# 两种不确定支持向量机分类性能的对比研究

刘成忠
(甘肃农业大学 信息科学技术学院,甘肃 兰州 730070)

**摘 要**:为了克服支持向量机方法对于噪声或孤立野值点敏感的问题,通过引入模糊理论与粗糙集方法,可以分别得到两种不确定支持向量机模型。文中通过分析和比较模糊支持向量机和粗糙支持向量机分类模型构造方法,解释了这两种不确定支持向量机模型克服噪声影响的原理。同时通过一个合成数据集和一组标准数据集对这两种不确定支持向量机的泛化性能进行了对比验证。实验结果表明,相比传统支持向量机,两种不确定支持向量机都能不同程度地提高分类精度,并且模糊支持向量机算法整体表现出了更好的泛化性能。

**关键词**:支持向量机;模糊理论;粗糙集
**中图分类号**:TP31 **文献标识码**:A **文章编号**:1673-629X(2011)11-0156-04

# Comparative Study on Classification Performances of Two Indeterminate Support Vector Machines

LIU Cheng-zhong
(College of Information Science & Technology, Gansu Agricultural University, Lanzhou 730070, China)

**Abstract**: In order to overcome the problem that support vector machine is sensitive to the noise and isolated outliers, introduce fuzzy theory and rough set theory into support vector machine to get two kinds of indeterminate support vector machines. Through analysis and comparison of the construction method of fuzzy support vector machine and that of rough support vector machine, the principles of the two indeterminate methods reducing the outliers are explained. At the same time, generalization performances of the two indeterminate support vector machines are comparatively verified through a synthetic data set and a set of standard data. Experiment results show that the two indeterminate methods have better performances of reducing outliers than traditional support vector machine, that they can significantly improve the classification accuracy, and that fuzzy support vector machine has a better generalization performance on the whole.

**Key words**: support vector machine; fuzzy theory; rough set

## 0 引 言

支持向量机(Support Vector Machine,简称 SVM)是一种基于统计学习理论的专门研究有限样本预测的机器学习方法,它是建立在结构风险最小化基础上,比传统的学习方法具有较好的学习性能和泛化能力[1]。但由于支持向量机中最优超平面是依靠靠近分类超平面的少数几个支持向量来确定,当这几个支持向量包含有噪声或孤立野值点时,那么依靠这几个支持向量所建立的超平面就不是最优超平面,常常会发生错误的分类结果[2,3]。

为了克服支持向量机方法对于噪声或孤立野值点敏感的问题,通过引入模糊理论与粗糙集方法,可以分别得到两种不确定支持向量机模型:模糊支持向量机与粗糙支持向量机[4,5]。模糊支持向量机是根据不同输入样本对分类的贡献差异,赋以不同的隶属度,从而消弱噪声或孤立野值点对分类性能的影响。粗糙支持向量机是利用粗糙集思想,对传统支持向量机中的分类间隔进行拓展,分别定义分类间隔的上近似与下近似,对分布在不同区域的样本采用不同的惩罚,从而达到控制噪声的目的。

文中首先比较了两种不确定支持向量机模型的构造方法,分析这两种不确定模型在降低噪声对分类器影响方面的区别与联系。其次通过实验对两种不确定支持向量机方法进行分类性能对比。实验结果表明相比传统支持向量机,两种不确定支持向量机的分类精度都有不同程度的提高,而模糊支持向量机整体表现出了更好的泛化性能。

**收稿日期**:2011-03-07;**修回日期**:2011-06-15
**基金项目**:甘肃省自然科学基金(096RJZA004);甘肃省教育科研基金(0902-04);甘肃省科技支撑计划(1011NKCA058)
**作者简介**:刘成忠(1969-),男,甘肃天祝人,副教授,研究方向为智能决策支持系统。

## 1　传统支持向量机(vSVM)

为了进行类别划分,对于给定的一组样本集 $\{x_i, y_i\}, i = 1, 2, \cdots, l$,这里 $y_i = 1$ 或 $-1$,SVM 依据结构风险最小化原则,将其学习过程转化为如下所示的优化问题:

$$\min \frac{1}{2} \| w \|^2 - v\rho + \frac{1}{l} \sum_{i=1}^{l} \xi_i \tag{1}$$

$$\text{s.t } y_i(w \cdot z_i + b) \geqslant \rho - \zeta_i \quad \zeta_i \geqslant 0, \rho \geqslant 0, i = 1, \cdots, l$$

