

带拒识能力的双层支持向量模型分类器

胡正平, 张 晔

(哈尔滨工业大学信息工程系, 黑龙江哈尔滨 150001)

摘 要: 本文构造了一种带拒识能力的双层支持向量模型分类器. 在训练学习过程中, 首先对各类样本特征空间求取最小的包含球形边界, 得到各类样本的球形支持向量域表示. 这样对于输入的非目标样本即可利用各类的支持向量域进行拒识或接受处理; 然后针对接受的样本再利用基于超平面分割的 SVM 训练器进行分类判决. 无论是在第一层求取边界的优化问题中, 还是在第二层的分类超平面优化过程中, 都采用相乘性更新迭代规则直接求解, 优化速度与最小二乘支持向量机 (LS SVM) 相当. 仿真实验表明本文提出的通过引入拒绝层和判决层的新支持向量模型策略是合理可行的, 在实际模式识别领域具有广阔的应用前景.

关键词: 支持向量分类器; 核函数; 支持向量域描述; 拒识性能

中图分类号: TP301 文献标识码: A 文章编号: 0372-2112 (2005) 07-1200-04

A Two-Layer Support Vector Classifier with Rejection Feature

HU Zheng ping, ZHANG Ye

(Department of Information Engineering, Harbin Institute of Technology, Harbin, Heilongjiang 150001, China)

Abstract: A two layer support vector classifier model with rejection feature is proposed in this paper. Firstly the sphere support vectors of each class to describe the distribution of the sample were obtained by searching all the sphere boundaries containing the samples of each class. Then the input pattern of non object classes could be rejected by the first support vector domain description (SVDD). If a pattern is accepted by the first SVDD, the second layer of support vector classifier (SVC) with maximum margin between two classes will be used for classification. In addition, Instead of the traditional quadratic programming, multiplicative iterative updates rule is used to solve the optimizing problems in SVDD of the first layer and the SVC in second layer. Compared to the tradition algorithm of the support vector machine, the new method improves greatly the computational speed of optimization. Experimental results demonstrate that the method of two Layer support vector classifier with Rejection Feature is feasible and it could be applicable in many real pattern recognition fields.

Key words: support vector classifier; kernel function; support vector data description; rejection performance

1 引言

支持向量分类器 (Support Vector Classifier, SVC) 因为其优越的性能成为近年研究的热点, 它在逼近与分类应用领域取得了极大的成功. SVC 算法是 Vapnik^[1] 提出的一种有监督学习的两类分类器, 在训练时需要知道两类样本的类别. 但是经典的 SVC 分类器总是假定测试样本属于两类之一, 然而在很多分类识别问题中, 往往存在许多非目标样本, 这时需要分类器带有拒识判决的能力. 这样一方面可以加快识别的速度, 另一方面可以增加识别系统的可靠性.

David M J Tax^[2] 等针对单类数据提出了支持向量域描述 (Support Vector Domain Description, SVDD), 主要用来进行数据描述和剔除奇异点. SVDD 的基本思想就是计算包含一组数据

的最小超球体边界来对数据的分布范围进行描述^[3]. 该方法给我们的启发是: 如果能够知道一部分样本的分布范围, 就可以利用它构造带拒识能力的分类器. SVDD 具有良好的数据选择能力, 但是缺乏对比鉴别分类能力 (因为它仅仅针对一类数据本身); 而 SVC 缺乏数据分布描述能力, 却具有最优的对比鉴别能力. 如何同时吸取二者的优点构造一种新的具有良好拒识能力的分类器是本文工作的出发点.

