

可变惩罚因子的支持向量数据描述算法

刘富,侯涛,刘云,张潇

(吉林大学通信工程学院,长春 130022)

摘要:支持向量数据描述(SVDD)是一种有效的数据描述算法。该算法中,作为定值的惩罚因子决定了数据描述的精度。然而实践中惩罚因子的选择是极其困难的,尤其是在训练数据含有噪声的情况下。为了解决这个问题,本文提出了一种可变惩罚因子的支持向量数据描述(VT-SVDD)算法。该算法根据样本点在核空间的位置分布,为每个样本计算一个惩罚因子,然后基于这种可变惩罚因子求解一个凸约束二次规划,即可以得到对数据集的球形域描述。为了验证所提的 VT-SVDD 的性能,在 UCI 数据集上进行了无噪声、有噪声两类训练数据的仿真实验。实验结果表明,VT-SVDD 能有效提高传统 SVDD 的精确度和稳健性。

关键词:计算机应用;支持向量数据描述(SVDD);惩罚因子;稳健性

中图分类号:TP391.4 **文献标志码:**A **文章编号:**1671-5497(2014)02-0440-06

DOI:10.13229/j.cnki.jdxbgxb201402026

A variable trade-off parameter support vector domain description

LIU Fu, HOU Tao, LIU Yun, ZHANG Xiao

(College of Communications Engineering, Jilin University, Changchun 130022, China)

Abstract: Despite the factor that Support Vector Domain Description (SVDD) model is an effective method for describing a set of training data, one inherent drawback is that the description is very sensitive to the selection of the trade-off parameter, which is hard to estimate in practice. To solve the difficulty, we proposed a novel Variable Trade-off parameter Support Vector Domain Description (VT-SVDD). In the proposed VT-SVDD, first, we assigned a position-based variable trade-off parameter to each data point. Then we computed a convex constrained quadratic programming based on the variable trade-off parameters. Finally, we can obtain a spherical data domain description for the training data. Experimental results demonstrate that the VT-SVDD can significantly improve the accuracy and robustness on UCI data sets.

Key words: computer application; support vector domain description (SVDD); trade-off parameter; robustness

收稿日期:2013-02-25.

基金项目:国家自然科学基金项目(51105170);吉林大学创新基金项目(20121101).

作者简介:刘富(1968-),男,教授,博士生导师.研究方向:计算机视觉及模式识别. E-mail:liufu@jlu.edu.cn

通信作者:侯涛(1980-),女,博士研究生.研究方向:模式识别及系统生物学. E-mail:ht_happy@126.com

0 引言

SVDD(Support vector data description)是Tax等人以支持向量分类器为基础提出的一种球形数据描述算法^[1-2]。该算法的主要思路是在核空间中寻找一个能够包含所有训练样本的最小封闭超球。在进行分类判断时,若一个新数据点在特征空间中的像落入该超球体内,则这个数据点被视为一个正常点,否则,若新数据点在特征空间中的像落入该超球体外,则这个数据点被视为一个异常点。作为一种基本的数据域描述工具,目前,SVDD算法已在很多领域中得到了应用和推广,如人脸识别^[3-4]、语音识别^[5]、医学成像检测^[6]和不同纹理背景的织物样本检测^[7]等。同时,基于SVDD算法也开发了一些改进的数据描述算法,如:Lee等^[8]提出了一种改进的基于密度诱导的支持向量数据描述方法(D-SVDD);Guo等^[9]基于SVDD算法提出一种异常边界检测算法,这种算法通过对SVDD边界的调整,以获得一个更紧致的数据描述;Zhang等^[10]提出了一种基于SVDD的模糊多类分类器,这种分类器改进了C-means聚类方法(Possibilistic C-means,PCM);Wang等^[11]提出了一种改进的基于核空间位置分布的支持向量数据描述方法(PSVDD),这种方法首先对每一个目标数据点的核空间位置进行估计,然后用所估计的距离进行加权,再将权重引入SVDD方法,进行计算,从而实现对目标数据点的空间域描述。

