

最小方差支撑向量数据域描述

王晓明^{1*}, 王士同², 彭 宏¹

(1. 西华大学 数学与计算机学院, 成都 610039; 2. 江南大学 数字媒体学院, 江苏 无锡 214122)

(* 通信作者电子邮箱 wxmwm@yahoo.cn)

摘 要: 支撑向量数据域描述(SVDD)是一种已经得到了广泛应用的核方法,但是其在构建超球时没有充分考虑数据分布信息。针对此问题,首先等价改写了SVDD算法优化问题,然后重新定义了该优化问题中的距离定义形式,进而提出了最小方差支撑向量数据域描述(MVSVDD)算法。该算法充分考虑数据的分布信息。实验结果表明,相对于传统SVDD算法,MVSVDD在泛化能力上得到了较为明显的提高,体现出了更好的描述数据域的能力。

关键词: 支撑向量数据域描述;核方法;例外点检测;最小类方差支撑向量机;数据分布

中图分类号: TP181 **文献标志码:** A

Minimum variance support vector data description

WANG Xiao-ming^{1*}, WANG Shi-tong², PENG Hong¹

(1. School of Mathematics and Computer Engineering, Xihua University, Chengdu Sichuan 610039, China;

2. School of Digital Media, Jiangnan University, Wuxi Jiangsu 214122, China)

Abstract: Support Vector Data Description (SVDD), which is one of the widely applied kernel methods, has not taken the information of data distribution into full consideration. Concerning this issue, the optimization of SVDD was first reformulated equivalently, and then the distance in the optimization was redefined. Finally, a new algorithm called Minimum Variance Support Vector Data Description (MVSVDD) was presented, which exploited the information of data distribution. The experimental results denote that, in contrast to SVDD, MVSVDD obtains clear enhancement in generalization performance, and has better ability of describing data.

Key words: Support Vector Data Description (SVDD); kernel method; outlier detection; minimum class variance support vector machine; data distribution

0 引言

近年来,核方法^[1]作为一种优秀的学习方法得到了广泛的研究和发展,并取得了显著成果。支撑向量机(Support Vector Machine, SVM)^[2]是核方法中一类典型的代表性算法。但是,传统的SVM建立决策超平面依据的是最大间隔距离原则,从而导致该方法仅仅考虑了每类数据边界上的样本,而并没有充分地利用数据的分布信息,因而在一定程度上制约了泛化能力的进一步提高。针对这个问题,最近出现了两种从不同角度提出的改进算法:最小类方差支撑向量机(Minimum Class Variance Support Vector Machine, MCVSVM)^[3]和最大最小间隔机(Maxi-Min Margin Machine, M⁴)^[4]支撑向量机。从距离定义的形式上讲,SVM主要采用了欧氏距离,而MCVSVM和M⁴则在构建超平面时采用了马氏距离^[4],通过改变距离定义的形式在这两种算法中达到了引入数据分布信息的目的。

支撑向量数据域描述(Support Vector Data Description, SVDD)^[5-7]是核方法中另一类重要的算法。该算法主要是在特征空间中通过构造一个超球来描述一类正常(normal)数据,在超球内的样本点被视为正常点,然而在超球外面的样本点则被认为是例外点(outlier)、反常(abnormal)点或新颖(novel)点。SVDD算法还被应用于支撑向量聚类(Support Vector Clustering, SVC)^[8]中。目前,SVDD算法另一个研究热

点方向则是通过Core-Sets来实现该算法的快速几何逼近求解,而且该技术也被应用于解决SVM算法的优化问题^[9]。

但是,传统SVDD算法在特征空间中构建超球时缺少考虑数据的分布信息。然而,数据分布信息对于刻画或描述数据本身有着重要的作用。因此,针对此问题,通过借鉴MCVSVM和M⁴算法针对传统SVM算法缺少考虑数据分布信息的问题而通过定义新的距离形式来引入数据分布信息的基本思想,本文提出了最小方差支撑数据域描述(Minimum Variance Support Vector Data Description, MVSVDD)算法。文中首先等价改写了SVDD算法的优化问题,然后通过重新定义向量间的距离形式来推导构建了MVSVDD算法的优化问题。文中详细讨论了MVSVDD优化问题的求解方式。相对于传统的SVDD算法,该算法充分利用了数据的分布信息,实验结果表明,其具有更好的泛化能力。

