2010 年 7 月 重庆师范大学学报(自然科学版) Jul. 2010
第 27 卷 第 4 期 Journal of Chongqing Normal University(Natural Science) Vol. 27 No. 4

DOI :10.3969/J. ISSN. 1672-6693. 2010. 04. 017

一种改进的最小二乘支持向量机算法*

万 辉[1], 魏 延[2]

(1. 重庆师范大学 科研处;2. 信息科学与工程学院, 重庆 400047)

摘要 :最小二乘支持向量机是标准支持向量机的一种扩展,它是支持向量机在二次损失函数下的一种形式。它用等式约束代替不等式约束,求解过程变为解一组等式方程,避免了求解耗时的二次规划问题,但同时也丧失了标准支持向量机的稀疏性,影响了二次学习的效率。针对上述问题,本文提出了一种改进的最小二乘支持向量机增量学习方法。改进的最小二乘支持向量机算法采用自适应剪枝方法对解进行稀疏,根据每次训练得到的分类器性能来设定剪枝阈值和样本增量的大小,如果得到的分类器性能好,剪枝阈值和样本增量就大,反之,剪枝阈值和样本增量就小,从而提高了最小二乘支持向量机训练效率,解决了稀疏性问题。最后,仿真实验表明该算法方案可行。

关键词 :最小二乘支持向量机 ;增量学习 ;稀疏性

中图分类号 :TP391.41 文献标识码 :A 文章编号 :1672-6693(2010)04-0069-04

支持向量机是 Vapnik[1] 等人提出的一类新型机器学习方法,已在很多领域中得到了成功应用[2]。但是该方法的具体实现算法在计算上存在着一些问题,包括训练算法速度慢、算法复杂而难以实现以及测试阶段运算量大等。为了减少支持向量机学习方法的计算复杂性,笔者利用支持向量机优化问题本身所具有的特性,提出了改进的支持向量算法[3-5]。

最小二乘支持向量机是标准支持向量机的一种扩展,它是支持向量机在二次损失函数下的一种形式。它用等式约束代替不等式约束,求解过程变为解一组等式方程,避免了求解耗时的二次规划问题,但同时也丧失了标准支持向量机的稀疏性。最小二乘支持向量机算法中几乎所有样本都成为支持向量,因此,在处理大样本数据情况时将无法得到满意的结果。故本文将采用增量的学习方法,用最小二乘支持向量机对大样本数据进行学习训练,有效地解决了处理大样本数据的问题。

1 最小二乘支持向量机

Suykens 在 1999 年提出了最小二乘支持向量机,其主要思想是把最小二乘的思想引入了支持向量机,但它与支持向量机都是要构造一个能够使间隔最大的超平面,使各类中离分类线最近的样本之间的距离最大。最小二乘支持向量机和支持向量机相比有不同的优化函数,这是因为它对每一个数据点都加入一个改正量 e_i,使不等式约束变为等式约束,故只需解线性等式方程组,计算量减少,最重要的是避免了支持向量机中的惩罚因子 C 值的选择问题。对于支持向量机在样本集中自动选取支持向量,随着参数的改变,支持向量也随着改变,并且只有支持向量对应的 Lagrange 乘子 α_i 不为零。而最小二乘支持向量机所有的样本都当作了支持向量,每一个样本对应一个 α_i,且 α_i 可正可负,这是由于目标函数采用了 2-范数 $\sum_{i=1}^{l} e_i^2$,支持值大小与训练点处的误差成正比 $\alpha_i = \gamma \cdot e_i$。

求解最小二乘支持向量机优化超平面,即为求解下面的凸优化问题的解:

$$\min J_{LS}(w,b,e) = \frac{1}{2}w^{\mathrm{T}}w + \frac{\gamma}{2}\sum_{i=1}^{l} e_i^2 \tag{1}$$

$$\text{s.t. } y_i(w^{\mathrm{T}}\varphi(x_i) + b) = 1 - e_i,\ i = 1,\cdots,l \tag{2}$$

定义 Lagrange 函数:

$$L(w,b,e;\alpha) = J_{LS} - \sum_{i=1}^{l}\alpha_i\left\{y_i 1\left[\ w^{\mathrm{T}}\varphi(x_i) + b\right] - 1 + e_i\right\} \tag{3}$$

