

# Research on Credit Risk Early-Warning for Listed Companies in Chengyu Economic Zone Based on Best Fuzzy Support Vector Machine

Kai Xu<sup>1,2</sup>, Zongfang Zhou<sup>1</sup>

<sup>1</sup>School of Tourism Economics and Management, Chengdu University, Chengdu, 610106, China

<sup>2</sup>School of Management and Economics, University of Electronic Science & Technology of China, Chengdu, 610054, China

## 基于最优 FSVM 的企业信用风险预警研究 ——来自成渝经济区上市公司的经验证据

徐凯<sup>1,2</sup>, 周宗放<sup>1</sup>

<sup>1</sup>成都大学旅游与经济管理学院, 成都 610106, 中国

<sup>2</sup>电子科技大学经济与管理学院, 成都 610054, 中国

### Abstract

Taking listed companies in Chengyu Economic Zone as an example, this paper introduces the fuzzy algorithm into support vector machine (SVM), constructing the model of fuzzy support vector machine (FSVM) for Credit risk early-warning, which based on four different kernel functions (linear, polynomial, sigmoid and Gauss radial basis) are compared as well as compared with traditional statistical models and other artificial intelligent models. The result of investigation illustrates that FSVM based on Gauss radial basis kernel function is not only superior to that based on other three kernel functions, but also better significantly than traditional statistical models and other artificial intelligent models.

**Keywords:** enterprise credit risk; FSVM; kernel function; Chengyu economic zone; early-warning

### 摘要

以我国成渝经济区上市公司为研究对象, 将模糊方法 (Fuzzy Approach) 引入支持向量机 (Support Vector Machine, SVM), 构建了信用风险预警的模糊支持向量机 (FSVM) 模型, 并对四种不同核函数 (线性、多项式、神经元

的非线性和 Gauss 径向基) 下的 FSVM 进行了性能对比研究, 同时, 也与传统统计模型和其他三种人工智能模型进行了性能对比。实证结果表明, Gauss 径向基核函数下的 FSVM 模型不仅较其他三种核函数作用下的 FSVM 模型具有更为优越的预测性能, 同时, 也显著优于传统统计模型和其他一些常见的人工智能模型。

**关键词:** 企业信用风险; FSVM; 核函数; 成渝经济区; 预警

### 1. 引言

随着全球经济和金融市场一体化进程的加快, 上市公司之间的竞争日益激烈, 各种金融衍生工具不断涌现, 使得信用风险爆发的可能性大大增加。而一旦爆发信用危机, 不仅严重威胁上市公司的生存与发展, 同时还严重损害投资者利益, 甚至引发金融危机, 给国家和世界经济带来灾难。因此, 开展信用风险预警研究, 以准确有效的预测、防范和化解信用危机, 对促进上市公司的健康发展、优化投资者的投资决策、推动国家和世界经济的稳定发展具有极其重要的现实意义。尤其是次贷危机、欧债危机等金融危机的相继爆发, 进一步说明信用风险是金融危机的始作俑者和推波助澜者, 使得信用风险预警研究成为了学术与实务界研究的热点问题之一。

信用风险预警模型至今已被国内外大量

学者广泛关注,较早的预警模型主要是以单变量分析[1]、多元判别分析(Multiple Discriminant Analysis, MDA)[2]、逻辑(Logit)回归[3]、Fisher 判别法[4]、SAS 判别法[5]以及分类树[6]等为代表的传统统计模型。然而,传统统计模型却存在比较明显的缺陷,如传统统计模型通常假设样本服从正态分布,实际上,获取的样本往往难以满足正态分布的前提假设;又如传统统计模型属于线性模型,但现实世界中的样本数据却往往呈现出非线性特征<sup>[7]</sup>,如果仍然使用传统统计模型进行线性建模,就很可能导致预测不准确。

