

# 基于模糊支持向量机的上市公司财务困境预测<sup>①</sup>

杨海军, 太 雷

(北京航空航天大学经济管理学院, 北京 100191)

**摘要:**支持向量机(SVM)已经成功地应用于财务困境预测问题的研究,且已证明优于多元线性判别分析(MDA)、逻辑回归(Logistic regression)和神经网络(NN)等方法。然而,传统 SVM 使用结构风险最小化的原则,这样可能导致错误分类的经验风险升高,特别是当样本点与最优超平面十分接近的时候,这种误分类的经验风险显著升高。另外,传统 SVM 还存在过拟合问题,所以对数据集中的外点或噪声十分敏感。因此,采用模糊支持向量机(FSVM)算法来改进上述不足。首先,建立一个适当的成员模型用于对整个数据集的模糊处理;然后通过外点侦察方法(ODM)来发现外点,其中 ODM 集成了模糊 C-均值算法(Fuzzy C-mean algorithm)和无监督神经网络中的自组织映射(SOM)。最后,为主体集和外点集中的样本点分配不同的权值。还将 FSVM 应用于上市公司财务困境预测的实证研究,实证结果表明 FSVM 与传统 SVM 相比,FSVM 能较好的解决经验风险升高和过度拟合问题,确实降低了外点的影响并提高了分类器的分类准确率。

**关键词:**公司财务困境;模糊支持向量机;预测

**中图分类号:**F830 **文献标识码:**A **文章编号:**1007-9807(2009)03-0102-09

## 0 引言

随着上市公司财务状况对金融机构的决策与收益影响越来越大,公司财务困境预测研究正成为一个热点问题,因此建立一个准确的财务困境预测模型成为解决该问题的关键。该问题的解决有利于上市公司及时预防和化解财务危机并提高危机预警管理的科学性;借贷者(银行)可以避免贷款的高风险;投资者可以获得财务风险的警示;政府监管机构可以更有效、更科学地进行市场监管,维护市场稳健运行。

当一个公司长期处于严重亏损状态或者公司资不抵债的时候,公司将发生财务危机。更广泛的定义为,引起公司财务危机还有其他的原因,如失败的经营管理、领导层的错误决策、市场运作困难等等。然而,公司财务危机的共同假设是公司财务

综合指标适当地反应了上述全部特征。一些分类技术建议使用财务比率数据来预测财务危机,例如,Beaver 最早提出公司财务预警分析模型,该模型利用单一的财务比率来预测公司的财务困境<sup>[1]</sup>。随着许多学者在该领域研究工作的不断深入,在研究方法上进行了完善和改进。其中,具有代表性的是 Beaver 和 Altman 提出的多元线性判别分析(MDA)模型<sup>[2,3]</sup>,该模型确定了资产营运资本率、资产留存收益率、资产报酬率、债务权益市价率和总资产周转率这 5 个变量作为判别变量,产生了一个综合判别公司财务状况恶化程度的概率值,即 Z-值。但是线性判别模型对预测变量有着严格的联合正态分布要求,或者要求协方差矩阵相等,然而大量实证结果表明多数财务比率并不满足这一假设要求。为克服这些局限性,

① 收稿日期:2007-10-22;修订日期:2009-03-16

基金项目:国家自然科学基金资助项目(70771002,70831001)。

作者简介:杨海军(1970-)男,天津人,博士,副教授。Email: navy@buaa.edu.cn

Martin 和 Ohlson 等提出了逻辑回归 (LR) 模型<sup>[4,5]</sup>, 与前述的模型相比逻辑回归模型很好的解决了上述问题, 因此具有更广泛的适用性, 然而逻辑回归模型仍存在自身的缺陷, 诸如线性、非线性的传统统计学的严格假设条件, 要求预测变量之间、函数形式关系之间相互独立, 然而在现实世界中这些函数关系以及预测变量都不是完全独立的, 而是彼此相关的. 为了开发一种更为精确更为通用预测方法来解决这个问题, 包括决策树、神经网络、模糊逻辑、遗传算法、支持向量机等方法的数据挖掘和机器学习技术被广泛引入到公司财务困境预测上来, 并取得了较好的结果. Odam 等在 1990 年运用人工神经网络模型进行财务困境预测的研究, 在 1998 年 Franco 和 Varetto 应用遗传算法在这方面进行了尝试<sup>[6]</sup>. 胡援成、田满文采用改进 BP 神经网络模型, 以求提高精度, 并借助多种模型比较来选择更好的预测方法<sup>[7]</sup>. Min 和 Lee 年采用 SVM 方法对上市公司财务困境问题进行预测, 并得出 SVM 方法优于神经网络、多元判别分析和 Logistic 回归的结果<sup>[8]</sup>. 汪寿阳等人在采用 SVM 方法预测外汇与国际原油价格波动方面做了大量研究, 均取得了较好的预测结果<sup>[9-11]</sup>. 诸克军等人提出了模糊 C-均值算法, 在给定预分类数的前提下实现最佳分类<sup>[12]</sup>. 程功等人研究了在噪音的信息环境下, 银行如何利用结构化模型来预测违约概率问题<sup>[13]</sup>. 华中生和王宇开发了一种集成二项判别规则的支持向量机模型, 并用于上市公司财务困境预测研究, 实证结果表明 IBDR 的性能优于传统支持向量机<sup>[14]</sup>. 支持向量机在经验风险最小化的前提下采用结构风险最小化的原则, 这样可以在很大程度上提高模型的泛化能力, 但是传统的支持向量机也同样存在错误分类经验风险升高的问题, 特别是当样本点与最优超平面十分接近的时候, 这种误分类的经验风险显著升高. 另外, 传统 SVM 存在一定过度拟合问题, 所以对外点或噪声十分敏感. 本文尝试在传统 SVM 的基础上采用模糊支持向量机 (FSVM) 来进行上市公司财务困境预测的实证研究, 用以克服上述方法的缺憾<sup>[15]</sup>. 本文根据不同数据点对产生最优超平面有不同重要性的原则来模糊处理数据以求提高预测精度. FSVM 使模糊化的成本函数惩罚项最小化, 然后重新构造约束

