的参数有正则化参数 λ 、截断参数 s 和 SCAD 惩罚参数 α 。对于参数 s 和 α ，实验分别取 -1 ^[11]和 3.7 ^[10]，并且其他算法也同样采用这两个取值。而对于剩下的参数 λ ，本文采用 10 折交叉验证（10-fold Cross Validation）^[17]来确定其最优取值，最终确定参数 $\lambda = 1$ （ $T-2$ 模型）和 $\lambda = 2$ （ $T-3$ 模型）。考虑到算法对比的有效性问 题，同样利用 10 折交叉验证来确定其他三个算法的 λ 取值，最后确定的参数值如下：L1-SVM 取 $\lambda = 4$ 和 $\lambda = 6$ ，SCAD-SVM 取 $\lambda = 0.5$ 和 $\lambda = 0.5$ ，TSVM 取 $\lambda = 1$ 和 $\lambda = 8$ ，其中各算法前面的 λ 值为 $T-2$ 模型的，后面的为 $T-3$ 模型的。

第一类实验是将 2010 年和 2011 年样本公司的财务数据作为训练样本，分别用于建立 L1-SVM、SCAD-SVM、TSVM 和 SCAD-TSVM 算法的 $T-2$ 模型和 $T-3$ 模型，主要为了进行模

型稀疏性、指标相关性以及模型预测分类效果的对比分析。上述算法建立的各种模型对应的系数值如图 1 所示，并可得训练样本经由各算法变量选择后的相关系数情况（见表 2）。建模后，将 2012 年样本公司的财务数据分别代入上述各模型，可以得到各算法在测试样本上的预报精度和第 I、II 类错误（其中第 I 类错误是指将 ST 公司判别为正常公司，第 II 类错误是指将正常公司判别为 ST 公司）（见表 3）。

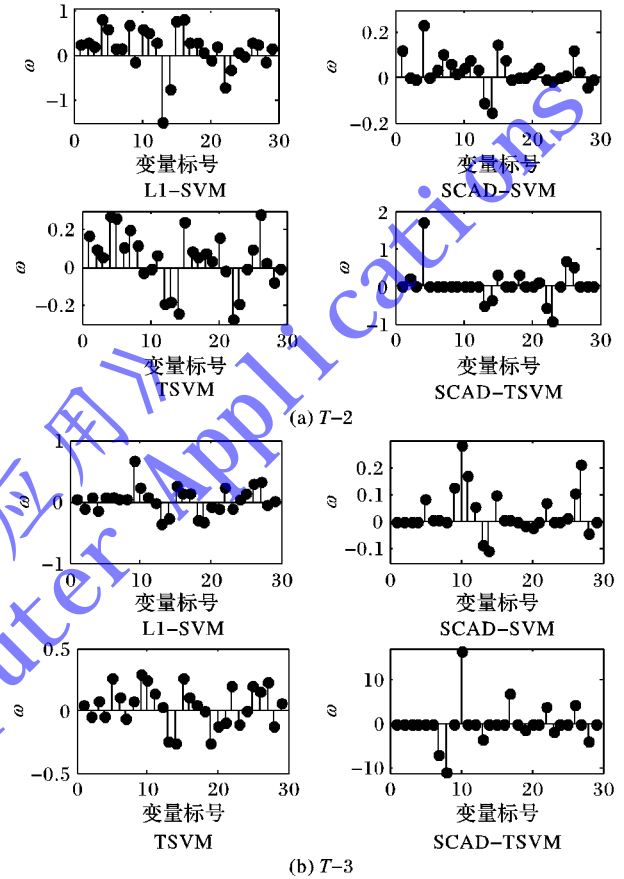


图 1 4 种算法的 $T-2$ 和 $T-3$ 模型所对应的系数值

表 2 训练样本变量选择后的相关系数情况
(相关系数值不小于 0.5 的比例) %

| 算法 | $T-2$ 模型 | $T-3$ 模型 |
|--------------|----------|----------|
| L1-SVM 算法 | 19.21 | 10.34 |
| SCAD-SVM 算法 | 19.21 | 6.49 |
| TSVM 算法 | 17.95 | 10.83 |
| SCAD-TSVM 算法 | 10.91 | 8.89 |

