

支持向量机在企业技术创新绩效评估中的应用

王露琰, 刘瑞霞, 吴 琼, 毛雪莲

(西安工程大学 管理学院, 陕西 西安 710048)

摘 要: 技术创新绩效评估对于企业来说至关重要, 构建一种有效的评估模型非常关键。引入支持向量机, 建立了一种新的评估模型——基于支持向量机的企业技术创新绩效评估模型。并通过实证研究, 证明了该方法在技术创新绩效评估中的有效性及优越性。

关键词: 技术创新; 绩效评价; 支持向量机

中图分类号: F091.354

文献标识码: A

文章编号: 1001- 7348(2007)09- 0072- 02

0 前 言

企业技术创新绩效评估, 是对企业的创新效果和效率进行检验和测评。通过创新绩效评估, 可使企业科学地认识自身的技术创新状态, 分析影响企业技术创新绩效的主要因素, 从而为企业技术创新绩效改善提供依据。目前广泛采用的企业技术创新绩效评价方法主要包括层次分析法、模糊综合评判法、数据包络分析法和人工神经网络分析法等。鉴于以上方法都存在一些局限性, 比如主观性过强等^[1], 本文引入支持向量机 SVM, 并将其应用到企业技术创新绩效评估中, 取得了较好的效果。

1 支持向量机的原理^[2 3 4]

支持向量机 (SVM, Support Vector Machine) 是统计学习理论中最新的内容, 支持向量机最初的研究是针对模式识别中的线性可分的两类问题。但 SVM 要求分割线不仅能将两类样本正确分开, 还要使分类间隔最大。这样的分割线被称为最优分类线(高维即为最优分类面), 使分类间隔最大实际上就是使结构风险最小化。

对于线性可分的训练样本集 $D=\{(x_i, y_i), i=1, 2, \dots, l\}$, $x_i \in R^n$ 存在 (w, b) 使 $(W \cdot x_i + b) > 0, \forall x_i \in \text{Class1}, (W \cdot x_i + b) < 0, \forall x_i \in \text{Class2}$ 分类的目的是寻求 (w, b) 使最优分类面满足分类间隔最大。并且对 (w, b) 进行如下约束: $\min_{i=1, 2, \dots, n} |w \cdot x_i + b| = 1$ 。满足上式的超平面成为典型超平面, 典型超平面到最近数据点的距离为 $\frac{1}{w}$, 这样求最优分类面问题转化为下

面的最优化问题:

$$\min \phi(w) = \frac{1}{2} \|w\|^2 \quad \text{s.t. } y_i [\langle w, x_i \rangle + b] \geq 1 \quad i=1, 2, \dots, l \quad (1)$$

$$\text{其对偶问题为: } \max Q(a) = \sum_{i=1}^l a_i - \frac{1}{2} \sum_{i,j=1}^l a_i a_j y_i y_j (x_i \cdot x_j)$$

$$\text{约束条件为 } \min a_i \geq 0, \sum_{i=1}^l y_i a_i = 0, (i=1, 2, \dots, l) \quad (2)$$

其中: a_i 为 Lagrange 乘子。

问题转化为在一个不等式约束条件下二次函数寻优的问题存在唯一解。解中只有一部分(通常是少部分) a_i 不为 0, 对应的样本就是支持向量。求解上述问题, 得到最优分类函数, 有:

$$f(x) = \text{sgn}\{ (wx) + b \} = \text{sgn}\{ \sum_{i=1}^l a_i y_i (x_i \cdot x) + b \} \quad (3)$$

根据 KT 条件。对于大多数样本来说, $a_i = 0$ 。对应 $a_i = 0$ 的样本被称为支持向量。式中的求解可通过任选一个支持向量机, 由(4)中的约束条件 $y_i (wx_i + b) + \xi_i = 1$ (此时取等号) 求出。

