文章编号 1004-924X(2012)01-0140-08

基于边界的最大间隔模糊分类器

刘忠宝[1,2*],王士同[1]
(1. 江南大学 数字媒体学院,江苏 无锡 214122;
2. 山西大学商务学院 信息学院,山西 太原 030031)

摘要:对利用超平面、超(椭)球等几何形状实现数据分类的基于边界的主流分类方法进行了研究,在此基础上,提出了一种将空间点作为分类依据的最大间隔模糊分类器(MFC)。该方法首先在模式空间中找到一个模糊分类点 c,c 点距离两类样本要尽可能近且类间夹角间隔尽可能大。然后,测试样本通过 c 与训练样本间的最大化夹角间隔实现分类。最后,利用 MFC 的核化对偶式与最小包含球(MEB)的等价性,将 MFC 的应用范围从二类推广到单类。与主流分类方法的比较实验表明,MFC 具有优良的分类性能和抗噪能力,其分类最高精度可达 98.8%。

关 键 词:模式分类;模糊分类器;模糊分类点;抗噪能力;单类问题
中图分类号:TP391.4 **文献标识码**:A **doi**:10.3788/OPE.20122001.0140

Maximum-margin fuzzy classifier based on boundary

LIU Zhong-bao[1,2*], WANG Shi-tong[1]

(1. *School of Digital Media, Jiangnan Univerisity, Wuxi* 214122, *China*;
2. *School of Information, Business College of Shanxi University, Taiyuan* 030031, *China*)
* *Corresponding author, E-mail: liu_zhongbao@hotmail.com*

Abstract: Several kinds of current boundary classification methods based on hyperplane, hypersphere or ellipsoid were studied, and a novel classification method called Maximum-margin Fuzzy Classifier (MFC) was proposed by using a space point as the classification criterion. By the proposed method, a fuzzy classified point c was chosen in the pattern space firstly, which should be as close to two classes as possible. Moreover, the angle between the two classes should be also as large as possible. Then, the testing points could be classified in terms of the maximum angular margin between c and all the training points. Finally, the applications of the MFC were popularized from two-class classification to one-class classification according to the kernel dual of MFC equivalent to the Minimum Enclosed Ball (MEB). Comparative experiments with current classification methods verify that the MFC has good classification performance and noise resistance ability and its classification accuracy has been reached 98.9%.

Key words: pattern classification; fuzzy classifier; fuzzy classified point; noise resistance; one-class classification

收稿日期:2011-07-19;**修订日期**:2011-09-19.
基金项目:国家 863 高技术研究发展计划资助项目(No. 2007AA1Z158, No. 2006AA10Z313);国家自然科学基金资助项目(No. 60773206, No. 60704047)

1 引　言

分类是数据挖掘、机器学习和模式识别的重要方法之一，目前已广泛应用于字符识别、计算机辅助诊断、语音识别、签名认证、地学分析、文件检索等领域。分类技术的核心是构造分类器。构造分类器的过程一般分为训练和测试两步。在训练阶段，分析训练数据集的特点，为每个类别产生一个相应数据集的准确描述或模型；在测试阶段，利用类别的描述或模型对新进数据进行分类并测试其分类精度[1]。目前许多分类方法被机器学习、专家系统、统计学和神经生物学方面的研究者提出，如决策树、统计方法、机器学习方法、神经网络方法等。其中基于边界的方法[2-4]应用最为广泛。该方法将目标数据中的高密度区域映射到一个正半空间或者封闭的超(椭) 球里，同时保证包含大部分目标数据且上述几何形状体积最小，以达到最佳分类效果。经典的 SVM 及其变种[2,3,5-10]在空间内寻找一个超平面将两类分开；支持向量数据描述 (Support Vector Data Description, SVDD)[3]采用最小体积球约束目标数据达到剔除奇异点的目的。Wei 等人[5]利用超椭球代替了 SVDD 中的超球以考虑数据的结构信息，此外最小体积包含椭球(Minimum Volume Enclosing Ellipsoid, MVEE)[11]以及核最小体积覆盖椭球 (Kernel Minimum Volume Covering Ellipsoid, KMVCE)[12]均是通过优化椭球体积来寻找最小超椭球。

上述方法在分类决策时认为所有样本具有相同的作用。然而当训练样本中含有噪声点和孤立点时，上述方法的分类性能受到很大影响。基于上述分析，本文提出一种新颖的最大间隔模糊分类器 (Maximum-margin Fuzzy Classifier, MFC)，该方法试图在模式空间中找到一个模糊分类点将两类样本分开。已有的边界分类方法利用超平面、超(椭)球等几何形状作为分类依据，MFC 与其最大区别在于采用空间点作为分类模型。此外，模糊技术的引入保证 MFC 分类时对样本区别对待，减小或消除奇异点的影响，有效提高了分类效率。MFC 具有如下优势：(1)优良的分类性能；(2)同时解决二类分类问题和单类分类问题；(3)良好的抗噪能力。

2 背景知识

2.1 模糊理论

模糊理论是一种处理不精确性和不确定性信息的理论工具。采用模糊技术进行模式识别时，某特征属于某集合的程度由 0 与 1 之间的隶属度来描述。把一个具体的元素映射到一个合适的隶属度，由隶属度函数实现。常见的隶属度函数有：