优化问题和约束拉格朗日函数. 最后, 通过对偶形式解出原问题中的最优超平面.

## 1 模糊支持向量机数学模型

传统 SVM 的训练过程对那些远离自己所在类的数据点十分敏感. 假设存在线性可分的两个类, 无论这两个类多么紧密, SVM 都能找到最优超平面将它们分开. 然而, 对于线性不可分的两个类, 如何适当选取自由参数  $C$  的值是至关重要的.

在传统 SVM 的优化目标函数  $\Phi(\mathbf{w}, \xi) = \frac{1}{2} \mathbf{w}^T \mathbf{w} +$

$C \sum_{i=1}^n \xi_i$  中,  $C \sum_{i=1}^n \xi_i$  是成本函数, 即惩罚项.  $C$  越大意味着对分类误差的惩罚越大, 误分类的数据点数目越少. 反之,  $C$  越小意味着忽略那些似是而非的误分类数据点, 同时得到更宽的分类间隔. 无论  $C$  的值是大是小, 在 SVM 的训练过程中这个自由参数是个常值. 也就是说, 在 SVM 的训练过程中训练数据点被同等对待. 这样必然存在对特殊数据点 (外点或者噪声) 的敏感性. 因此, FSVM 的核心内容就是根据数据点对所在类的不同重要性分配给每个数据点不同的权重. 当每个数据点  $x_i$  都分配权重  $\mu_i$ , 则训练集  $S$  变化为模糊化的训练集  $S_f$ .

定义

$$S_f = \{x_i, y_i, \mu_i\}_{i=1}^n \quad (1)$$

当正类 ( $y_i = +1$ ) 时, 权重值被定义为  $\mu_i^+$ ; 当负类 ( $y_i = -1$ ) 时, 权重被定义为  $\mu_i^-$ . 它们之间是相互独立的. 假设模糊化的训练数据集已经得到  $S_f$ , 接下来给出 FSVM 的数学模型<sup>[8,15]</sup>. 从成本函数构造出发, FSVM 在最大化分离间隔的同时最小化分类误差, 从而达到更好的泛化能力. 与 SVM 中的惩罚项不同, FSVM 使用模糊化的惩罚项来降低不重要数据点的影响. FSVM 中的惩罚项是权重值的函数. FSVM 的约束最优化问题的数学模型:

$$\min \Phi(\mathbf{w}, \xi, \mu) = \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^n \mu_i \xi_i \quad (2)$$

$$\text{s.t. } y_i (\mathbf{w}^T x_i + b) \geq 1 - \xi_i \quad (3)$$

$$\xi_i \geq 0 \quad i = 1, 2, \dots, n \quad (4)$$

其中,  $m$  影响成本函数中惩罚项的模糊度.

原问题的拉格朗日函数为

$$\Phi(w, b, \xi, \alpha, \beta, \mu) = \frac{1}{2}w^T w + C \sum_{i=1}^n \mu_i^m \xi_i - \sum_{i=1}^n \alpha_i [y_i (w^T x_i + b) - 1 + \xi_i] - \sum_{i=1}^n \beta_i \xi_i \quad (5)$$

其中,  $\alpha_i$  和  $\beta_i$  为非负拉格朗日乘子.

