紧密类超带模糊支持向量机

张亚普 孟相如 赵卫虎 张 立
(空军工程大学电讯工程学院 西安 710077)

摘 要 提出一种紧密类超带模糊支持向量机(Affinity Class-Hyperparallel Fuzzy Support Vector Machine,ACHFSVM),其以获得较好的抗噪性和泛化能力。该方法在摒弃样本集球形分布假设的同时,纳入对样本紧密度的考量,用类内超平面取代类中心,通过二次规划的方法在特征空间中寻找最小类超带,以其带宽表征样本紧密度,构造S型隶属度函数。基于UCI数据集的仿真结果表明该方法较同类算法具有更好的抗噪和分类性能。

关键词 模糊支持向量机,紧密度,模糊隶属度,分类
中图法分类号 TP181 **文献标识码** A

Affinity Class-hyperparallel Fuzzy Support Vector Machine

ZHANG Ya-pu MENG Xiang-ru ZHAO Wei-hu ZHANG Li
(Telecommunication Engineering Institue, Air Force Engineering University, Xi'an 710077, China)

Abstract An Affinity Class-Hyperparallel Fuzzy Support Vector Machine was proposed to get better classification result. This method not only takes the advantage of the affinity, but also abandons the estimation that the samples obey spherical-shape distribution. Instead of the cluster center, a hyperplane within the class is used to find a hyperparallel with the minimum distance while containing the maximum samples by the way of quadratic programming. The membership is achieved through a new S-function based on the distance of the hyperparallel which reflects the affinity of the samples. The simulation on UCI shows that the ACHFSVM is more robust and has better classification accuracy.

Keywords Fuzzy support vector machine, Affinity, Fuzzy membership, Classification

1 引言

支持向量机(Support Vector Machine,SVM)[1]是Vapnik于20世纪90年代在统计学习理论基础上提出的一种机器学习方法,它较好地解决了小样本、高维非线性数据的优化问题,最小化结构风险的引入提高了模型的泛化能力。在标准SVM中,所有训练样本对分类面的贡献相同,若训练样本中含有噪声或野值,分类面的选取将偏离最优解。针对上述情况,文献[2]提出模糊支持向量机(Fuzzy Support Vector Machine,FSVM),其在构造目标函数时,利用惩罚系数区分样本的不同贡献,通过对噪声或野值样本赋予较小的权值来减小影响,得到了较好的分类结果。

隶属度函数的设计是FSVM的关键。近年来随着对FSVM研究的深入,提出了许多新的隶属度设计方法。文献[3]针对非球形分布样本集,提出一种基于样本到类内超平面距离的隶属度设计方法,从而降低了隶属度对样本集几何形状的依赖,提高了FSVM的泛化能力;文献[4]在样本到类中心距离的基础上,考虑了类内样本间的紧密度,改善了FSVM的抗造性和分类能力;文献[5]用K近邻法对样本间关系进行考量,提出了一种新的基于紧密度的FSVM,从而简化了隶属度计算,提高了算法效率。上述隶属度设计方法不同程度地改善了FSVM性能,但如何客观准确地反映样本集实际情况,进一步提高分类效果,仍是难点问题。本文提出了一种紧密类超带模糊支持向量机,兼顾样本集分布和紧密度等方面设计出了更为合理的隶属度函数,并在UCI数据集上验证了该方法的有效性。

2 紧密类超带隶属度函数设计

2.1 基本设计思想

确定SVM最优分类面的支持向量位于类边缘,野值或噪声常常也位于类边缘,如果无法将有效样本与野值或噪声进行正确的区分,则求出的分类面不是真正的最优分类面。FSVM通过隶属度函数给出合理的惩罚系数,对样本在分类面选取中的贡献加以区分。传统方法以样本到类中心的距离来度量其贡献大小,Zadeh[6]定义的S型函数将其表述为样本与类中心距离的一种非线性关系。文献[4]基于紧密度对FSVM展开研究,以特征空间中包含样本的最小超球半径度量样本紧密度,并结合样本到类中心距离设计隶属度函数。

由图1可知,点a,b,c到分类面距离相等,故其对分类面的确定贡献相同,实际隶属度应满足:$S_a=S_b=S_c$,而文献[4]的隶属度设计隐含了样本球形分布假设,当样本实际分布非球形时,会得出与实际偏差较大的隶属度值:$S_a>S_b>S_c$。