2.1.1 基于距离的隶属度函数

基于距离的隶属度[13]用样本到类中心之间的距离来衡量样本对所在类的贡献。设类中心为 $\bar{\boldsymbol{x}}$，样本点为 $\boldsymbol{x}_i$，类半径为 R，则 $R=\max_i \| \boldsymbol{x}_i-\bar{\boldsymbol{x}} \|$。类中各样本的隶属度函数为

$$s(\boldsymbol{x}_i)=1-\frac{\| \boldsymbol{x}_i-\bar{\boldsymbol{x}} \|}{R}+\delta,$$

其中 δ 为很小的正数，它保证 $s(\boldsymbol{x}_i)>0$。

2.1.2 基于紧密度的隶属度函数

基于紧密度的隶属度确定方法[14]在确定样本的隶属度时，既要考虑样本到所在类中心的距离，还要考虑样本与类中其他样本的关系，而样本与类中其他样本之间的关系通过类中样本的紧密度来反映。设正类中心为 $\bar{\boldsymbol{x}}_+$，负类中心为 $\bar{\boldsymbol{x}}_-$，正、负类的半径分别为 $R_+=\max_i \| \boldsymbol{x}_i-\bar{\boldsymbol{x}}_+ \|$ 和 $R_-=\max_i \| \boldsymbol{x}_i-\bar{\boldsymbol{x}}_- \|$，两类中心的距离为 $T=\| \bar{\boldsymbol{x}}_+-\bar{\boldsymbol{x}}_- \|$，则每个正类样本到正类中心的距离为 $d_i^+=\| \boldsymbol{x}_i-\bar{\boldsymbol{x}}_+ \|$，每个负类到负类中心的距离为 $d_i^-=\| \boldsymbol{x}_i-\bar{\boldsymbol{x}}_- \|$；$\varepsilon$ 为半径控制因子，满足 $\varepsilon>0$，有 $T\varepsilon<R_+$ 和 $T\varepsilon<R_-$。则隶属度函数定义为：

$$s_i^+=\begin{cases}\dfrac{\delta+D_i^+}{R_+} & D_i^+\leqslant T\varepsilon\\ \delta & D_i^+>T\varepsilon\end{cases},\quad s_i^-=\begin{cases}\dfrac{\delta+D_i^-}{R_+} & D_i^-\leqslant T\varepsilon\\ \delta & D_i^->T\varepsilon\end{cases},$$

其中 δ 为很小的正数，它保证 $s_i>0$。

2.2 MEB 及 CCMEB 问题

2.2.1 MEB 问题

最小包含球 (Minimum Enclosed Ball, MEB)[15]用于单类分类问题。MEB 设法找一个以 c 为中心，R 为半径能够包含所有样本的最小超球体。求最小超球的半径就是求解以下的二次规划问题：

$$\min R^2$$
$$s.t. \| \boldsymbol{c}-\varphi(\boldsymbol{x}_i) \|^2\leqslant R^2 \quad i=1,\cdots,N, \quad (1)$$

其中 $\boldsymbol{c}$ 为超球体球心，R 为超球体半径，$\varphi(\boldsymbol{x})$ 表示从原始样本空间到高维特征空间的映射。

由 Lagrangian 定理可将原问题转化为如下对偶形式：

$$\max_{\boldsymbol{a}} \boldsymbol{a}^{\mathrm{T}} diag(\boldsymbol{K}) - \boldsymbol{a}^{\mathrm{T}}\boldsymbol{K}\boldsymbol{a}, \tag{2}$$

$$s.t. \quad \boldsymbol{a}^{\mathrm{T}}\boldsymbol{1}=1; \boldsymbol{a}\geqslant 0, \tag{3}$$

其中 $\boldsymbol{a}=[a_1,\cdots,a_N]^{\mathrm{T}}$，$\boldsymbol{1}=[1,\cdots,1]^{\mathrm{T}}$，核函数 $\boldsymbol{K}=[k(x_i,x_j)]=[\varphi(\boldsymbol{x}_i)^{\mathrm{T}}\varphi(\boldsymbol{x}_j)]$，$\boldsymbol{0}=[0,\cdots,0]^{\mathrm{T}}$。

2.2.2 CCMEB 问题

中心受限最小包含球（Center-Constrained MEB，CCMEB）[8] 是 MEB 问题的扩展。设 $\delta_i\in\boldsymbol{R}$，将原核空间的样本点扩展为 $\begin{bmatrix}\varphi(\boldsymbol{x}_i)\\ \delta_i\end{bmatrix}$，将原球心扩展为 $\begin{bmatrix}\boldsymbol{c}\\ 0\end{bmatrix}$，则非线性 CCMEB 问题可表示为：

$$\min R^2$$
$$s.t. \ \| \boldsymbol{c}-\varphi(\boldsymbol{x}_i)\|^2+\delta_i^2\leqslant R^2 \quad i=1,\cdots,N, \tag{4}$$