原问题的对偶问题为

$$\max Q(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j x_i^T x_j \quad (6)$$

$$s.t. \sum_{i=1}^n \alpha_i y_i = 0 \quad (7)$$

$$0 \leq \alpha_i \leq C \mu_i^m, \quad i = 1, 2, \dots, n \quad (8)$$

FSVM s 的 KT 条件为

$$\alpha_i [y_i (w^T x_i + b) - 1 + \xi_i] = 0 \quad (9)$$

$$(C \mu_i^m - \alpha_i) \cdot \xi_i = 0, \quad i = 1, 2, \dots, n \quad (10)$$

比较 SVM s 与 FSVM s 唯一的不同之处就是拉格朗日乘子  $\alpha_i$  的上确界不同. 在 SVM s 中, 乘子  $\alpha_i$  的上确界是自有参数  $C$ , 而在 FSVM s 中, 乘子  $\alpha_i$  上确界是随着权重值函数变化而动态变化的.

## 2 权重模型

在 FSVM s 中, 权重模型分两部分实现 (如图 1). 首先, 外点应该从训练集中识别出来. 然后, 分别模糊化外点集和主体集 (去除外点后的训练集), 从而到模糊化的训练集  $S_f$ .

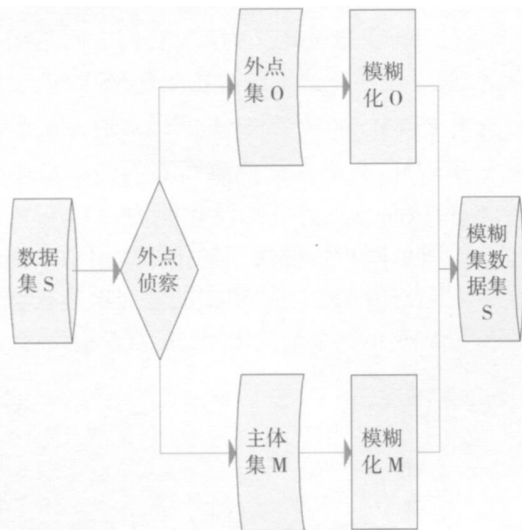


图 1 FSVM s 的权重模型

Fig. 1 Weighted model of FSVM s

### 2.1 外点侦察

从训练集中搜寻外点之前, 先给出外点的两个重要特征. 第一, 外点远离主体集; 第二, 外点的数量远远小于正常数据点的数量. 依据这两个特征, Huang 和 Liu 集成自组织映射 (SOM) 和模糊 c-均值算法 (FCM) 开发了智能外点侦察模型 (ODM)<sup>[15]</sup>. 这个模型有两个优点: 首先, ODM 可以在一维空间完成外点侦察任务, 并且可以通过这个一维空间观察到训练数据点的分布. 其次, ODM 可以自动确定外点集和主体集.

设  $X_c = \{x_i\}_{i=1}^p$  为  $S$  中的正类或负类,  $x_i \in R^N$ , 则构造 ODM 的程序如下:

步骤 1 无监督的特征映射<sup>[16]</sup>.

设自组织映射 SOM 为一维自组织映射, 然后利用 Kohonen's feature map 的学习规则来更新输入与输出层之间的权重向量, 每个训练数据点  $x_i \in R^N$  映射成一个标量. 这个标量由基于最小化的欧几里得距离相似性匹配函数所确定:

$$y_i = \arg \min_j \|x_i - w_j\|, \quad j = 1, 2, \dots, l \quad (11)$$

$t$  时刻权重的更新规则为

$$w_j(t+1) = w_j(t) + \eta(t)h(t)[x(t) - w_j(t)] \quad (12)$$

其中,  $\eta$  为学习率,  $h$  为邻近函数,  $l$  为输出结点的个数. 通过无监督特征映射, 生成标量集合  $Y_c = \{y_i\}_{i=1}^p$  且  $y_i \in R$ .

步骤 2 将一个虚拟集合合并到训练集中.

设  $Y_v$  为虚拟集,  $Y_c \in R, |Y_v| = |Y_c| = p$  和  $Y_c \cap Y_v = \emptyset$ .  $Y_c$  与  $Y_v$  满足下面两个条件:

- 分离条件: 集合  $Y_c$  与集合  $Y_v$  之间的分离间隔充分大;

- 紧密条件: 集合  $Y_v$  中元素的紧密程度充分大. 于是分别定义分离指数  $SI$  和紧密指数  $CI$

当  $y_i \in \{Y_c, Y_v\}$  且  $p' = 2p$  时

$$SI = \frac{1}{p} \sum_{i=1}^{p'} \mu(d(y_i, \bar{y})) \quad (13)$$

当  $y_i \in \{Y_v\}$  且  $p' = p$  时

$$CI = \frac{2}{p} \sum_{i=1}^{p'} \mu(d(y_i, \bar{y})) \quad (14)$$

其中,  $\bar{y}$  为  $Y$  的均值. 设  $d$  为标准化的距离

$$d(y_i, \bar{y}) = \frac{|y_i - \bar{y}|}{2 \cdot \max \|y_j - \bar{y}\|}$$

$$0 \leq d \leq 1/2 \quad (15)$$

集合  $A$  的权重函数定义为  $(\text{semi} - \pi)$  权重函数

当  $0 \leq d \leq 1/2$  时,

$$\mu_A = 1 - 2d^2 \quad (16)$$

其他情况

$$\mu_A = 0 \quad (17)$$

显然地,  $SI$  的值越大, 集合间的分离程度越大;  $CI$  值越小, 集合的紧密程度越大. 以  $SI$  和  $CI$  这两个指数为标准, 得到满足上述条件的  $Y_i$ . 至此, 可以为下一步建立候选集合  $Y = \{Y_1, Y_2\} = \{y_1, \dots, y_b, \dots, y_2\}$ .