针对这个问题不少学者提出了合理有效的解决问题的思路. 文献[4] 提出基于 (Support vector representation and discrimination machine) 支持向量表示-鉴别机 (SVRDM) 的具有拒识能力的分类方法, 该方法的优点是一步实现可拒识分类器, 缺点是控制拒绝能力的门限参数选择比较复杂, 不具有自适应能力. 文献[5] 提出另外一种类似于本文的两步实现策略, 首先

将两类归于一类, 第一步决定是否拒绝, 第二步进行分类. 该方法缺点是第一步需要知道非目标样本并进行训练, 一方面计算量巨大, 另一方面在实际应用中往往不容易收集到全部的非目标样本. 而本文提出的方法不需要全部非目标样本, 只需要知道部分两类目标样本本身. 文献[6]提出了基于模板等预处理结合 SVC 的分类方法, 预处理的过程就是数据粗选择过程, 该方法对于可模板化描述的问题是合理有效的, 缺陷是对于不能模板化的数据无能为力, 同时门限选择困难.

常规 SVM 算法大都采用标准性二次优化技术解决凸二次规划问题. 它首先需要计算和存储核矩阵, 需要较大的存储空间, 多数情况下寻优算法占用了时间的主要部分. 近年来人们针对 SVM 方法本身提出许多改进方法, 例如 SMO^[7] (Sequential Minimal Optimization) 算法, 其改进的思路是采用循环迭代: 将原来的问题分解为若干子问题, 按照某种迭代策略, 反复求解子问题, 最终结果收敛到原问题的最优解. Suykens 和 Vandewal^[8] 等提出了 LS SVM (Least Square Support Vector Machine) 最小二乘支持向量机, 只有等式约束而没有 G SVM 的不等式约束, 从而将二次规划问题转化为线性方程组的求解, 大大简化了计算的复杂度. 本文提出的两层分类器中采用相乘性更新规则^[9] 求解二次优化问题. 该方法不需要仔细选择每一步的学习步长, 不需要固定工作子集, 提供了一种新的直接优化方法.

2 系统框图

构造的可拒识双层支持向量模型分类器如图 1 所示, 该系统主要由两个部分组成: 基于 SVDD 的数据选择器和基于最大间隔的超平面鉴别分类器, 各自的支持向量通过样本训练而得到.

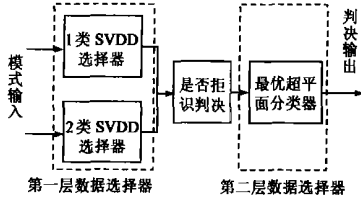


图 1 带拒识能力双层支持向量模型分类器框图

3 带拒识能力的双层支持向量模型分类器

3.1 支持向量域数据选择器

SVDD^[2] 通过寻找包含一组数据的最小半径的超球体对其进行描述, 可见它可用于对一组数据的分布进行描述. 对于奇异的数据点, 它应该位于超球体的外面. 为了减少奇异点数据的影响, 引入松弛变量 ξ 进行描述 (即允许存在错误), 求最小超球体体积可以转化为下面的二次规划问题:

$$\min [R^2 + C \sum_i \xi_i] \quad (1)$$

其约束条件为

$$(x_i - a)^T (x_i - a) \leq R^2 + \xi_i \quad (2)$$

其中

$$\xi_i \geq 0$$

上面的优化问题通过引入拉格朗日系数变为

$$\min L = R^2 + C \sum_i \xi_i - \beta_i \sum_i \xi_i - \sum_i \alpha_i (R^2 + \xi_i - (x_i - a)^T (x_i - a)) \quad (3)$$

其中 $\alpha_i, \beta_i \geq 0$

求式(3)的最小值可转化为如下的对偶问题

$$\max F = \sum_i \alpha_i K(x_i, x_i) - \sum_i \sum_j \alpha_i \alpha_j K(x_i, x_j) \quad (4)$$

其中, $\sum_i \alpha_i = 1, 0 \leq \alpha_i \leq C, K(\cdot)$ 为核函数. 通常情况下, 即使去除奇异点数据, 数据依然不会呈现球状分布, 这时可以通过引入核函数将它映射到更高维空间进行讨论.

对于两类问题 (也可扩展到多类问题中) 的情况, 设有训练样本 $(x_i, y_i) \quad i = 1, 2, \dots, n, y_i \in \{-1, +1\}$, 分别对于 $y_i = 1$ 和 $y_i = -1$ 的样本建立超球体, 通过求解式(4)分别计算其拉格朗日系数对应的支持向量.