到稿日期:2010-07-25 返修日期:2010-11-23 本文受陕西省自然科学基金项目(SJ08F14,2009JQ8008)资助。
张亚普(1985—),男,硕士,主要研究方向为宽带通信网络技术,E-mail:zhi_shui_gu_mei@163.com;孟相如 教授,博士生导师;赵卫虎 硕士;张 立 博士。

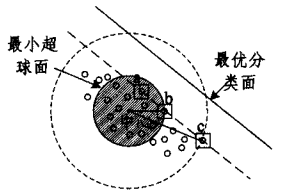

图 1　基于球形分布假设的隶属度示意

针对上述不足，本文以类内超平面取代类中心，以最小类超带带宽表征紧密度，设计了一种新的隶属度函数，从而在加入紧密度的同时，减小了其对样本几何分布的依赖。该方法的核心思想是：在特征空间中寻找一条包围样本集的最小类超带，当样本集中不存在噪声或野值样本时，则寻找一个能够包围所有样本的最小类超带（类超带带宽最小）；当样本集中含有噪声或野值样本时，可以允许一小部分样本位于带外，寻找一个能够包围尽可能多样本的最小类超带，如图 2 所示。

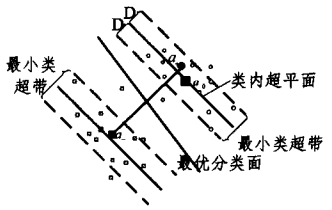

图 2　最小类超带示意

2.2　最小类超带确定

最优分类面存在于两类样本之间，故以最优分类面的法向量为类超带的法向量能客观反映样本对分类面的贡献特征。在最优分类面未知的情况下，可以以两类样本几何中心连线来近似该法向量，进而寻找最小类超带。如图 2 所示，以正、负类样本均值点表征样本集的几何中心 a_+，a_-，则法向量 $\omega=a_+-a_-$。

以正类为例，设 a_0 在类内超平面上，其到类超带两边界的距离相等，则类内超平面可表示为：

$$\omega^T(x-a_0)=0 \tag{1}$$

任意样本点到超平面距离为：

$$d_i=\frac{|\omega^T(x_i-a_0)|}{\|\omega\|} \tag{2}$$

由于噪声和野值的存在，引入松弛变量 $\xi_i\geqslant 0(i=1,2,\cdots,l)$，从而允许一部分样本位于类超带外侧，则最小类超带的求解即为如下最优化问题：

$$\begin{aligned}&\min\ D^2+C\sum_{i=1}^{l}\xi_i\\&\text{s. t. }\ d_i^2\leqslant D^2+\xi_i\\&\qquad\xi_i\geqslant 0\end{aligned} \tag{3}$$

式中，D 为包围样本集的最小类超带带宽的一半，惩罚因子 C 用以调整类超带宽度与允许带外存在样本的数量。C 越大，惩罚越大，允许的带外样本就越少。为求式(3)，定义 Lagrange 函数：

$$L(D,a_0,\xi_i,\alpha,\beta)=D^2+C\sum_{i=1}^{l}\xi_i-\sum_{i=1}^{l}\alpha_i(D^2+\xi_i-d_i^2)-\sum_{i=1}^{l}\beta_i\xi_i \tag{4}$$

式中，$\alpha_i\geqslant 0,\beta_i\geqslant 0$ 均为 Lagrange 系数。

由 Karush-Kuhn-Tucker 定理可得：

$$\sum_{i=1}^{l}\alpha_i=1 \tag{5}$$

$$a_0=\sum_{i=1}^{l}\alpha_i x_i \tag{6}$$

$$C-\alpha_i-\beta_i=0 \tag{7}$$

对式(4)进行展开合并得：

$$L(D,a_0,\xi_i,\alpha,\beta)=D^2(1-\sum_{i=1}^{l}\alpha_i)+\sum_{i=1}^{l}\xi_i(C-\alpha_i-\beta_i)+\sum_{i=1}^{l}\alpha_i d_i^2 \tag{8}$$

将式(5)、式(7)带入式(8)化简整理：

$$\begin{aligned}L(D,a_0,\xi_i,\alpha,\beta)&=\sum_{i=1}^{l}\alpha_i d_i^2=\sum_{i=1}^{l}\alpha_i\left(\frac{|\omega^T(x_i-a_0)|}{|\omega|}\right)^2\\&=\sum_{i=1}^{l}\alpha_i(x_i,x_i)-2a_0\sum_{i=1}^{l}\alpha_i x_i+a_0^2\sum_{i=1}^{l}\alpha_i\end{aligned} \tag{9}$$