步骤 3 对集合执行 FCM<sup>[17]</sup>.

在此 FCM 算法用于找出簇的中心  $v_i$  和权重  $\mu_{ij} \in [0, 1]$ .  $\mu_{ij}$  描述了元素  $y_j$  属于第  $j$  簇的程度.

$$v_i = \frac{1}{\sum_{j=1}^{2p} (u_{ij})^e} \sum_{j=1}^{2p} (u_{ij})^e y_j \quad (18)$$

$$u_{ij} = \frac{(1/\|y_j - v_i\|^2)^{1/(e-1)}}{\sum_{k=1}^c (1/\|y_j - v_k\|^2)^{1/(e-1)}} \quad (19)$$

$$i = 1, 2, \dots, c; j = 1, 2, \dots, 2p$$

其中,  $e$  称为指数权重, 描述了  $u_{ij}$  的模糊程度. 对  $u_{ij}$  有两个约束条件

$$\sum_{i=1}^c u_{ij} = 1, j = 1, 2, \dots, 2p \quad (20)$$

$$0 < \sum_{j=1}^{2p} u_{ij} < 2p, i = 1, 2, \dots, c \quad (21)$$

$c$  为簇的个数. 根据<sup>[10]</sup>, 最优划分数 (BPN), 即  $c_{best}$  可以通过最小化划分熵  $H(U, c)$  来确定.

$$H(U, c) = \frac{1}{2p} \sum_{j=1}^{2p} \sum_{i=1}^c |u_{ij} \ln u_{ij}| \quad (22)$$

于是

$$c_{best} = \arg \min_{c=2}^{2p-1} \left\{ \min_{U \in \Omega_c} [H(U, c)] \right\} \quad (23)$$

其中,  $\Omega_c$  为给定  $c$  时最解集.

步骤 4 确定最大化主体集.

$$\text{设 } P_m^1 = \arg \min_{i=1}^{c_{best}} |P_i|, \text{ 其中 } P_m \in P,$$

$$P = \{P_i \mid i = 1, 2, \dots, c_{best} - 1\} \quad (24)$$

步骤 5 寻找主体集和外点集.

$$\text{设 } P_T = \arg \min_{i=1}^{c^*} |P_i|,$$

$$P_o = \{P_o^b \mid b = 0, \dots, r\} \text{ 且 } |P_o| = 0$$

$$P_m = \{P_m^a \mid a = 1, \dots, q\}$$

寻找主体集和外点集的程序参见附录.

## 2.2 权重函数

分配权重值的规则依赖于训练数据点对于所在类的相对重要程度. 为了获得模糊训练集  $S_f$ , 分别给出如下两种情况的权重函数:

当  $x_j \in M$  时, 权重函数

$$\mu_i = 1 - \frac{\|x_i - \bar{x}\|}{\max_{j \in M} \|x_j - \bar{x}\|} + \varepsilon \quad (25)$$

$$x \in R^2$$

其中,  $\|\cdot\|$  为欧几里得距离,  $\varepsilon$  为充分小的正数,  $\bar{x}$  为  $M$  中所有数据点的均值.

当  $\forall x_i \in O$  时, 权重函数

$$\mu_i = \sigma \quad (26)$$

其中,  $\sigma, 0 < \sigma < \varepsilon$   $\varepsilon$  为充分小的正数.

## 3 实证研究

本文以沪深两市按行业板块分类的上市公司作为研究对象, 数据来源于 Wind 资讯. 将公司因财务状况异常而被特别处理 (Special Treatment) 作为企业陷入财务困境的标志, 选取 2007 年以前的 ST 公司作为财务困境公司样本. 由于证监会是根据上市公司前两年的年报公布的业绩判断其是否出现财务状况异常并决定是否要对其进行特别处理的, 所以采用上市公司前两年的年报预测公司是否会被 ST 显然会夸大模型的预测能力. 因此, 本文选择在上市公司被 ST 前三年的数据进行预测, 判断其最终是否会陷入财务困境, 即如果某上市公司在 2006 年被特别处理, 本文采用 2003 年的年报数据进行预测. 为了剔除不同年份、行业和资产规模因素对财务困境预测的影响, 国内学者们都默认采用 1:1 的配比原则来选择财务健康的上市公司作为配对样本. 然而, 这种情况表明国内的研究没有意识到样本比对模型效率影响的问题. 在本文中, 本文根据相关文献<sup>[18]</sup> 的研究结果——认为以 1:3 为配比比率比较适合我国的情况, 同时采用 FSVM 方法来对我国上市公司财务困境进行预测.