由于传统统计模型存在上述缺陷,因而以人工神经网络(Artificial Neural Networks, ANNs)[8]和支持向量机(Support Vector Machine, SVM)[9]为代表的人工智能模型应运而生。ANNs 能够有效地克服传统统计模型存在的缺陷,对样本数据没有过多的前提要求,被研究学者大量运用于金融风险预警研究中,取得了较好的效果。但 ANNs 仍然存在不足,如容易陷入局部最优、可调整参数和阈值过多,从而对模型的预测精度产生了较大的影响。而与 ANNs 基于经验风险最小化建模不同, SVM 基于结构风险最小化进行建模,能够有效获得全局最优而避免陷入局部最优。同时, SVM 的可调整参数较 ANNs 更少,因而在危机预警中能够取得比 ANNs 更为优异的预测效果[10]。值得注意的是,当 SVM 处理包含较多噪声样本的数据时,容易受到噪声的干扰,从而导致预测精度较低。为了减少这些噪声样本点对 SVM 预测性能的干扰, Lin, Wang(2002)和 Huang, Liu(2002)将模糊方法(Fuzzy Approach)引入到了 SVM 中,提出了一个新颖的 SVM 模型——模糊 SVM 模型(FSVM)。FSVM 通过对每个样本点赋予不同的隶属度(噪声样本点被赋予较低的隶属度,而非噪声样本点被赋予较高的隶属度),从而减少了噪声样本点对 SVM 预测性能的影响,进而能够有效地提升 SVM 的预测精度。另外, SVM 模型的预测性能是否优越与核函数的选择息息相关。目前, SVM 的核函数主要包括线性、多项式、Gauss 径向基和神经元的非线性作用四种核函数。然而,究竟选择上述何种核函数能够使得 SVM 预警模型取得最优的预测效果,目前还没有定论<sup>[16]</sup>。尽管有部分学者通过实证研究证明了基于 Gauss 径向基核函数的 SVM 模型的预测效果最优,然而,

由于研究对象的非一致性,使得在进行成渝经济区上市公司信用风险预警时,基于 Gauss 径向基核函数的 SVM 模型不一定能达到最优预测效果。

迄今为止,国内外已有众多学者运用 SVM 对信用风险预警和评估进行了研究。Niklis et al.(2014)将包括 SVM 的机器学习理论、期权方法和会计数据相结合,用于信用风险的评估之中,构建了一个准确有效的信用风险评估模型,取得了较好的效果。Haoming Zhong et al.(2014)基于上市公司的财务数据,比较了 SVM、BPNN、ELM、I-ELM 四种模型,研究表明 SVM 模型效果突出,具有更好的预测精度。李丽和周宗放(2013)将 AdaBoost 算法与 SVM 相结合,建立了 SVM 集成分类器,实证表明 SVM 集成分类器较单一 SVM 方法具有更高的分类准确率。张日和周宗放(2009)基于沪深股市的财务数据,运用所提出的基于多目标规划和支持向量机的企业信用评估模型,并与传统 SVM 和 Logit 模型进行对比,研究表明可以提高模型的准确率。他们的研究都取得了较好的效果。

成渝经济区作为长江流域三大经济区之一,既担负着承接中东部地区产业转移的重要任务,又肩负着国家振兴西部的重要使命,对于促进区域经济的协调发展起着至关重要的作用。而成渝经济区内的上市公司,作为成渝经济区发展与建设的重要经济载体,在维持整个区域经济的稳定与发展中显得尤为重要。因此,对成渝经济区内上市公司的信用风险预警进行研究,以保证成渝经济区上市公司能够稳定运行与繁荣发展,从而促进西部地区区域经济协调发展,具有十分重要的现实意义。

