

## 具有磁场效应的大间隔支持向量机

陶剑文\*<sup>①②</sup> 王士同<sup>①</sup>

<sup>①</sup>(江南大学信息工程学院 无锡 214122)

<sup>②</sup>(浙江工商职业技术学院信息工程学院 宁波 315012)

**摘要:** 为了提升模式分类泛化性能, 该文提出一种新颖的具有磁场效应的大间隔支持向量机(MFSVM)。为了学习最优分割超平面, MFSVM 通过引入最小化的  $q$ -磁场带, 使得一类(或正常类)被包含其中, 而另一类(或异常类)与该  $q$ -磁场带的间隔尽可能地大, 从而实现类内内聚性的提高和类间间隔的增大, 增强 SVM 学习泛化能力。在人造和实际数据集上实验结果显示, MFSVM 分别在二类和一类模式分类上的性能均优于或等同于相关方法。

**关键词:** 模式分类; 磁场效应; 新奇检测; 支持向量机; 核方法

中图分类号: TP181

文献标识码: A

文章编号: 1009-5896(2011)05-1055-07

DOI: 10.3724/SP.J.1146.2010.00896

## Maximal Margin Support Vector Machine with Magnetic Field Effect

Tao Jian-wen<sup>①②</sup> Wang Shi-tong<sup>①</sup>

<sup>①</sup>(School of Information Engineer, Southern Yangtze University, Wuxi 214122, China)

<sup>②</sup>(School of Information Engineer, Zhejiang Business Technology Institute, Ningbo 315012, China)

**Abstract:** In this paper, a novel maximal margin Support Vector Machine with Magnetic Field effect (MFSVM) is proposed in allusion to the improvement of the generalization performance of pattern classification issue. By introducing a minimum  $q$ -magnetic field tape, the basic idea of MFSVM is to find an optimal hyper-plane with magnetic field effect such that one class (or normal patterns) can be enclosed in the  $q$ -magnetic field tape due to the magnetic attractive effect, while at the same time the margin between the  $q$ -magnetic field tape and the other class (or abnormal patterns) is as large as possible due to magnetic repulsion, thus implementing both maximum between-class margin and minimum within-class volume so as to improve the generalization capability of the proposed method. Experimental results obtained with benchmarking and synthetic datasets show that the proposed algorithm is effective and competitive to other related methods in such cases as two-class and one-class pattern classification respectively.

**Key words:** Pattern classification; Magnetic field effect; Novelty detection; Support Vector Machine (SVM); Kernel approach

### 1 引言

模式分类旨在通过有限的训练样本学习一个分类器, 且该分类器须对未来数据具有良好的泛化能力<sup>[1]</sup>。类间间隔和类内聚类性是影响分类器分类性能的两种重要因素, 增大类间间隔和提高类内聚类性有利于分类能力的提高<sup>[2]</sup>。根据统计学习理论(SLT)<sup>[3,4]</sup>, 构建一个线性或非线性的SVM(Support Vector Machine)满足如下VC 维复杂度<sup>[4,5]</sup>:

$$VC \leq \min \left\{ \left\lceil R_{\max}^2 / \Delta_{\min}^2 \right\rceil, m \right\} + 1$$

其中  $R$  指覆盖所有样本的最小超球的半径,  $\Delta$  指两类间的最小间隔,  $m$  指样本维数。因此, 为了使得 VC 维尽量最小化, 须同时使得  $R$  最小化(即类内体积缩小, 等价于类内内聚性的最大化)和  $\Delta$  最大化, 前者导出超球支持向量机(SVM)(或称 gap-tolerant 分类机)<sup>[5,6]</sup>, 而后者则导出超球 SVM 的对偶版本线性 SVM<sup>[3,4,7]</sup>。

已有多种用于模式分类的方法提出, 其中建立在结构风险最小化(SRM)<sup>[4,7]</sup>理论基础上的线性支持向量机(SVM)及其相关变体仍然是目前实现模式分类的主流方法之一<sup>[8]</sup>, 其中,  $v$ -SVM<sup>[9]</sup>是经典 SVM 的一个最著名的拓展变体, 其引入一个新参数  $v$  来控制支持向量数下界和训练误差上界, 进一步提升 SVM 的泛化能力。近来, 关于超球 SVM 的研究,

2010-08-24 收到, 2010-12-10 改回

国家自然科学基金(60975027, 60903100)和宁波市自然科学基金(2009A610080)资助课题

\*通信作者: 陶剑文 jianwen\_tao@yahoo.com.cn

文献[1]提出一种大间隔最小体积球形支持向量机(MMHSVM),运用 SVM 基本理论和 SVDD<sup>[10]</sup>超球结构,构造两个大小不一的同心超球,小超球将正类样本包裹其中,大超球将负类样本排斥在外。Ye 等人<sup>[11]</sup>综合 SVDD 中的小球体和 SVM 中的大间隔思想,提出一种用于新奇检测的小球体和大间隔方法(SSLM),该方法在训练样本中利用了少量异常样本信息。