### 3.1 基于基准数据集的性能比较

为了清晰地显示 FSVMs 比传统 SVMs 具有更好的性能, 本文先用基准数据集 (R 语言中的

Mixture data) 做实证比较. 在试验中, 本文先建立 1-D 自组织映射. 一些参数的设定为: 输出结点数为 200 输入结点数为 2 邻近函数的初始值为 100 学习率为 0.9 图 2 显示了 1-D 自组织映射在权重分布, 其中实心点代表输出点的权重值, 加号代表数据点. 图 3 展现了 1-D 自组织映射将数据集映射成一个标量集, 并显示了这个标量集的分布情况.

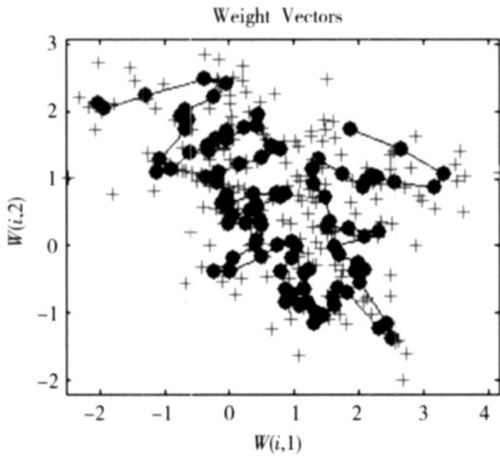


图 2 1-D SOM 的输出结点  
Fig. 2 Output nodes of 1-D SOM

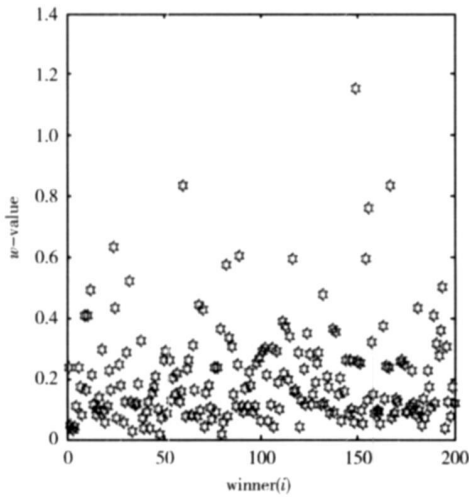


图 3 标量点沿一维空间的分布  
Fig. 3 Distribution of scalar in 1-D

图 4 和表 1 给出了传统 SVMs 的分类结果, 其中一共运算了 443 步, 每 100 步的运算结果在表中给出. 最后的分类结果为: 有 23 个数据点被错误分类, 分离间隔为 55.32 支持向量的个数为 151, 正确率为 88.5%. 与传统的 SVMs 相对应, 图 5 和表 3 展现了 FSVMs 的分类结果, 其中一种运算了 600 步, 每 100 步的运算结果在表中给出. 最后的

分类结果为: 有 17 个数据点被误分类, 分离间隔为 38.87, 支持向量的个数为 145, 正确率为 91.5%. 比较这两种方法, FSVMs 的正确率比传统的 SVMs 高出 3%, 是因为 FSVMs 误分类了 17 个数据点而传统 SVMs 有 23 个误分类点. 同时 FSVMs 的支持向量个数少于传统 SVMs 产生的个数. 然而, FSVMs 的算法步数多于传统的 SVMs 故计算时间消耗较多, 并且分离间隔也小于传统的 SVMs 但总体来看, FSVMs 比 SVMs 的性能更好.