目前虽然对SVDD算法已有一些研究,然而该算法仍存在一个本质缺陷,就是作为定值的惩罚因子决定着SVDD的数据描述精度,然而实践中惩罚因子的选择是极其困难的。尤其在实际应用中,采集到的训练数据往往包含噪声,这种给所有的训练数据(包含正常点和异常点)都设置相同惩罚因子的方法,并没有考虑不同数据在训练集中的差异,这在一定程度上影响了SVDD的数据描述能力。为了解决这个问题,本文提出了一种可变惩罚因子的稳健支持向量数据描述方法(VT-SVDD)。这种方法基于样本点在核空间的位置分布,为每个样本计算一个惩罚因子,然后基于这种可变惩罚因子求解一个凸约束二次规划,就可以得到对数据集的球形域描述。为了验证所提的VT-SVDD方法的性能,本文在UCI数据库中收集了8个连续数据集,在这些数据集上进行

了无噪声、有噪声两方面仿真实验,实验结果表明,VT-SVDD模型提高了传统SVDD模型的精确度和稳健性。

1 传统SVDD方法

对于一个给定数据集 $\mathbf{X}, \mathbf{X} = \{x_1, x_2, \dots, x_N\}$, 其中 $x_i = (x_{i1}, x_{i2}, \dots, x_{iM}) \in \mathbf{R}^M$ 是一个 M 维的数据点, 传统 SVDD 方法的主要思路是求解式(1)所示的凸约束二次规划获得一些支持向量, 通过这些支持向量, 可以在核空间找到一个 o^* 为球心, r 为球半径的最小超球 $S(o^*, r)$, 使得该最小超球包含尽可能多的该类数据点。

$$\begin{cases} \min_{c, r, \xi} & r^2 + C \sum_{i=1}^N \xi_i \\ \text{s. t.} & \| \varphi(x_i) - o^* \|^2 \leq r^2 + \xi_i \\ & \forall x_i \in \mathbf{X}, \quad i = 1, 2, \dots, N \end{cases} \quad (1)$$

式中: $\|\cdot\|$ 为 2 范数; $\varphi(\cdot)$ 为对称的非线性映射函数; $\xi_i \geq 0$ 为一个松弛变量; C 为惩罚因子, 用来表示取舍数据点的意愿, C 的取值越大, 表示数据集 \mathbf{X} 中的那些数据点是正常点的可能性越大, 舍去它们的意愿就越弱, 求解式(1)获得的最小超球 $S(o^*, r)$ 的容积就越大, C 的取值越小, 表示数据集 \mathbf{X} 中的数据点是异常点的可能性越大, 舍去这些点的意愿就越强, 求解式(1)获得的最小超球 $S(o^*, r)$ 的容积就越小^[1]。

综上所述, 在传统 SVDD 算法中, 作为定值的惩罚因子 C 决定着球形数据描述的大小, 也决定着 SVDD 对数据集描述的精度。

2 可变惩罚因子 SVDD 方法

2.1 定义中值向量

首先需要计算数据集 \mathbf{X} 的中值向量 μ 。对于含有 N 个数据点的样本集 $\mathbf{X}, \mathbf{X} = \{x_1, x_2, \dots, x_N\}$, 每个数据点 $x_i = (x_{i1}, x_{i2}, \dots, x_{iM}) \in \mathbf{R}^M$ 是一个 M 维的向量。令在样本集 \mathbf{X} 中, 第 j 个元素的次序统计量为 $x_{:,j}(1) \leq x_{:,j}(2) \leq \dots \leq x_{:,j}(N)$, $j \in [1, M]$, 则对第 j 个元素, 当 N 是奇数时, 样本集的中位数为 $\mu_j = x_{:,j}(\frac{N+1}{2})$; 当 N 是偶数时, 中位数为 $\mu_j = \frac{1}{2}[x_{:,j}(\frac{N}{2}) + x_{:,j}(\frac{N}{2}+1)]$ 。当 j 从 1 取到 M 时, 可以获得一个 M 维的中值向量 $\mu = (\mu_1, \mu_2, \dots, \mu_M)$ 。