磁场效应理论(MFET)源自于物理学领域,缘于其磁场吸附效应, MFET 也被应用于社会学、经济学等领域以说明集聚效应。一般来说,在类球形数据分布(或某个核函数映射后呈球形数据分布)情况下,超球 SVM 方法较无约束线性 SVM 具有更严格的界<sup>[5]</sup>,但是数据呈球形分布本身是一个较为严格的约束,大多情况下,即使是经过某个核函数映射后,输入数据的整体分布形态也并非完全球形分布,利用一个最小包含超球很难实现一类数据的类内内聚性的最大化。受 MFET 和 SSLM 等思想的启发,结合超球 SVM 的类内内聚性和传统线性 SVM 的大间隔思想,提出一种新颖的具有磁场效应的大间隔 SVM(MFSVM),旨在学习一个具有磁场效应的最优超平面  $Y$ ,通过引入一个最小化的  $q$ -磁场带,使得一类(或正常类)被包含其中,而另一类(或异常类)与该  $q$ -磁场带的间隔尽可能的大,从而实现类内内聚性的提高和类间间隔的增大,增强线性 SVM 的泛化性能。MFSVM 的分割超平面犹如一个磁场强度为  $q$  的磁场,对一类模式进行吸引而对另一类模式进行排斥,这种磁场效应包括了吸附效应、筛选效应(排斥效应)和自我加强效应(内聚效应),吸附效应和自我加强效应使得一类模式类内内聚性增强,筛选效应使另一类模式因排斥而最大化远离超平面,从而使得 MFSVM 具有较强的模式分类泛化能力。

## 2 MFSVM

### 2.1 问题描述

对于一个包含  $N$  个模式的二类分类问题, 设给定训练集  $T = \{(x_1, y_1), \dots, (x_N, y_N)\}$ , 其中  $x_i \in \chi \subset R^d (1 \leq i \leq m_1 + m_2 = N)$  为输入数据,  $y_i \in \{+1, -1\}$  为类标签, 且当  $1 \leq i \leq m_1$  时,  $y_i = 1$ ; 当  $m_1 + 1 \leq i \leq N$  时,  $y_i = -1$ , 又设其中一类(或正常类)含有  $m_1$  个模式  $\{x_i, y_i\}_{i=1}^{m_1}$ , 另一类(或异常类)含有  $m_2$  个模式  $\{x_i, y_i\}_{i=m_1+1}^N$ 。两类模式间的最小距离称为  $\rho$ -间隔。为了更好地描述 MFSVM 问题, 首先引入如下概念:

**定义 1(硬  $q$ -磁场带)** 设 MFSVM 的磁场强度  $q > 0$ , MFSVM 的超平面  $Y = (w \cdot x) + b$  的硬  $q$ -磁

场是指距离该超平面的间隔为  $q$  的左右两超平面所包含的区域:  $\{(x_i, y_i) | |(w \cdot x_i) + b| \leq q, i = 1, 2, \dots, m_1\}$ 。为了方便, 若无特别说明, 下文简称硬  $q$ -磁场带为  $q$ -磁场带。

**定义 2( $q$ -磁场带超平面)** 对于给定的 MFSVM 磁场强度  $q > 0$ , 如果某个超平面  $y = (w \cdot x) + b$  的  $q$ -磁场带包含了输入模式中一类所有训练样本, 则称该超平面为 MFSVM 的  $q$ -磁场带超平面, 即超平面  $Y = (w \cdot x) + b$  满足:  $|(w \cdot x_i) + b| \leq q, i = 1, 2, \dots, m_1$ 。

**定义 3( $q$ -磁场带左右边界支持向量集)** 对于 MFSVM 的某个  $q$ -磁场带  $\{(x_i, y_i) | -q \leq (w \cdot x_i) + b \leq q\}, i = 1, 2, \dots, m_1$ , 令左右等式成立的向量集  $\{(x_i, y_i) | (w \cdot x_i) + b = -q\}$  和  $\{(x_i, y_i) | (w \cdot x_i) + b = q\}$  称为  $q$ -磁场带的左右边界支持向量集。

由定义 1-定义 3 可知, MFSVM 旨在构建一个  $q$ -磁场带超平面, 使得一类包含于该  $q$ -磁场带内, 而另一类被排斥在距离该  $q$ -磁场带间隔为  $\rho$  的区域内, 从而实现二类模式分割。图 1 为对应于 1 维空间  $R^1$  上的 MFSVM 问题的一个例子。图中“ $\times$ ”形数据点为一类(或正类)训练点, “-”形数据点为另一类(或负类)训练点, 实线表示超平面(直线)  $Y = (w \cdot x) + b$ , 两虚线之间的区域是该超平面的  $q$ -磁场带, 所有一类(或正类)数据点均在该  $q$ -磁场带内, 故该超平面为 MFSVM 的  $q$ -磁场带超平面。

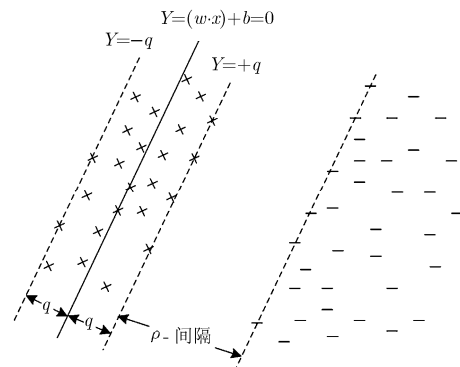


图 1 MFSVM 决策平面示意

MFSVM 通过最小化磁场强度  $q$  值, 即  $q$ -磁场带最小化, 使得一类类内内聚性最大化, 同时最大化类间最小间隔  $\rho$ , 实现类间间隔最大化, 引进松弛向量  $\xi$  和不平衡训练数据<sup>[1]</sup>惩罚参数  $C_i = 1/(v_i m_i)$ , 其中  $v_i$  为可调参数 ( $i = 1, 2$ ), 软  $q$ -磁场带 MFSVM 的优化问题描述如下:

$$\min_{w, q, \rho, \xi, b} \frac{1}{2} \|w\|^2 + \mu q - \nu \rho + C_1 \sum_{i=1}^{m_1} (\xi_i + \xi_i^*) + C_2 \sum_{j=m_1+1}^N \xi_j \tag{1}$$

s. t.