其中 α_i 为 Lagrange 乘子,可正可负。根据 Karush-Kuhn-Tucker(KKT) 条件

* 收稿日期 :2009-09-11 修回日期 :2009-10-27
作者简介 :万辉,男,助理研究员,硕士,研究方向为机器学习、图像处理。

$$\begin{cases}\dfrac{\partial L}{\partial w}=0\Rightarrow w=\sum\limits_{i=1}^{l}\alpha_i y_i\varphi(x_i)\\ \dfrac{\partial L}{\partial b}=0\Rightarrow\sum\limits_{i=1}^{l}\alpha_i y_i=0\\ \dfrac{\partial L}{\partial e_i}=0\Rightarrow\alpha_i=\gamma e_i\\ \dfrac{\partial L}{\partial \alpha_i}=0\Rightarrow y_i(w^{\mathrm{T}}\cdot\varphi(x_i)+b)-1+e_i=0\end{cases}\tag{4}$$

容易得出矩阵方程

$$\begin{bmatrix} I & 0 & 0 & -Z^{\mathrm{T}}\\ 0 & 0 & 0 & -Y^{\mathrm{T}}\\ 0 & 0 & \gamma\cdot I & -I\\ Z & Y & I & 0\end{bmatrix}\begin{bmatrix} w\\ b\\ e\\ \alpha\end{bmatrix}=\begin{bmatrix}0\\0\\0\\ \vec{1}\end{bmatrix}\tag{5}$$

其中

$$Z=[\varphi(x_1)^{\mathrm{T}}y_1,\cdots,\varphi(x_l)^{\mathrm{T}}y_l]$$
$$Y=[y_1,\cdots,y_l]$$
$$e=[e_1,\cdots,e_l]$$
$$\alpha=[\alpha_1,\cdots,\alpha_l]$$
$$\vec{1}=[1,\cdots,1]$$

忽略无须计算的 w,e 可得到

$$\begin{bmatrix}0 & Y^{\mathrm{T}}\\ Y & Z\cdot Z^{\mathrm{T}}+\dfrac{I}{\gamma}\end{bmatrix}\begin{bmatrix} b\\ \alpha\end{bmatrix}=\begin{bmatrix}0\\ \vec{1}\end{bmatrix}\tag{6}$$

将(6)式展开成线性方程组的形式,得

$$\begin{cases}\sum\limits_{i=1}^{l}\alpha_i y_i=0\\ \sum\limits_{i=1}^{l}\left[\alpha_i y_i y_j\varphi(x_i)\varphi(x_j)+by_i+\dfrac{\alpha_i}{\gamma}\right]=1 \quad j=1,\cdots,l\end{cases}\tag{7}$$

根据 Mercer 条件,(7)式可变换为

$$\begin{bmatrix}0 & y_1 & \cdots & y_l\\ y_1 & y_1y_lK(x_1,x_l)+\dfrac{1}{\gamma} & \cdots & y_1y_lK(x_1,x_l)\\ \square & \square & \square & \square\\ y_l & y_ly_1K(x_l,x_1) & \cdots & y_ly_lK(x_l,x_l)+\dfrac{1}{\gamma}\end{bmatrix}\cdot\begin{bmatrix}b\\ \alpha_1\\ \square\\ \alpha_l\end{bmatrix}=\begin{bmatrix}0\\1\\ \square\\ 1\end{bmatrix}\tag{8}$$

分析(8)式知,如果训练样本集包含 l 个样本点,那么(8)式有 $l+1$ 个未知数,$l+1$ 个方程以及 $(l+1)^2$ 个乘数项系数。可以看到,样本点的数量直接决定了(8)式的系数矩阵。

将 Mercer 条件应用于 $\Omega=ZZ^{\mathrm{T}}$,而

$$\Omega_j=y_iy_j\varphi(x_i)^{\mathrm{T}}\varphi(x_j)=y_iy_jK(x_i,x_j)\tag{9}$$

由此可见,(1)式的分类问题可以通过求解(6)式的线性方程组获得,从而避免了求解二次规划问题。当(6)式中的系数矩阵满秩时,KKT 系统是一个平方系统,有唯一解。

最终可以求得最小二乘支持向量机分类器

$$y(x)=\mathrm{sign}\left[\sum\alpha_iy_iK(x,x_i)+b\right]\quad i=1,\cdots,l\tag{10}$$

其中 α,b 为方程组(6)的解,$K(\cdot)$ 为核函数。

2 增量学习方法[6-7]