其中训练样本 $x_i$ 被函数 $z_i = \varphi(x_i)$ 映射到高维特征空间,$w \in R^N$ 是超平面的系数向量,$b \in R$ 为阈值,$\xi_i$ 为松弛变量,$0 \leqslant v \leqslant 1$ 用来控制支持向量的数目。

采用拉格朗日乘子法把上述最优分类面问题转换为其对偶问题:

$$\min(\alpha) = \frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} \alpha_i \alpha_j y_i y_j K(x_i, x_j) \tag{2}$$

$$\text{s.t} \sum_{i=1}^{l} \alpha_i y_i = 0, 0 \leqslant \alpha_i \leqslant \frac{1}{l}, \sum_{i=1}^{l} \alpha_i \geqslant v$$

于是相应的分类决策函数为:

$$f(x) = \text{sgn}\left( \sum_{i=1}^{m} \alpha_i^* y_i K(x_i \cdot x) + b^* \right) \tag{3}$$

其中 $\alpha_i^*$ 为对应 $\alpha_i \neq 0$ 的向量,称为支持向量,$m(m < l)$ 为支持向量的数目,$b^*$ 为与 $\alpha_i^*$ 对应的阈值,$K(x_i, x) = \varphi(x_i) \cdot \varphi(x)$ 为满足 Mercer 条件的核函数[6]。常用的三种核函数如表 1 所示。

表 1　常用核函数

| Kernel function | Expression |
|---|---|
| Linear kernel | $x_i^T x$ |
| Polynomial kernel | $(1 + x_i^T x)^d$ |
| RBF kernel | $\exp(- \| x - x_i \|^2 / \sigma^2)$ |

## 2　不确定支持向量机

模糊理论与粗糙集方法是两种处理不确定性知识的有效数学工具。针对支持向量机方法对于噪声或孤立野值点敏感的问题,引入模糊理论与粗糙集方法,可以分别得到两种不确定支持向量机模型:模糊支持向量机与粗糙支持向量机。

### 2.1　模糊支持向量机(F-vSVM)

模糊支持向量机的主要构造思想是对分类贡献不同输入样本,分别赋予不同的隶属度。假设给定一组训练样本 $\{(x_1, y_1, s_1), \cdots, (x_l, y_l, s_l)\}$,其中 $s_i, i = 1, \cdots, l$ 就是样本 $x_i$ 属于所在类的隶属度,则模糊支持向量机模型如下所示:

$$\min \frac{1}{2} \| w \|^2 - v\rho + \frac{1}{l} \sum_{i=1}^{l} s_i \xi_i \tag{4}$$

$$\text{s.t } y_i(w \cdot \varphi(x_i) + b) \geqslant \rho - \xi_i \quad \zeta_i \geqslant 0, \rho \geqslant 0, i = 1, \cdots, l$$

注意式(4)中松弛项由传统支持向量机中的单一松弛因子变成了带有不同权重的松弛因子,对式(4)构造拉格朗日函数

$$L(w, b, \xi, \rho, \alpha, \beta, \lambda) = \frac{1}{2} \| w \|^2 - v\rho + \frac{1}{l} \sum_{i=1}^{l} s_i \xi_i - \sum_{i=1}^{l} \alpha_i (y_i(w \cdot z_i + b) - \rho + \xi_i) - \sum_{i=1}^{l} \beta_i \xi_i - \lambda\rho \tag{5}$$

其中 $\alpha_i \geqslant 0, \beta_i \geqslant 0, \lambda \geqslant 0, i = 1, \cdots, l$ 均为拉格朗日乘子,由于在鞍点处的 $w, b, \xi, \rho$ 的偏导数为零,则

$$\frac{\partial L}{\partial w} = w - \sum_{i=1}^{l} \alpha_i y_i \varphi(x_i) = 0 \tag{6}$$

$$\frac{\partial L}{\partial b} = - \sum_{i=1}^{l} \alpha_i y_i = 0 \tag{7}$$

$$\frac{\partial L}{\partial \xi_i} = \frac{1}{l} s_i - \alpha_i - \beta_i = 0 \tag{8}$$

$$\frac{\partial L}{\partial \rho} = - v + \sum_{i=1}^{l} \alpha_i - \lambda = 0 \tag{9}$$

将式(6)到式(9)代入式(5)得到模糊支持向量机的对偶描述:

$$\min(\alpha) = \frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} \alpha_i \alpha_j y_i y_j K(x_i, x_j) \tag{10}$$

$$\text{s.t. } \sum_{i=1}^{l} \alpha_i y_i = 0, 0 \leqslant \alpha_i \leqslant \frac{1}{l} s_i, \sum_{i=1}^{l} \alpha_i \geqslant v,$$