另对式(6)做如下变换：

$$a_0\sum_{i=1}^{l}\alpha_i=\sum_{i=1}^{l}\alpha_i x_i \tag{10}$$

将式(10)带入式(9)，得二次规划问题：

$$\begin{aligned}&\max\ \sum_{i=1}^{l}\alpha_i(x_i,x_i)-\sum_{i,j=1}^{l}\alpha_i\alpha_j(x_i,x_j)\\&\text{s. t. }\ \sum_{i=1}^{l}\alpha_i=1\\&\qquad 0\leqslant\alpha_i\leqslant C,i=1,\cdots,l\end{aligned} \tag{11}$$

由式(11)可求出样本所对应的 Lagrange 系数 $\alpha_i(i=1,2,\cdots,l)$，另由式(6)可知，最小类超带中心为所有样本的线性组合。当 $\alpha_i>0$ 时，对应的样本称为类边界样本；当 $\alpha_i=C$ 时，对应的样本位于带外，视为野值或噪声；当 $0<\alpha_i<C$ 时，对应的样本位于类超带边界内侧附近，可用来确定类超带。因此，以 $0<\alpha_i<C$ 对应的样本与类内超平面之间的距离均值来确定最小类超带的带宽：

$$D=E\left(\frac{|\omega^T(x_i-a_0)|}{\|\omega\|}\right) \tag{12}$$

式中，x_i 为 $0<\alpha_i<C$ 对应的所有类边界样本。

当样本非线性可分时，引入映射 $\phi(x)$ 将原始空间样本映射到特征空间，则利用核技巧将原始空间中的最小类超带寻优式(11)表述为：

$$\begin{aligned}&\max\ \sum_{i=1}^{l}\alpha_i K(x_i,x_i)-\sum_{i,j=1}^{l}\alpha_i\alpha_j K(x_i,x_j)\\&\text{s. t. }\ \sum_{i=1}^{l}\alpha_i=1\\&\qquad 0\leqslant\alpha_i\leqslant C,i=1,\cdots,l\end{aligned} \tag{13}$$

特征空间中两类样本的中心转化为：$a_+=\frac{1}{l_+}\sum_{i=1}^{l_+}\phi(x_i)$ 和 $a_-=\frac{1}{l_-}\sum_{i=1}^{l_-}\phi(x_i)$，其中 l_+，l_- 分别为正负样本的数量；最小类超带带宽表述形式不变，如式(12)所示，则特征空间中对其求解如下：

$$\begin{aligned}\omega^T(\phi(x_i)-a_0)&=\left(\frac{1}{l_+}\sum_{j=1}^{l_+}\phi(x_j)-\frac{1}{l_-}\sum_{j=1}^{l_-}\phi(x_j)\right)^T\cdot\left(\phi(x_i)-\sum_{\kappa=1}^{l_+}\alpha_\kappa\phi(x_\kappa)\right)\\&=\frac{1}{l_+}\sum_{j=1}^{l_+}\phi(x_j)\cdot\phi(x_i)-\frac{1}{l_+}\sum_{j=1}^{l_+}\phi(x_j)\cdot\sum_{\kappa=1}^{l_+}\alpha_\kappa\phi(x_\kappa)-\frac{1}{l_-}\sum_{j=1}^{l_-}\phi(x_j)\cdot\phi(x_i)+\frac{1}{l_-}\sum_{j=1}^{l_-}\phi(x_j)\cdot\sum_{\kappa=1}^{l_+}\alpha_\kappa\phi(x_\kappa)\\&=\frac{1}{l_+}\sum_{j=1}^{l_+}K(x_j\cdot x_i)-\frac{1}{l_+}\sum_{j=1}^{l_+}\sum_{\kappa=1}^{l_+}\alpha_\kappa K(x_\kappa\cdot x_j)-\frac{1}{l_-}\sum_{j=1}^{l_-}K(x_j\cdot x_i)+\frac{1}{l_-}\sum_{j=1}^{l_-}\sum_{\kappa=1}^{l_+}\alpha_\kappa K(x_\kappa\cdot x_j)\end{aligned}$$

$$\|\omega\|=\sqrt{\omega^T\omega}=[(\frac{1}{l_+}\sum_{i=1}^{l_+}\phi(x_i)\frac{1}{l_-}\sum_{i=1}^{l_-}\phi(x_i))^T\cdot(\frac{1}{l_+}\sum_{i=1}^{l_+}\phi(x_i)-\frac{1}{l_-}\sum_{i=1}^{l_-}\phi(x_i))]^{\frac{1}{2}}$$

