文章编号:1001 - 9081(2009)03 - 0877 - 03

一种广义最小二乘支持向量机算法及其应用

吴宗亮,窦　衡
(电子科技大学 电子工程学院,成都 610054)
(nanyuanfeixue@ sina. com)

摘　要:最小二乘支持向量机(LS-SVM)是处理不可分样本集情况下模式分类的有效工具,但是该算法在处理很多实际分类问题时,表现出了一定的局限性。为了进一步增强最小二乘支持向量机的推广能力,提出一种通用的广义最小二乘支持向量机算法,并且把这种新算法首先应用到雷达一维距离像的识别中,实验表明新的算法能取得更好的识别效果。

关键词:最小二乘支持向量机;不可分样本集;雷达一维距离像
中图分类号: TP181;TP391.4　　**文献标志码**:A

Generalized least squares support-vector-machine algorithm and its application

WU Zong-liang, DOU Heng
(*College of Electronic Engineering, University of Electronic Sicence and Technology of China, Chengdu Sichuan* 610054, *China*)

Abstract: Least Squares Support-Vector-Machines (LS-SVM) algorithm is an efficient project about pattern classification on unclassifiable sample set condition. While dealing with many factual pattern classification problems, this algorithm reflects certain limitation. A generalized LS-SVM algorithm was introduced to further improve the applicability of LS-SVM. This new method was applied to radar range profile's recognition. The experimental results show that this new method can achieve better recognition effect.

Key words: Least Squares Support Vector Machine (LS-SVM); unclassifiable sample set; radar range profile

0　引言

最小二乘支持向量机[1-3] (Least Squares Support Vector Machine,LS-SVM)是处理不可分样本集情况下模式分类的有效工具,但是该算法在处理各类样本数据分布范围相差很大、各类样本数量悬殊很大以及各类样本错分所造成的危害程度不同等情况时具有一定的局限性。针对上述情况,本文提出了一种通用的广义最小二乘支持向量机算法,以提高 LS-SVM 的实际推广能力。把新算法应用到雷达一维距离像的识别中,确实取得了更好的识别效果。

1　广义最小二乘支持向量机

设训练样本 $\boldsymbol{x}_i \in \mathbf{R}^n, i = 1,2,\cdots,l$ 属于两类,用 $y_i \in \{-1,1\}$ 作为标记,其中正样本子集和负样本子集的样本数分别为 l_1 和 l_2,并且样本总数 $l = l_1 + l_2$。通过一个非线性映射 $\boldsymbol{\Phi}$,将输入的训练样本 $\boldsymbol{x}$ 映射到一个高维特征空间 F 中,即:$\boldsymbol{\Phi}:\mathbf{R}^n \to F,\ \boldsymbol{x} \to \boldsymbol{\Phi}(\boldsymbol{x})$。此时,输入的训练样本由原来的 $\boldsymbol{x}$ 变为 $\boldsymbol{\Phi}(\boldsymbol{x})$。$n$ 维空间中的样本集 $\{\boldsymbol{x}_i\}_{1 \leqslant i \leqslant l}$,由非线性映射到高维特征空间 F 后,得到高维样本集 $\{\boldsymbol{\Phi}(\boldsymbol{x}_i)\}_{1 \leqslant i \leqslant l}$。

1.1　最小二乘支持向量机简介

支持向量机的目标就是寻找一个最优分类超平面[4],尽可能地使得不同类别的样本能够正确分开,其数学表述就是构造一个如下形式的分类器:

$$f(\boldsymbol{x}) = \operatorname{sign}\left(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{\Phi}(\boldsymbol{x}) + b\right) \tag{1}$$

使得样本 $\boldsymbol{x}$ 能够被函数 $f(\boldsymbol{x})$ 正确分类。利用最小二乘支持向量机算法寻找最优分类超平面就是求解下面的优化问题:

$$\min_{w,e} J(w, e) = \frac{1}{2}\boldsymbol{w}^{\mathrm{T}}\boldsymbol{w} + \frac{r}{2}\sum_{i=1}^{l} e_i^2 \tag{2}$$