由 Lagrangian 定理易得上述问题的对偶形式：

$$\max_{\boldsymbol{a}} \boldsymbol{a}^{\mathrm{T}} diag(\boldsymbol{K}+\boldsymbol{\Delta}) - \boldsymbol{a}^{\mathrm{T}}\boldsymbol{K}\boldsymbol{a}, \tag{5}$$

$$s.t. \quad \boldsymbol{a}^{\mathrm{T}}\boldsymbol{1}=1; \boldsymbol{a}\geqslant 0, \tag{6}$$

其中 $\boldsymbol{a}=[a_1,\cdots,a_N]^{\mathrm{T}}$，$\boldsymbol{K}=[k(x_i,x_j)]=[\varphi(\boldsymbol{x}_i)^{\mathrm{T}}\varphi(\boldsymbol{x}_j)]$，$\boldsymbol{\Delta}=[\delta_1^2,\cdots,\delta_N^2]^{\mathrm{T}}\geqslant 0$，$\boldsymbol{0}=[0,\cdots,0]^{\mathrm{T}}$，$\boldsymbol{1}=[1,\cdots,1]^{\mathrm{T}}$。

对于任意的常数 $\eta\in\boldsymbol{R}$ 有：

$$\max_{\boldsymbol{a}} \boldsymbol{a}^{\mathrm{T}} diag(\boldsymbol{K}+\boldsymbol{\Delta}-\eta\boldsymbol{1}) - \boldsymbol{a}^{\mathrm{T}}\boldsymbol{K}\boldsymbol{a}$$
$$s.t. \quad \boldsymbol{a}^{\mathrm{T}}\boldsymbol{1}=1; \boldsymbol{a}\geqslant 0, \tag{7}$$

由于 η 与 α 无关，易知(7)式与式(5)同解。任何形如(7)式且 $\boldsymbol{\Delta}\geqslant 0$，均可视为 MEB 问题[8]。

3 最大间隔模糊分类器 MFC

3.1 概述

MFC 最初是针对二类分类问题提出的，但将 MFC 分别经过对偶变换、核变换后得到的核化对偶式等价于 MEB 问题，表明 MFC 也可解决单类分类问题。本文算法整体结构如图 1 所示。

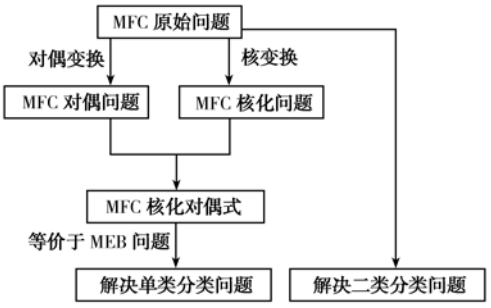

图 1　本文算法结构

Fig. 1　Structure of proposed algorithm

3.2 原始优化问题

对于包含 N 个模式二类划分问题，设模糊训练集合 $T=\{(\boldsymbol{x}_1,y_1,s_1),\cdots,(\boldsymbol{x}_N,y_N,s_N)\}$，其中 $\boldsymbol{x}_i\in\boldsymbol{R}^d(1\leqslant i\leqslant N_+ + N_- = N)$ 为输入数据，$y_i\in\{1,-1\}$ 为类别标签，s_i 为模糊隶属度。规定：$1\leqslant i\leqslant N_+$ 时，$y_i=1$；$N_+ +1\leqslant i\leqslant N$ 时，$y_i=-1$。假设第一类含有 N_+ 个模式 $\{\boldsymbol{x}_i,\boldsymbol{y}_i\}_{i=1}^{N_1}$，第二类含有 N_- 个模式 $\{\boldsymbol{x}_j,\boldsymbol{y}_j\}_{i=N_++1}^{N_1}$。

MFC 的目标是在模式空间找到一个模糊分类点 $\boldsymbol{c}$，确保两类分开且两类间隔最大。该优化问题可描述为：

$$\min_{c,\rho,\xi_i} \ \frac{1}{N}\sum_{i=1}^{N}\| \boldsymbol{x}_i-\boldsymbol{c}\|^2-\upsilon\rho+C\sum_{j=1}^{N} s_i\xi_i, \tag{8}$$

$$s.t. \quad y_i\frac{\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{c}}{\|\boldsymbol{x}_i\|\,\|\boldsymbol{c}\|}\geqslant\rho-\xi_i, \tag{9}$$

$$\xi_i\geqslant 0 \quad i=1,\cdots,N, \tag{10}$$

其中 $\boldsymbol{c}$ 为分类点；ρ 表示样本与分类点夹角距离意义上的间隔，称为“类间夹角间隔”；υ 为可调参数且满足 $\upsilon>0$；C 为惩罚因子，用于控制对错分样本惩罚的程度；s_i 为隶属度，表明样本属于某类的可靠程度，MFC 通过对不同样本赋予不同的隶属度，达到对不同样本采用不同程度惩罚的作用；$\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{c}/\|\boldsymbol{x}_i\|\,\|\boldsymbol{c}\|$ 表示样本 $\boldsymbol{x}_i$ 与分类点 $\boldsymbol{c}$ 的夹角距离；ξ_i 为松弛因子。