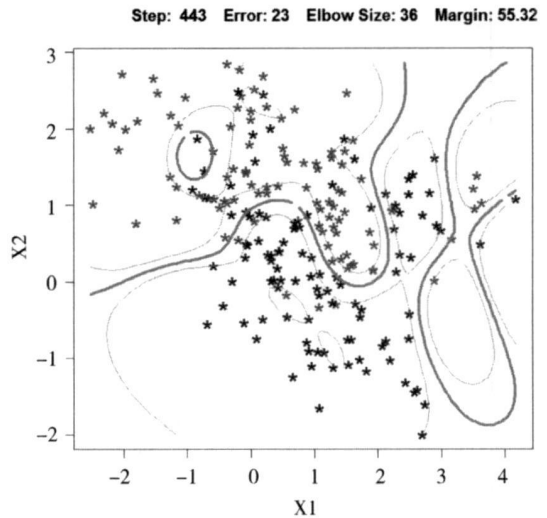


图 4 传统 SVMs 的分类结果  
Fig. 4 Classification results of SVMs

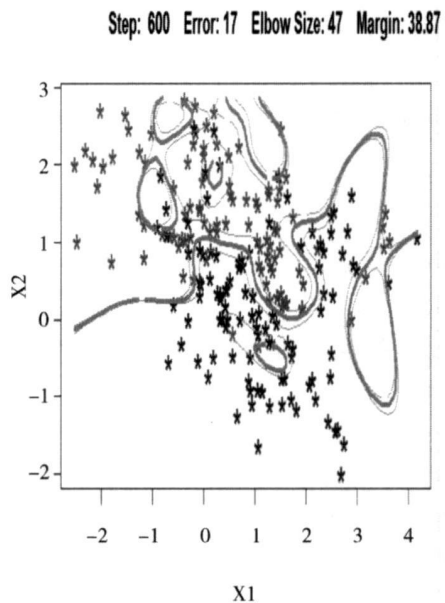


图 5 FSVMs 的分类结果  
Fig. 5 Classification results of FSVMs

表 1 传统 SVM 的输入结果

Table 1 Output results of conventional SVM

steps	Lambda	Error	Elbow	Support	Margin
1	20.166803	45	2	200	127.16631
100	6.862707	41	13	188	99.85790
200	2.188731	36	17	172	84.91902
300	0.481014	32	27	165	72.10625
400	0.038890	28	31	154	61.29164
443	0.009902	23	36	151	55.31863

表 2 FSVM 的输出结果

Table 2 Output results of FSVM

steps	Lambda	Error	Elbow	Support	Margin
1	20.166803	45	2	200	127.16631
100	6.862707	41	13	188	99.85790
200	2.188731	36	17	172	84.91902
300	0.481014	32	27	165	72.10625
400	0.038890	28	31	154	61.29164
500	0.001871	24	41	146	49.99257
600	0.000099	17	47	145	38.87135

图 6 图 7 和表 3 说明了在 FSVM 中,应用径向核函数比应用其他核函数在分类正确率上有更好的性能,正确率可达到 94% 以上,同时支持向量的个数也相对较少。而线性核函数在性能上表现最差,正确率仅有 73%,支持向量多达 171

个。多项式核函数则介于两者之间,正确率为 81%,支持向量机的个数为 153 因此,本文使用基于径向核函数的 FSVM 来对上市公司财务困境进行实证研究。

Step: 120 Error: 54 Elbow Size: 3 Margin: 122.98

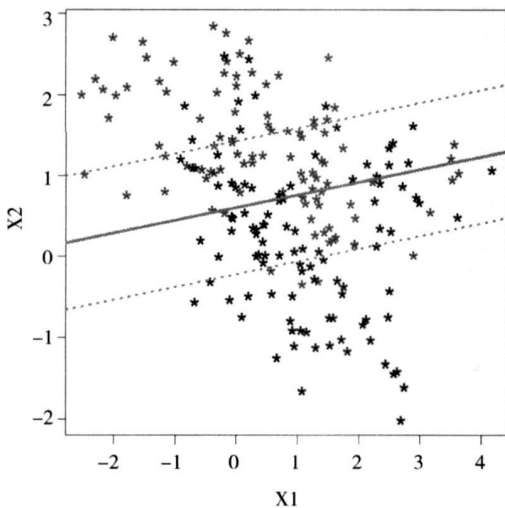


图 6 基于线性核函数的 FSVMs 分类结果

Fig. 6 Classification results of FSVMs based on linear kernel function

Step: 360 Error: 51 Elbow Size: 11 Mar

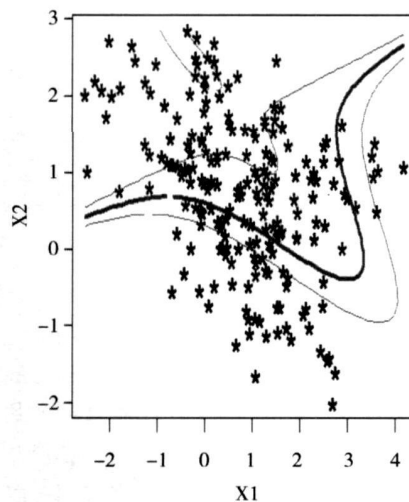


图 7 基于多项式核函数 FSVMs 的分类结果 (h = 3)

Fig. 7 Classification results of FSVMs based on polynomial kernel function (h = 3)

表 3 基于三种核函数 FSVM s的分类结果

Table 3 Classification results of FSVM s with three kernel functions

Kernel	Steps	Lambda	Error E low	Support	Margin	Correctness
Linear	120	0.26	54	171	122.98	73%
Poly	360	0.00	51	148	109.80	74.5%
Radial	600	0.00	12	137	34.78	94%

### 3.2 基于实际数据的上市公司财务困境预测

在财务困境预测这个问题上, 由于长期以来缺少基本理论的支持, 没有形成一套标准的建模指标体系, 因此, 一般按照研究者自己拟定的标准选取, 以尽可能地提高准确率. 尽管研究者选取的指标不尽相同, 但总体来说基本思路是一致的: 把

财务指标分成几大类, 分别从各类中选取一个或几个指标代表该类的财务特点, 经过反复试验最后确定指标. 本文采用国内文献的财务指标体系, 即从盈利能力、偿债能力、运营能力、经营前景、资本结构等五大类指标中各选取若干个指标.