并且满足等式约束:

$$y_i[\boldsymbol{w}^{\mathrm{T}}\boldsymbol{\Phi}(\boldsymbol{x}_i) + b] = 1 - e_i;\ i = 1, \cdots, l \tag{3}$$

其中 $e_i \in \mathbf{R}$ 是分类误差,$r > 0$ 为惩罚系数,其目的主要是调整分类误差。如果训练数据中含有较大的噪声,应该选择较小的 r。

1.2　广义最小二乘支持向量机的推导过程

标准的最优分类超平面到两类之间的分类间隔是相等的,都为 $1/\|\boldsymbol{w}\|$,这样可以使得结构风险最小化[5],正负两类可以很平等地获得两类之间的区域。但是模式分类经常会遇到各类样本数据分布范围相差很大、各类样本数量悬殊很大以及各类样本错分所造成的危害程度不同等情况,此时如果还应用标准的最优分类超平面就会出现较大的偏差。如果可以适当调整最优分类超平面的分类间隔,就可以较好地解决以上几种分类问题。基于此,本文提出了一种广义最小二乘支持向量机算法,下面给出其推导过程。

在这里,把等式约束(3)修改为:

$$y_i[\boldsymbol{w}^{\mathrm{T}}\boldsymbol{\Phi}(\boldsymbol{x}_i) + b] = \begin{cases} \lambda - e_i,\ y_i = 1 \\ 1 - e_i,\ y_i = -1 \end{cases} \tag{4}$$

此时正负两类的分类间隔为:

$$\frac{\lambda}{\|\boldsymbol{w}\|} + \frac{1}{\|\boldsymbol{w}\|} = \frac{1 + \lambda}{\|\boldsymbol{w}\|} \tag{5}$$

优化问题也转变为:

$$\min_{\boldsymbol{w},\boldsymbol{e}} J(\boldsymbol{w}, \boldsymbol{e}) = \frac{1}{1 + \lambda}\boldsymbol{w}^{\mathrm{T}}\boldsymbol{w} + \frac{r}{2}\sum_{i=1}^{l} e_i^2 \tag{6}$$

收稿日期:2008 - 09 - 22;**修回日期**:2008 - 11 - 11。

作者简介:吴宗亮(1979 -),男,陕西宝鸡人,硕士研究生,主要研究方向:信号处理、模式识别、雷达目标识别;　窦衡(1964 -),男,四川乐山人,副教授,主要研究方向:信号处理、模式识别、微波成像。

其中 $\lambda > 0$，是事先确定的常数，目的就是控制分类超平面与正负两类之间的间隔比例。当 $0 < \lambda < 1$ 时，分类超平面靠近正样本类；当 $\lambda > 1$ 时，分类超平面靠近负样本类；当 $\lambda = 1$ 时，即为标准的 LS-SVM。

为了求目标函数 J 的最小值，构造 Lagrange 函数 L 如下：

$$\begin{aligned} L = & J(\boldsymbol{w}, \boldsymbol{e}) - \sum_{y_i=1} a_i\left\{y_i\left[\boldsymbol{w}^{\mathrm{T}}\boldsymbol{\Phi}(\boldsymbol{x}_i) + b\right] + e_i - \lambda\right\} - \\ & \sum_{y_i=-1} a_i\left\{y_i\left[\boldsymbol{w}^{\mathrm{T}}\boldsymbol{\Phi}(\boldsymbol{x}_i) + b\right] + e_i - 1\right\} = \\ & J(\boldsymbol{w}, \boldsymbol{e}) - \sum_{y_i=1} a_i\left\{y_i\left[\boldsymbol{w}^{\mathrm{T}}\boldsymbol{\Phi}(\boldsymbol{x}_i) + b\right] + e_i\right\} - \\ & \sum_{y_i=-1} a_i\left\{y_i\left[\boldsymbol{w}^{\mathrm{T}}\boldsymbol{\Phi}(\boldsymbol{x}_i) + b\right] + e_i\right\} + \sum_{y_i=1}\lambda a_i + \sum_{y_i=-1} a_i = \\ & J(\boldsymbol{w}, \boldsymbol{e}) - \sum_{i=1}^{l} a_i\left\{y_i\left[\boldsymbol{w}^{\mathrm{T}}\boldsymbol{\Phi}(\boldsymbol{x}_i) + b\right] + e_i\right\} + \sum_{y_i=1}\lambda a_i + \\ & \sum_{y_i=-1} a_i \end{aligned} \tag{7}$$