2.2 计算可变惩罚因子

为了计算每个数据点的惩罚因子,需要计算在核空间中每个训练数据与中值 μ 的空间距离向量 $d, d = [d_l | l = 1, 2, \dots, N]$

$$d_l = \| \varphi(x_l) - \varphi(\mu) \|^2 = K(x_l, x_l) + K(\mu, \mu) - 2K(x_l, \mu) \quad (2)$$

$$\forall l = 1, 2, \dots, N$$

式中: $K(\cdot)$ 为核函数,且 $K(a, b) = (\varphi(a), \varphi(b))$ 。

惩罚因子 $w_i (\forall i = 1, 2, \dots, N)$ 的计算公式为

$$w_i = 1 - \frac{d_i - \min_{1, \dots, N} \{d_l\}}{\max_{1, \dots, N} \{d_l\} - \min_{1, \dots, N} \{d_l\}} \quad (3)$$

本文定义每个数据点 x_i 的惩罚因子 w_i 反比例于该数据点 x_i 与中值向量 μ 在核空间的距离 d_i , 并且将其归一化到区间 $[0, 1]$ 。中值向量 μ 在特征空间中的像 $\varphi(\mu)$ 设为数据集 X 的核中心。通过定义惩罚因子 w_i 来衡量每个数据点 x_i 在数据集 X 中是异常点的可能性。在核空间中, 如果一个数据点 x_i 距离核中心 $\varphi(\mu)$ 越远, 这个数据点 x_i 就越有可能是一个异常点, 对其舍弃的意愿越强烈, 那么分配给该数据点 x_i 的惩罚因子就越小; 反之, 如果一个数据点 x_i 距离核中心 $\varphi(\mu)$ 越近, 这个数据点 x_i 就越有可能是一个正常点, 对其舍弃的意愿越弱, 那么分配给数据点 x_i 的惩罚因子就越大。

2.3 基于可变惩罚因子的 SVDD 模型

对数据集 X 中的每个数据点 x_i 引入它的惩罚因子 w_i , 让可变的惩罚因子 w_i 取代传统 SVDD 模型中的定值惩罚因子 C 来描述数据集 X 的特征空间域。通过解凸约束二次规划以获得包含最多数据点的最小超球:

$$\begin{cases} \min_{c, r, \xi} & r^2 + \sum_{i=1}^N w_i \xi_i \\ \text{s. t.} & \| \varphi(x_i) - o^* \|^2 \leq r^2 + \xi_i \end{cases} \quad (4)$$

$$\forall x_i \in X \quad i = 1, 2, \dots, N$$

在式(4)的约束下, 构建该公式的拉格朗日函数:

$$L(o^*, r, \alpha_i, \beta_i, \xi_i) = r^2 + \sum_{i=1}^N w_i \xi_i + \sum_{i=1}^N \alpha_i [\| \varphi(x_i) - o^* \|^2 - r^2 - \xi_i] - \sum_{i=1}^N \beta_i \xi_i \quad (5)$$

在拉格朗日乘子 $\alpha_i \geq 0$ 和 $\beta_i \geq 0$ 的条件下, 对拉格朗日函数中的每个变量求偏导, 并令其等于 0,

对 $\forall i = 1, 2, \dots, N$, 新的约束条件变为

$$\begin{cases} \sum_{i=1}^N \alpha_i = 1 \\ o^* = \sum_{i=1}^N \alpha_i \varphi(x_i) \\ \alpha_i = w_i - \beta_i \end{cases} \quad (6)$$

因为 $\alpha_i \geq 0$ 和 $\beta_i \geq 0$, 将变量 β_i 从式(6)的第三个约束中移除, 则 $\alpha_i = w_i - \beta_i$ 变成 $w_i \geq \alpha_i \geq 0$ 这样一个约束条件。