$$=[\frac{1}{l_+}\sum_{i=1}^{l_+}\phi(x_i)\cdot\frac{1}{l_+}\sum_{j=1}^{l_+}\phi(x_j)-\frac{1}{l_+}\sum_{i=1}^{l_+}\phi(x_i)\cdot\frac{1}{l_-}\sum_{j=1}^{l_-}\phi(x_j)-\frac{1}{l_-}\sum_{i=1}^{l_-}\phi(x_i)\cdot\frac{1}{l_+}\sum_{j=1}^{l_+}\phi(x_j)+\frac{1}{l_-}\sum_{i=1}^{l_-}\phi(x_i)\cdot\frac{1}{l_-}\sum_{j=1}^{l_-}\phi(x_j)]^{\frac{1}{2}}$$

$$=[\frac{1}{l_+^2}\sum_{i=1}^{l_+}\sum_{j=1}^{l_+}K(x_i\cdot x_j)-\frac{2}{l_+\times l_-}\sum_{i=1}^{l_+}\sum_{j=1}^{l_-}K(x_i\cdot x_j)+\frac{1}{l_-^2}\sum_{i=1}^{l_-}\sum_{j=1}^{l_-}K(x_i\cdot x_j)]^{\frac{1}{2}}$$

则特征空间中最小类超带带宽为：

$$D=E(\frac{|\omega^T(x_i-a_0)|}{\|\omega\|})=E\left(\frac{\left|\frac{1}{l_+}\sum_{j=1}^{l_+}K(x_j\cdot x_i)-\frac{1}{l_+}\sum_{j=1}^{l_+}\sum_{k=1}^{l_+}\alpha_kK(x_k\cdot x_j)-\frac{1}{l_-}\sum_{j=1}^{l_-}K(x_j\cdot x_i)+\frac{1}{l_-}\sum_{j=1}^{l_-}\sum_{k=1}^{l_+}\alpha_kK(x_k\cdot x_j)\right|}{\sqrt{\frac{1}{l_+^2}\sum_{i=1}^{l_+}\sum_{j=1}^{l_+}K(x_i\cdot x_j)-\frac{2}{l_+\times l_-}\sum_{i=1}^{l_+}\sum_{j=1}^{l_-}K(x_i\cdot x_j)+\frac{1}{l_-^2}\sum_{i=1}^{l_-}\sum_{j=1}^{l_-}K(x_i\cdot x_j)}}\right)\tag{14}$$

注：隶属度函数与SVM训练函数应选取相同的核函数，以保证特征空间的统一。

2.3 紧密类超带隶属度

本文利用最小类超带带宽设计基于样本紧密度的S型隶属度函数。为使此隶属度函数能客观反映样本的不确定性，首先对S型函数做如下设计：前半段为凹函数，以减小隶属度变化率，使边界样本仍具有较大隶属度；后半段为凸函数，使隶属度在远离类内超平面时急剧下降，从而使同样远离类内超平面的边界样本和野值有了较大的区分，如式(15)所示。

$$S_i(d_i,a,b,c)=\begin{cases}1, & d_i\leqslant a\\ 1-\dfrac{25(d_i-a)^2}{12(c-a)^2}, & a\leqslant d_i\leqslant b\\ \dfrac{50(d_i-c)^2}{27(c-a)^2}, & b\leqslant d_i\leqslant c\\ 0, & d_i\geqslant c\end{cases}\tag{15}$$

式中，$d_i=\dfrac{|\omega^T(\phi(x_i)-a_0)|}{\|\omega\|}$，为任意样本点到类内超平面的距离。

当边界样本与类内超平面距离较小时，无论 b 值如何，边界样本都具有较大隶属度。当此距离较大时，若 b 值较小，边界样本隶属度可能由后段S型函数末端决定，被赋予很小值；相反，若 b 值较大，则被赋较大值。综上，本文以最小类超带带宽的一半为最佳 b 值，从而对位于最小类超带边界附近的正常样本和噪声加以有效区分。参数 a,b,c 满足下式：