本文以成渝经济区上市公司为研究对象,基于四种不同的核函数构建不同的 FSVM 信用风险预警模型,并对研究出具有最优预测性能的一类 FSVM 预警模型。本文与上述研究文献存在以下较为明显的差异:(1)就所掌握的文献而言,尚未发现有文献对成渝经济区上市公司进行信用风险预警研究;(2)本文不仅构建了 FSVM 信用危机预警模型,而且又通过实证研究进一步探讨了四种核函数下 FSVM 的预测性能,并从中获得了最优的 FSVM 模型;(3)将(2)中获得的最优 FSVM 模型与 BP 神经网络(BPNN)、决策树(DT)、Logit 回归和 Probit 回归等四种常见模型进行对比,验证了 Gauss

径向基核函数下的FSVM模型的优越性。

本文的逻辑结构安排如下：第2节是成渝经济区上市公司信用风险预警模型的构建，第3节是实证分析，第4节是本文的结论。

2. 模型构建

2.1. 企业信用风险预警的FSVM模型

假设成渝经济区上市公司的样本集为  $(y_i, x_i, s_i)$ ，其中， $i = 1, 2, \dots, n$ ，表示上市公司的样本数量为  $n$ ， $y_i \in \{+1, -1\}$  为状态指标变量，表示  $i$  公司未来是否会发生信用危机，其中，“+1”表示发生信用危机，“-1”表示未发生危机， $x_i = (x_{i1}, x_{i2}, \dots, x_{ip})$  是  $P$  维特征指标变量，表示上市公司  $i$  拥有  $P$  个信用危机预警指标， $s_i = (s_{i1}, s_{i2}, \dots, s_{in})$  为每家上市公司样本所对应的模糊隶属度，且  $0 \leq s_i \leq 1$ 。

为了达到对上市公司信用危机进行预测的目的，本文将其中的一部分样本  $(y_k, x_k, s_k)$  划分为训练集进行训练，从而得到如式(1)的风险预警模型：

$$y_k = \text{sgn}(f(x_k, s_k)) \quad (1)$$

其中， $k = 1, 2, \dots, j$ ， $\text{sgn}(x)$  是符号函数， $f(x)$  是一个与训练集中样本点的特征指标变量有关的决策函数。同时，又将另一部分样本  $(y_h, x_h, s_h)$  划分为测试集，并对预警模型进行性能测试与评价。其中， $h = j+1, j+2, \dots$ 。

于是，通过样本  $(y_k, x_k, s_k)$  来建立风险预警模型，就需要在  $y_k((w^T x_k) + b) + \xi_k \geq 1$  的条件下求解以下最优问题：

$$\min \frac{1}{2} \|w\|^2 + C \sum_{k=1}^j s_k \xi_k \quad (2)$$

其中， $w$  是可调权值向量， $b$  为偏置向量， $\xi_k$  是非负的松弛变量， $C$  为惩罚因子。为了求解上述最优问题，运用拉格朗日乘法将上述最优问题转化为对偶问题：

$$\max_{\alpha} -\frac{1}{2} \sum_{q=1}^j \sum_{q=1}^j y_k y_q \alpha_k \alpha_q K(x_k, x_q) + \sum_{q=1}^j \alpha_q \quad (3)$$

$$\text{s.t.} \sum_{k=1}^j y_k \alpha_k = 0, \quad 0 \leq \alpha_k \leq s_k C, k = 1, 2, \dots, j \quad (4)$$

其中， $\alpha$  为拉格朗日乘子， $(x_q, y_q)$  为  $\alpha$  的一个正分量  $\alpha_q$  所对应的样本点，

$K(x_k, x_q)$  是 SVM 的核函数。目前常用的核函数主要有以下四种：

(1) 线性核函数 (Linear Kernel)：

$$K(x, x') = x^T x' ;$$

(2) 多项式核函数 (Polynomial Kernel)：

$K(x, x') = (ax + b)^c$ ，其中  $a, b, c$  为参数。特别地，当  $a, c = 1, b = 0$  时，它就成为了线性核函数；