由 KKT 互补条件可得:

$$\begin{cases} \beta_i \xi_i = 0 \\ \alpha_i (\| \varphi(x_i) - o^* \|^2 - r^2 - \xi_i) = 0 \end{cases} \quad (7)$$

重写式(5)可以获得式(4)的对偶函数如下:

$$\begin{cases} \max L = \sum_{i=1}^N \alpha_i K(x_i, x_i) - \sum_{i,j=1}^N \alpha_i \alpha_j K(x_i, x_j) \\ \text{s. t.} \quad w_i \geq \alpha_i \geq 0, \sum_{i=1}^N \alpha_i = 1, \forall i \end{cases} \quad (8)$$

值得注意的是, 拉格朗日乘子 $\alpha_i, \forall i = 1, 2, \dots, N$ 的上边界不再被相同的惩罚因子 C 决定, 而是每个 α_i 被对应的可变惩罚因子 w_i 控制着。基于可变惩罚因子的一个负面影响就是增强了核参数对分类结果的影响, 因为不仅计算惩罚因子, 而且拉格朗日乘子的上边界都依赖于所选的核参数。

根据拉格朗日乘子 α_i 的值和其对应的惩罚因子 $w_i, \forall i = 1, 2, \dots, N$, 数据集 X 中的数据点可分为以下三类:

(1) 球内向量(IP): $\alpha_i = 0$, 则数据点 x_i 在超球面内部。因为已知 $\alpha_i = w_i - \beta_i$ 且 $\beta_i \xi_i = 0$, 又因为 $\alpha_i = 0$, 可推导出 $\xi_i = 0$ 。因此, $\| \varphi(x_i) - o^* \|^2 \leq r^2 + \xi_i = r^2$ 。

(2) 支持向量(SV): $0 < \alpha_i < w_i$, 则数据点 x_i 在超球面上。因为已知 $\alpha_i = w_i - \beta_i$ 且 $\beta_i \xi_i = 0$, 由 $\alpha_i < w_i$ 可推导 $\xi_i = 0$ 。在 $0 < \alpha_i$ 的条件下, 要满足 $\alpha_i (\| \varphi(x_i) - o^* \|^2 - r^2 - \xi_i) = 0$ 。因此, $\| \varphi(x_i) - o^* \|^2 = r^2 + \xi_i = r^2$ 。

(3) 球外向量(OP): $\alpha_i = w_i$, 则数据点 x_i 在超球面外部。由于 $\alpha_i = w_i - \beta_i$ 且 $\beta_i \xi_i = 0$, 由 $\alpha_i = w_i$ 可推导出 $\xi_i \geq 0$ 。又根据 $\alpha_i (\| \varphi(x_i) - o^* \|^2 - r^2 - \xi_i) = 0$ 且 $\alpha_i = w_i \geq 0$ 可得 $\| \varphi(x_i) - o^* \|^2 = r^2 + \xi_i \geq r^2$ 。

根据计算所获得的这些支持向量, 就可以获得一个包含最多正常数据的超球 $S(o^*, r)$,

$S(o^*, r) = \{x_i \mid \|\varphi(x_i) - o^*\|^2 \leq r^2, \forall x_i \in X\}$ 。

半径 r 可定义为

$$r = \text{mean}\{\|\varphi(x_i) - o^*\| \mid x_i \in SV\}$$

2.4 核参数选择

为获得较好的分类结果,本文选择高斯核函数 $K_G(x_i, x_j) = \exp[-q(x_i - x_j)^2]$ 来验证,对数据集 $X = \{x_1, x_2, \dots, x_N\}$, 取 $q_{\min} = 1/\max_{i,j, i \neq j} \|x_i - x_j\|^2$, $q_{\max} = 1/\min_{i,j, i \neq j} \|x_i - x_j\|^2$ 。在区间 $[q_{\min}, q_{\max}]$ 中,用网格搜索策略 $[q_1, q_2, \dots, q_M]$ 来寻找能够取得最优分类结果的 q 参数。