$$(w \cdot x_i) + b \leq q + \xi_i, \quad 1 \leq i \leq m_1 \quad (2)$$

$$(w \cdot x_i) + b \geq -q - \xi_i^*, \quad 1 \leq i \leq m_1 \quad (3)$$

$$(w \cdot x_j) + b \geq q + \rho - \xi_j, \quad m_1 + 1 \leq j \leq m_1 + m_2 \quad (4)$$

$$\xi \geq 0, \rho > 0, q > 0 \quad (5)$$

**定理 1** MFSVM 的  $q$ -磁场带超平面存在且不唯一。

**证明** 由于 MFSVM 训练集中一类(或正常类)只有有限( $m_1$ )个训练点, 所以当  $q$  充分大时,  $q$ -磁场带超平面总是存在, 且该  $q$  值大于或等于优化问题式(1)-式(5)中的最优值  $q^*$ 。显然, 对于给定的  $q > 0$ , 当  $q \geq q^*$  时,  $q$ -磁场带超平面存在且不唯一; 当  $q < q^*$  时,  $q$ -磁场带超平面不存在。证毕

### 2.2 MFSVM 算法

为导出原始问题式(1)-式(5)的对偶问题, 引入 Lagrange 函数:

$$\begin{aligned} L(w, q, \rho, \xi, b, \alpha, \beta, \gamma) = & \frac{1}{2} \|w\|^2 + \mu q - \nu \rho \\ & + C_1 \sum_{i=1}^{m_1} (\xi_i + \xi_i^*) + C_2 \sum_{j=m_1+1}^N \xi_j \\ & + \sum_{i=1}^{m_1} \alpha_i ((w \cdot x_i) + b - q - \xi_i) \\ & + \sum_{i=1}^{m_1} \alpha'_i (-q - \xi_i^* - (w \cdot x_i) - b) \\ & - \sum_{j=m_1+1}^N \alpha_j ((w \cdot x_j) + b - q - \rho + \xi_j) \\ & - \sum_{k=1}^N \beta_k \xi_k - \sum_{k=1}^{m_1} \gamma_k \xi_k^* \end{aligned} \quad (6)$$

其中  $\alpha_i \geq 0, \alpha'_i \geq 0, \alpha_j \geq 0, \beta_i, \gamma_k \geq 0$  分别为 Lagrange 乘子向量。这样, 可得以下定理:

**定理 2** 最优化问题

$$\min_{\alpha} \frac{1}{2} \|w^*\|^2 \quad (7)$$

s. t.

$$0 < \alpha_i < C_1, 0 < \alpha'_i < C_1, i = 1, \dots, m_1 \quad (8)$$

$$0 < \alpha_j < C_2, j = 1, \dots, m_2 \quad (9)$$

$$\sum_{j=1}^{m_2} \alpha_j = \nu \quad (10)$$

$$\sum_{i=1}^{m_1} \alpha_i = \nu + \frac{1}{2} \mu \quad (11)$$

$$\sum_{i=1}^{m_1} \alpha'_i = \frac{1}{2} \mu \quad (12)$$

其中

$$w^* = \sum_{i=1}^{m_1} (\alpha'_i - \alpha_i) x_i + \sum_{j=1}^{m_2} \alpha_j x_j \quad (13)$$

是原始问题式(1)-式(5)的对偶问题。

**证明** 在 Lagrange 函数  $L(w, q, \rho, \xi, b, \alpha, \beta, \gamma)$  中分别对  $w, q, \rho, \xi, b$  等原始变量求偏导数并令所得方程为零可得

$$\frac{\partial L}{\partial w} = w + \sum_{i=1}^{m_1} \alpha_i x_i - \sum_{i=1}^{m_1} \alpha'_i x_i - \sum_{j=1}^{m_2} \alpha_j x_j = 0 \quad (14)$$

$$\frac{\partial L}{\partial \rho} = -\nu + \sum_{j=1}^{m_2} \alpha_j = 0 \quad (15)$$

$$\frac{\partial L}{\partial q} = \mu - \sum_{i=1}^{m_1} \alpha_i - \sum_{i=1}^{m_1} \alpha'_i + \sum_{j=1}^{m_2} \alpha_j = 0 \quad (16)$$

$$\left. \begin{aligned} \frac{\partial L}{\partial \xi_i} = C_1 - \alpha_i - \beta_i = 0 \\ \frac{\partial L}{\partial \xi_i^*} = C_1 - \alpha'_i - \gamma_k = 0 \end{aligned} \right\} \quad (17)$$

$$\frac{\partial L}{\partial \xi_j} = C_2 - \alpha_j - \beta_j = 0 \quad (18)$$

$$\frac{\partial L}{\partial b} = \sum_{i=1}^{m_1} \alpha_i - \sum_{i=1}^{m_1} \alpha'_i - \sum_{j=1}^{m_2} \alpha_j = 0 \quad (19)$$