(3) Gauss 径向基核函数 (Radical Basis Function, RBF)：

$$K(x, x') = \exp(-\gamma \|x - x'\|^2), \quad \text{其中 } \gamma \text{ 为参数；}$$

(4) 神经元的非线性作用核函数

(Sigmoid Kernel)：

$K(x, x') = \frac{1}{1 + \exp(-ax_i - bx_j)}$ ，其中  $a, b$  为参数。

于是，通过求解上述对偶问题就能得到最终的预警模型：

$$f(x) = \text{sgn} \left( \sum_{k=1}^j \alpha_k^* y_k K(x_k, x_q) + b^* \right) \quad (5)$$

然而，从前文分析可知，要得到上述最终预警模型，关键还在于模糊隶属度  $s_i$  的确定，因此，下文将进一步探讨模糊隶属度  $s_i$  的求解方法。

2.2. 模糊隶属度的构造

目前，针对模糊隶属度的构造，已诞生出众多方法，但迄今为止，还没有一个可遵循的普遍准则。然而，就所掌握的研究文献而言，运用最多的方法是通过样本点到类中心的距离来构造模糊隶属度。因此，本文也考虑运用该方法来构造模糊隶属度  $s_i$ 。

首先，根据每个样本对应的状态指标变量，将训练集  $(y_k, x_k, s_k)$  划分为两类，一类是正类样本集  $(y_k^+, x_k^+)$ ，另一类是负类样本集  $(y_k^-, x_k^-)$ ，定义正类样本集的样本中心为  $(y_{center}^+, x_{center}^+)$ ，负类样本集的样本中心为  $(y_{center}^-, x_{center}^-)$ ，则各类样本点到各类样本中心的最大距离分别为：

$$d^+ = \max \|x_{center}^+ - x_k^+\| \quad (6)$$

$$d^- = \max \|x_{center}^- - x_k^-\| \quad (7)$$

于是，模糊隶属度  $s_i$  就可以通过如下公式进行定义：

$$s_i = \begin{cases} 1 - \frac{\|x_{center}^+ - x_i\|}{(d^+ + \sigma)} & x_i \in x_k^+ \\ 1 - \frac{\|x_{center}^- - x_i\|}{(d^- + \sigma)} & x_i \in x_k^- \end{cases} \quad (8)$$

其中， $\sigma$  为一个任意小的正数。因此，将计算得出的模糊隶属度  $s_i$  代入式 (2) 中，就能够计算得到最终的 FSVM 预警模型，并进一步基于测试样本  $(y_h, x_h, s_h)$ ，对训练得到的 FSVM 预警模型开展性能测试与评价。

### 3. 实证结果与分析

#### 3.1. 样本和特征指标选取

本文以成渝经济区上市公司（共 75 家）为研究对象，并以这些公司因财务状况异常而被特别处理（Special Treatment, ST）作为公司陷入信用危机的标志，从而将这些公司分为信

用危机样本公司和信用正常样本公司两类。为了尽可能多地获得信用危机样本公司，本文将曾经被 ST 的公司都作为信用危机样本公司，以其最近一次发生 ST 事件的前一年财务指标作为特征指标，而剩余的其他上市公司作为信用正常样本公司，以其 2012 年度的财务指标作为特征指标，故最终划分出 23 家信用危机样本公司和 52 家信用正常样本公司。此外，为了尽可能地选取能够全面反映上市公司信用状况的特征指标，本文借鉴李丽，周宗放（2013）、朴文慧（2015）、Zhang, Zhou（2012）[11]，遵循指标选取的客观性和完整性以及可操作性，选取出能够反映上市公司盈利能力、发展能力、偿债能力和营运能力的共 20 项信用风险特征指标（见表 1）。样本公司的数据来源于华泰联合炒股软件以及深圳国泰安信息技术有限公司。