上述优化问题中，$\frac{1}{N}\sum_{i=1}^{N}\|\boldsymbol{x}_i-\boldsymbol{c}\|^2$ 保证样本距离分类点最近，尽量避免奇异点对分类的影响；$-\upsilon\rho$ 保证类间夹角间隔最大；$C\sum_{i=1}^{N} s_i\xi_i$ 允许存在误差，在一定程度上提高了算法的泛化能力。为了推导方便，将式(9)改写为式(11)，则原优化问题转化为：

$$\min_{c,\rho,\xi_i} \ \frac{1}{N}\sum_{i=1}^{N}\|\boldsymbol{x}_i-\boldsymbol{c}\|^2-\upsilon\rho+C\sum_{j=1}^{N} s_i\xi_i$$
$$s.t. \quad y_i\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{c}\geqslant\rho-\xi_i, \tag{11}$$
$$\xi_i\geqslant 0 \quad i=1,\cdots,N$$

3.3 对偶问题

定理 1：MFC 原始优化问题的对偶问题为：

$$\max_{\alpha} \ -\frac{4}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i y_i \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{x}_j-\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i a_j y_i y_j \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{x}_j, \tag{12}$$

$$s.\ t. \quad \sum_{i=1}^{N}\alpha_i=\upsilon, \tag{13}$$

$$0\leqslant\alpha_i\leqslant s_i\mathrm{C} \quad i=1,\cdots,\mathrm{N}. \tag{14}$$

证明：根据 Lagrangian 定理，上述原始问题的 Lagrangian 方程为：

$$L(c,\rho,\alpha,\xi_i)=\frac{1}{N}\sum_{i=1}^{N}\|\boldsymbol{x}_i-\boldsymbol{c}\|^2-\upsilon\rho+\mathrm{C}\sum_{i=1}^{N}s_i\xi_i-\sum_{i=1}^{N}\alpha_i(y_i\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{c}-\rho+\xi_i)-\sum_{i=1}^{N}\beta_i\xi_i, \tag{15}$$

其中 Lagrangian 乘子 $\alpha_i\geqslant 0$。

$L(\boldsymbol{c},\rho,\alpha,\xi_i)$ 分别对 $\boldsymbol{c}$、ρ、ξ_i 等变量求偏导，并令各偏导方程等于 0，可得：

$$\frac{\partial L}{\partial\rho}=-\upsilon+\sum_{i=1}^{N}\alpha_i=0\Rightarrow\sum_{i=1}^{N}\alpha_i=\upsilon, \tag{16}$$

$$\frac{\partial L}{\partial\boldsymbol{c}}=-\frac{2}{N}\sum_{i=1}^{N}(\boldsymbol{x}_i-\boldsymbol{c})-\sum_{i=1}^{N}\alpha_i y_i\boldsymbol{x}_i=0$$

$$\Rightarrow\boldsymbol{c}=\sum_{i=1}^{N}\left(\frac{1}{N}+\frac{1}{2}\alpha_i y_i\right)\boldsymbol{x}_i, \tag{17}$$

$$\frac{\partial L}{\partial\xi_1}=s_i\mathrm{C}-\alpha_i-\beta_i=0\Rightarrow\alpha_i+\beta_i=s_i\mathrm{C};$$

$$0\leqslant a_i\leqslant s_i\mathrm{C} \quad i=1,\cdots,\mathrm{N}, \tag{18}$$

将式(16)～(18)代入式(15)中，定理成立。

3.4 核化问题

在非线性情况下，通过一个满足 Mercer 条件的核函数对输入样本进行高维映射，并在高维空间中进行模式分类。非线性 MFC 表示如下：

$$\min_{c,\rho,\xi_i} \ \frac{1}{N}\sum_{i=1}^{N}\|\varphi(\boldsymbol{x}_i)-\boldsymbol{c}\|^2-\upsilon\rho+\mathrm{C}\sum_{j=1}^{N}s_i\xi_i, \tag{19}$$

$$s.\ t. \quad y_i\varphi(\boldsymbol{x}_i)^{\mathrm{T}}\boldsymbol{c}\geqslant\rho-\xi_i, \tag{20}$$

$$\xi_i\geqslant 0 \quad i=1,\cdots,N$$

其中映射函数 $\varphi:\mathbf{R}^d\rightarrow\mathbf{R}^D(D\gg d)$ 将原始样本空间映射到高维特征空间。

定理 2：非线性 MFC 对偶问题为：

$$\max_{\alpha} \ -\frac{4}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i y_i k(\boldsymbol{x}_i,\boldsymbol{x}_j)-\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i a_j y_i y_j k(\boldsymbol{x}_i,\boldsymbol{x}_j), \tag{21}$$

$$s.\ t. \quad \sum_{i=1}^{N}\alpha_i=\upsilon$$

$$0\leqslant\alpha_i\leqslant s_i\mathrm{C} \quad i=1,\cdots,N$$

其中核函数 $k(\boldsymbol{x}_i,\boldsymbol{x}_j)=\varphi(\boldsymbol{x}_i)^{\mathrm{T}}\varphi(\boldsymbol{x}_j)$。