最终决策函数为:

$$f(x) = \text{sgn}\left( \sum_{i=1}^{m} \alpha_i^* y_i K(x_i \cdot x) + b^* \right)$$

在支持向量机训练过程中,噪声和异常样本往往会产生很大的拉格朗日乘子,从而主导了决策函数[7]。而模糊支持向量机正是针对这个问题,对那些噪声或异常样本指定较小的隶属度,抑制噪声对分类器的影响,提高模型的抗干扰能力。

### 2.2　粗糙支持向量机(RS-vSVM)

粗糙支持向量机就是利用粗糙集思想对传统支持向量机中的分类间隔进行拓展,分别定义分类间隔的上近似与下近似,对分布在不同区域的样本采用不同的惩罚[8~12]。拓展的分类间隔如图 1 所示。

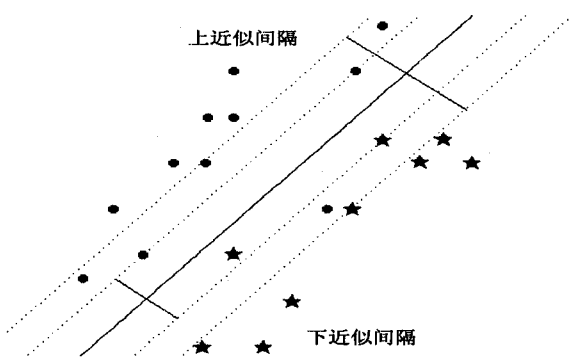


图 1　拓展的分类间隔

与传统支持向量机求解类似,粗糙支持向量机也

是通过最大化分类间隔来得到最优分类面，具体模型如式(11)所示。

$$\min \frac{1}{2}\|w\|^2 - v\rho_l - v\rho_u + \frac{1}{l}\sum_{i=1}^{l}\xi_i + \frac{\delta}{l}\sum_{i=1}^{l}\xi_i' \tag{11}$$

$$\text{s.t. } y_i(w\cdot\varphi(x_i)+b) \geqslant \rho_u - \xi_i - \xi_i'$$

$$0 \leqslant \xi_i \leqslant \rho_u - \rho_l, \xi_i' \geqslant 0, \rho_l \geqslant 0, \rho_u \geqslant 0$$

同理采用拉格朗日乘子法对问题(11)进行求解，

$$L(\cdot) = \frac{1}{2}\|w\|^2 - v\rho_l - v\rho_u + \frac{1}{l}\sum_{i=1}^{l}\xi_i + \frac{\delta}{l}\sum_{i=1}^{l}\xi_i' - \sum_{i=1}^{l}\alpha_i(y_i(w\cdot z_i+b) - \rho_u + \xi_i + \xi_i') - \sum_{i=1}^{l}\beta_i\xi_i - \sum_{i=1}^{l}\lambda_i(\rho_u - \rho_l - \xi_i) - \sum_{i=1}^{l}\eta_i\xi_i' - \mu_1\rho_l - \mu_2\rho_u \tag{12}$$

其中 $\alpha_i \geqslant 0, \beta_i \geqslant 0, \lambda_i \geqslant 0, \eta_i \geqslant 0, \mu_1 \geqslant 0, \mu_2 \geqslant 0$

$$\frac{\partial L}{\partial w} = w - \sum_{i=1}^{l}\alpha_i y_i \varphi(x_i) = 0 \tag{13}$$

$$\frac{\partial L}{\partial b} = -\sum_{i=1}^{l}\alpha_i y_i = 0 \tag{14}$$

$$\frac{\partial L}{\partial \xi_i} = \frac{1}{l} - \alpha_i - \beta_i + \lambda_i = 0 \tag{15}$$

$$\frac{\partial L}{\partial \xi_i'} = \frac{\delta}{l} - \alpha_i - \eta_i = 0 \tag{16}$$

$$\frac{\partial L}{\partial \rho_l} = -v + \sum_{i=1}^{l}\lambda_i - \mu_i = 0 \tag{17}$$

$$\frac{\partial L}{\partial \rho_u} = -v + \sum_{i=1}^{l}\alpha_i - \sum_{i=1}^{l}\lambda_i - \mu_2 = 0 \tag{18}$$

得到粗糙支持向量机的对偶形式：

$$\min(\alpha) = \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i\alpha_j y_i y_j K(x_i, x_j) \tag{19}$$

$$\text{s.t. } \sum_{i=1}^{l}\alpha_i y_i = 0, 0 \leqslant \alpha_i \leqslant \frac{\delta}{l}, \sum_{i=1}^{l}\alpha_i \geqslant 2v$$