表 1. 信用风险特征指标

一级指标	序号	二级指标	一级指标	序号	二级指标
偿债能力	$x_1$	流动比率	盈利能力	$x_{12}$	主营业务毛利率
	$x_2$	速动比率		$x_{13}$	主营业务利润率
	$x_3$	现金比率		$x_{14}$	成本费用利润率
	$x_4$	资产负债率		$x_{15}$	净资产收益率
	$x_5$	负债权益比率		$x_{16}$	每股收益
	$x_6$	股东权益比率		$x_{17}$	总资产收益率
营运能力	$x_7$	应收账款周转率		发展能力	$x_{18}$
	$x_8$	存货周转率	$x_{19}$		总资产增长率
	$x_9$	流动资产周转率	$x_{20}$		主营业务收入增长率
	$x_{10}$	固定资产周转率			
	$x_{11}$	总资产周转率			

#### 3.2. 数据预处理

由于信用风险预警样本的特征指标众多，而各类指标之间的量纲却并不相同，为了避免较大量纲的指标对于较小量纲指标的影响，本文采用标准差分法对样本数据进行无量纲化处理，方法如下：

$$x'_i = \frac{(x_i - \bar{x}_i)}{SD_i} \quad (9)$$

其中， $\bar{x}_i$  为指标  $x_i$  的均值， $SD_i$  为指标  $x_i$  的标准差。通过对样本数据进行上述无量纲化处理，就得到新的样本集  $(y_i, x'_i, s_i)$ ，从而进一步开展 FSVM 的构建工作。

#### 3.3. 训练样本与测试样本的划分

由于信用风险预警模型的构建需要通过训练

与测试两个步骤来完成，因而在构建 FSVM 预警模型之前，就需要提前划分训练样本与测试样本。借鉴刘碧森等（2007）、李云飞等（2010）、宋新平，丁永生（2008），本文将所有的信用危机样本公司和信用正常样本公司的 70% 划分为训练样本（共包含 52 家上市公司，其中，信用正常样本公司有 36 家，信用危机样本公司有 16 家），剩余的 30% 划分为测试样本（共包含 23 家上市公司，其中，信用正常样本公司有 16 家，信用危机样本公司有 7 家）。

#### 3.4. 实验结果与分析

本文使用 MATLAB2016b 进行建模，SVM 的参数  $C$  设置为 1，其他参数采取默认设置。因此，四种核函数下的 FSVM 与 SVM 模型的对比结果见表 2。从表 2 可以清楚地看到，不论是基于何种核函数，FSVM 模型的预测精度都大大高于单一 SVM，说明模糊方法能够有效地减少样

表 2. 不同核函数下的准确率对比结果

核函数	FSVM	SVM
线性核函数	65.25%	56.55%
多项式核函数	78.36%	39.33%
Gauss 径向基核函数	86.96%	69.67%
神经元的非线性作用核函数	73.91%	61.87%

本中奇异点或野点对 SVM 预测性能造成的影响,从而大幅度提升 SVM 的预测精度。此外,从表 2 还能够明显地发现,不同核函数对 FSVM 模型预测性能会产生不同的影响,在 Gauss 径向基核函数下,FSVM 的预测精度高达 85%以上,大大高于其他三种核函数下的 FSVM 的预测精度,从而说明在成渝经济区上市公司的信用危机预测上,Gauss 径向基核函数下的 FSVM 模型具有最为优越的预测性能。

另外,为了更为充分地展示 FSVM 模型优越的预测性能,本文还将 BPNN、DT、Logit 和 Probit 四种常见模型与 Gauss 径向基核函数下的 FSVM 进行对比研究。结果见表 3。从表 3

径向基核函数下的 FSVM 模型具有最为优越的预测性能,同时,还显著优于传统的统计模型和其他一些人工智能预警模型。

#### 4. 研究结论

本文以我国成渝经济区上市公司为研究对象,将模糊方法引入 SVM,构建了 FSVM 模型,并基于四种不同的核函数进行了性能对比研究,同时,还与传统统计模型和其他三种人工智能模型也进行了性能对比研究。实证结果表明,Gauss 径向基核函数下的 FSVM 模型较其他三种核函数下的 FSVM 模型具有最为优越的预测性能,同时,还显著优于传统统计模型和其他