表 4 上市公司财务指标

Table 4 Financial index of listed corporation

指标大类	指标变量 (%)
盈利能力	销售净利率、主营业务利润率、总资产收益率、主营业务利润增长率
偿债能力	流动比率、速动比率、利息保障倍数、股东权益与固定资产比率
运营能力	总资产周转率、股东权益周转率、总资产增长率 固定资产周转率、存货周转率、应收账款周转率
经营前景	税后利润增长率、净资产增长率、营业利润增长率
资本结构	资产负债比率、固定资产比率、股东权益比率

表 5 FSVM s模型的训练和预测结果

Table 5 Training and predicting results of FSVM s model

Model	train			prediction			
	sv	n	n <sub>1</sub>	n <sub>2</sub>	accuracy	I error	II error
SVM s	67	13	4	9	87%	96%	91%
FSVM s	60	8	3	5	91%	97%	95%

表 6 SVM s与 FSVM s性能比较

Table 6 Performance of SVM s and FSVM s

year	accuracy		I error		II error	
	SVM s	FSVM s	SVM s	FSVM s	SVM s	FSVM s
t-3	87%	91%	96%	97%	91%	95%
t-4	72%	80%	76%	85%	80%	77%
t-5	51%	65%	66%	70%	53%	64%

为了选择对 ST 公司和非 ST 公司区分能力最强的指标变量以及剔除过多的指标带来的多重共线性问题, 对数据分别进行正态性检验、参数及非参数检验及指标变量之间的多重共线性检验, 以此来筛选并确定本文进入模型的变量. 其流程图如图 8 所示.

将上述财务指标进行统计处理后, 得到 9 个

财务比率指标 (净资产收益率、流动比率、速动比率、总资产周转率、净资产增长率、总资产增长率、资产负债比率、税前利润增长率、利息保障倍数), 则分别代入到传统 SVM s 与 FSVM s 中, 综合这两种模型研究结果得表 6 可以看出, FSVM s 模型优于 SVM s 模型, 其中第 1 类、第 2 类错误以及预测准确率分别高出 1%、4% 和 4%.

同时, 为了检验模型的长期预警能力和稳定性, 本文将所选样本的  $(t-4)$  年、 $(t-5)$  年的财务数据采用同样的方法依次进行判定分析, 所得结果如表 7 所示. FSVM 和 SVM 两个预警模型在公司被特别处理前三年预警正确率分别 91% 和 87%, 而前四年和前五年的预警正确率则分别迅速下降到 80%、72% 和 65%、51%. 可见, 预警模型的长期预警能力并不十分理想, 这可能是因为部分 ST 公司的经营失败具有“突变”性质, 从而影响了预警模型的判别效果和提前预警能力. 因

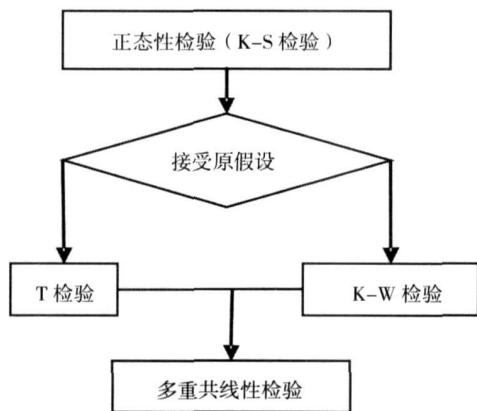


图 8 筛选财务比率变量流程图

Fig. 8 Flow chart of filtering financial ratio

此, 本文有必要结合国内上市公司的实际情况, 采用以定量模型为主、定性分析为辅的方法来进一步修正和完善财务困境预测.

## 4 结论与展望

在本文中, 本文根据不同数据点对产生最优超平面有不同重要性的原则, 应用模糊支持向量机模糊处理数据, 并最小化模糊成本函数惩罚项, 然后用凸二次规划求解约束优化问题的对偶形式, 最后得到原问题的最优超平面. 解决了传统支持向量机在分类 (两类) 过程中的过度拟合以及外点敏感问题, 并且提高了支持向量机的分类性能 (如正确率、支持向量个数、分离间隔等). 在实证研究中, 本文成功地将模糊支持向量机应用于上市公司财务困境预测, 试验结果表明 FSVM 比传统的 SVM 有更好的预测精度. 因此, FSVM 更适合作为财务困境的预测模型, 应用该方法对上市公司财务困境进行预测是成功的和有意义的. 在下一步的工作中, 本文考虑将 FSVM 应用于多类分类器, 并使用不同的隶属度函数来展开进一步的研究工作.