表 3 4 种算法在测试样本上的预测分类结果 %

| 算法 | $T-2$ 模型 | | | $T-3$ 模型 | | |
|-----------|----------|----------|-------|----------|----------|-------|
| | 第 I 类错误 | 第 II 类错误 | 准确率 | 第 I 类错误 | 第 II 类错误 | 准确率 |
| L1-SVM | 15.38 | 0 | 92.31 | 38.46 | 38.46 | 61.54 |
| SCAD-SVM | 0 | 15.38 | 92.31 | 30.77 | 38.46 | 65.38 |
| TSVM | 15.38 | 0 | 92.31 | 30.77 | 38.46 | 65.38 |
| SCAD-TSVM | 7.69 | 0 | 96.15 | 30.77 | 23.08 | 73.08 |

从图 1 可以发现：无论 L1-SVM、SCAD-SVM、TSVM 和 SCAD-TSVM 算法是在构建 $T-2$ 模型还是 $T-3$ 模型，SCAD-TSVM 算法的模型稀疏性最好（即实现了变量选择）且明显优

于其他 3 种算法的，而且 SCAD-SVM 算法的次之，L1-SVM 算法和 TSVM 算法的最差。因此，SCAD-TSVM 算法的模型能很好地同时实现变量选择和系数估计，而且 SCAD 惩罚函数具

有一定的优越性且更适合作为正则化项。还有,从表2可知:对于 $T-2$ 模型,训练样本经由 SCAD-TSVM 算法变量选择后的相关系数值不小于0.5的比例是这4种算法中最小的且明显优于其他算法,TSVM 算法的比例次小,但是 SCAD-TSVM 算法的比例要比 TSVM 算法的少7个多百分点;对于 $T-3$ 模型,SCAD-SVM 算法的比例最小,而 SCAD-TSVM 算法的比例仅比 SCAD-SVM 算法的大不到3%。可见,SCAD-TSVM 算法对于各种时间段的指标数据都具有较强的适应性。

不仅如此,从表3可以看出:无论是 $T-2$ 模型还是 $T-3$ 模型,SCAD-TSVM 算法都具有比其他算法更高的预报准确率且至少高出大约4个百分点。还有利用年份越近的财务数据建立预警模型,模型的预报精度越高,而且在 $T-2$ 模型中 SCAD-TSVM 算法的预报精度就高达96.15%。对于 $T-2$ 模型,SCAD-TSVM 算法的第I类错误虽比 SCAD-SVM 算法的高,但是其第II类错误为0%,远比 SCAD-SVM 算法的低,而且 SCAD-TSVM 算法的第I、II类错误不会比 L1-SVM 算法和 TSVM 算法的高;对于 $T-3$ 模型,SCAD-TSVM 算法的第I、II类错误不会高于其他3种算法的,而且其预报准确率明显优于其他算法的。结合 $T-2$ 模型和 $T-3$ 模型的结果可知,SCAD-SVM 算法的分类效果仅次于 SCAD-TSVM 算法的且不会比其他两种算法的分类效果差,可见 SCAD 惩罚函数有助于提高机器学习算法的预测分类效果。

第二类实验主要是为了比较 L1-SVM、SCAD-SVM、TSVM 和 SCAD-TSVM 算法在不同训练样本数目下的预测分类效果。首先从2010年和2011年公司的样本数据中依次随机抽取20,22,26,32,40和50个训练样本(ST公司数目与正常公司数目之比为1:1)建立4种算法的 $T-2$ 模型和 $T-3$ 模型,然后将2012年公司的财务数据用于构建好的模型并获得预报精度。对于每个训练样本数都分别重复进行10次实验,最终获得每个算法在各训练样本数目下平均预报准确率(见图2)。

从图2可以看出,无论是 $T-2$ 模型还是 $T-3$ 模型,SCAD-TSVM 算法的平均预测准确率都要高于其他3种算法的。而且随着样本数目增大,其平均预测准确率都大致呈现优于其他算法更加明显,同时其在训练样本数目较小时就具有相对较高的预报精度。对于 $T-2$ 模型,4种算法的平均预测准确率都大致呈现出递增的趋势;对于 $T-3$ 模型,SCAD-