表 3. 不同预警模型准确率对比结果

模型	FSVM	BPNN	DT	Logit	Probit
预测精度	86.96%	52.37%	27.19%	81.61%	81.61%

可以看出,FSVM 具有最高的预测精度,表明 FSVM 不仅优于传统的预警模型,如 Logit、Probit,也优于其他的智能预警模型,如 BPNN、DT,从而再次证明了 Gauss 径向基核函数下的 FSVM 模型具有最为优越的预测性能。

由于仅仅从分类准确率来评价模型的预测性能还缺少类似于数理统计检验所具有的科学性,因此,本文还将引入 McNemar 检验方法来比较各预警模型的预测结果是否存在显著差异,从而更为科学准确地评价各模型的预测性能。结果见表 4。从表 4 可以看出,FSVM 与其余四种常见模型的 McNemar 检验  $P$  值皆小于 1%,表明通过显著性水平为 1%的显著性检验,从而证明 FSVM 与其余四种模型的预测性能之间存在明显的差异。

综上所述,在四种核函数下,FSVM 模型的预测性能优于单一 SVM 模型,尤其是 Gauss

一些常见的人工智能模型。

根据以上实证结果,Gauss 径向基核函数下的 FSVM 模型有效地预测了我国成渝经济区上市公司的信用危机,这对于维护成渝经济区和西部地区经济的稳定发展能够起到良好的促进作用。为简化问题的分析,本研究模型仅针对信用风险的二分类问题,模型惩罚因子采用固定值,对于多分类问题和因子动态取值有待于今后进一步研究。

#### Acknowledgements

This study was supported by National Natural Science Foundation of China (No.71271043).

#### 致谢

本研究得到了国家自然科学基金项目(71271043)的资助。

表 4. 不同预警模型的 McNemar 检验结果 ( $P$  值)

	BPNN	DT	Logit	Probit
FSVM	2.8962e-04	2.1518e-05	7.9623e-04	7.9623e-04
BPNN		0.1356	0.1138	0.1138
DT			0.0190	0.0190
Logit				1

## Risk Analysis and Crisis Response in Big Data Era (RAC-16)

### 参考文献

- [1] Beaver W. Financial Ratios as Predictors of Failure, Empirical Research in Accounting: Selected Studies. *Journal of Accounting Research*, pp.71-111, 1966.
- [2] Altman E.I. Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *Journal of Finance*, pp.589-609, 1968.
- [3] Ohlson J. Financial Ratios and the Probabilistic Prediction of Bankruptcy. *Journal of Accounting Research*, pp.109-130, 1980.
- [4] 何树红, 王善民. 基于变量双重检验的 Fisher 信用风险度量模型. *系统工程*, 25(08):15-20, 2007.
- [5] 李君艺, 梁智城. SAS 判别分析在商业银行信用风险评估中的应用. *计算机安全*, (07):13-16, 2011.
- [6] 张维, 李玉霜, 王春峰. 递归分类树在信用风险分析中的应用. *系统工程理论与实践*, (03):50-55, 2000.
- [7] Gordy M B.A Comparative anatomy of credit risk models. *Journal of Banking and Finance*, pp.119- 149, 2000.
- [8] Tam, Kiang. Predicting bank failures: A neural network approach. *Management Science*, pp.926-947, 1992.
- [9] Vapnik V. N. The nature of statistical learning theory. *New York: Spring-Verlag*, U.S.A. 1995.
- [10] 杨毓, 蒙肖莲. 用支持向量机 (SVM) 构建企业破产预测模型. *金融研究*, (10):65-75, 2006.
- [11] Zhang M, Zhou Z. A Credit Rating Model for Enterprises Based on Projection Pursuit and K-Means Clustering Algorithm. *Journal of Risk Analysis & Crisis Response*, pp.131-138, 2012.