其中 a_i 是 Lagrange 乘子，因为是等式约束，所以其值可正可负。在这里，函数 L 首先对 b 求偏导，且令其偏导数等于0，可得：

$$\frac{\partial L}{\partial b} = 0 \Rightarrow \sum_{i=1}^{l} a_i y_i = 0 \Rightarrow \sum_{y_i=1} a_i = \sum_{y_i=-1} a_i \tag{8}$$

所以可以得到：

$$\sum_{i=1}^{l} a_i = \sum_{y_i=1} a_i + \sum_{y_i=-1} a_i = 2\sum_{y_i=1} a_i = 2\sum_{y_i=-1} a_i \tag{9}$$

把式(9)代入式(7)可以得到：

$$\begin{aligned} L = & J(\boldsymbol{w}, \boldsymbol{e}) - \sum_{i=1}^{l} a_i\left\{y_i\left[\boldsymbol{w}^{\mathrm{T}}\boldsymbol{\Phi}(\boldsymbol{x}_i) + b\right] + e_i\right\} + \\ & \frac{1+\lambda}{2}\sum_{i=1}^{l} a_i \end{aligned} \tag{10}$$

然后，函数 L 分别对 $\boldsymbol{w}$、e_i 以及 a_i 求偏导，且令偏导数等于0，可得：

$$\frac{\partial L}{\partial \boldsymbol{w}} = 0 \Rightarrow \boldsymbol{w} = \frac{1+\lambda}{2}\sum_{i=1}^{l} a_i y_i \boldsymbol{\Phi}(\boldsymbol{x}_i) \tag{11}$$

$$\frac{\partial L}{\partial e_i} = 0 \Rightarrow r e_i = a_i \Rightarrow e_i = \frac{a_i}{r} \tag{12}$$

$$\frac{\partial L}{\partial a_i} = 0 \Rightarrow y_i\left[\boldsymbol{w}^{\mathrm{T}}\boldsymbol{\Phi}(\boldsymbol{x}_i) + b\right] + e_i - \frac{1+\lambda}{2} = 0 \tag{13}$$

其中 $i = 1, \cdots, l$，把式(11)(12)代入式(13)，消掉 $\boldsymbol{w}$ 和 e_i，并且利用满足 Mercer 条件的内积核函数[6]，即：$k(\boldsymbol{x}_i, \boldsymbol{x}_j) = (\boldsymbol{\Phi}(\boldsymbol{x}_i)\cdot\boldsymbol{\Phi}(\boldsymbol{x}_j)) = \boldsymbol{\Phi}(\boldsymbol{x}_i)^{\mathrm{T}}\boldsymbol{\Phi}(\boldsymbol{x}_j)$，$1 \leqslant i, j \leqslant l$。式(13)经过整理转变为：

$$\sum_{j=1}^{l} a_j y_i y_j k(\boldsymbol{x}_i, \boldsymbol{x}_j) + \frac{2}{r(1+\lambda)}a_i + \frac{2}{1+\lambda}b y_i = 1 \tag{14}$$

其中 $i = 1, \cdots, l$，把式(8)经过适当的变形，可以得到：

$$\frac{2}{1+\lambda}\sum_{i=1}^{l} a_i y_i = 0 \tag{15}$$

把式(14)、(15)改写成如下矩阵形式：

$$\begin{bmatrix} 0 & \boldsymbol{Y}^{\mathrm{T}} \\ \boldsymbol{Y} & \boldsymbol{A} \end{bmatrix}\begin{bmatrix} b \\ \boldsymbol{a} \end{bmatrix} = \begin{bmatrix} 0 \\ \vec{1} \end{bmatrix} \tag{16}$$