相应的决策函数为：

$$f(x) = \text{sgn}\left(\sum_{i=1}^{m}\alpha_i^* y_i K(x_i \cdot x) + b^*\right)$$

粗糙支持向量机也是采用了对分布在不同区域的样本给予不同的惩罚来降低噪声对最优分类面的影响。下面结合KKT条件，给出具体说明：

① 当 $0 \leqslant \alpha_i^* < \frac{1}{l}$，对应样本满足 $y_i(w\cdot\varphi(x_i)+b) \geqslant \rho_u$，这时 $\xi_i$ 与 $\xi_i'$ 都为零，因此不做任何惩罚。

② 当 $\frac{1}{l} \leqslant \alpha_i^* < \frac{\delta}{l}$，由 $\frac{\delta}{l} - \alpha_i^* - \eta_i = 0$ 与 $\xi_i'\eta_i = 0$，得到 $\xi_i' = 0$，对应样本满足 $y_i(w\cdot\varphi(x_i)+b) = \rho_u - \xi_i$。同时，由于 $0 \leqslant \xi_i \leqslant \rho_u - \rho_l$，因此 $\rho_l \leqslant y_i(w\cdot\varphi(x_i)+b) \leqslant \rho_u$，也就是对应样本分布在上近似间隔与下近似间隔之间，对这类样本指定较小的惩罚。

③ 当 $\alpha_i^* = \frac{\delta}{l}$，由 $\frac{1}{l} - \alpha_i^* - \beta_i + \lambda_i = 0$ 可以得到 $\lambda_i > \beta_i \geqslant 0$，又根据 $\lambda_i(\rho_u - \rho_l - \xi_i) = 0$ 得到 $\rho_u - \rho_l = \xi_i$。因此对应样本满足 $y_i(w\cdot\varphi(x_i)+b) = \rho_l - \xi_i'$，分布在下近似间隔以内，对这类样本则指定较大的惩罚。

## 3 实验结果与分析

为了对比两种不确定支持向量机模型的分类性能，文中分别用一个合成数据集和一组标准数据集进行测试。

### 3.1 合成数据集

本部分实验是对分布在二维平面上的一组含有野值的样本进行分类，分别采用 vSVM、RS-vSVM 和 F-vSVM。

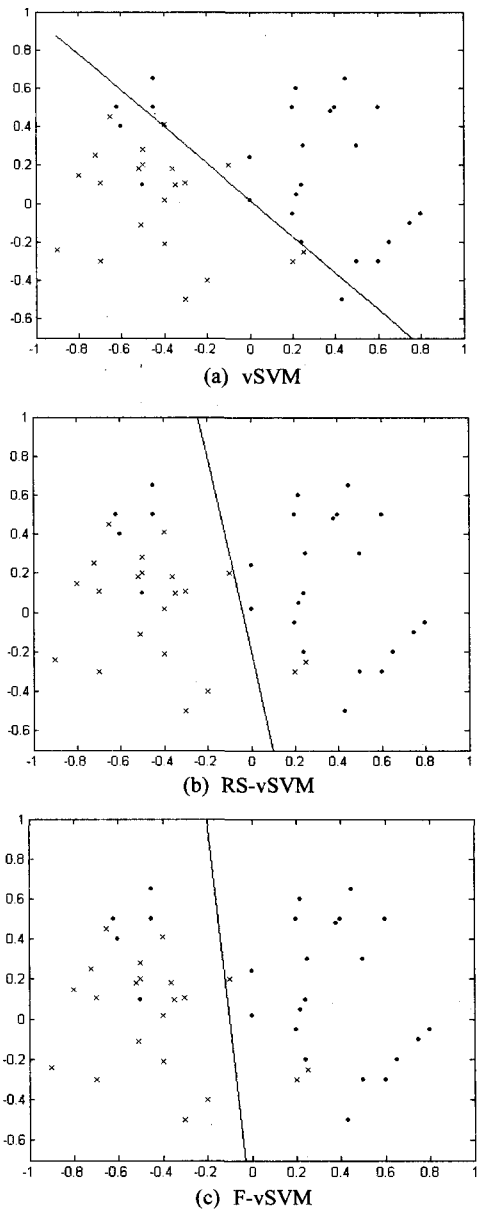


图2 不同模型的分类结果

由图2(a)可以看出，在传统 v 支持向量机中，由于最优分类面受到野值样本的影响，使得推广能力严重下降。而两种不确定支持向量机(b 与 c)能够有效

区分野值样本和正常样本,区分样本进行不同惩罚的方法明显降低了噪声对最优分类面的影响。

### 3.2 标准数据集

本部分实验数据来自 UCI 机器学习数据库,数据集具体描述见表 2。

表 2　数据集信息

| Dataset Name | Samples | Features | Classes |
|---|---|---|---|
| breast | 683 | 9 | 2 |
| heart | 296 | 13 | 2 |
| live | 345 | 6 | 2 |