2.5 决定函数

用符号函数 f 来决定是否数据点 x_i 属于数据集 X 。

$$f = \text{sign}(r^2 - \|\varphi(x_i) - o^*\|^2) \quad (9)$$

这里有两种情况:①如果新数据点在核空间的像 $\varphi(x_i)$ 落在超球内部或超球面上,即是 $f = 1$ 或 $f = 0$,则认为新数据点 x_i 是正常点,属于数据集 X 。②如果新数据点在核空间的像 $\varphi(x_i)$ 落在超球外部,即是 $f = -1$,则认为新数据点 x_i 是异常点,不属于数据集 X 。

3 实验结果与比较

3.1 实验数据

为了验证 VT-SVDD 算法的性能,本文从 UCI 数据库^[12]下载了 8 个连续数据集,然后从这 8 个数据集中挑选一些数据进行实验,具体的实验数据描述见表 1。在这些数据集上,进行了两个方面仿真实验来验证分类模型的性能:一个实验是在 UCI 数据库收集的原始数据集上进行;另一个实验是在原始的 UCI 数据集中加入噪声,用含有噪声的数据集验证分类模型的稳健性。在这两个实验中,用分类精度作为衡量分类模型稳健性的标准,分类精度越高,代表该分类模型越稳健。

表 1 数据集描述

Table 1 Summary of data sets

数据集	类别数	实验样本数	特征数
wine	3	160	13
WDBC	2	500	30
WPBC	2	180	33
iris	3	140	4
ionosphere	2	320	34
sonar	2	200	60
diabetes	2	760	8
glass	5	140	9

在每个实验中,将 VT-SVDD 与 4 种分类方法(PSVDD、SVDD、K-NN 和 SVMs)进行了比较。其中 VT-SVDD、PSVDD、SVDD 和 SVMs 所选核函数均为高斯核函数,核参数 q 及 SVDD、SVMs 的惩罚因子 C 通过交叉验证的方法,寻找让每个方法取得最优识别率的参数^[2,13]。

3.2 不含噪声数据的实验结果

首先,用 UCI 数据库中收集的原始数据集验证 VT-SVDD 算法的性能。以 wine 数据为例,先将其平均分成 10 份,选择其中的 9 份作为训练样本,余下的 1 份作为测试样本。然后用本文所提的 VT-SVDD 判断测试样本的分类类别。将这一过程重复进行 10 次,以保证 wine 中每一个数据都可以作为测试样本进行分类测试。最后,统计 10 次预测的精度,求取其平均值作为 wine 数据集的最终分类精度。数据集用各种不同的分类模型得到的分类精度见表 2。

表 2 原始数据集的分类精度

Table 2 Classification accuracies of raw data sets %

数据集	VT-SVDD	PSVDD	SVDD	K-NN	SVMs
wine	93.8	93.1	89.4	91.2	90.6
WDBC	91.6	93	89.4	92	90.8
WPBC	68.3	73.3	65.9	68.1	66.3
iris	90.7	92.9	90.7	92.3	89.8
ionosphere	78.8	81.3	81.9	81.8	80.2
sonar	89.5	81.5	83.3	81.1	83.1
diabetes	71.3	70.7	70.5	71.3	67
glass	90	90	86	90.5	88.5
Average	84.3	84.5	82.1	83.5	82

从实验结果可以看出,在 8 个实验数据集中,VT-SVDD 在 wine 和 sonar 数据集上的分类精度高于其他分类方法;在 diabetes 数据集上,VT-SVDD 和 K-NN 共同取得 71.3% 的最高分类精度;PSVDD 在 WDBC、WPBC 和 iris 数据集上取得的分类精度高于其他分类方法;K-NN 在 glass 数据集上取得了最高的分类精度。总体而言,本文所提的 VT-SVDD 方法和 PSVDD 方法都能取得高于 84% 的分类精度,K-NN 其次,能取得 83.5% 的分类精度,经典 SVDD 和 SVMs 的实验结果较保守,分别取得了 82.1% 和 82% 的分类精度。