1 支撑向量数据域描述算法

假定有一个含有 N 个样本的训练数据集 $D = \{x_i | x_i \in \mathbf{R}^d, i = 1, 2, \dots, N\}$,首先通过一个非线性映射函数 $\varphi: X \rightarrow H$ 把输入空间 X 的数据映射到一个高维特征空间 H ,形成了特征空间中样本集 $D' = \{\varphi(x_i) | x_i \in \mathbf{R}^d, i = 1, 2, \dots, N\}$;然后SVDD算法再在特征空间中构建一个包含所有正常样本点的最小超球。为此,SVDD算法定义了如下的优化问题:

收稿日期: 2011-07-21; **修回日期:** 2011-09-19。 **基金项目:** 国家自然科学基金资助项目(61103168); 西华大学重点科学研究基金资助项目(Z1012617); 四川省网络智能信息处理高校重点实验室资助项目(SCXZD1002-10)。

作者简介: 王晓明(1977-),男,四川简阳人,博士,主要研究方向:模式识别、图像处理; 王士同(1964-),男,江苏扬州人,教授,博士生导师,主要研究方向:人工智能、模糊系统; 彭宏(1966-),男,四川乐山人,教授,主要研究方向:图像处理。

$$\min R^2 + C \sum_{i=1}^N \xi_i \quad (1)$$

$$\text{s. t. } \|\varphi(\mathbf{x}_i) - \mathbf{a}\|^2 \leq R^2 + \xi_i, \xi_i \geq 0, i = 1, 2, \dots, N$$

其中: \mathbf{a} 表示超球中心, R 便是超球半径, C 正则化系数(或惩罚系数)。

2 数据分布信息驱动的支撑向量数据域描述

2.1 支撑向量数据域描述优化问题的等价形式

引理1 考虑如下优化问题:

$$\min \mathbf{a}^T \mathbf{a} - s + C \sum_{i=1}^N \xi_i \quad (2)$$

$$\text{s. t. } s - 2\mathbf{a}^T \varphi(\mathbf{x}_i) + \varphi^T(\mathbf{x}_i) \varphi(\mathbf{x}_i) - \xi_i \leq 0, \xi_i \geq 0, \\ i = 1, 2, \dots, N$$

并且假定 $s = \mathbf{a}^T \mathbf{a} - R^2$, 则 $\{\mathbf{a}^*, s^*, \xi^*\}$ 为优化问题(2)的最优解的充分必要条件是 $\{\mathbf{a}^*, R^*, \xi^*\}$ 为优化问题(1)的最优解(详细推导和证明请参考笔者前期的工作^[10])。

依据引理1, 可以把 SVDD 算法原始优化问题(1)的分析和研究转变为对其等价优化问题(2)的分析和研究。在 SVDD 算法实际应用中, 常采用的是高斯核, 所以本文讨论主要基于高斯核情况。高斯核可以表示为 $k(\mathbf{u}, \mathbf{v}) = \exp(-(\mathbf{u} - \mathbf{v})^T (\mathbf{u} - \mathbf{v}) / (2\sigma^2))$, 其中 σ 为核参数。显然, $\varphi^T(\mathbf{x}_i) \varphi(\mathbf{x}_i) = k(\mathbf{x}_i, \mathbf{x}_i) = 1$ 。因此, 高斯核情况下 SVDD 算法等价优化问题(2)又可以进一步改写为

$$\min \mathbf{a}^T \mathbf{a} - s + C \sum_{i=1}^N \xi_i \quad (3)$$

$$\text{s. t. } s - 2\mathbf{a}^T \varphi(\mathbf{x}_i) + 1 - \xi_i \leq 0, \xi_i \geq 0, i = 1, 2, \dots, N$$