$\boldsymbol{Y} = \frac{2}{1+\lambda}[y_1, y_2, \cdots, y_l]^{\mathrm{T}}$，$\boldsymbol{a} = [a_1, a_2, \cdots, a_l]^{\mathrm{T}}$ 以及 $\vec{1} = [1, 1, \cdots, 1]^{\mathrm{T}}$，在这里重点交代一下 $\boldsymbol{A}$ 的组成：

$$\boldsymbol{A} = \boldsymbol{\Omega} + \frac{2}{r(1+\lambda)}\boldsymbol{I} \tag{17}$$

其中 $\boldsymbol{I}$ 是 l 阶单位矩阵，$\boldsymbol{\Omega} = \left(y_i y_j k(\boldsymbol{x}_i, \boldsymbol{x}_j)\right)$，$1 \leqslant i, j \leqslant l$。在这里，可以发现广义最小二乘支持向量机与标准的最小二乘支持向量机的形式完全相同，不同的是向量 $\boldsymbol{Y}$ 和矩阵 $\boldsymbol{A}$ 的构成有所不同。如果在这里令 $\lambda = 1$，那么广义最小二乘支持向量机就退化成了标准的最小二乘支持向量机。

求解方程组(16)可以得到：

$$b = \frac{\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{A}^{-1}\vec{1}}{\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{Y}},\ \boldsymbol{a} = \boldsymbol{A}^{-1}(\vec{1} - \boldsymbol{Y}b)。$$

由于 $\boldsymbol{A}$ 是一个半正定对称矩阵，一般情况下其逆是存在的。

最后给出两种广义最小二乘支持向量机的分类函数。第一种为：

$$f(\boldsymbol{x}) = \mathrm{sign}\left[\frac{1+\lambda}{2}\sum_{i=1}^{l} a_i y_i k(\boldsymbol{x}_i, \boldsymbol{x}) + b\right] \tag{18}$$

在这种情况下，广义最小二乘支持向量机的关键参数 λ 的初始值设置相当重要，要视具体问题而定。若各类样本的类内离散度相差很大，即各类样本的分布范围相差很大，则可根据类内离散度来确定 λ 的值；若各类样本数目相差很大，则可根据正负样本数的比值来确定 λ 的值；若实际中考虑到各类样本错分所造成的损失及危害，则可根据危害程度来确定 λ 的值。

针对各类样本分布范围相差很大的情况，这里给出一种快速简捷的 λ 的估值方法。在高维特征空间中，令正类的训练样本集为 $\{\boldsymbol{\Phi}(\boldsymbol{x}_i)\}$，$i = 1, 2, \cdots, l_1$，那么其中心点为：

$$\boldsymbol{\Phi}_s = \frac{1}{l_1}\sum_{i=1}^{l_1}\boldsymbol{\Phi}(\boldsymbol{x}_i) \tag{19}$$

正样本集的一个点到其中心点的期望平方距离 d_1 可表示为：

$$\begin{aligned} d_1 = & \frac{1}{l_1}\sum_{s=1}^{l_1}\|\boldsymbol{\Phi}(\boldsymbol{x}_s) - \boldsymbol{\Phi}_s\|^2 = \\ & \frac{1}{l_1}\sum_{s=1}^{l_1}k(\boldsymbol{x}_s, \boldsymbol{x}_s) + \frac{1}{l_1^2}\sum_{i,j=1}^{l_1}k(\boldsymbol{x}_i, \boldsymbol{x}_j) - \frac{2}{l_1^2}\sum_{i,s=1}^{l_1}k(\boldsymbol{x}_i, \boldsymbol{x}_s) = \\ & \frac{1}{l_1}\sum_{s=1}^{l_1}k(\boldsymbol{x}_s, \boldsymbol{x}_s) - \frac{1}{l_1^2}\sum_{i,j=1}^{l_1}k(\boldsymbol{x}_i, \boldsymbol{x}_j) \end{aligned} \tag{20}$$