整个实验过程使用三类核函数,分别为 Linear、Polynomial、RBF,同时为了方便比较分别给定核参数 $d=2$ 和 $\sigma=1$,并给定 $\nu=0.1$。对于模糊 v 支持向量机中隶属度的计算采用文献[4]给出的方法,粗糙 v 支持向量机中参数 $\delta$ 在文献[1,5]上得到,步长设定为 0.5。

实验结果见表 3 ~ 表 5。

表 3　使用核函数为 Linear 的结果

| Dataset | vSVM | RS-vSVM | F-vSVM |
|---|---|---|---|
| breast | 0.9474 | 0.9603 | 0.9649 |
| heart | 0.8108 | 0.8212 | 0.8541 |
| live | 0.6603 | 0.6811 | 0.6999 |

表 4　使用核函数为 Polynomial 的结果

| Dataset | vSVM | RS-vSVM | F-vSVM |
|---|---|---|---|
| breast | 0.9474 | 0.9603 | 0.9649 |
| heart | 0.7432 | 0.7432 | 0.7802 |
| live | 0.6603 | 0.6820 | 0.7011 |

表 5　使用核函数为 RBF 的结果

| Dataset | vSVM | RS-vSVM | F-vSVM |
|---|---|---|---|
| breast | 0.9676 | 0.9691 | 0.9711 |
| heart | 0.8370 | 0.8481 | 0.8427 |
| live | 0.6686 | 0.6829 | 0.7011 |

从表 3 的实验结果可以看出,两种不确定支持向量机方法在 3 个数据集上的分类性能得到了不同程度的提升。因此可以得出结论,对样本采用差异化惩罚的学习策略,可以有效降低噪声对最优分类面的影响。模糊支持向量机整体表现出更好的泛化性能。

## 4　结束语

文中分析和比较了模糊支持向量机和粗糙支持向量机分类模型的构造方法,同时通过一个合成数据集和一组标准数据集对这两种不确定支持向量机的泛化性能进行了对比验证。实验结果表明,相比传统支持向量机方法,这两种不确定支持向量机方法的分类性能都得到了不同程度的提升,而模糊支持向量机表现出了更好的泛化性能。

**参考文献:**


[1] 李国正,王　猛. 支持向量机导论[M]. 北京:电子工业出版社,2005.

[2] 李红莲,王春花,袁保宗. 一种改进的支持向量机 NN-SVM[J]. 计算机学报,2003, 26(8):1015-1019.

[3] Ke Haixin, Zhang Xuegong. Editing support vector machines[C]//In: Proceedings of International Joint Conference on Neural Networks. Washington, USA: [s. n.], 2001: 1464-1467.

[4] Lin C F, Wang S D. Fuzzy support vector machines[J]. IEEE Trans on Neural Networks, 2002, 13(2): 464-471.

[5] Zhang Junhua, Wang Yuanyuan. A rough margin based support vector machine[J]. Information Science, 2008, 178: 2204-2214.

[6] Hsu Chih-Wei, Chang Chih-Chung, Lin Chih-Jen. A Practical Guide to Support Vector Classification[EB/OL]. [2003-07]. http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf.

[7] 王　凯,张永祥,姚晓山,等. 支持向量机惩罚参数的自适应调整方法[J]. 计算机工程与应用,2008,44(26):45-47.

[8] Asharaf S, Shevade S K, Narasimha M. Rough support vector clustering[J]. Pattern Recognition, 2005, 38: 1779-1783.

[9] 汤义强,毛军军,李　侠,等. 基于粗集属性约简的电力供应量 SVM 回归预测[J]. 计算机技术与发展,2010,20(9):48-51.

[10] 冯利军,李书全,宋连友. 利用粗糙集理论提高 SVM 预测系统的实时性[J]. 计算机技术与发展,2006,16(9):30-34.

[11] 张　芬,陶　亮,孙　艳. 基于混合核函数的 SVM 及其应用[J]. 计算机技术与发展,2006,16(2):176-178.

[12] 陶秀凤,周鸣争. 基于支持向量机的多传感器信息融合算法[J]. 计算机技术与发展,2006,16(6):177-183.