令 $t = s - 1$, 并去掉目标函数中的常数1, 上面的优化问题进一步又可以变换为

$$\min \mathbf{a}^T \mathbf{a} - t + C \sum_{i=1}^N \xi_i \quad (4)$$

$$\text{s. t. } t - 2\mathbf{a}^T \varphi(\mathbf{x}_i) - \xi_i \leq 0, \xi_i \geq 0, i = 1, 2, \dots, N$$

2.2 最小方差支撑向量数据域描述优化问题的定义

高斯核情况下 SVDD 等价优化问题(4)中体现距离的主要是内积运算: $\mathbf{a}^T \mathbf{a}$ 和 $\mathbf{a}^T \varphi(\mathbf{x}_i)$, 而这实质上可以视为欧氏距离度量准则下的两个向量之间的距离。显然, 这种距离缺少考虑数据的分布信息, 然而事实上数据分布信息对于数据的刻画和描述有重要的作用。

针对此问题, 为在 SVDD 中利用数据分布信息, 首先定义核空间中数据的散度矩阵:

$$\mathbf{S} = \sum_{i=1}^N (\varphi(\mathbf{x}_i) - \mathbf{m})(\varphi(\mathbf{x}_i) - \mathbf{m})^T \quad (5)$$

其中 $\mathbf{m} = (1/N) \sum_{i=1}^N \varphi(\mathbf{x}_i)$ 为均值向量。散度矩阵在一定程度上反映了数据的总体分布信息。然后, 针对 SVDD 高斯核情况下其等价优化问题(4)定义了如下优化问题:

$$\min \mathbf{a}^T \mathbf{S}^{-1} \mathbf{a} - t + C \sum_{i=1}^N \xi_i \quad (6)$$

$$\text{s. t. } t - 2\mathbf{a}^T \mathbf{S}^{-1} \varphi(\mathbf{x}_i) - \xi_i \leq 0, \xi_i \geq 0, i = 1, 2, \dots, N$$

其中: $\mathbf{a}^T \mathbf{S}^{-1} \mathbf{a}$ 和 $\mathbf{a}^T \mathbf{S}^{-1} \varphi(\mathbf{x}_i)$ 为马氏距离。令 $\mathbf{a} = \mathbf{S} \mathbf{b}$, 并注意到 \mathbf{S} 为对称矩阵, 所以 $\mathbf{a}^T \mathbf{S}^{-1} \mathbf{a} = \mathbf{b}^T \mathbf{S} \mathbf{S}^{-1} \mathbf{S} \mathbf{b} = \mathbf{b}^T \mathbf{S} \mathbf{b}$, $\mathbf{a}^T \mathbf{S}^{-1} \varphi(\mathbf{x}_i) = \mathbf{b}^T \mathbf{S} \mathbf{S}^{-1} \varphi(\mathbf{x}_i) = \mathbf{b}^T \varphi(\mathbf{x}_i)$, 从而上面的优化问题(6)可进一步改写, 得到 MVSVD 的优化问题定义如下:

$$\min \mathbf{b}^T \mathbf{S} \mathbf{b} - t + C \sum_{i=1}^N \xi_i \quad (7)$$

$$\text{s. t. } t - 2\mathbf{b}^T \varphi(\mathbf{x}_i) - \xi_i \leq 0, \xi_i \geq 0, i = 1, 2, \dots, N$$

注意, 上面导出的 MVSVD 优化问题中并没有散度矩阵

\mathbf{S} 逆矩阵, 尽管其推导过程中出现了逆矩阵。事实上, 这里可以用伪逆矩阵来代替上面的推导, 可以得到同样的结果。所以, 散度矩阵 \mathbf{S} 是否可逆并不影响上面的推导过程和结果, 但是却直接影响了优化问题的求解方式。下面将讨论在散度矩阵 \mathbf{S} 为非奇异性情况下的求解方法。

对于优化问题(7), 可以把其转化为对偶优化问题来求解, 其拉格朗日式为

$$L = \mathbf{b}^T \mathbf{S} \mathbf{b} - t + C \sum_{i=1}^N \xi_i + \sum_{i=1}^N \alpha_i [t - 2\mathbf{b}^T \varphi(\mathbf{x}_i) - \xi_i] - \sum_{i=1}^N \beta_i \xi_i \quad (8)$$