3.3 含噪声数据的实验结果

为了进一步对比各种分类模型的稳健性,本文将原始的 UCI 数据集加入噪声,用含有噪声的数据集来验证 VT-SVDD 方法的稳健性。同样以 wine 数据集为例,先将其平均分成 10 份,选择其

中的 9 份数据集并将其污染成含有 $a\%$ 的类噪声的数据集,余下的 1 份作为测试样本。用 VT-SVDD、PSVDD、SVDD、K-NN 和 SVMs 分类方法对测试数据进行分类预测。将这一过程重复进行 10 次,以保证 wine 数据集中每一个样本都可以作为测试样本进行分类测试。这样就可以得到在噪声水平为 $a\%$ 的环境下对数据集进行分类的

实验结果。统计 10 次预测的分类精度,求取其平均值作为 wine 数据集的最终分类精度。

本文对 8 个数据集分别添加 5%、10% 和 20% 的噪声水平,然后测试各种分类模型的稳健性。数据集在不同噪声水平下的分类精度结果见表 3。

表 3 显示,在训练数据添加噪声的情况下,

表 3 噪声数据集的分类精度

Table 3 Classification accuracies of noisy data sets

数据集	噪声水平	VT-SVDD	PSVDD	SVDD	K-NN	SVMs	%
wine	5	93.1	90.6	81.9	90	85.6	
	10	92.5	90	75.6	87.5	81.3	
	20	92.5	90	69.4	86.3	77.5	
WDBC	5	91.2	91.6	82.8	91.6	87	
	10	91.2	90.8	80.2	90.6	82.4	
	20	90.6	89	77.8	89.2	78.6	
WPBC	5	67.2	67.2	64.4	66.7	64.4	
	10	67.8	67.2	63.3	66.1	63.3	
	20	67.2	65	60.6	61.1	60.6	
iris	5	91.4	90.7	87.9	91.4	85.7	
	10	91.4	90	86.4	91.4	79.3	
	20	91.4	80.7	86.4	91.4	74.3	
ionosphere	5	77.8	77.5	80	81.6	79.4	
	10	77.5	75	74.7	79.4	76.9	
	20	75.6	72.5	69.4	77.5	72.2	
sonar	5	80	77.5	80.5	79.5	80.5	
	10	74.5	75	74	75	74.5	
	20	69.5	72.5	70.5	73.5	70.5	
diabetes	5	70.7	69.3	69.2	71.2	64.6	
	10	69.9	68.8	69.2	70.3	63.2	
	20	70.3	68.7	67.8	69.7	61.3	
glass	5	85	88.6	90	84.3	80.7	
	10	85	88.6	88.6	84.3	77.9	
	20	83.6	80.7	88.6	76.4	75.7	
Average	—	81.1	80	76.6	80.3	74.9	

VT-SVDD 仍能够在 wine、WDBC、WPBC、iris 和 diabetes 数据集上取得最好的分类精度;K-NN 在 iris、ionosphere、diabetes 数据集上取得最好的分类精度;PSVDD 和传统 SVDD 次之,分别在 WDBC、sonar 和 glass 数据集上取得最好的分类精度。总体而言,本文所提的 VT-SVDD 能取得最高的分类精度为 81.1%;PSVDD、K-NN 次之,分别取得 80% 和 80.3% 的分类精度;传统 SVDD 和 SVMs 也分别取得了 76.6% 和 74.9% 的分类精度。

将表 3 与表 2 的实验结果进行比较可以看出,训练数据添加噪声后,所有分类方法的分类精度都有所下降,其中 VT-SVDD、K-NN 的分类精度下降最为缓慢,均为 3.2%;PSVDD、SVDD 次之,添加噪声后分类精度分别下降了 4.5% 和