由式(14)可得式(13), 由式(15)可得式(10), 再由式(10), 式(16), 式(19)可分别得到式(11), 式(12)。由式(17)得到式(8), 由式(18)得到式(9)。将式(14)-式(19)代入式(6), 命题可得证。证毕

**定理 3** 最优化问题式(7)-式(13)是一个凸二次规划问题, 其所得优化解为全局最优解<sup>[12]</sup>。

**证明** 详细证明略。

当正负两类样本在输入空间不为线性 MFSVM 可分时, 可引入非线性映射  $\phi$ <sup>[7]</sup>, 将输入空间的模式样本映射到高维甚至无限维特征空间  $H$  中, 实现 MFSVM 可分, 在特征空间  $H$  中两个向量  $\phi(x_i), \phi(x_j)$  的内积可利用 Mercer 核函数  $K(x_i, x_j)$  映射来表示计算, 即有  $K(x_i, x_j) = \phi(x_i) \cdot \phi(x_j)$ 。

由 KKT 条件和 MFSVM 模型可知:

当  $0 < \alpha_i < C_1$  时,  $\xi_i = 0, (w^* \cdot x_i) + b = q, \alpha_i$  所对应的  $x_i$  属于  $q$ -磁场带右边界支持向量集  $SV_1 = \{x_i \mid 0 < \alpha_i < C_1, 1 \leq i \leq m_1\}$ ;

当  $0 < \alpha'_i < C_1$  时,  $\xi_i^* = 0, (w^* \cdot x_i) + b = -q, \alpha'_i$  所对应的  $x'_i$  属于  $q$ -磁场带左边界支持向量集  $SV_2 = \{x'_i \mid 0 < \alpha'_i < C_1, 1 \leq i \leq m_1\}$ 。  $SV_1$  和  $SV_2$  称为一类(或正类)的支持向量集。

当  $0 < \alpha_j < C_2$  时,  $\xi_j = 0, (w^* \cdot x_j) + b = q + \rho, \alpha_j$  所对应的  $x_j$  属于另一类(或负类)的支持向量集  $SV_3 = \{x_j \mid 0 < \alpha_j < C_2, 1 \leq j \leq m_2\}$ 。

**定理 4** 设  $(\alpha_i, \alpha'_i, \alpha_j)$  为二次优化问题式(7)-式(13)的 Lagrange 乘子的解, 则  $(w^*, q^*, \rho^*, b^*)$  为原

始优化问题式(1)-式(5)的优化解,且满足:

$$w^* = \sum_{i=1}^{m_1} (\alpha'_i - \alpha_i) x_i + \sum_{j=1}^{m_2} \alpha_j x_j \quad (20)$$

$$b^* = -\frac{1}{2} \left( \frac{1}{n_1} \sum_{x_i \in SV_1} (w^* \cdot x_i) + \frac{1}{n_2} \sum_{x_i \in SV_2} (w^* \cdot x'_i) \right) \quad (21)$$

$$q^* = \frac{1}{2} \left( \frac{1}{n_1} \sum_{x_i \in SV_1} (w^* \cdot x_i) - \frac{1}{n_2} \sum_{x_i \in SV_2} (w^* \cdot x'_i) \right) \quad (22)$$

$$\rho^* = \frac{1}{n_3} \sum_{x_j \in SV_3} (w^* \cdot x_j) + b^* - q^* \quad (23)$$

其中  $n_1 = |SV_1|$ ,  $n_2 = |SV_2|$ ,  $n_3 = |SV_3|$ ,  $|\cdot|$  表示集合基数。

**证明** 由式(14)得式(20)。下面分别求类间间隔  $\rho$ 、磁场强度  $q$  和截距  $b$  的优化解。

根据 KKT 条件,对于输入  $x_i \in SV_1 (1 \leq i \leq m_1)$  式(2)变成一个所有松弛变量等于 0 的等式。

$$(w^* \cdot x_i) + b^* = q^*, \quad 1 \leq i \leq m_1 \quad (24)$$

对于输入  $x'_i \in SV_2 (1 \leq i \leq m_1)$ , 式(3)变成一个所有松弛变量等于 0 的等式。

$$(w^* \cdot x'_i) + b^* = -q^*, \quad 1 \leq i \leq m_1 \quad (25)$$

同理,对于输入  $x_j \in SV_3 (1 \leq j \leq m_2)$ , 式(4)也变成一个等式。

$$(w^* \cdot x_j) + b^* = q^* + \rho^*, \quad 1 \leq j \leq m_2 \quad (26)$$

由式(24), 式(25)分别可得

$$b^* = -\frac{1}{2} \left( \frac{1}{n_1} \sum_{x_i \in SV_1} (w^* \cdot x_i) + \frac{1}{n_2} \sum_{x_i \in SV_2} (w^* \cdot x'_i) \right) \quad (27)$$

$$q^* = \frac{1}{2} \left( \frac{1}{n_1} \sum_{x_i \in SV_1} (w^* \cdot x_i) - \frac{1}{n_2} \sum_{x_i \in SV_2} (w^* \cdot x'_i) \right) \quad (28)$$

由式(26)可得

$$\rho^* = \frac{1}{n_3} \sum_{x_j \in SV_3} (w^* \cdot x_j) + b^* - q^* \quad (29)$$