如果一测试样本 z 满足如下条件之一, 则这个点被接受, 进行后面的鉴别分类, 否则拒识判决.

$$K(z, z) - 2 \sum_i \alpha_i^{(1)} K(x_i, z) + \sum_{i,j} \alpha_i^{(1)} \alpha_j^{(1)} K(x_i, x_j) \leq R_1^2$$

(第一类数据选择器)

$$K(z, z) - 2 \sum_i \alpha_i^{(2)} K(x_i, z) + \sum_{i,j} \alpha_i^{(2)} \alpha_j^{(2)} K(x_i, x_j) \leq R_2^2$$

(第二类数据选择器) 这里的核参数及其核函数类型可以不同.

3.2 最大间隔超平面分类鉴别器

设 $x_i \{i = 1, \dots, N\}$, P 维特征矢量, $y_i \{+1, -1\}$ 类标记. 如果输入数据线性可分, 决策函数为

$$D(x) = w^T g(x) + b \quad (5)$$

$g(x)$ 为一非线性映射函数, 将 x (P 维) 映射到 l 维空间, w 为 l 维矢量, b 为标量.

对于线性可分的数据, 决策函数满足

$$y_i (w^T g(x_i) + b) \geq 1 \quad i = 1, \dots, N \quad (6)$$

如果训练数据线性不可分, 引入松弛变量 $\xi_i (> 0)$, 决策函数变为

$$y_i (w^T g(x_i) + b) \geq 1 - \xi_i, \quad i = 1, \dots, N$$

$$\xi_i \geq 0 \quad i = 1, \dots, N \quad (7)$$

一阶范数软间隔优化问题对应的拉格朗日函数为

$$\max L = \frac{1}{2} \langle w, w \rangle + C \sum_i \xi_i - \sum_i \beta_i \xi_i - \sum_i \alpha_i [y_i (\langle g(x_i), w \rangle + b) - 1 + \xi_i] \quad (8)$$

式(8)中第一项控制分类器的复杂度 (VC 维), 第二项控制错误分类的经验误差, C 为常数, $\alpha_i, \beta_i \geq 0$. 对偶表示可以通过求对应于 w, ξ, b 偏导, 置零并带入式(8)得到.

$$\max L = \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} y_i y_j \alpha_i \alpha_j \langle g(x_i), g(x_j) \rangle \quad (9)$$

引入核函数, 上式变为

$$\max L = \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} y_i y_j \alpha_i \alpha_j K(x_i, x_j) \quad (10)$$

其中 $\sum_i \alpha_i y_i = 1$ 且 $0 \leq \alpha_i \leq C$

判决输出函数为

$$f(x) = \text{sign}(\sum_i y_i \alpha_i K(x_i, x) + b) \quad (11)$$

3.3 相乘性更新规则求解二次规划问题

上面式(4)和式(10)的非负二次规划优化问题可以归纳为下面的标准形式^[9]

$$\min F(v) = \frac{1}{2} v^T A v + b^T v \quad (12)$$

约束条件为: $v_i \geq 0 \quad \forall i$, 况且矩阵 A 为对称和半正定的. $F(v)$ 的优化为凸二次优化. 由于非负的限制条件, 可以构造一迭代方法求解全局最优.

在式(12)中设 A^+ 和 A^- 表示如下定义的非负矩阵:

$$A_{i,j}^{\pm} = \begin{cases} A_{i,j}, & A_{i,j} > 0 \\ 0, & \text{otherwise} \end{cases} \quad A_{i,j}^{\mp} = \begin{cases} |A_{i,j}|, & A_{i,j} < 0 \\ 0, & \text{otherwise} \end{cases} \quad (13)$$

于是有 $A = A^+ - A^-$, 根据上面定义的非负矩阵, 迭代规则定义为

$$v_i \leftarrow v_i \left[\frac{-b_i + \sqrt{b_i^2 + 4(A^+ v)_i (A^- v)_i}}{2(A^+ v)_i} \right] \quad (14)$$

通过分析式(14)右式可见整个过程不会脱离非负限制. 并且其迭代规则是收敛的, 它可以单调下降到全局最小.