$$\begin{cases}a=\dfrac{2}{3}D\\ b=D\\ c=\dfrac{3}{2}D\end{cases}\tag{16}$$

S型隶属度函数与样本分布的关系如图3所示。

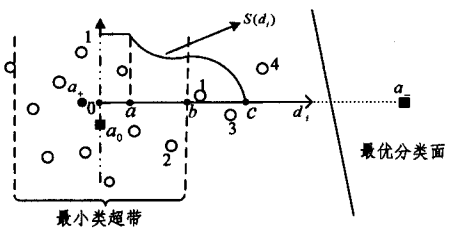

图3 S型隶属度函数示意

由图3中样本分布可知，样本点1、2在最小类超带的边界附近，故其属于正常样本且对最优分类面的确定作用较大，应赋予较大的隶属度值；点3、4虽距最优分类面较近，但远离最小类超带，故其为噪声或野值的可能性很大，应赋予很小隶属度值。观察图3中隶属度曲线 $S(d_i)$，易知其较好地反映了上述分析，有效区分了类超带边界附近的正常样本和野值(或噪声)。

3 紧密类超带模糊支持向量机

依据上文设计的隶属度函数构造紧密类超带模糊支持向量机。设原始训练样本集为 $T=\{(x_i,y_i)\mid(x_i,y_i)\in R^N\times\{\pm1\},i=1,\cdots,l\}$，特征空间映射为 $\phi(x)$，由式(15)得第 i 个样本的模糊隶属度 $S_i(0<S_i\leqslant1,i=1,2,\cdots,l)$，则样本集变为 $T'=\{(\phi(x_1),y_1,S_1),(\phi(x_2),y_2,S_2),\cdots,(\phi(x_l),y_l,S_l)\}$，ACHFSVM表述为：

$$\begin{aligned}&\min\ \frac{1}{2}\|\omega\|^2+C\sum_{i=1}^{l}S_i\xi_i\\ &\text{s.t. }\ y_i[\omega\cdot\phi(x_i)+b]\geqslant1-\xi_i\\ &\quad\xi_i\geqslant0,i=1,2,\cdots,l\end{aligned}\tag{17}$$

式中，CS_i 表示对训练样本错分程度的惩罚，使分类间隔与错分率之间达到一种平衡。

根据Karush-Kuhn-Tucker定理和拉格朗日乘数法，式(17)的对偶问题为：

$$\begin{aligned}&\max\ Q(\alpha)=\sum_{i=1}^{l}\alpha_i-\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i\alpha_jy_iy_jK(x_i,x_j)\\ &\text{s.t. }\ \sum_{i=1}^{l}\alpha_iy_i=0\\ &\quad0\leqslant\alpha_i\leqslant S_iC,i=1,2,\cdots,l\end{aligned}\tag{18}$$

式中，$K(x_i,x_j)=\phi(x_i)\cdot\phi(x_j)$ 为核函数。

可见，ACHFSVM与C-SVM的区别仅在于对样本 S_i 的赋值不同，当对于任意样本满足 $S_i=1$ 时，ACHFSVM便转化为C-SVM。在C-SVM中参数 C 是一个自定义的惩罚因子，它控制对错分样本惩罚的程度，寻求样本错分率与模型泛化性的平衡。C 越大，对错分样本的惩罚程度越大，得到分类面的间隔越小，模型泛化性越差；反之，会使分类面间隔变大，错分率提高，分类结果较差。显然，C-SVM对所有样本的错分均以 C 进行惩罚，缺少灵活性。由式(17)可知，ACHFSVM对错分样本的惩罚为 CS_i，因此对于噪声点可通过减小其 S_i 值来达到降低错分惩罚的目的，减小其对目标函数的贡献，对非噪声点，则赋予较大 S_i，从而克服了C-SVM错分惩罚灵活性不够的缺点，实现了错分惩罚的自适应，得到了更好的分类结果。

当 $\alpha_i>0$ 时，相应的样本 x_i 为支持向量。若 $0\leqslant\alpha_i\leqslant S_iC$，支持向量 x_i 位于分类面附近；若 $\alpha_i=S_iC$，支持向量 x_i 为错误分类样本。若噪声点位于分类面附近，C-SVM则要寻找对

噪声点也能尽量正确分类的最优分类面；而 ACHFSVM 则通过给噪声点赋予较小的 S_i，使得 $\alpha_i = S_i C$，以错分噪声点为代价，寻找最优分类面。噪声点在分类面确定时参考价值不高，过分重视对它的分类正确率会影响模型的泛化性，因此，通过对噪声的模糊处理，ACHFSVM 能得到更具泛化性的分类模型。