其中  $b^*$  和  $q^*$  分别由式(27)和式(28)求得。证毕

根据定理 4,为了测试一个新模式  $x \in R^d$  的类别, MFSVM 判断其是否属于 MFSVM 的  $q$ -磁场带,即 MFSVM 的非线性决策函数为

$$f(x) = \text{sgn}(q^* - |w^{*T} \phi(x) + b^*|) \quad (30)$$

其中  $(\cdot)^T$  表示向量(或矩阵)的转置。

### 3 MFSVM 模型性能分析

作为 SSLM 的一个带约束的线性 SVM 对偶版本的 MFSVM 通过目标函数  $\min(1/2) \|w\|^2 + \mu q - \nu \rho$  来同时最大化类内内聚性和类间间隔,为了说明 SVM, SSLM 和 MFSVM 3 种方法的不同的泛化性能,通过构造一个最小包含球(MEB)以包含 SVM 和 MFSVM 的所有训练样本,而对于 SSLM, MEB 仅包含正常样本。图 2(a)-图 2(c)分别显示这 3 种情况的分类模型,其中“ $\times$ ”形数据为一类(或正类),“-”形数据为另一类(或负类)。图 2(a)显示 SVM 仅通过类间间隔最大化来提升分类性能,图 2(b)中 MFSVM 通过同时最小化  $q$ -磁场带(即一类类内内聚最大化)和最大化  $\rho$ -间隔(即最大化类间间隔),明显提升分类性能,图 2(c)显示 SSLM 与 MFSVM 的模型比较,虽然 SSLM 能实现一类(正常类)的类内内聚性,但对于类线性数据分布, MFSVM 提升一类(正常类)类内内聚性明显高于 SSLM,从而使得 MFSVM 比 SSLM 具有更严格的分类边界,由此,可得如下定理。

**定理 5** 设包含 SSLM 一类(正常类)的 MEB 半径为  $R$ , MFSVM 的最小化  $q$ -磁场带的磁场强度为  $q^*$ , 则 MFSVM 对于一类(正常类)的最大化类内内聚性  $\max C_{\text{MFSVM}}$  不低于 SSLM 一类(正常类)的最大化类内内聚性  $\max C_{\text{SSLM}}$ , 且当  $q^* = R$  时,  $\max C_{\text{MFSVM}} = \max C_{\text{SSLM}}$ 。

**证明** 对于一个有限( $N$ 个)样本集  $T = \{(x_1, y_1), \dots, (x_N, y_N)\}$ , 其中,  $x_i$  为训练样本,  $y_i \in \{+1, -1\}$  为类标签。根据文献[11], 总能寻求一个包含  $T$

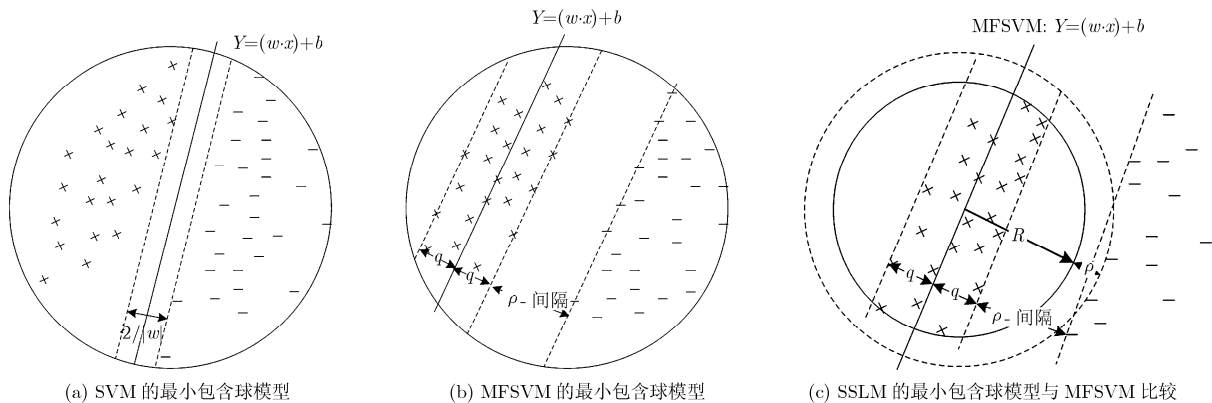


图 2 SVM, MFSVM, SSLM 3 种方法性能比较

的最小包含球(MEB)  $B(a, R)$ , 其中  $a, R$  分别指 MEB 的球心和半径, 设该 MEB 体积为  $\Omega_{\text{SSLM}}(R)$ 。

根据定理 1, 对于 SSLM 的一个 MEB, 总能对应找到 MFSVM 的一个最小化  $q$ -磁场带, 从而能得到 MFSVM 的一个最小化  $q$ -磁场带超平面  $Y = (w \cdot x) + b$ , 不失一般性, 设该超平面穿过 SSLM 的 MEB 球心(如图 2(c)所示)。设 MFSVM 的  $q$ -磁场带所包含样本的体积为  $\Omega_{\text{MFSVM}}(q)$ , 由定理 1 可知,  $\Omega_{\text{MFSVM}}(q^*) \leq \Omega_{\text{MFSVM}}(q)$ , 其中  $q^*$  为优化问题式(1)-式(5)的解。从图 2(c)可看出, 对于给定的类间间隔, MFSVM 超平面  $q$ -磁场带所包含样本的体积最多为 SSLM 的 MEB 的体积, 即  $\Omega_{\text{MFSVM}}(q^*) \leq \Omega_{\text{MFSVM}}(q) \leq \Omega_{\text{SSLM}}(R)$ , 且当  $q^* = q = R$  时, 等号成立, 从而可得  $\max C_{\text{MFSVM}} \geq \max C_{\text{SSLM}}$ , 当  $q^* = R$  时, 等号成立。证毕