对于式(4)来说, 它是式(12)的特殊情况, 只需令 $b_i = K(x_i, x_i)$, $A_{i,j} = K(x_i, x_j)$, 则式(4)求解迭代更新规则为

$$a_i \leftarrow a_i \left[\frac{-K(x_i, x_i) + \sqrt{K(x_i, x_i)^2 + 4(A^+ a)_i (A^- a)_i}}{2(A^+ a)_i} \right] \quad (15)$$

同样对于式(10)来说, 令 $b_i = -1$, $A_{i,j} = \gamma \delta_j K(x_i, x_j)$, 则式(10)求解迭代更新规则为

$$a_i \leftarrow a_i \left[\frac{-1 + \sqrt{1 + 4(A^+ a)_i (A^- a)_i}}{2(A^+ a)_i} \right] \quad (16)$$

利用上面的式(15)(16)相乘性迭代更新规则求解上面的二次优化问题, 避免了常规 SVM 求解中的矩阵运算, 优化速度得到较大提高, 算法实现简单明了.

4 可拒识分类器选择及其判决规则

本文构造的双层可拒识支持向量分类器最终规则如下,

对于输入检测模式 x , 如果

[1] $x \in \{\text{class1} \cap \text{class2}\}$, 则接受该输入模式, 进入第二层鉴别分类器.

[2] $x \notin \{\text{class1} \cup \text{class2}\}$ 则拒识该输入模式.

[3] $x \in \{\text{class1}\}$ 或 $x \in \{\text{class2}\}$ 但 $x \notin \{\text{class1} \cap \text{class2}\}$, 如果第二层鉴别分类器判决结果与第一层相同, 则确认, 否则拒绝该结果.

5 仿真实验

为了验证本文提出的算法的性能, 本文进行了两组实验, 一组对随机产生的满足高斯分布的三类数据(一类作为非目

标样本)进行实验, 第二组对自然形状的三种树叶(一组代表非目标样本)进行分类实验.

(1) 高斯分布样本点分类实验

随机产生了 3 类 3 维样本的数据点, 每类样本数目 800 点, 满足多元正态分布, 样本点分布参数如表 1 所示.

表 1 合成样本点参数分布

类别	均值 1	均值 2	均值 3	标准偏差
1	0.2	1.0	2.5	1
2	1.0	1.8	0.5	1
3(非目标)	0.6	1.4	1.6	1

实验时, 每一类别中选择 500 点作为训练样本, 300 点作为测试样本. 同时将本文提出的方法与 LS-SVM 算法进行对比, 实验中采用高斯核函数(0.5), 错误惩罚平衡因子 C 等于 8. 对比实验结果如表 2 所示.

表 2 对比实验结果

方法	正确分类率	正确拒识率	训练时间
LS-SVM	81.4%	无	269(s)
拒识分类器 (最小二乘训练)	80.3%	90.5%	341(s)
本文方法	80.6%	90.2%	328(s)

(2) 自然形状树叶分类识别实验

实验选取三种树叶作为实验对象(见图 2). 每类树叶 50 幅, 30 幅作为训练样本, 20 幅作为测试样本, 非目标样本 20 幅(树叶 1 为非目标样本). 本文采用广义 Hough 变换抽取其轮廓特征作为输入特征, 实验结果如表 3 所示.

从上面的实验可以看出, 本文提出的带拒识能力分类器无论对于合成数据还是真实数据既具有较好的拒识性能又具有良好的分类鉴别性能.