在线性不可分情况下, 引入非负松弛变量 ξ_i 和惩罚参数 C , 则有 $y_i (wx_i + b) + \xi_i = 1, (i=1, 2, \dots, l)$, 将目标改为 $\varphi(w, \xi) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l \xi_i$

$$\text{约束条件: } y_i (wx_i + b) + \xi_i = 1 \quad i=1, 2, \dots, l; \xi_i \geq 0 \quad i=1, 2, \dots, l, \text{ 就得到最优分类超平面。} \quad (4)$$

广义最优超平面的对偶问题与线性可分情况下几乎完全相同, 只是约束条件变为: $0 \leq a_i \leq C \quad i=1, 2, \dots, l$ 。

对于线性不可分问题, 通过引入核函数技术, 将低维的输入空间通过非线性变换映射到高维的特征空间, 在特

收稿日期: 2006- 09- 04

作者简介: 王露琰(1981-), 女, 汉族, 山东临沂人, 西安工程大学管理学院管理科学与工程专业硕士研究生, 研究方向为技术创新和工业工程; 刘瑞霞(1963-), 女, 汉族, 山西河津人, 西安工程大学副教授, 研究方向为技术创新、工业工程和生产管理; 吴琼(1982-), 女, 汉族, 湖北随州人, 西安工程大学硕士研究生, 研究方向为金融工程; 毛雪莲(1982-), 女, 汉族, 安徽宿州人, 西安工程大学硕士研究生, 研究方向为技术创新、工业工程。

征空间求解分类问题。Vapnik 将这种非线性映射函数称为核函数。引入核函数后,以上各式中的内积都可以用核函数代替。此时 SVM 的决策函数为:

$$f(x) = \text{sgn}\left\{\sum_{i=1}^l a_i y_i K(x_i, x) + b\right\} \quad (5)$$

2 基于SVM的企业技术创新绩效评估的实证研究

2.1 指标体系的建立

X_1 : 技术创新费用投入强度=技术创新费用/产品销售收入; X_2 : R & D 经费投入强度=R&D 经费支出/产品销售收入; X_3 : 科技活动人员比重=科技活动人员/从业人员年平均数; X_4 : 科技活动人员中高中级技术职称人员比重=高中级技术职称人员/科技活动人员; X_5 : 工程技术人员比重=工程技术人员数/从业人员年平均数; X_6 : 获取专利个数: 当年获取专利个数; X_7 : 设备水平先进程度=微电子控制设备原价/生产经营用机器设备原价; X_8 : 新产品销售比重=创新产品销售收入/产品销售收入; X_9 : 新产品出口比重=创新产品出口收入/创新产品销售收入^[9]。

2.2 样本来源

在采集样本时,考虑到企业的经营活动和业务范围,为了使各项指标具有可比性,本文选择了 150 家专用设备制造业的上市公司。实验数据以《中国科技统计年鉴-2005》中“各地区大中型工业企业(股份有限公司)(2004)”的数据为基础,对属于专用设备制造企业的股份有限公司技术创新绩效进行评估。在所选取的企业中,80 家企业的技术创新绩效良好,其余的 70 家企业的技术创新绩效较差。本文随机抽取 110 家作为训练样本集,用于构造 SVM 模型;其余的 40 家作为测试样本集,用于检验。

2.3 构造 SVM 模型的步骤^{[3][4][9]}

本文构造的 SVM 模型的内积核函数采用最常用的径向基函数 $K(x, x_i) = \exp(-\|x - x_i\|^2 / 2\sigma^2)$ 综合考虑最小错分样本数和最大分类间隔,在高维空间中构造软间隔 σ^2 、C 等参数,采用交叉验证方法确定 $\sigma^2=100$ 、 $C=10^7$ 。

(1) 根据实际统计值及专家经验建立训练样本集 (X_i, Y_i) , $Y_i \in \{-1, 1\}$; X_i 为 n 维向量, n 为评价体系指标个数;

(2) 选择径向基核函数 $K(x, x_i) = \exp(-\|x - x_i\|^2 / 2\sigma^2)$ 并将输入样本正规化;