## 参考文献:

- [1] Beaver W. Financial ratios as predictors of failure: empirical research in accounting: Selected studies [J]. *Journal of Accounting Research*, 1966, Supp(5): 71-111.
- [2] Altman E. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy [J]. *The Journal of Finance*, 1968, 23(4): 589-609.
- [3] Altman E, Haldeman R, Narayanan P. Zeta analysis: A new model to identify bankruptcy risk of corporations [J]. *Journal of Banking and Finance*, 1977, (1): 29-54.
- [4] Martin D. Early warning of bank failure: A logistic regression approach [J]. *Journal of Banking and Finance*, 1977, (1): 249-276.
- [5] Ohlson J. Financial ratios and the probabilistic prediction of bankruptcy [J]. *Journal of Accounting Research*, 1980, 18 (1): 109-131.
- [6] Franco V. Genetic algorithms applications in the analysis of insolvency risk [J]. *Journal of Banking & Finance*, 1998 (22): 1421-1439.
- [7] 胡援成, 田满文. 上市公司财务困境预测模型的再比较 [J]. *经济学 (季刊)*, 2005, (4): 173-188.  
Hu Yuancheng, Tian Manwen. A comparison of models for the prediction of financial distress in Chinese stock market [J]. *China Economic Quarterly*, 2005, (4): 173-188 (in Chinese).
- [8] Min JH, Lee YC. Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters [J]. *Expert Systems with Applications*, 2005, (28): 603-614.
- [9] Lai KK, Yu L, Zhou L, et al. Credit Risk Evaluation with Least Square Support Vector Machines [J]. *RSKR*, 2006, LNAI



- (4062): 490—495.
- [10] Xie W, Yu L, Xu S *et al*. A New Method for Crude Oil Price Forecasting Based on Support Vector Machines. ICCS 2006, Part IV, LNCS (3994): 444—451.
- [11] Huang W, Nakanoria Shouyang Wang. Forecasting stock market movement direction with support vector machine [J]. Computers & Operations Research, 2003 (32): 2513—2522.
- [12] 诸克军, 成金华, 郭海湘. 模糊软分类中最佳聚类数的确定 [J]. 管理科学学报, 2005, 8(3): 8—14.  
Zhu Kejun, Cheng Jinhua, Guo Haixiang. Optimal number of clusters in fuzzy soft clustering [J]. Journal of Management Sciences in China, 2005, 8(3): 8—14 (in Chinese).
- [13] 程 功, 张 维, 熊 熊. 信息噪音结构化模型与银行违约概率度量 [J]. 管理科学学报, 2007, 10(4): 38—48.  
Cheng Gong, Zhang Wei, Xiong Xiong. Noisy information structural model and bank evaluation of default probability [J]. Journal of Management Sciences in China, 2007, 10(4): 38—48 (in Chinese).
- [14] Hua Z, Wang Y. Predicting corporate financial distress based on integration of support vector machine and logistic regression [J]. Expert Systems with Applications, 2007 (33): 434—440.
- [15] Huang H., Liu Y. Fuzzy support vector machines for pattern recognition and data mining [J]. International Journal of Fuzzy Systems, 2002 (4): 1—12.
- [16] Kohonen T. The self-organizing map [J]. Proc. IEEE, 1990 (9): 1464—1480.
- [17] Kanevskii M, Pozdnukhov A, Canu S *et al*. Advanced spatial data analysis and modeling with support vector machines [J]. International Journal of Fuzzy Systems, 2002 4(1): 606—615.
- [18] 石晓军, 肖远文, 任若恩. Logistic 违约率模型的最优化样本配比与分界点研究 [J]. 财经研究, 2005, (9): 39—49.  
Shi Xiaojun, Xiao Yuanwen, Ren Ruoen. Optimal sample pairing and critical value of logistic default risk modeling: The China case [J]. Economic Research, 2005, (9): 39—49. (in Chinese).

## Predicting financial distress of listed corporations based on fuzzy support vector machine

YANG Hai-jun, TAI Lei

School of Management and Economics, Beihang University, Beijing 100191, China

**Abstract** The support vector machine (SVM) has been applied to the problem of bankruptcy prediction, and proved to be superior to competing methods such as the linear multiple discriminant approaches, logistic regression and the neural network. However, conventional SVM employs the structural risk minimization principle, thus empirical risk of misclassification may be higher, especially when a point to be classified is close to the optimization hyper-plane. Furthermore, SVMs are very sensitive to outliers or noises because of overfitting problem. In this paper, FSVM was proposed to deal with the corporate financial distress prediction. First, a proper membership model was used to fuzzy all the training data of positive/negative class. Second, the outliers were detected by the proposed outlier detection method (ODM). The ODM was a hybrid method based on the fuzzy c-means (FCM) algorithm cascaded with an unsupervised neural network, called self-organizing map (SOM). Finally, FSVM was applied to corporate financial distress prediction, and experical results indicated that the proposed FSVMs actually reduced the effect of outliers and yielded higher classification rate than SVMs.

**Key words** corporate financial distress; fuzzy support vector machines; prediction