4 仿真实验分析

4.1 Abalone 数值实验

选取 UCI[7] 数据库中 abalone 9 类(abalone-9)和 10 类(abalone-10)样本的第 3 维和第 6 维数据组成二维实验数据集，将实验数据集分为两组：训练数据集和测试数据集，如表 1 所列。

表 1 实验数据集结构

数据集 \ 数据类	abalone-9/个	abalone-10/个
实验数据集	240	260
训练数据集	100(含 10 个噪声点)	100
测试数据集	140	160

以该实验数据集为对象，比较 C-SVM，HFSVM[3]，AFSVM[4] 和 ACHFSVM 的抗噪性和分类性能。

如图 4 所示，10 个 abalone-9 噪声点处在类边缘且靠近 abalone-10 样本，图 4(a)中噪声点分布于 C-SVM 分类面周围，而其它 3 个图分类面较图 4(a)向左偏移至噪声区域边缘附近。由于 C-SVM 未对噪声进行处理，且噪声点靠近 abalone-10 样本，对其错分将产生较大代价 $C\xi_i$，从而使目标函数 $\frac{1}{2}\|\omega\|^2+C\sum_{i=1}^{l}\xi_i$ 寻优时，倾向于使错分代价最小；其它 3 种方法利用隶属度 S_i 对噪声点和正常点加以区分，较小的 S_i 能降低对噪声的错分代价 $CS_i\xi_i$，从而使目标函数 $\frac{1}{2}\|\omega\|^2+C\sum_{i=1}^{l}S_i\xi_i$ 寻优时，实现最大化最小间隔和最小化错分代价的平衡。分析可知：隶属度函数通过影响错分代价来干预分类面的选取。

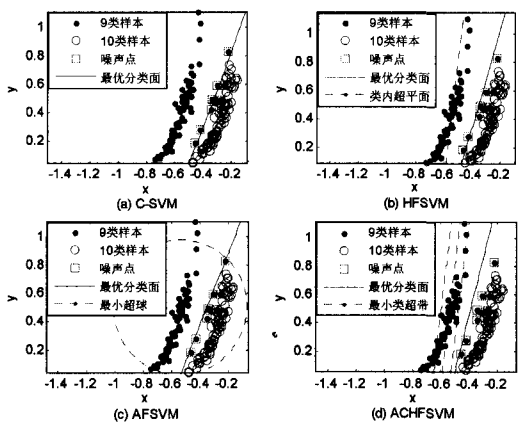

图 4 训练数据集实验结果

图 4(b)，(c)和(d)的分类面依次向左偏移，说明噪声对目标函数的贡献越来越小，因此其倾向于寻找能对正常样本正确分类的最优分类面。图中虚线分别为各自隶属度函数的示意，在 HFSVM 中，10 个噪声点均被赋予较大的 S_i，且与正常点区分不明显；AFSVM 基于类中心的隶属度设计只对分布于超球边缘附近的噪声点赋予较小的 S_i，仍无法有效区分超球内的噪声；在 ACHFSVM 中，噪声点远离类超带，S 型隶属度函数对噪声点赋予极小的 S_i，有效识别了噪声点。可见，紧密类超带隶属度函数噪声识别能力最强，故 ACHFSVM 目标函数受噪声影响最小。

对未知数据的正确预测来体现分类模型泛化性，以图 4 所得的 4 个分类模型对测试数据进行实验，结果如图 5 所示。由图 5 可知，ACHFSVM 得到了最佳的分类结果。因为在模型训练过程中，ACHFSVM 的隶属度函数对噪声进行了有效识别，目标函数以保证对正常样本的分类正确率为目标来寻优分类面，所以得到了泛化性较好的分类模型。可见，紧密类超带隶属度函数有效地提高了 ACHFSVM 分类模型泛化性。

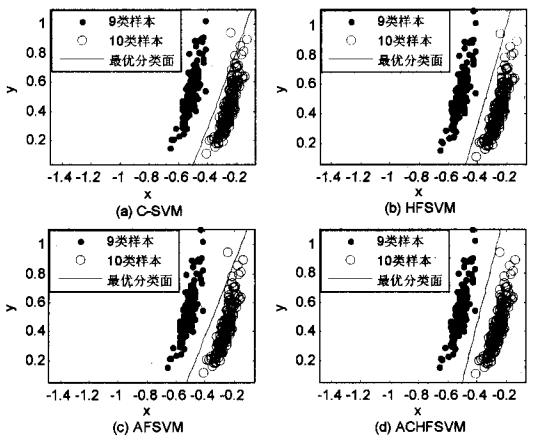

图 5 测试数据集实验结果

4.2 UCI 数据集实验仿真

在 UCI 机器学习数据库中选择 4 个数据集：glass，mushroom，splice，german，其结构特征如表 2 所列。

表 2 UCI 数据集结构特征

数据集	样本个数	样本维数	类别种类
Glass	214	10	7
Mushroom	8124	22	2
Splice	3190	62	3
German	1000	20	2