3.5 类间夹角间隔 ρ 的求解

由 KKT 条件可知：对于支持向量，式(20)等号成立，即：

$$\rho=y_i\varphi(\boldsymbol{x}_i)^{\mathrm{T}}\boldsymbol{c}+\xi_i. \tag{22}$$

设支持向量集为：$\mathrm{S}=\{\boldsymbol{x}_i\mid a_i>0, i=1,\cdots,\mathrm{N}\}$。将每个 $\boldsymbol{x}_i\in\mathrm{S}$ 代入式(22)并求平均可得类间夹角间隔 ρ：

$$\rho=\frac{1}{|\mathrm{S}|}\sum_{x_i\in\mathrm{S}}y_i\sum_{j=1}^{N}\left(\frac{1}{N}+\frac{1}{2}\alpha_i y_i\right)k(\boldsymbol{x}_i,\boldsymbol{x}_j)+\xi_i.$$

3.6 决策函数

MFC 的决策函数如下：

$$f(\boldsymbol{x})=\mathrm{sgn}(\varphi(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{c}-\rho)=\mathrm{sgn}\left[\sum_{j=1}^{N}\left(\frac{1}{N}+\frac{1}{2}\alpha_i y_i\right)k(\boldsymbol{x}_i,\boldsymbol{x})-\rho\right],$$

若 $f(\boldsymbol{x})>0$ 则 $\boldsymbol{x}$ 属于第一类；若 $f(\boldsymbol{x})<0$ 则 $\boldsymbol{x}$ 属于第二类。将上述决策函数称为“夹角差决策函数”。

4 理论分析

4.1 可调参数 υ 性质

性质 1：用 MFC 对样本进行分类，若所得的类间夹角间隔 $\rho>0$，则有：

(1)若记错分样本数为 p，则 $\upsilon\geqslant ps(\boldsymbol{x}_i)\mathrm{C}$；

(2)若记支持向量数为 q，则 $\upsilon\leqslant qs(\boldsymbol{x}_i)\mathrm{C}$。

证明：

(1)由于错分样本 $\boldsymbol{x}_i$ 对应的松弛因子 $\varepsilon^*>0$，有 $\alpha_i^*=s(\boldsymbol{x}_i)\mathrm{C}$，因此 $\upsilon=\sum_{j=1}^{N}\alpha_i^*\geqslant ps(\boldsymbol{x}_i)\mathrm{C}$。

(2)由于支持向量 $\boldsymbol{x}_i$ 对应的 $\varepsilon_i^*>0$，由约束条件式(13)可知：$\upsilon=\sum_{j=1}^{N}\alpha_i^*\leqslant qs(\boldsymbol{x}_i)\mathrm{C}$。

上述性质说明可调参数 υ 具有边界性，这为 υ 值的选取提供了重要依据。

4.2 单类问题

定理 3：非线性 MFC 等价于 MEB 问题。

证明：令 $\beta_i=\frac{1}{\upsilon}\alpha_i$ 并将其带入非线性 MFC 对偶式有：

$$\max_{\beta} \ \sum_{i=1}^{N}\sum_{j=1}^{N}\beta_i y_i\frac{-4}{N\upsilon}k(\boldsymbol{x}_i,\boldsymbol{x}_j)-\sum_{i=1}^{N}\sum_{j=1}^{N}\beta_i\beta_j y_i y_j k(\boldsymbol{x}_i,\boldsymbol{x}_j), \tag{23}$$

$$s.t. \quad \sum_{i=1}^{N}\beta_i=1$$

$$0\leqslant\beta_i\leqslant\frac{s_iC}{\upsilon} \quad i=1,\cdots,N.$$

上式等价于：

$$\max_{b} \quad \boldsymbol{b}^{\mathrm{T}}[diag(\widetilde{\mathbf{K}})+\boldsymbol{\Delta}-\eta\mathbf{1}]-\boldsymbol{b}^{\mathrm{T}}\widetilde{\mathbf{K}}\boldsymbol{b}$$

$$s.t. \quad \boldsymbol{b}^{\mathrm{T}}\mathbf{1}=1;\boldsymbol{b}\geqslant 0, \tag{26}$$

其中 $\boldsymbol{b}=[b_1,\cdots,b_N]^{\mathrm{T}}$，$\mathbf{0}=[0,\cdots,0]^{\mathrm{T}}$，$\mathbf{1}=[1,\cdots,1]^{\mathrm{T}}$，$\widetilde{\mathbf{K}}=[y_iy_jk(x_i,x_j)]$，$\boldsymbol{\Delta}=-\mathrm{diag}(\widetilde{\mathbf{K}})-\frac{4}{N\upsilon}y_i\sum_{j=1}^{N}k(x_i,x_j)+\eta\mathbf{1}$。为了推导方便，特将式(25)写成式(26)。当 η 取值足够大时，总能保证 $\Delta\geqslant 0$，则 MFC 等价于 MEB 问题，这说明 MFC 可解决单类问题。

5 实验分析

通过与 C-SVC、ν-SVC、KNN 等主流分类器比较，验证 MFC 的有效性。实验环境为 3 GHz Pentium4 CPU、256M RAM、Windows XP 及 Matlab7.0。实验选取的核函数为高斯核函数：

$$k(\boldsymbol{x}_i,\boldsymbol{x}_j)=\exp\left(-\|\boldsymbol{x}_i-\boldsymbol{x}_j\|^2\Big/2\delta^2\right)$$

5.1 实验参数设置

目前参数选择的主流方法有：单一验证估计、留一法、k 倍交叉验证法以及基于样本相似度的方法等。除测试样本外，本文将训练样本分为 4 份训练集和 1 份验证集进行 5 倍交叉验证获取实验参数。