5.5%;SVMs 的分类精度下降最多,为 7.1%。

图 1 展示了随着训练数据集中噪声的增加,

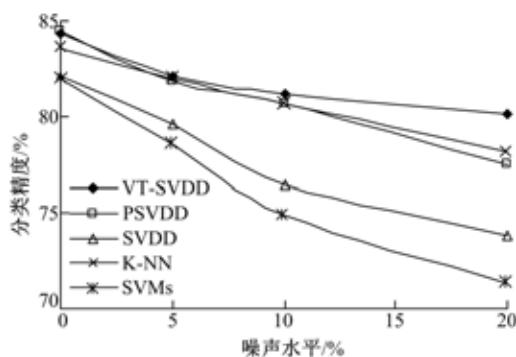


图 1 噪声水平对分类精度的影响

Fig. 1 Influence of noise levels on performance of classifiers

各种分类方法所取得的分类精度的变化趋势。当训练集中的噪声从0增加到20%时,VT-SVDD方法的分类精度变化最小,对噪声的干扰表现最为稳健;其次是K-NN和PSVDD;传统SVDD和SVMs对噪声较为敏感。总之,本文提出的VT-SVDD方法不仅改进了传统SVDD方法的稳健性,与其他常用分类方法相比也是比较稳健的。

4 结束语

在传统SVDD方法中,作为定值的惩罚因子决定了数据描述的精度,然而实践中惩罚因子的选择是极其困难的,为了解决这一矛盾,本文提出了一种可变惩罚因子的稳健支持向量数据描述方法(VT-SVDD)。这种方法基于样本点在核空间的位置分布,为每个样本点计算一个惩罚因子,然后基于这种可变惩罚因子求解一个凸约束二次规划,就可以得到对数据集的球形域描述。在UCI数据集上进行了无噪声、有噪声两个方面的仿真实验,结果表明,VT-SVDD方法提升了传统SVDD方法的精确度和稳健性。

参考文献:

- [1] Tax D M, Duin R P. Support vector data description [J]. Machine Learning, 2004, 54(1):45-66.
- [2] Tax D M, Duin R P. Support vector domain description[J]. Pattern Recognition Letters, 1999, 20(11):1191-1199.
- [3] Lee S-W, Park J, Lee S-W. Low resolution face recognition based on support vector data description [J]. Pattern Recognition, 2006, 39(9):1809-1812.
- [4] Seo J, Ko H. Face detection using support vector domain description in color images[C]// Acoustics, Speech, and Signal Processing(ICASSP04), Montreal, Quebec, Canada, 2004.
- [5] Dong X, Zhaojun W, Wanfeng Z. Support vector domain description for speaker recognition [C]// Proceedings of the 2001 IEEE Signal Processing Society Workshop, North Falmouth, MA, USA, 2001.
- [6] Sjöstrand K, Hansen M S, Larsson H B, et al. A path algorithm for the support vector domain description and its application to medical imaging[J]. Medical Image Analysis, 2007, 11(5):417-428.
- [7] Bu H-g, Wang J, Huang X-b. Fabric defect detection based on multiple fractal features and support vector data description[J]. Engineering Applications of Artificial Intelligence, 2009, 22(2):224-235.
- [8] Lee K, Kim D-W, Lee D, et al. Improving support vector data description using local density degree [J]. Pattern Recognition, 2005, 38(10):1768-1771.
- [9] Guo S-M, Chen L-C, Tsai J S-H. A boundary method for outlier detection based on support vector domain description[J]. Pattern Recognition, 2009, 42(1):77-83.
- [10] Zhang Y, Chi Z-X, Li K-Q. Fuzzy multi-class classifier based on support vector data description and improved PCM[J]. Expert Systems with Applications, 2009, 36(5):8714-8718.
- [11] Wang C-D, Lai J. Position regularized support vector domain description [J]. Pattern Recognition, 2013, 46(3):875-884.
- [12] Blake C L, Merz C J. UCI repository of machine learning databases[EB/OL]. [2012-11-10]. <http://www.ics.uci.edu/mlearn/MLRepository.html>
- [13] Vapnik V. The Nature of Statistical Learning Theory[M]. Berlin: Springer, 1999.