为了验证 ACHFSVM 对含噪声数据的适应性，分别在每个数据集中加入一定比例的噪声，使每个纯净的 UCI 数据集派生出 5 个噪声数据集(噪声比例以 5%递增)。实验中核函数取径向基核，并使用十折交叉法确定参数，实验结果如图 6 所示。

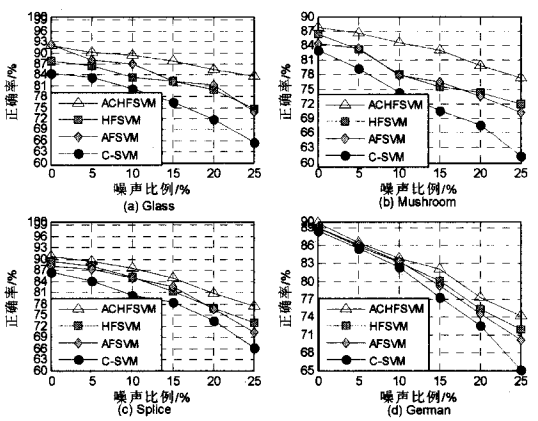

图 6 正确率对比图

(下转第 278 页)

表 2　不同比特率下压缩算法重建性能

	分组 1	分组 2	分组 3	分组 4	分组 5	分组 6
预测参考帧	22 波段	78 波段	111 波段	146 波段	163 波段	179 波段
0.1bpp	23.02dB	21.59dB	40.43dB	35.19dB	50.30dB	36.85dB
0.25bpp	38.78dB	25.14dB	55.87dB	43.76dB	43.68dB	41.81dB
0.5bpp	33.73dB	28.38dB	50.12dB	41.88dB	49.48dB	49.41dB
1.0bpp	36.11dB	34.39dB	59.25dB	45.27dB	57.05dB	51.08dB
1.5bpp	41.29dB	36.44dB	61.48dB	49.81dB	59.30dB	54.52dB

表 3　重建图像性能比较($\overline{PSNR}$)

	0.1bpppb	0.25bpppb	0.5bpppb	1bpppb	1.5bpppb
本文算法	**34.56dB**	**39.84dB**	**42.17dB**	**47.19dB**	**50.47dB**
3D-SPIHT	32.96dB	37.45dB	41.78dB	45.12dB	48.31dB
算法[9]	33.85dB	38.37dB	41.26dB	46.56dB	49.27dB

结束语　实验表明，本文提出的结合自适应波段分组与码率预分配的高光谱图像压缩算法依据高光谱图像谱间相关性的不同，利用吸引力传播聚类算法进行自适应波段分组，继而对不同分组内的高光谱图像采用分段预测算法去除谱间冗余，同时根据预测残差信息量的大小对空间压缩算法进行自适应码率分配。在可接受的计算复杂度下，算法保证了高光谱图像的重建质量。在不同比特率下的测试结果表明其压缩质量优于对比算法，具有一定的应用前景。

参 考 文 献

[1] Pearlman W A. A new fast and efficient image codec based on set partitioning in hierarchical trees [J]. IEEE Trans on Circuits system Video Technonlogy, 1996, 5(9): 243-250

[2] Dragotti P L, Poggi G, Ragozini A R P. Compression of Multispectral Mages by Three-dimensional SPIHT Algorithm [J]. IEEE Transaction on Geoscience and Remote Sensing, 2000, 38(1): 416-428

[3] 吴铮，何明一. 基于小波变换和分段 DPCM 混合编码的多光谱遥感图像压缩算法[J]. 电子与信息学报，2003，25(6)：747-250

[4] Frey F, Dueck D. Mixture modeling by affinity propagation[J]. Neural Information Processing Systems, 2005, 18: 379-386

[5] Frey B J, Dueck D. Clustering by passing messages between data points [J]. Science, 2007: 972-976