其中: $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \dots, \alpha_N]^T$ 和 $\boldsymbol{\beta} = [\beta_1, \beta_2, \dots, \beta_N]^T$ ($\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbf{R}^N$) 是拉格朗日乘子。对于拉格朗日式(8), 分别对 $\mathbf{b}, t, \boldsymbol{\beta}$ 求导得, 并根据 Karush-Kuhn-Tucker (KKT) 条件^[11], 可以得到始优化问题(7)的对偶优化问题为

$$\min \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j H_{ij} \quad (9)$$

$$\text{s. t. } \sum_{i=1}^N \alpha_i = 1, 0 \leq \alpha_i \leq C, i = 1, 2, \dots, N$$

其中: $H_{ij} = \varphi^T(\mathbf{x}_i) \mathbf{S}^{-1} \varphi(\mathbf{x}_j)$ 。假定这个问题的最优解为 $\boldsymbol{\alpha}^* = [\alpha_1^*, \alpha_2^*, \dots, \alpha_N^*]^T$, 可以得到原始优化问题(7)的 \mathbf{b} 的最优解为:

$$\mathbf{b}^* = \mathbf{S}^{-1} \sum_{i=1}^N \alpha_i^* \varphi(\mathbf{x}_i) \quad (10)$$

对于 $t, 0 \leq t \leq C$, 类似于 SVDD, 称 \mathbf{x}_i 为支撑向量。用 D_{SV} 来表示支撑向量集, N_{SV} 表示支撑向量个数。此时, 根据 KKT 条件有 $t - 2\mathbf{b}^T \varphi(\mathbf{x}_i) - \xi_i = 0$, 可以据此得到 t 的最优解。但是为了使计算更可靠, 通常对于 t 的最优解取所有支撑向量的均值, 即

$$t^* = \frac{2}{N_{SV}} \sum_{i=1}^{N_{SV}} \sum_{j=1}^N \alpha_j^* H_{ij}; \mathbf{x}_i \in D_{SV} \quad (11)$$

2.3 优化问题的求解

高斯核情况下 MVSVD 的优化问题(7)中散度矩阵矩阵 \mathbf{S} 通常为奇异性矩阵, 因为高斯核所对应的特征空间是无穷维空间。下面讨论该问题的处理方法。

首先设定 $\mathbf{X}^\circ = [\varphi(\mathbf{x}_1), \varphi(\mathbf{x}_2), \dots, \varphi(\mathbf{x}_N)]$, 则根据文献[12]可知特征空间的散度矩阵 \mathbf{S} 可以表示为

$$\mathbf{S} = \mathbf{X}^\circ \mathbf{L} (\mathbf{X}^\circ)^T \quad (12)$$

其中: $\mathbf{L} = \mathbf{I} - \frac{1}{N} \mathbf{e} \mathbf{e}^T$, \mathbf{I} 为单位矩阵, $\mathbf{e} = [1, 1, \dots, 1]^T$ 。根据核空间的再生理论^[1]有

$$\mathbf{b} = \sum_{i=1}^N \eta_i \varphi(\mathbf{x}_i) \quad (13)$$

这里 $\eta_i (i = 1, 2, \dots, N)$ 为待定的系数。把上式代入优化问题(7), 即得

$$\min \boldsymbol{\eta}^T \mathbf{K} \mathbf{L} \mathbf{K} \boldsymbol{\eta} - t + C \sum_{i=1}^N \xi_i \quad (14)$$

$$\text{s. t. } t - 2\boldsymbol{\eta}^T \mathbf{k}_i - \xi_i \leq 0, \xi_i \geq 0, i = 1, 2, \dots, N$$