为了进一步说明 MFSVM 的泛化能力, 提出如下定理。

**定理 6** MFSVM 算法能同时实现最大化类间间隔和正负类类内内聚性。

**证明** MFSVM 的目标函数:

$$\min \frac{1}{2} \|w\|^2 + \mu q - \nu \rho$$

等价于  $\min q$ ,  $\min \|w\|^2$  和  $\max \rho$ , 其中  $\min q$  等价于  $q$ -磁场带的最小化, 提高了一类(或正类)的内聚性;  $\min \|w\|^2$  实现类间间隔误差最大化;  $\max \rho$  实现类间间隔的最大化。下面证明 MFSVM 使另一类(或负类)类内内聚性最大化。

从直观上来看, MFSVM 模型在增大类间间隔  $\rho$  时可以减小另一类(或负类)的体积, 但从理论上讲, 有限的  $\rho$  值变化并不能影响另一类(或负类)外部体积的正无穷大的变化, 故不能从  $\max \rho$  来直观判断另一类(或负类)的体积最小化。为了说明类间间隔  $\rho$  与另一类(或负类)体积间的变化关系, 可以通过同构的方式<sup>[1]</sup>, 将另一类(或负类)的区域同构于一个  $1/(2\rho)$ -磁场带区域。

设 MFSVM 的超平面为  $Y = (w \cdot x) + b$ ,  $q$ -磁场带为

$$\{(x_i, y_i) \mid (w \cdot x_i) + b \leq q, i = 1, 2, \dots, m_1, y_i = 1\} \quad (31)$$

$q$ -磁场带的右超平面为

$$y_q = (w \cdot x) + b - q \quad (32)$$

另一类(或负类)区域表示为

$$\{x_j, y_j \mid (w \cdot x_j) + b \geq q + \rho, 1 \leq j \leq m_2, y_j = -1\} \quad (33)$$

另一类(或负类)区域中任一点  $x_j$  ( $1 \leq j \leq m_2$ ) 到  $q$ -磁场带的右超平面的距离为

$$D_{x_j} = (w \cdot x_j) + b - q$$

由式(33)可得

$$D_{x_j} = (w \cdot x_j) + b - q \geq \rho \quad (34)$$

设  $x_j$  到  $q$ -磁场带的右超平面的垂直投影点为  $x'_j$ , 连接  $x_j$  和  $x'_j$ , 并在该连接线  $l$  上选取一点  $x''_j$ , 使得  $x''_j$  到  $q$ -磁场带的右超平面的距离

$$0 \leq D_{x''_j} = (w \cdot x''_j) + b - q = \frac{1}{D_{x'_j}} \leq \frac{1}{\rho} \quad (35)$$

则点  $x_j$  与  $x''_j$  是一一对应的, 即存在双射  $f: x_j \leftrightarrow x''_j$ , 而由于  $x_j$  的任意性, 式(35)表示一个  $1/(2\rho)$ -磁场带, 其左右超平面分别为  $Y = 0, Y = 1/(2\rho)$ , 从而使得式(33)所表示的区域与式(35)所表示的  $1/(2\rho)$ -磁场带区域同构。  $\max \rho$  等价于  $\min 1/(2\rho)$ , 即最小化  $1/(2\rho)$ -磁场带, 而  $1/(2\rho)$ -磁场带又与式(33)所表示的另一类(或负类)区域同构, 故  $\max \rho$  使得另一类(或负类)体积最小化。由此可断言, MFSVM 既能实现类间间隔最大化, 又能实现正、负类类内体积的最小化。证毕

从定理 5, 定理 6 可以看出, MFSVM 方法具有增强的模式分类泛化能力。

按照文献[1]中术语, 本文称对应于 Lagrange 乘子  $\alpha_i > 0$  的训练样本  $x_i$  ( $1 \leq i \leq N$ ) 为支持向量(SV), 对应于松弛变量  $\xi_j > 0$  的训练样本  $x_j$  ( $1 \leq j \leq N$ ) 称为间隔误差(ME), 则有如下定理。

**定理 7**(参数定理) 设  $m_+, m_+, m_-$  分别指 MFSVM 的  $q$ -磁场带左右超平面对应的间隔误差数和负类的间隔误差数,  $s_+, s_+, s_-$  分别指 MFSVM 的  $q$ -磁场带左右支持向量集基数和负类的支持向量集基数, 则 MFSVM 的参数  $\nu, \mu$  能同时控制正负类支持向量集的下界和间隔误差集的上界, 即满足:

$$C_1 m_+ \leq (\nu + (1/2)\mu) \leq C_1 s_+ \quad (36)$$

$$C_1 m_+^* \leq (1/2)\mu \leq C_1 s_+^* \quad (37)$$

$$C_2 m^- \leq \nu \leq C_2 s_- \quad (38)$$

**证明** 根据文献[1,6]中方法可证, 证明略。

定理 7 对本文实验参数的选择具有一定的指导意义。

## 4 实验结果及分析

### 4.1 实验设置

通过与其他几个核学习方法进行比较来评价 MFSVM 算法的性能, 且重点关注模式识别中二类 and 一类分类问题, 对于多类问题, 可通过一对多和一对一等方法化为二类分类问题来解决。对于二类模式分类问题, 重点比较 SVM,  $\nu$ -SVM 与 MFSVM 算法的分类精度, 对于一类模式分类(新奇检测)问题<sup>[1]</sup>, 重点比较 OCSVM<sup>[4]</sup>, SSLM<sup>[11]</sup>和