[6] He Ming-yi, Bai Lin, Dai Yu-chao, et al. Hyperspectral Image Lossless Compression Algorithm Based on Adaptive Band Regrouping[C]// Proc. SPIE Volume 7455 Satellite Data Compression, Communication and Processing V. San Diego (USA), 2009

[7] CCSDS. Lossless Data Compression[S]. CCSDS 121. 0-B-1 Blue Book, 1997

[8] CCSDS. Image Data Compression[S]. CCSDS 122. 0-B-1 Blue Book, 2005

[9] 孙蕾，罗建书，谷德峰. 基于谱间预测和码流预分配的高光谱图像压缩算法[J]. 光学精密工程，2008，16(4)

(上接第 254 页)

由图 6 可知，各正确率曲线的最高点都在 0 处，最低点都在 25 处，且曲线随噪声比例的增加呈递降趋势。噪声是上述现象的主要原因，当噪声比例为 0%时，所有样本均正常，有足够的点供 SVM 训练从而得到正确的分类面；当噪声比例逐渐增加时，正常样本减少，样本能提供的正确分类信息减少，而干扰信息随噪声比例的增加而增多，从而使分类效果逐渐下降，故正确率曲线都在 25%处得到最低值。由此可知，噪声的存在将导致 SVM 分类效果下降。

观察图 6 发现，从 C-SVM，HFSVM，AFSVM 到 ACHFSVM 曲线的降速逐渐变缓，且 ACHFSVM 的曲线高于其它 3 条。由于 C-SVM 未对噪声点进行处理，其训练不仅受到正常样本减少的影响，而且无法排除噪声点的干扰信息，故其受噪声影响最大，曲线下降速度最快；其它 3 条曲线的结果主要受各自隶属度函数设计好坏的影响，ACHFSVM 的隶属度函数准确反应了样本的不确定性，从而有效地排除了噪声点的干扰，提高了分类效果，故曲线降速最缓。可见，紧密类超带隶属度函数能有效减弱噪声对分类效果的影响。

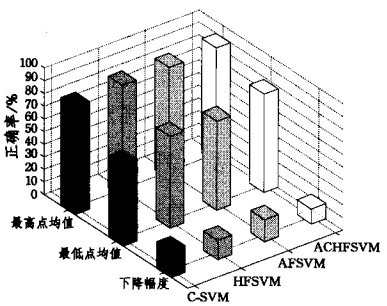

图 7　柱状分析图

下面对正确率曲线进行进一步量化分析，分别计算各曲线的最高点均值(0 处均值)、最低点均值(25 处均值)和下降幅度(最高点均值与最低点均值之差)，并绘制柱状分析图，如图 7 所示。

易知，ACHFSVM 的分类正确率在 0%和 25%处的均值都是最高的，而正确率下降幅度最小。可见，紧密类超带隶属度函数不仅给噪声点赋予较小的 S_i，还给正常点赋予了较大的 S_i，加大了正常样本和噪声的相对区分度，使 ACHFSVM 得到了较好的分类效果。

结束语　针对传统隶属度函数设计中存在的问题，本文提出了一种紧密类超带模糊支持向量机，其在特征空间中引入最小类超带描述样本集的紧密度信息，并在其带宽的基础上构造 S 型隶属度函数，减小了函数对样本几何分布的依赖，有效提高了正常样本和噪声的相对区分度。基于 UCI 数据的仿真表明，该方法明显改善了 FSVM 的抗噪性和泛化性。

参 考 文 献

[1] Cortes C, Vapnik V. Support Vector Networks [J]. Machine Learning, 1995, 20: 273-297

[2] Lin C F, Wan Sh D. Fuzzy support vector machines[J]. IEEE Transactions on Neural Networks, 2002, 13(2): 464-471

[3] 杜喆，刘三阳，齐小刚. 一种新隶属度函数的模糊支持向量机[J]. 系统仿真学报，2009，21(7)：1901-1903

[4] 张翔，肖小玲，徐光祐. 基于样本之间紧密度的模糊支持向量机方法[J]. 软件学报，2006，17(5)：951-958

[5] 唐浩，廖与禾，孙峰. 具有模糊隶属度的模糊支持向量机算法[J]. 西安交通大学学报，2009，43(7)：40-43

[6] 边肇祺，张学工. 模式识别[M]. 北京：清华大学出版社，2000：135-136

[7] UC Irvine Machine Learning Repository[OL]. http: // www. ics. uci. edu/~mlearn/ML Repository. html