其中: $\mathbf{K} = [\mathbf{k}_1, \mathbf{k}_2, \dots, \mathbf{k}_N]$ 为核矩阵, $\mathbf{k}_i = [k(\mathbf{x}_i, \mathbf{x}_1), k(\mathbf{x}_i, \mathbf{x}_2), \dots, k(\mathbf{x}_i, \mathbf{x}_N)]^T$, $\boldsymbol{\eta} = [\eta_1, \dots, \eta_N]^T$ 。记 $\mathbf{H} = \mathbf{K} \mathbf{L} \mathbf{K}$, 则优化问题(14)可以转化为

$$\min \boldsymbol{\eta}^T \mathbf{H} \boldsymbol{\eta} - t + C \sum_{i=1}^N \xi_i \quad (15)$$

$$\text{s. t. } t - 2\boldsymbol{\eta}^T \mathbf{k}_i - \xi_i \leq 0, \xi_i \geq 0, i = 1, 2, \dots, N$$

应该注意, 矩阵 $\mathbf{H} = \mathbf{K} \mathbf{L} \mathbf{K}$ 是 \mathbf{R}^N 空间中以 $\{\mathbf{k}_i (i = 1, 2, \dots, N)\}$ 为样本的散度矩阵, 因此上述优化问题(15)实际上

是以 K 为输入数据的线性 MVSVD 算法的原始优化问题 (7)。对于上面的优化问题 (15), 有如下结论:

定理 1 假定优化问题 (15) 中的矩阵 H 为奇异, Ψ 表示 H 非零特征值对应的特征向量张成的空间, Π 表示零特征值对应的特征向量所张成的空间, 空间 Ψ 和空间 Π 为互补空间并一起组成了 H 的特征向量空间, 则对于任意 $\eta \in \mathbf{R}^N$, η 都可以表示成 $\eta = v + \gamma^{[13]}$, 这里 $v \in \Psi, \gamma \in \Pi$, 从而优化问题 (15) 可以等价转化为

$$\min v^T H v - t + C \sum_{i=1}^N \xi_i \quad (16)$$

$$\text{s. t. } t - 2v^T k_i - \xi_i \leq 0, \xi_i \geq 0, i = 1, 2, \dots, N$$

证明 首先应该注意, 矩阵 H 是 \mathbf{R}^N 空间中以 $\{k_i (i = 1, 2, \dots, N)\}$ 为样本的散度矩阵; 那么, $\gamma \in \Pi \Rightarrow \gamma^T H \gamma = 0$ 时, 说明输入变量 $k_i (i = 1, 2, \dots, N)$ 向 γ 方向投影后总体散度为 0。因此, 当 $k_i \neq k_j$ 时有 $\gamma^T k_i = \gamma^T k_j$, 这意味着所有的样本向 γ 方向投影后是同一个数据点, 为一个常量, 所以可以令 $\gamma^T k_i = r$, 则有:

$$\eta^T k_i = (v + \gamma)^T k_i = v^T k_i + r \quad (17)$$

$$\eta^T H \eta = (v + \gamma)^T H (v + \gamma) = v^T H v + v^T H \gamma + \gamma^T H v + \gamma^T H \gamma = v^T H v \quad (18)$$

成 (18) 立是由于 $v^T H \gamma = 0$ 和 $\gamma^T H v = 0$ 。所以对于优化问题 (7) 的拉格朗日式子 (8) 有:

$$L = \eta^T H \eta - t + C \sum_{i=1}^N \xi_i + \sum_{i=1}^N \alpha_i (t - 2\eta^T k_i + \xi_i) - \sum_{i=1}^N \beta_i \xi_i = v^T H v - t + C \sum_{i=1}^N \xi_i + \sum_{i=1}^N \alpha_i (t - 2v^T k_i + \xi_i) - \sum_{i=1}^N \beta_i \xi_i + 2r \quad (19)$$

注意, $2r$ 为常量, 在式 (19) 中可以去掉, 这样式 (19) 正是优化问题 (16) 的拉格朗日式子。假定 $v^* \in \Psi$ 是它的最优解, 则 η 的最优解可以表示成 $\eta^* = v^* + \gamma^*$ 。由式 (19) 有

$$\left. \frac{\partial L}{\partial \eta} \right|_{\eta=\eta^*} = \left. \frac{\partial L}{\partial v} \right|_{v=v^*} = 0 \Leftrightarrow 2Hv^* + 2 \sum_{i=1}^N \alpha_i k_i = 0 \quad (20)$$