令 $\boldsymbol{\Psi} = \{\boldsymbol{\Phi}(\boldsymbol{x}_1), \boldsymbol{\Phi}(\boldsymbol{x}_2), \cdots, \boldsymbol{\Phi}(\boldsymbol{x}_{l_1})\}$，那么：

$$\boldsymbol{K} = \boldsymbol{\Psi}^{\mathrm{T}}\boldsymbol{\Psi} \tag{21}$$

称 $\boldsymbol{K}$ 为核矩阵，所以 d_1 本质上是核矩阵对角元素的平均值减去所有元素的平均值。d_1 越大，正样本集越分散，d_1 越小，正样本集越集中。负样本集的一个点到其中心点的期望平方距离 d_2 的计算方式同上，此时 λ 初始值可以设为 d_1/d_2 的平方根。λ 的初始值设定也可以有其他方法，不过上述方法简捷，容易计算。当然具体在分类识别的时候可以不断调整 λ 的值，以达到最佳的识别效果。

第二种分类函数为：

$$f(\boldsymbol{x}) = \mathrm{sign}\left[\sum_{i=1}^{l} a_i y_i k(\boldsymbol{x}_i, \boldsymbol{x}) + b\right] \tag{22}$$

这种分类函数套用了标准最小二乘支持向量机的分类函数，主要是为了通过调整 λ 达到调整参数向量 $\boldsymbol{a}$ 以及最优分类超平面法向量 $\boldsymbol{w}$ 的目的，从而取得更好的识别效果。在这种情况下，参数 λ 的取值以1为界线，左右调整，以达到最佳的识别效果。

需要说明的是,在该文中采取高斯径向基核函数:

$$k(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp\left(-\|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2/2\sigma^2\right) \tag{23}$$

2 实验结果

本文实验所采取的实测数据是 ISAR 雷达对空中的两种飞机(a, b)所成的距离像。ISAR 交替发射窄带(脉冲时宽为 1 μs)和宽带两种波形:窄带系统主要用于跟踪目标和产生宽带本振定时信号;宽带信号的带宽为 400 MHz(理论上距离分辨率为0.375 m),采样点数为256(其经过 FFT 后所得一维距离像的像点数也为 256)。每种飞机录取了 7 段数据,每段数据含 26000 个宽、窄带信号(相邻间隔 2.5 ms)。宽带信号为全去斜后的正交双通道信号(其 FFT 即一维距离像),每段数据含 260 个宽带正交双通道信号,即每段数据有 260 幅距离像。实验数据为两种飞机各取一段的 130 幅距离像(总数为 260 幅)作训练样本,另外 130 幅为测试样本(总数为 260 幅)。在对实测数据训练前,作如下两步预处理。

1)归一化。将每一幅图像用其总能量归一。

2)距离对准。利用 Fourier 变换的平移不变性,将目标一维距离像作 Fourier 变换即可对齐,同时根据实数 Fourier 变换的共轭对称性,可取一维距离像 Fourier 变换的一半(128 维)作为识别输入矢量进行实验。

在下面实验中,参数设置为:$r = 100$, $\sigma^2 = 0.2$。

实验一　利用广义最小二乘支持向量机的第一种分类函数进行实验,当 $\lambda = 1.0$ 时为标准的最小二乘支持向量机,表 1 给出其实验结果。

表 1　第一种分类函数的实验结果　　%

λ 取值	a 正确识别率	b 正确识别率	平均正确识别率
1.0	96.15	97.69	96.92
3.1	96.15	99.23	97.69
3.2	96.15	100.00	98.08
3.3	96.15	100.00	98.08

实验二　利用广义最小二乘支持向量机的第二种分类函数进行实验,当 $\lambda = 1.0$ 时为标准的最小二乘支持向量机,表 2 给出其实验结果。

在这里,令对a类飞机的正确识别率为$f_a(\lambda)$,对b类飞机的正确识别率为$f_b(\lambda)$,最好的参数 λ 就是$f_a(\lambda) + f_b(\lambda)$的最大值所对应的 λ。当然,本文给出了 λ 值的初始化方法,但是在实验中,还需要不断对其进行调整,以达到最佳的识别效果。