MFSVM 算法的精度。

所有实验样本首先归一化为 $[-1,1]$ 。实验算法均采用高斯核函数(RBF)  $K(x,y) = \exp(-1/\gamma \|x-y\|^2)^{[1]}$ , 其中  $\gamma$  值在集合  $\{\sigma^2/8, \sigma^2/4, \sigma^2/2, \sigma^2, 2\sigma^2, 4\sigma^2\}$  中选取, 其中  $\sigma$  为训练样本平均范数的平方根;

对于 SVM, 惩罚参数  $C$  在集合  $\{0.01, 0.05, 0.1, 0.5, 1, 5, 10\}$  中选取;

对于  $\nu$ -SVM, 参数  $\nu$  在集合  $\{0.01k, 0.1k\}$  中选取, 其中  $k$  为 1-10 间整数;

对于 OCSVM, 参数  $\nu$  在集合  $\{0.01, 0.02, 0.04, 0.06, 0.08, 0.1, 0.2, 0.4, 0.6, 0.8, 1\}$  中选取;

对于 SSLM, 参数  $\nu$  在集合  $\{10, 30, 50, 70, 90\}$  中选取,  $\nu_1, \nu_2$  在集合  $\{0.001, 0.01\}$  中选取<sup>[11]</sup>;

对于 MFSVM, 根据参数定理, 参数  $\nu$  在集合  $\{1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60\}$  中选取, 参数  $\mu$  在集合  $\{1, 5, 10, 15, 20, 30\}$  中选取, 参数  $\nu_1, \nu_2$  在集合  $\{0.001, 0.01, 0.1\}$  中选取。

## 4.2 实验结论

本文所有实验均通过执行 5 重交叉验证来搜索优化参数值<sup>[8]</sup>, 每个数据集实验独立执行 10 次, 计算 10 次实验的平均值和标准差, 并记录为实验结果。

**4.2.1 二元模式分类** 为了评价所提方法的二元分类性能, 选取 14 个 UCI 实际数据集作为评测数据<sup>[11]</sup>, 数据集详细信息如表 1 所示。表 2 记录了 MFSVM, SVM 和  $\nu$ -SVM 3 种方法相应的平均分类精度和标准差。从表中数据可看出, 本文所提方法对于所有实际数据集均具有优于或可比较的二元分类性能。

表 1 二元模式分类 UCI 数据集

Datasets	#pos	#neg	$m_1$	$m_2$	$d$
Sonar	111	97	111	97	60
Imports	71	88	71	88	25
periplasm	284	52	284	52	7
Cancer	151	47	151	47	33
Breast	458	241	458	241	9
Heart	120	150	120	150	13
Abalone	1407	2770	1407	2000	10
Balance	288	337	288	337	4
Glass	76	138	76	138	9
vehicle	17	197	17	197	9
Arrhythmia	237	183	237	183	278
Ionosphere	225	126	225	126	34
Waveform	300	600	300	600	21
Vowel	48	480	48	480	10

表 2 UCI 数据集二类模式分类平均精度与标准差(%)

Dataset	SVM	$\nu$ -SVM	MFSVM
Sonar	<b>52.57±0.13</b>	<b>52.57±0.13</b>	<b>52.6±1.25</b>
Imports	<b>63.79+0.01</b>	<b>63.79+0.01</b>	<b>63.8±0.32</b>
periplasm	<b>87.3±0.77</b>	83.2±10.14	<b>87.4±0.31</b>
Cancer	<b>86.4±1.52</b>	<b>86.4±1.52</b>	<b>86.4±0.12</b>
Breast	<b>96.94±1.04</b>	96.5±2.12	<b>97.2±1.02</b>
Heart	<b>37.69±1.58</b>	36.92±3.45	<b>83.8±5.17</b>
Abalone	<b>78.3±3.46</b>	76.6±4.26	79.1±0.01
Balance	<b>29±15.27</b>	<b>29±15.32</b>	71.8±0.03
Glass	<b>84±7.54</b>	<b>84±7.54</b>	90.2±3.21
vehicle	<b>93.2±0.04</b>	<b>93.2±0.04</b>	<b>93.2±0.1</b>
Arrhythmia	<b>68.9±4.76</b>	<b>68.9±4.76</b>	<b>68.9±0.2</b>
Ionosphere	<b>16.66±10.64%</b>	<b>16.67±9.34%</b>	<b>84.7±0.18%</b>
Waveform	<b>75.2±0.14%</b>	<b>75.2±0.14%</b>	<b>74.9±1.13%</b>
Vowel	<b>88.17±0.02%</b>	<b>88.18±0.02%</b>	<b>88.2±0.42%</b>