TSVM 算法的平均预测准确率在样本数目40之前大致呈现递增的趋势,而且整体下降幅度比其他算法的都小。可见,SCAD-TSVM 算法具有更高的预报精度和更强的适应性。

第三类实验是为了检验 L1-SVM、SCAD-SVM、TSVM 和 SCAD-TSVM 算法在进行多次随机重复实验的稳定性。首先将三年的财务数据合在一起,然后按照1:1的比例随机分配训练样本和测试样本并用于模型构建和测试,分别如此重复进行10,20,30和40次实验,最后得到4种算法对于两个数据集(即 $T-2$ 年和 $T-3$ 年的财务数据)在测试样本上的预报精度的均值和标准差(见表4)。从表4可以看出:无论是 $T-2$ 模型还是 $T-3$ 模型,SCAD-TSVM 算法在测试样本上的预报准确率的均值都要比其他3种算法的高,而其标准差还比其他算法的都小。同时在多数情况下,SCAD-SVM 算法的平均预测准确率仅次于最优的 SCAD-TSVM 算法的。可见,SCAD-TSVM 算法在预报精度、稳定性等方面要优于其他3种算法,并且 SCAD 惩罚项有助于改善算法的分类效果。

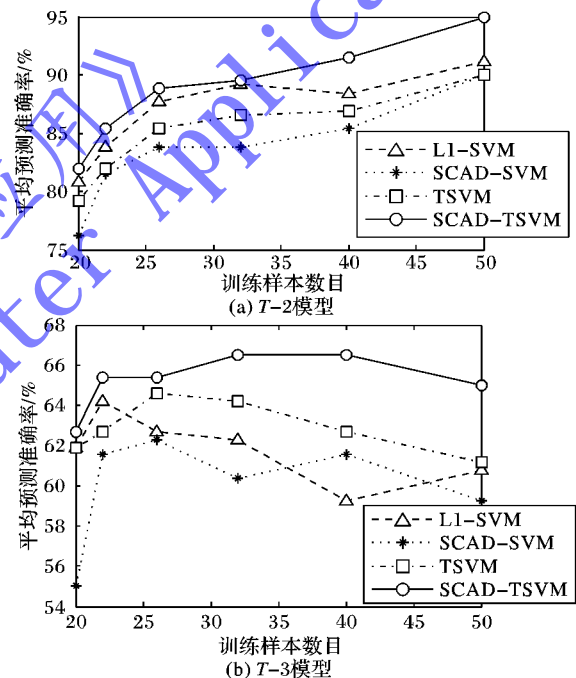


图2 4种算法在不同训练样本数目下的平均预测分类效果对比

表4 4种算法在不同次数下预报精度的均值和标准差对比

| 次数 | $T-2$ 模型 | | | | $T-3$ 模型 | | | |
|----|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| | L1-SVM | SCAD-SVM | TSVM | SCAD-TSVM | L1-SVM | SCAD-SVM | TSVM | SCAD-TSVM |
| 10 | 82.37 ± 4.81 | 84.74 ± 4.26 | 84.21 ± 5.82 | 85.26 ± 4.15 | 61.58 ± 6.93 | 60.26 ± 7.99 | 61.05 ± 5.38 | 63.16 ± 5.11 |
| 20 | 82.76 ± 5.16 | 82.63 ± 5.69 | 83.29 ± 5.49 | 85.13 ± 4.77 | 60.00 ± 5.51 | 63.95 ± 5.61 | 63.95 ± 5.54 | 64.21 ± 5.29 |
| 30 | 82.37 ± 5.13 | 83.77 ± 5.49 | 83.33 ± 5.23 | 84.12 ± 4.34 | 63.95 ± 5.99 | 60.88 ± 7.34 | 63.60 ± 6.03 | 64.12 ± 5.37 |
| 40 | 81.58 ± 5.20 | 83.62 ± 5.45 | 82.76 ± 5.39 | 83.82 ± 4.68 | 63.09 ± 6.43 | 63.62 ± 6.84 | 60.66 ± 6.74 | 65.20 ± 6.24 |
| 平均 | 82.27 | 83.69 | 83.40 | 84.58 | 62.16 | 62.18 | 62.32 | 64.17 |