参数通过网格搜索策略[16]选择。高斯核函数的方差 δ 在网格 $\{\bar{x}/2\sqrt{2},\bar{x}/2,\bar{x}/\sqrt{2},\bar{x},\sqrt{2}\,\bar{x},2\bar{x},2\sqrt{2}\,\bar{x}\}$ 中搜索选取，其中 $\bar{x}$ 为训练样本平均范数的平方根；C-SVC 中，惩罚因子 C 在网格 {0.01，0.05，0.1，0.5，1，5，10} 中搜索选取；ν-SVC 中参数 ν 在网格 {0.1，0.5，1，5，10} 中搜索选取；KNN 中参数 K 在网格 {1，3，5，7，9} 中搜索选取；MFC 中可调参数 ν 在网格 {0.1，0.5，1，5，10} 中搜索选取，惩罚因子 C 在网格 {0.1，0.5，1，5，10} 中搜索选取。

5.2 二类模式分类

为了验证 MFC 二类模式分类的有效性，选择 UCI 数据集作为实验数据集(见表 1)。表 1 中 # Total 表示样本总数，# Class1 表示第一类的样本数，# Class2 表示第二类的样本数，Dim 表示样本维数。

表 1 二类模式分类实验数据集

Tab.1 Datasets used in two-class classification experiments

Datasets	# Total	# Class1	# Class2	Dim
Wine	125	55	70	13
Iris	100	50	50	4
Liver	345	145	200	4
Heart	190	145	45	13
Spectf	225	190	35	44
Ecoli	125	75	50	7
Glass	145	70	75	9
Pima	765	265	500	8

通过 5 倍交叉验证取得最佳参数后，依次在实验数据集上运行 C-SVC、ν-SVC、KNN 以及 MFC。MFC 分别取基于距离的隶属度函数和基于紧密度的隶属度函数，得到 MFCD 和 MFCC。实验参数及实验结果依次记录于表 2 和表 3。

表 2 二类模式分类实验参数

Tab.2 Experimental parameters in two-class classification

Datasets	C-SVC	ν-SVC	KNN	MFCD	MFCC
Wine	$C=0.01$ $\delta=\bar{x}/2\sqrt{2}$	$\nu=0.1$ $\delta=\bar{x}/2\sqrt{2}$	$K=9$	$C=0.5$ $\nu=5$ $\delta=\bar{x}$	$C=0.5$ $\nu=5$ $\delta=\sqrt{2}\,\bar{x}$
Liver	$C=0.01$ $\delta=\sqrt{2}\,\bar{x}$	$\nu=0.5$ $\delta=\sqrt{2}\,\bar{x}$	$K=3$	$C=0.1$ $\nu=1$ $\delta=\bar{x}/\sqrt{2}$	$C=0.1$ $\nu=5$ $\delta=2\sqrt{2}\,\bar{x}$
Heart	$C=0.01$ $\delta=2\sqrt{2}\,\bar{x}$	$\nu=0.1$ $\delta=\bar{x}/2\sqrt{2}$	$K=7$	$C=5$ $\nu=0.1$ $\delta=2\sqrt{2}\,\bar{x}$	$C=0.1$ $\nu=1$ $\delta=22\sqrt{2}\,\bar{x}$
Spectf	$C=1$ $\delta=\bar{x}/2$	$\nu=0.1$ $\delta=\bar{x}/2$	$K=7$	$C=1$ $\nu=0.5$ $\delta=\sqrt{2}\,\bar{x}$	$C=0.1$ $\nu=10$ $\delta=2-\bar{x}$
Ecoli	$C=0.01$ $\delta=2\sqrt{2}\,\bar{x}$	$\nu=0.1$ $\delta=2\sqrt{2}\,\mathrm{x}$	$K=7$	$C=5$ $\nu=5$ $\delta=2\sqrt{2}\,\bar{x}$	$C=5$ $\nu=1$ $\delta=2\sqrt{2}\,\bar{x}$
Glass	$C=0.5$ $\delta=\bar{x}/\sqrt{2}$	$\nu=0.1$ $\delta=\bar{x}/\sqrt{2}$	$K=3$	$C=0.1$ $\nu=5$ $\delta=\bar{x}/2\sqrt{2}$	$C=0.1$ $\nu=5$ $\delta=\bar{x}/\sqrt{2}$
Pima	$C=0.01$ $\delta=\bar{x}/2\sqrt{2}$	$\nu=0.1$ $\delta=\bar{x}/2$	$K=7$	$C=0.1$ $\nu=1$ $\delta=\bar{x}/\sqrt{2}$	$C=0.1$ $\nu=5$ $\delta=\bar{x}$

由表 3 可以看出：在 Liver、Spectf、Glass、Pima 数据集上，MFCD 具有最优的分类精度；在 Wine、Ecoli 数据集上，MFCC 具有最优的分类精度；在 Heart 数据集上，MFCD 和 MFCC 具有相同的分类精度且优于其他三种方法。综上，对于二类模式分类问题，MFC 具有优于 C-SVC、ν-SVC 和 KNN 等方法的分类性能。

表 3　二类模式分类结果

Tab. 3　Results of two-class classification (%)