**4.2.2 一元模式分类** 为了评价 MFSVM 在一类分类(OCC)问题中的分类性能, 实验采用 9 个 UCI 数据集<sup>[11]</sup>: Ball, Delft, Spectf0, Liver, Biomed, Housing, Hepatitis, Diabetes, Balance, 其中, 正常类对应于 UCI 数据集中的健康、正常或良性等情况, 异常类对应于 UCI 数据集中疾病、异常或恶性等情况, 按照文献[11]中实验设置, 对数据集分别随机抽取 70%正常样本和一小部分异常样本用于训练, 以使得 95%训练样本属于正常类(One-Class), 仅有 5%的训练样本属于异常类(Outlier), 余下样本用作测试。9 个数据集的详细信息如表 3 所示。表 4 记录了 3 种方法对 8 个实际数据集的 OCC 分类精度和标准差。结果显示本文所提方法对于所有数据集均具有优于或可比较的一类分类性能, 甚至在某些数据集(如 Spectf0, Housing)上取得了明显的性能优势。

## 5 结束语

受磁场效应理论的启发, 本文提出了一种具

表 3 一类模式分类 UCI 数据集

Datasets	Normal examples	Abnormal examples	Target( $m_1$ )	$m_2$	$d$
Ball	913	3237	640	40	32
Delft	376	1124	263	14	64
Spectf0	95	254	67	5	44
Liver	200	145	140	7	6
Biomed	127	67	90	5	5
Housing	458	48	300	15	13
Hepatitis	123	32	90	5	19
Diabetes	500	268	200	10	8
Balance	288	337	200	10	4

表4 UCI数据集一类模式分类平均精度与标准差(%)

Dataset	OCSVM	SSLM	MFSVM
Ball	78.98±5.53	<b>96.9±0.85</b>	<b>96.3±0.56</b>
Delft	77.3±13.12	<b>80.7±2.76</b>	<b>87.7±3.61</b>
Spectf0	69.37±13.13	<b>72.73±6.96</b>	<b>90.2±4.49</b>
Liver	51.43±4.12	<b>69.5±6.23</b>	<b>69.5±8.18</b>
Biomed	65.35±10.06	<b>81.1±6.09</b>	84.7±6.57
Housing	13.2±7.09	<b>83.2±3.34</b>	<b>92.1±0.72</b>
Hepatitis	61.32±1.33	<b>60.1±8.47</b>	<b>72.9±1.04</b>
Diabetes	52.1±5.78	<b>62.9±7.82</b>	<b>68.9±1.60</b>
Balance	61.54±12.26	<b>78.7±4.22</b>	<b>81.6±2.19</b>

有磁场效应的大间隔核学习机 MFSVM, 其融合了二类模式分类和新奇检测算法的思想, 使得其同时具备二类模式分类和一类模式分类(新奇检测)的优势。在 Mercer 核诱导的特征空间, MFSVM 构建一个具有磁场效应的  $q$ -磁场带超平面, 通过磁场效应使得一类(正类)因吸引而被包含于该  $q$ -磁场带内, 而另一类因排斥而尽量远离该超平面, 从而同时实现类内内聚性和类间间隔的最大化, 显著增强线性 SVM 的模式分类性能, 人工数据和实际数据实验结果均显示 MFSVM 具有相较于其他相关方法明显或相比较的性能优势。如何进一步提高 MFSVM 模式分类的鲁棒性是一个有待进一步研究的问题。

### 参 考 文 献

- [1] 文传军, 詹永照等. 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.  
Wen Chuan-jun and Zhan Yong-zhao, *et al.*. Maximal-margin minimal-volume hypersphere support vector machine[J]. *Control and Decision*, 2010, 25(1): 79-83.
- [2] Shivaswamy P and Jebara T. Maximum relative margin and data-dependent regularization [J]. *Journal of Machine Learning Research*, 2010, 11(3): 747-788.
- [3] Vapnik V. The Nature of Statistical Learning Theory [M]. New York: Springer Verlag, 1995: 160-195.
- [4] Burges C J C. A tutorial on support vector machines for pattern recognition [J]. *Data Mining and Knowledge Discovery*, 1998, 2(2): 955-974.
- [5] Chung Fu-Lai, Deng Zhao-hong, and Wang Shi-tong. From minimum enclosing ball to fast fuzzy inference system training on large datasets [J]. *IEEE Transactions on Fuzzy System*, 2009, 17(1): 173-184.
- [6] Wang Di, Zhang Bo, and Zhang Peng, *et al.*. An online core vector machine with adaptive MEB adjustment [J]. *Pattern Recognition*, 2010, 43(1): 3468-3482.
- [7] Müller, K R, Mika S, Ratsch G, Tsuda K, and Schölkopf B. An introduction to kernel-based learning algorithms [J]. *IEEE Transactions on Neural Networks*, 2001, 12(2): 181-201.
- [8] Orabona F, Castellini C, and Caputo B, *et al.*. On-line independent support vector machines [J]. *Pattern Recognition*, 2010, 43(1): 1402-1412.
- [9] Schölkopf B, Smola A J, Williamson R, and Bartlett P L. New support vector algorithms [J]. *Neural Computation*, 2000, 12(5): 1207-1245.
- [10] Tax D M J and Duin R P W. Support vector data description [J]. *Machine Learning*, 2004, 54(1): 45-66.
- [11] Wu Mingrui and Ye Jieping. A small sphere and large margin approach for novelty detection using training data with outliers [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009, 31(11): 2088-2092.
- [12] 邓乃扬, 田英杰. 支持向量机——理论与、算法与拓展[M]. 北京: 科学出版社, 2009: 5-19, 81-132, 170-180.

陶剑文: 男, 1973年生, 副教授, 博士生, 研究方向为模式识别、Web挖掘。

王士同: 男, 1964年生, 教授, 博士生导师, 研究领域为人工智能、机器学习等。