因此, 优化问题 (15) 的最优解 η^* 只与优化问题 (16) 的最优解 v^* 有关, 而 γ^* 可以是 Π 空间中的任意向量。所以式 (15) 和 (16) 定义的优化问题是等价的。

定理 1 说明, 假定有 n 个非零特征值, 则 Ψ 空间就是 n 维。根据线性代数理论, 其必然和 n 维的 \mathbf{R}^n 空间同构^[13]。所以存在标准正交的 $P \in \mathbf{R}^{N \times n}$ 使得 $v = Pu$ 成立, 这里 $v \in \Psi, u \in \mathbf{R}^n, P$ 为 H 非零特征值所对应的特征向量组成, 这样式 (16) 就可转换为:

$$\min u^T \bar{H} u - t + C \sum_{i=1}^N \xi_i \quad (21)$$

$$\text{s. t. } t - 2u^T \bar{k}_i - \xi_i \leq 0, \xi_i \geq 0, i = 1, 2, \dots, N$$

其中: $u \in \mathbf{R}^n, \bar{k}_i = P^T k_i, \bar{H} = P^T H P$ 。注意, 这里 $\bar{H} = P^T H P$ 必为可逆矩阵, 其为 H 矩阵的非零特征值构成的对角矩阵。这样, 该优化问题的形式和求解方式类似于优化问题 (7) 矩阵 S 非奇异情况。假定式 (21) 的最优解为 $\{u^*, t^*\}$, 则此时得到的决策函数为

$$f(x) = \text{sign}((u^*)^T \hat{x} - t^*) \quad (22)$$

其中: $\hat{x} = P^T k$ 为测试样本 x 所对应的特征空间样本在转换空间中对应的数据点。这里 $k = [k(x_1, x), k(x_2, x), \dots, k(x_n, x)]^T$ 。这样, 对于测试样本 x , 如果 $f(x) = -1$ 则该样本判别为例外点, 否则认为该样本点为正常样本点。

3 实验验证

为了测试文中提出的 MVSVD 算法的可行性及泛化能

力, 在本章中将进行实验研究和分析。由于本文方法主要是针对传统 SVDD 算法的不足而提出, 所以在实验中主要将其与传统 SVDD 算法进行了比较。实验中, 从 UCI^[14] 选取 5 个数据集来进行测试, 表 1 给出这些数据的基本信息。在进行实验前, 首先对所有数据进行了预处理, 即把其每一维特征处理为均值为 0、方差为 1。

表 1 实验中 UCI 数据基本信息描述

数据集	样本数	特征维数	类别数
Heart	270	13	2
Wdbc	569	30	2
Ionosphere	351	34	2
Newthyroid	215	5	3
Wine	178	13	3

在单分类数据域描述实验中, 正常类样本被分类为正常类称为真正 (True Positive, TP), 正常类样本被分类为异常类称为假负 (False Negative, FN), 异常类样本被分类为正常类称为真正 (False Positive, FP), 异常类样本被分类为异常类称为真负 (True Negative, TN)。单分类的评价指标^[15]主要包括: $\text{precision} = (TP + 1) / (TP + FP + 1)$, $\text{recall} = TP / (TP + FN)$, 以及 $F\text{-value} = \frac{2 \times \text{recall} \times \text{precision}}{\text{recall} + \text{precision}} \times 100\%$ 。F-value 更

加全面地反映了测试结果。一个更好的例外点检测算法应该得到更高的 F-value。所以在本实验中采用了 F-value 来作为单分类器的性能评价指标。

实验中把数据集的某一类数据作为目标类, 而其他类数据作为例外点。训练数据集从目标数据集中随机选出 80%, 剩下的目标数据点和例外点一起组成了测试数据集。实验中采用了高斯核函数, 即核函数为 $\exp\left(-\frac{(u-v)^T(u-v)}{2\sigma^2}\right)$, 其中 σ 为核参数。对于算法中参数 σ 和 C , 通过 10 重交叉验证^[15]来进行选择, 把使得 F-value 最大的参数作为算法的最佳参数。