表 3 树叶识别性能比较

方法	分类正确率	正确拒识率
LS-SVM	100%	无
拒识分类器 (最小二乘训练)	100%	95%
本文方法	100%	95%

6 结论

基于 SVDD 的球形支持向量具有很好的数据分布描述能力, 能够较好的提供每一类数据分布的紧凑描述; SVM 能够以

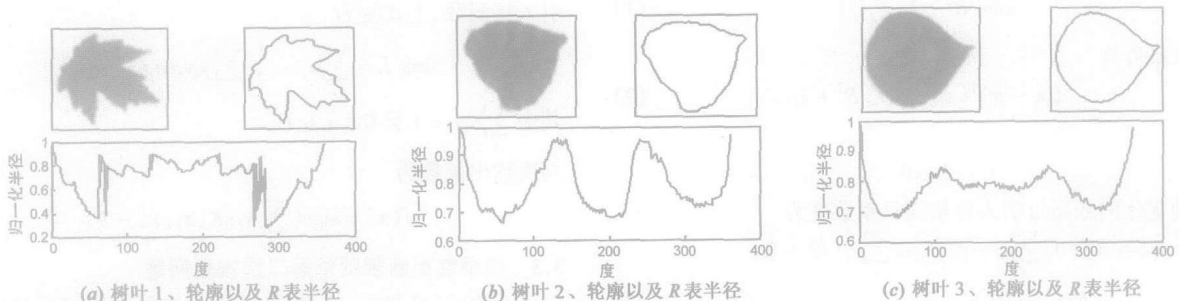


图 2 三种类型树叶及其轮廓表示

最大间隔提供两类数据的分类鉴别能力,其缺点是不能提供数据的分布范围描述. 本文将二者的优点相结合提出了带拒识能力的两层支持向量模型,并且采用相乘性更新迭代规则直接求解第一层和第二层支持向量的解. 由于避免了大量的矩阵运算,算法速度与最小二乘支持向量机(LS-SVM)相近. 本文提出的方法对支持向量机在实际模式识别的具体应用具有一定的指导意义.

参考文献:

- [1] Vapnik. The Nature of Statistical Learning Theory[M]. New York, NY: Wiley, 1998. chapter 5.
- [2] David M J Tax, Robert P W Duin. Support vector data description[J]. Machine Learning, 2004, 54: 45- 66.
- [3] 朱美琳, 刘向东, 陈世福. 用球结构的支持向量机解决多分类问题[J]. 南京大学学报, 2003, 39(2): 153- 158.
Zhu Mei lin, Liu Xiang dong, Chen Shi fu. Solving the problem of multi-class pattern recognition with sphere-structured support vector machines[J]. Journal of Nanjing University, 2003, 39(2): 153- 158(in Chinese).
- [4] Chao Yuan, David Casasent. A novel support vector classifier with better rejection performance[A]. Proceeding of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition[C]. Madison, WI, United States: Institute of Electrical and Electronics Engineers Computer Society, 2003. 1063- 1069.
- [5] S A Rizvi, T N Saasawi, N M Nasrabadi. A clutter rejection technique for FLIR imagery using region based principal component analysis[J]. Pattern Recognition, 2000, 33(11): 1931- 1933.
- [6] Yang G Z, Huang T S. Human face detection in a complex background [J]. Pattern Recognition, 1994, 27(1): 58- 63.

- [7] J C Platt. Fast algorithm for training support vector machines using Sequential minimal optimization[A]. Advance in Kernel Methods support Vector Learning[C]. Massachusetts USA: MIT Press, 1999, 185- 208.
- [8] Suykens J A K, Vandewalle J. Least squares support vector machine classifiers[J]. Neural Processing Letters, 1999, 9(3): 293- 300.
- [9] 陆从德, 张太猛, 胡金燕. 基于乘性规则的支持向量分类器[J]. 计算机学报, 2004, 27(5): 690- 694.
Lu Cong de, Zhang Tai yi, Hu Jin yan. Support vector domain classifier based on multiplicative updates[J]. Chinese Journal of Computers, 2004, 27(5): 690- 694(in Chinese).

作者简介:



胡正平 男, 1970 年出生于四川仪陇, 哈尔滨工业大学博士研究生, 燕山大学教师, 研究方向为统计学习理论与模式识别. E-mail: tnpochw@263.net.



张 晔 男, 1964 年出生于辽宁北镇县, 教授, 博士生导师, 研究方向为遥感图像处理、模式识别.