4 结语

本文提出了 SCAD 惩罚截断 Hinge 损失 SVM (SCAD-TSVM) 的财务危机预警模型。实证结果表明:相比 L1-SVM、SCAD-SVM 和 TSVM 算法建立的模型,SCAD-TSVM 算法构建的模型稀疏性更强,变量选择后训练样本指标相关性较低,并且对于新的测试样本具有更高的预报精度。还有,在各种训练样本数目下,SCAD-TSVM 算法的平均预测准确率都要高于

其他3种算法,而且其在样本数较小时就有较好的预测分类效果。在未来的进一步研究中,将考虑非线性的 SCAD-TSVM 以及引入公司治理、市场信息等非财务指标变量到模型中,以期构建在预报精度、鲁棒性等方面更好的预警模型。

参考文献:

- [1] LI F, DU Z, LI L. The model of predicting finance distress of enterprise: theory evolution and its review [J]. Management Review, 2011, 23(9): 144-151. (李帆, 杜志涛, 李玲娟. 企业财务预警

- 模型: 理论回顾及其评论[J]. 管理评论, 2011, 23(9): 144 - 151.)
- [2] ZHANG L, WANG Y, LIU J. Forewarning model of listed companies against financial crisis based on Bayesian discrimination [J]. *Commercial Research*, 2009(4): 112 - 114. (张立军, 王瑛, 刘菊红. 基于贝叶斯判别分析的上市公司财务危机预警模型研究[J]. 商业研究, 2009(4): 112 - 114.)
- [3] ZHU Y, YAO Y, ZHANG Y. Research on financial crisis pre-warning model of listed company based on PCA and Logistic regression [J]. *Journal of Zhejiang University of Technology*, 2012, 40(6): 692 - 694, 689. (朱永忠, 姚焯, 张艳. 基于主成分分析和 Logistic 回归的上市公司财务困境预警模型的研究[J]. 浙江工业大学学报, 2012, 40(6): 692 - 694, 698.)
- [4] YANG S, WANG L. Research on financial warning for listed companies by using BP neural networks and panel data [J]. *Systems Engineering—Theory & Practice*, 2007, 27(2): 61 - 67. (杨淑娥, 王乐平. 基于 BP 神经网络和面板数据的上市公司财务危机预警[J]. 系统工程理论与实践, 2007, 27(2): 61 - 67.)
- [5] LIU Z, HUANG Z, YAN F, *et al.* Financial failure prediction using support vector machine with Q -Gaussian kernel [J]. *Journal of Computer Applications*, 2013, 33(6): 1767 - 1770. (刘遵雄, 黄志强, 晏峰. 等. Q -高斯核支持向量机的财务危机预报[J]. 计算机应用, 2013, 33(6): 1767 - 1770.)
- [6] ZHANG H, LIU W. The application of combining forecasts in the early-warning of listed company's financial risk [J]. *Friends of Accounting*, 2011(2): 101 - 102. (张红梅, 刘文蕊. 组合预测在上市公司财务预警中的应用[J]. 会计之友, 2011(2): 101 - 102.)
- [7] MIN J H, LEE Y. Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters [J]. *Expert Systems with Applications*, 2005, 28(4): 603 - 614.
- [8] WANG B, LI N. Research of predicting financial distress [J]. *Journal of Nanjing University of Aeronautics and Astronautics: Social Sciences*, 2007, 9(3): 61 - 64. (王宝富, 李南. 财务困境的预测研究[J]. 南京航空航天大学学报: 社会科学版, 2007, 9(3): 61 - 64.)
- [9] ZHANG H H, AHN J, LIN X, *et al.* Gene selection using support vector machines with non-convex penalty [J]. *Bioinformatics*, 2006, 22(1): 88 - 95.
- [10] FAN J, LI R. Variable selection via nonconcave penalized likelihood and its oracle properties [J]. *Journal of the American Statistical Association*, 2001, 96(456): 1348 - 1360.
- [11] WU Y, LIU Y. Robust truncated hinge loss support vector machines [J]. *Journal of the American Statistical Association*, 2007, 102(479): 974 - 983.
- [12] ZHU J, ROSSET S, HASTIE T, *et al.* 1-norm support vector machines [C]// *Advances in Neural Information Processing Systems*. Cambridge, MA: MIT Press, 2004: 49 - 56.
- [13] TORII Y, ABE S. Decomposition techniques for training linear programming support vector machines [J]. *Neurocomputing*, 2009, 72(4/5/6): 973 - 984.
- [14] LIU Y, SHEN X, DOSS H. Multicategory ψ -learning and support vector machine: computational tools [J]. *Journal of Computational and Graphical Statistics*, 2005, 14(1): 219 - 236.
- [15] ZHENG D, HUANG K. CSL: a comprehensive sparse learning package [EB/OL]. [2013-06-20]. <http://www.enm.bris.ac.uk/staff/xkh/CSL1.0.rar>.
- [16] CHANG C, LIN C. LIBSVM: a library for support vector machines [J]. *ACM Transactions on Intelligent Systems and Technology*, 2011, 2(3): 1 - 27.
- [17] DUDA R O, HART P E, STORK D G. *Pattern classification* [M]. 2nd ed. Hoboken, NJ: Wiley, 2000.