重复上述实验过 20 次, 并把 20 次的测试结果的平均值作为最后的测试结果, 如表 2 所示。从表中可发现, 相对于传统 SVDD 算法, MVSVD 在多数情况下能够获得更高的 F-value, 其在泛化能力上总体上得到了较为明显的提高。这也说明了在数据域描述算法中考虑利用数据的分布信息将有助于提高算法的性能。这正是本文算法 MVSVD 的出发点。

表 2 UCI 测试数据集上的 F-value

数据集	目标类	F-value/%	
		SVDD	MVSVD
Heart	1	70.319	73.428
	2	68.273	75.779
Wdbc	1	76.785	76.243
	2	71.032	78.396
Ionosphere	1	88.780	90.683
	2	77.881	84.644
Newthyroid	1	91.729	95.707
	2	74.418	73.628
	3	70.967	74.839
Wine	1	80.000	85.632
	2	63.157	68.421
	3	76.595	78.511

- tor machines: the cascade SVM [J]. *Neural Information Processing Systems*, 2005, 17: 521–528.
- [4] LEE Y-J, MANGASARIAN O L. RSVM: Reduced support vector machines [EB/OL]. [2010-02-20]. <http://citeseer.ist.psu.edu/viewdoc/summary?doi=10.1.1.102.3640>.
- [5] 朱方, 顾军华, 杨欣伟, 等. 一种新的支持向量机大规模训练样本集缩减策略[J]. *计算机应用*, 2009, 29(10): 2736–2741.
- [6] CHANG F, GUO Y-C, LIN X-R, *et al.* Tree decomposition for large-scale SVM problems [J]. *Journal of Machine Learning Research*, 2010, 11: 2935–2972.
- [7] 文贵华, 向君, 丁月华. 基于商空间粒度理论的大规模 SVM 分类算法[J]. *计算机应用研究*, 2008, 25(8): 2229–2301.
- [8] PLATT J C. Fast training of support vector machines using sequential minimal optimization [C]// *Advances in Kernel Methods—Support Vector Learning*. Cambridge: MIT Press, 1999: 185–208.
- [9] TSANG I W-H, 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.
- [10] SZEDMAK S, SHAWE-TAYLOR J. Multiclass learning at one-class complexity [R]. Southampton: University of Southampton, School of Electronics and Computer Science, 2005.
- [11] TSANG I W-H, KWOK J T, LAI K T. Core vector regression for very large regression problems [C]// *ICML '05: Proceedings of the 22nd International conference on Machine learning*. New York: ACM, 2005: 913–920.
- [12] WANG DI, ZHANG BO, ZHANG PENG, *et al.* An online core vector machine with adaptive MEB adjustment [J]. *Pattern Recognition*, 2010, 43(10): 3468–3482.
- [13] BÂDOIU M, CLARKSON K L. Optimal core-sets for balls [J]. *Computational Geometry: Theory and Applications*, 2008, 40(1): 14–22.
- [14] SMOLA A J, SCHÖLKOPF B. Sparse greedy matrix approximation for machine learning [C]// *ICML '00: Proceedings of the Seventeenth International Conference on Machine Learning*. Stanford: Morgan Kaufmann, 2000: 911–918.
- [15] TSANG I W-H, KWOK J T. LibCVM: The LibCVM Toolkit is a C++ implementation of the improved core vector machine [CP/OL]. [2010-03-04]. <http://c2inet.sce.ntu.edu.sg/ivor/cvm.html>.
- [16] CHANG C-C, LIN C-J. LIBSVM: A Library for support vector machines [EB/OL]. [2010-01-04]. <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>.
- [17] TSANG I W-H, KWOK J T-Y, CHEUNG P M. Generalized core vector machines [J]. *IEEE Transactions on Neural Networks*, 2006, 17(5): 1126–1140.
- [18] KUMAR P, MITCHELL J S B, YILDIRIM E A. Approximate minimum enclosing balls in high dimensions using corsets [J]. *ACM Journal of Experimental Algorithmic*, 2003, 8: 1–29.
- [19] PANIGRAHY R. Minimum enclosing polytope in high dimensions [C/OL]// *Proceedings of CoRR*, 2004 [2010-02-12]. <http://arxiv.org/abs/cs/0407020>.
- [20] TSANG I W-H, KWOK J T, CHEUNG P-M. Very large SVM training using core vector machines [C]// *Proceedings of the Tenth International Workshop on Artificial Intelligence and Statistics*. Barbados: The Society for Artificial Intelligence and Statistics, 2005: 349–356.
- [21] ASHARAF S, MURTY M N, SHEVADE S K. Multiclass core vector machine [C]// *Proceedings of the 24th International Conference on Machine Learning*. New York: ACM, 2007: 41–48.