通过表 1 可以看出,广义最小二乘支持向量机算法在处理实际分类问题时,确实具有较大的优势。从表 2 的实验结果可以看出,随着 λ 由小到大取值,某一类的识别率在上升,而另一类的识别率恰恰在下降,而最优的识别效果并不是在 $\lambda = 1.0$(即标准的最小二乘支持向量机)处,从而进一步印证了广义最小二乘支持向量机的优越性。

表 2　第二种分类函数的实验结果　　%

λ 取值	a 正确识别	b 正确识别率	平均正确识别率
0.90	94.62	100.00	97.31
0.95	96.15	99.23	97.69
1.00	96.15	97.69	96.92
1.10	96.15	96.92	96.54
1.20	96.92	96.15	96.54
1.30	97.69	94.62	96.16

3 结语

本文针对最小二乘支持向量机的不足之处,提出了广义最小二乘支持向量机算法,并对其进行了详尽的阐述。该方法完全可以应用到计算机或雷达图像处理、DNA 检测、故障诊断以及其他模式识别等领域。

参考文献:

[1] 姜静清. 最小二乘支持向量机算法及应用研究[D]. 长春: 吉林大学, 2007.

[2] 陈爱军. 最小二乘支持向量机及其在工业过程建模中的应用[D]. 杭州: 浙江大学, 2006.

[3] 朱家元, 陈开陶, 张恒喜. 最小二乘支持向量机算法研究[J]. 计算机科学, 2003, 30(7): 157－159.

[4] THEODORIDIS S, KOUTROUMBAS K. 模式识别[M]. 3 版. 李晶皎, 王爱侠, 张广渊, 等译. 北京: 电子工业出版社, 2006: 60－75.

[5] 边肇祺, 张学工. 模式识别[M]. 2 版. 北京: 清华大学出版社, 2000: 295－296.

[6] 田盛丰. 基于核函数的学习算法[J]. 北方交通大学学报, 2003, 27(2): 1－8.

(上接第 870 页)

综合来看,所有的最优值都出现在采用集成学习的方法中,只要根据问题的不同选择合适的集成方式和参数就可以获得较好的效果。

参考文献:

[1] KENNEDY J, EBERHART R C. Particle swarm optimization [C]// Proceedings of IEEE International Conference on Neural Networks. Piscataway, NJ: IEEE Service Center, 1995: 1942－1948.

[2] ENGELBRECHT A P. Computational intelligence: An introduction [M]. Chichester, England: John Wiley & Sons, 2002.

[3] SHARKEY A J C. Combining artificial neural nets: Ensemble and modular multi-net systems [M]. London: Springer-Verlag, 1999.

[4] WU JIAN-XIN, ZHOU ZHI-HUA, CHEN ZHAO-QIAN. Ensemble of GA based selective neural network ensembles [C]// Proceedings of the 8th International Conference on Neural Information Processing (ICONIP'01). Shanghai: Fudan University Press, 2001, 3: 1477－1482.

[5] WHITTAKER R. A fast algorithm for the greedy interchange for large-scale clustering and median location problems [J]. INFOR, 1983(21): 95－108.

[6] SHI Y, EBERHART R C. Empirical study of particle swarm optimization [C]// Proceedings of the 1999 Congress on Evolutionary Computation. Piscataway, NJ: IEEE Service Center, 1999: 1945－1950.

[7] SHI Y, EBERHART R. Fuzzy adaptive particle swarm optimization [C]// Proceedings of IEEE International Conference on Evolutionary Computation. Piscataway, NJ: IEEE Press, 2001: 101－106.

[8] XIE XIAO-FENG, ZHANG WEN-JUN, YANG ZHI-LIAN. Adaptive particle swarm optimization on individual level [C]// International Conference on Signal Processing (ICSP). [S. l.]: IEEE, 2002: 1215－1218.