(3) 规划问题: $\max Q(a) = \sum_{i=1}^l a_i - \frac{1}{2} \sum_{i,j=1}^l a_i a_j y_i y_j (x_i, x_j) \exp(-\|x_i - x_j\|^2 / 2\sigma^2)$ 约束条件 $\min a_i, \sum_{i=1}^l y_i a_i = 0 (i=1, 2, \dots, l)$, 求解拉格朗日系数 a_i ;

(4) 找出支持向量 SV, 求解分类超平面系数 b , 建立训练样本的最优决策超平面。

(5) 识别: 根据式 $f(x) = \text{sgn}\left\{\sum_{i=1}^l a_i y_i \exp(-\|x - x_i\|^2 / 2\sigma^2) + b\right\}$ 求解新输入数据 x 的决策值, 从而确定分类类型。

根据上述分析,构造样本集 (x, y) , 其中 x 是企业的相关数据, 其维数为 9, y 表示该企业技术创新绩效情况, $y=1$ 表示该企业技术创新绩效好; $y=-1$ 表示该企业技术创新绩效差。

2.4 结果的对比分析

本文使用 Matlab6.1 和 Osusvms 3.00 工具箱进行实验分析^[7]。为了验证该模型的实际效果, 本文做了对比研究。因为 BP 神经网络方法具有科学合理, 能减少主观因素的影响等优点, 近年来被广泛应用到技术创新绩效评估中, 附表对 SVM 和神经网络评估结果进行了比较。神经网络采用三层 BP 网络, 其中神经元激活函数为 Sigmoid 函数, 目标误差为 0.05, 隐层个数为 10。从表 2 中可以看出, SVM 在测试样本集中的准确率达到 86.36%, 识别能力明显优于 BP 神经网络模型的 82.73%。

对另外两种模型的预测能力进行比较。对于训练样本集,神经网络的准确率为 77.50%, SVM 则为 82.50%。在测试样本集中, 准确率都有了不同程度的下降, 神经网络下降了 5.23%, 而 SVM 则下降了 1.36%, 显然 SVM 的鲁棒性较好, 更能满足测试的要求。同时也说明模型的预测能力不是很好, 还需要进一步研究。

附表 SVM 和神经网络评估结果比较

模型	训练样本集(110)		测试样本集(40)	
	正确率	错误率	正确率	错误率
SVM 模型	86.36%	13.64%	85.00%	15.00%
BP 神经网络模型	82.73%	17.27%	77.50%	22.50%

3 结论

本文将 SVM 应用到企业技术创新绩效评估中, 使用真实的上市公司技术创新活动的数据进行了实证研究, 通过 SVM 与 BP 神经网络的比较研究, 发现 SVM 具有拟合度高, 分类简单、有效, 鲁棒性强等特点, 并且取得了很好的效果, 为企业技术创新绩效的评估提供了一条新途径。然而由于本文研究时间有限, 还有一些问题需要进一步加强研究。

参考文献:

- [1] 傅家骥. 技术创新学[M]. 北京: 清华大学出版社, 1998.
- [2] 张学工. 关于统计学习理论与支持向量机[J]. 自动化学报, 2000, 26(1): 32-41.
- [3] Cristianini N, Shawe-Taylor J. An Introduction to Support Vector Machines[M]. Cambridge University Press, 2000.
- [4] Vapnik V. N. The Nature of Statistical Learning Theory[M]. New York: Springer, 1998.
- [5] 杜栋, 庞庆华. 现代综合评价方法及案例精选[M]. 北京: 清华大学出版社, 2005.
- [6] 肖健华, 吴今培, 杨叔子. 基于 SVM 的综合评价方法研究[J]. 计算机工程, 2002, 28(8): 28-30.
- [7] 陆文聪. 支持向量机算法和软件 Chem SVM 介绍[J]. 计算机与应用化学, 2002, 19(6): 697-702.