(上接第418页)

4 结语

本文通过借鉴 MCVSVM 和 M^4 算法针对传统 SVM 算法存在的问题而通过定义新的距离形式来引入数据分布信息的基本思想,通过等价改写 SVDD 算法优化问题形式,从而在该算法中引入了数据分布信息,提出了 MVSVD 算法。相对于传统的 SVDD 算法, MVSVD 算法充分利用了数据的分布信息,体现出了更好的泛化能力。然而,文中工作主要针对的是高斯核情况,其他核情况下 MVSVD 算法优化问题的定义形式和性能需要进一步探索研究,这将是下一步的工作重点之一。

参考文献:

- [1] SCHOLKOPF B, SMOLA A. *Learning with kernels* [M]. Cambridge: MIT Press, 2002.
- [2] CRISTIANINI N, TAYLOR J S. *An introduction to support vector machines* [M]. Cambridge: Cambridge University Press, 2000.
- [3] ZAFEIRIOU S, TEFAS A, PITAS I. Minimum class variance support vector machines [J]. *IEEE Transactions on Image Processing*, 2007, 16(10): 2551–2564.
- [4] HUANG K Z, YANG H Q, KING I, *et al.* Maxi-min margin machine: Learning large margin classifiers locally and globally [J]. *IEEE Transactions on Neural Networks*, 2008, 19(2): 260–272.
- [5] TAX D M J, DUIN R P W. Support vector data description [J]. *Machine Learning*, 2004, 54(1): 45–66.
- [6] 何伟成, 方景龙. 基于信息熵的支持向量数据描述分类[J]. *计算机应用*, 2011, 31(4): 1114–1116.
- [7] 缪志敏, 潘志松, 袁伟伟, 等. 一种新的基于 SVDD 的多类分类算法[J]. *计算机科学*, 2009, 36(3): 65–68.
- [8] BEN-HUR A, HORN D, SIEGELMANN H T, *et al.* Support vector clustering [J]. *Journal of Machine Learning Research*, 2001, 2: 125–137.
- [9] TSANG I W, KWOK J T, ZURADA J M. Generalized core vector machines [J]. *IEEE Transactions on Neural Networks*, 2006, 17(5): 1126–1140.
- [10] WANG XIAOMING, CHUNG F-L, WANG SHITONG. Theoretical analysis for solution of support vector data description [J]. *Neural Networks*, 2011, 24(4): 360–369.
- [11] FLETCHER R. *Practical methods of optimization* [M]. 2nd ed. New York: Wiley, 1987.
- [12] HE XIAOFEL, YAN SHUICHENG, HU YUXIAO, *et al.* Face recognition using Laplacianfaces [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005, 17(3): 328–340.
- [13] YANG JIAN, YANG JING-YU. Why can LDA be performed in PCA transformed space? [J]. *Pattern Recognition*, 2003, 36(2): 563–566.
- [14] BLAKE C, MERZ C. UCI repository of machine learning databases [EB/OL]. [2011-04-16]. <http://www.ics.uci.edu/~ml-learn/MLRepository.html>.
- [15] 陈斌, 冯爱民, 陈松灿, 等. 基于单簇聚类的数据描述[J]. *计算机学报*, 2007, 30(8): 1325–1332.