Datasets	C-SVC	ν-SVC	KNN	MFCD	MFCC
Wine	91.7	93.3	96.3	95.0	98.3
Liver	63.5	65.9	62.4	73.5	68.8
Heart	78.1	75.6	67.1	88.8	88.8
Spectf	66.0	66.3	92.0	94.6	91.9
Ecoli	91.7	91.7	88.3	91.9	93.3
Glass	63.2	61.8	56.8	68.6	62.7
Pima	66.2	67.3	66.5	71.8	64.7

5.3　单类模式分类

为了验证 MFC 单类模式分类的有效性，选取 5 个 UCI 数据集作为实验数据集(见表 4)。表 4 中 # Normal 表示健康、正常或良性的样本数，# Abnormal 表示疾病、异常或恶性的样本数。

表 4　单类模式分类实验数据集

Tab. 4　Datasets used in one-class classification experiments

Datasets	# Normal	# Abnormal	Dim
Balance	288	30	4
Haberman	225	15	4
Abalone	689	30	8
Hayesroth	51	10	5
Iorosphere	165	25	34

通过 5 倍交叉验证取得最佳参数后，依次在实验数据集上运行 SVDD、OCSVM 以及 MFC。MFC 选取基于距离的隶属度函数。实验参数及实验结果依次记录于表 5 和表 6。

由表 6 可以看出：在 Balance、Haberman、Abalone、Iorosphere 数据集上，MFCD 具有最优的分类精度；在 Hayesroth 数据集上 MFCD 与 OCSVM 具有相同的分类精度且优于 SVDD。综上，对于单类模式分类问题，MFC 具有优于或相当于 SVDD 和 OCSVM 等方法的分类性能。

表 5　单类模式分类实验参数

Tab. 5　Parameters in one-class classification experiments

Datasets	SVDD	OCSVM	MFCD
Balance	$\delta=\bar{x}/2\sqrt{2}$	$\delta=2\sqrt{2}\bar{x}$	$C=5\ \ \nu=5\ \ \delta=\bar{x}/\sqrt{2}$
Haberman	$\delta=\bar{x}/2\sqrt{2}$	$\delta=\bar{x}$	$C=5\ \ \nu=5\ \ \delta=\bar{x}/2\sqrt{2}$
Abalone	$\delta=\bar{x}/2\sqrt{2}$	$\delta=\bar{x}$	$C=0.5\ \ \nu=0.1\ \ \delta=\bar{x}/2\sqrt{2}$
Hayesroth	$\delta=\bar{x}/2$	$\delta=\bar{x}/2$	$C=0.5\ \ \nu=0.1\ \ \delta=\bar{x}/2\sqrt{2}$
Iorosphere	$\delta=\bar{x}$	$\delta=\bar{x}$	$C=0.5\ \ \nu=1\ \ \delta=2\bar{x}$

表 6　单类模式分类结果

Tab. 6　Results of one-class classification experiments (%)

Datasets	SVDD	OCSVM	MFCD
Balance	89.8	91.8	93.9
Haberman	86.3	86.3	98.8
Abalone	75.9	68.3	76.9
Hayesroth	85.7	92.9	92.9
Iorosphere	80.0	75.0	85.0

5.4　抗噪性实验

为了验证 MFC 的抗噪能力，人工生成 100 个高斯数据(中心在(3，3)，标准差为 2)并随机产生 10 个噪声数据。实验数据如图 2 所示。实验参数及实验结果记录于表 7。

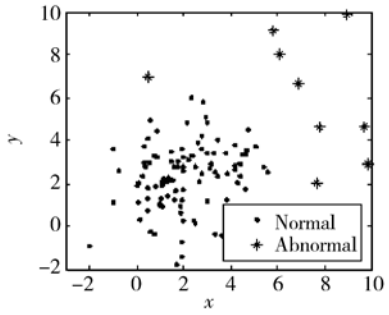

图 2　人工实验数据

Fig. 2　Man-made experimental datasets

由表 7 可以看出：在存在噪声的情况下，MFC 仍具有较高的分类精度，这说明 MFC 具有良好的抗噪性。

表 7　抗噪性能实验结果

Tab. 7　Results of noise resistance experiment

Algorithms	Parameters	Accuracy/(%)
MFCD	$C=0.5\ \ \nu=1\ \ \delta=\bar{x}$	97.0
MFCC	$C=1\ \ \nu=5\ \ \delta=\bar{x}/\sqrt{2}$	100

6 结　论

本文在分析现有的基于边界分类方法的基础上，提出一种新颖的模式分类方法 MFC。该方法将空间点作为分类依据，通过引入模糊技术来降低噪声点和孤立点对分类的影响。MFC 可解决二分类问题，且经过对偶变换和核变换后得到的核化对偶式等价于 MEB 问题，表明 MFC 也可解决单类分类问题。与主流分类方法比较实验可知：在二类模式分类方面，MFC 具有优于 C-SVC、ν-SVC 和 KNN 等方法的分类性能；在单类模式分类方面，MFC 具有优于或相当于 SVDD 和 OCSVM 等方法的分类性能；此外，在存在噪声的情况下，MFC 仍具有较高的分类精度。面向大规模数据分类是模式分类的研究热点。通过引入核心集向量机(Core Vector Machine，CVM)[8]将 MFC 的适用范围从中小规模数据集扩展到大规模数据集，这将是本文下一步的工作。