(上接第 872 页)

- [4] SHEHU A Y. Promoting financial inclusion for effective anti-money laundering and counter financing of terrorism (AML/CFT) [J]. *Crime, Law and Social Change*, 2012, 57(3): 305 - 323.
- [5] SHARMAN J C, CHAIKIN D. Corruption and anti-money-laundering systems: putting a luxury good to work [J]. *Governance*, 2009, 22(1): 27 - 45.
- [6] DOLAR B, SHUGHART W F, II. Enforcement of the USA patriot act's anti-money laundering provisions: have regulators followed a risk-based approach [J]. *Journal of Global Finance*, 2011, 22(1): 19 - 31.
- [7] VLCEK W. Global anti-money laundering standards and developing economies: the regulation of mobile money [J]. *Development Policy Review*, 2011, 29(4): 415 - 431.
- [8] DREEWSKI R, SEPIELAK J, FILIPKOWSKI W. System supporting money laundering detection [J]. *Digital Investigation*, 2012, 9(1): 8 - 21.
- [9] VANDANA P, LI J H, GAO P. A framework for preventing money laundering in banks [J]. *Information Management and Computer Security*, 2012, 20(3): 170 - 183.
- [10] PENG W B. Research on money laundering crime under electronic payment background [J]. *Journal of Computers*, 2011, 6(1): 147 - 154.
- [11] BAK J, JEDRZEJEK C, FALKOWSKI M. Application of an ontology-based and rule-based model to selected economic crimes: fraudulent disbursement and money laundering [C]// *Proceedings of the 2010 International Conference on Semantic Web Rules*. Berlin: Springer-Verlag, 2010: 210 - 224.
- [12] BYRNE J J. How IT auditing fights money laundering [J]. *Journal of Corporate Accounting and Finance*, 2011, 22(5): 63 - 67.
- [13] CAO D K, DO P. Applying data mining in money laundering detection for the Vietnamese banking industry [C]// *Proceedings of the 4th Asian Conference on Intelligent Information and Database Systems*, LNCS 7197. Berlin: Springer-Verlag, 2012: 207 - 216.
- [14] LE-KHAC N, MARKOS S, KECHADI M T. Towards a new data mining-based approach for anti-money laundering in an international investment bank [C]// *Proceedings of the First International ICST Conference on Digital Forensics and Cyber Crime*. Berlin: Springer-Verlag, 2010: 77 - 89.
- [15] OUYANG W. *Chinese anti-money laundering crime preliminary typology* [M]. Beijing: Law Press, 2007. (欧阳卫民. 中国反洗钱犯罪类型学初探[M]. 北京: 法律出版社, 2007.)
- [16] YU W, WANG J. Suspicious money laundering detection system based on eigenvector centrality measure of transaction network [J]. *Journal of Computer Applications*, 2009, 29(9): 2581 - 2585. (喻炜, 王建东. 基于交易网络特征向量中心度量的可疑洗钱识别系统[J]. 计算机应用, 2009, 29(9): 2581 - 2585.)