参考文献：

[1] 刘红岩，陈剑，陈国青. 数据挖掘中的数据分类算法综述[J]. 清华大学学报(自然科学版)，2002，42(6)：727-730.
LIU H Y，CHEN J，CHEN G Q. Review of classification algorithms for data mining [J]. *Journal of Tsinghua University (Science and Technology)*，2002，42(6)：727-730. (in Chinese)

[2] SCHOLKOPF B，PLATT J，SHAWE-TAYLOR J，*et al*.. Estimating the support of high-dimensional distribution [J]. *Neural Computation*，2001，13：1443-1471.

[3] TAX D，DUIN R，Support vector data description [J]. *Machine Learning*，2004 (54)：45-66.

[4] 冯爱民，薛晖，刘学军，等. 增强型单类支持向量机[J]. 计算机研究与发展，2008，45(11)：1858-1864.
FENG A M，XUE H，LIU X J，*et al*.. Enhanced one-class SVM [J]. *Journal of Computer Research and Development*，2008，45(11)：1858-1864. (in Chinese)

[5] WEI X K，HUANG G B，LI Y H. Mahalanobis ellipsoidal learning machine for one class classification [C]. *Proc. of the 6th Int. Conf. on Machine learning and cybernetics. Los Alamitos*：*IEEE Computer Society*，2007：3528-3533.

[6] VAPNIK V. *The Nature of Statistical Learning Theory* [M]. New York：Springer-Verlag，1995.

[7] MAHESH P，GILES M. Feature selection for classification of hyper spectral data by SVM [J]. *IEEE Trans. on Geoscience and Remote Sensing*，2010，48 (5)：2297-2307.

[8] TSANG I W，KWOK J T，CHEUNG P M. Core vector machines：fast svm training on very large data sets [J]. *Journal of Machine Learning Research*，2005，6：363-392.

[9] 文传军，詹永照，陈长军. 最大间隔最小体积球形支持向量机[J]. 控制与决策，2010，25(1)：79-83. (in Chinese)
WEN C J，ZHAN Y Z，CHEN C J. Maximal-margin minimal-volume hyper sphere support vector machine [J]. *Control and Decision*，2010，25 (1)：79-83. (in Chinese)

[10] SHIVASWAMY P，JEBARA T. Maximum relative margin and data-dependent regularization [J]. *Journal of Machine Learning Research*，2010 (11)：747-788.

[11] JUSZCZAK P. *Learning to recognize*：*A study on one-class classification and active learning* [D]. Delft：Delft University of Technology，2006.

[12] DOLIA A，HARRIS C，SHAWE-TAYLOR J. Kernel ellipsoidal trimming [J]. *Computational statistics and data analysis*，2007，52(1)：309-324.

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

[14] 孙名松，高庆国，王宣丹. 基于双隶属度模糊支持向量机的邮件过滤[J]. 计算机工程与应用，2010，46(2)：93-95.
SUN M S，GAO Q G，WANG X D. Mail filtering by dual membership fuzzy support vector machine [J]. *Computer Engineering and Application*，2010，46(2)：93-95. (in Chinese)

[15] TSANG I W，KWOK J T，ZURADA J M. Generalized core vector machines[J]. *IEEE Trans. on Neural Networks*，2006，17(5)：1126-1140.

[16] MULLER K R，MIKA S，RATSCH G，*et al*.. An introduction to kernel-based learning algorithms [J]. *IEEE Trans. on Neural Networks*，2001，12 (1)：181-201.

作者简介：

刘忠宝(1981－)，男，山西太谷人，博士研究生，2004 年于重庆交通大学获得学士学位，2007 年于北京工商大学获得硕士学位，主要从事模式识别、机器学习方面的研究。E-mail：liu_zhongbao@hotmail.com

王士同(1964－)，男，江苏扬州人，教授，博士生导师，主要从事人工智能与机器学习方面的研究。E-mail：wxwangst@yahoo.com.cn

---

●下期预告

控制力矩陀螺框架系统谐振抑制与高精度控制

徐向波，房建成，李海涛，陈彦鹏

（北京航空航天大学 仪器科学与光电工程学院，北京 100191）

控制力矩陀螺(CMG)是大型航天器姿态控制的关键执行机构，控制力矩陀螺框架系统的控制精度是影响其输出力矩精度的重要因素，系统中谐波减速器提高了框架系统的动态响应能力，但其产生的机械谐振大幅降低了系统的控制精度。为抑制框架系统的谐振并满足系统的控制精度，建立了框架系统动力学模型，根据系统动态性能要求选取合适阻尼系数来设计系统主导极点，使控制器产生的零点与机械谐振对应的极点重合形成偶极子，抑制了系统的机械谐振。仿真和实验结果显示：该系统有效地抑制了控制力矩陀螺框架系统谐振，0.175 rad/s 恒速控制精度为 0.002，0.175 sin(2πt) rad/s 正弦随动控制的幅值相对误差为 3.28%，相位差为 0.13 rad，很好地满足了控制力矩陀螺的高精度输出力矩需求。