对于传统的支持向量机,其对偶问题的最优解 $\alpha=[\alpha_1,\cdots,\alpha_l]$ 使每个样本 x 满足优化问题的 KKT 条件为

$$\alpha_i=0\Rightarrow y_if(x_i)>1\tag{11}$$
$$0<\alpha_i<c\Rightarrow y_if(x_i)=1\tag{12}$$
$$\alpha_i=c\Rightarrow y_if(x_i)\leqslant 1\tag{13}$$

其中当 Lagrang 乘子 $\alpha=0$ 对应的样本分布在分类器分类间隔之外,$0<\alpha<c$ 的样本位于分类间隔之上,$\alpha=c$ 位于分类间隔之内。非零的 α_i 对应的样本为支持向量,即 $f(x)=\pm 1$ 分类间隔边界上的样本为支持向量。

与传统的支持向量机相比,最小二乘支持向量机的 Lagrange 乘子 $\alpha_i=\gamma e_i$,α 中的所有元素都不为零,几乎所有样本都成为支持向量。为了得到稀疏的支持向量集,Suykens 在文献[8]中提出一种修剪算法,按照 α_i 的绝对值的大小决定训练集中数据向量(支持向量)的重要程度,绝对值大的就对应传统支持向量机中的支持向量即在分类间隔之上或是在分类间隔之内,而绝对值小的可以认为在分类器分类间隔之外的是可以去掉的。

增量学习的主要思想在传统支持向量机中可表述为,增量训练由初步支持向量样本和新的样本组成,再训练一次,所得的非支持向量样本点被抛弃。这种算法也可以看作是一种加强寻找支持向量的过程。该算法还可形式化描述为对一个支持向量机分类器 Ω,设初始训练集为 IS,增量训练集为 INS,那么重复学习可表示为 $Train(IS\cup INS)$,增量学习可以表示为 $Train(Sub(IS)\cup INS)$,其中 $Sub(IS)\subset IS$。

对于最小二乘支持向量机增量学习思想与传统

支持向量机相似,只是传统支持向量机的支持向量的 Lagrage 乘子 α_i 为非零,即为零的 α_i 对应的样本被抛弃;而最小二乘支持向量机增量学习中,被抛弃的是对应 Lagrage 乘子 α_i 绝对值较小的样本,即在分类中重要程度较小的支持向量。

3 最小二乘支持向量机改进算法

为了减少支持向量机学习方法的计算复杂性,利用支持向量机优化问题本身所具有的特性,提出了一些改进的支持向量算法。例如 Vapnik 提出的 Chunking 方法,即所谓的"块算法";Osuna 等人提出的固定样本数方法;Platt 提出的著名的 SMO(Sequential minimal optimization)算法。Chunking 算法中其循环迭代次数较少,但其每次循环求解二次规划的规模较大,而且其规模不能控制,因此其训练时间较长。而在固定训练样本集方法(Osuna 法及 SMO 方法)中,过分地强调了每次循环所求解的二次规划的规模,因此其每次循环所需要的时间较短,但总的循环次数大大增加,其总的训练时间也较长。

Suyken 针对支持向量机能够对大样本情况进行学习,提出了最小二乘支持向量机方法,但相比传统支持向量机,最小二乘支持向量机丧失了解的稀疏性,这势必影响样本点再次学习的效率。为此,Suyken 等人提出了一种简单易行的剪枝方法使最小二乘支持向量机重现传统支持向量机的稀疏性。

受上述方法的启示,本文提出了如下一种改进的最小二乘支持向量机增量学习方法,基本思想是:采用自适应方法确定剪枝阈值和样本增量的大小,如果得到的分类器性能好,剪枝阈值和样本增量就大,反之,剪枝阈值和样本增量就小。

3.1 改进的最小二乘支持向量机增量学习算法

1)对给定大小为 n 的训练样本集 N,取大小为 n_1 的样本集 N_1 作为初始样本。

2)应用最小二乘支持向量机算法对初始样本 N_1 进行训练,得到分类器 Γ_1 和样本对应的 Lagrange 乘子 α_1。

3)把剩下的 $N-N_1$ 个样本作为 Γ_1 的测试样本进行训练,得到违反 KKT 条件(错分)的样本。

4)计算出训练样本集 N_1 对应的 Lagrange 乘子中值 α_1,根据错分样本占测试样本的比例 l_1,选取剪枝域值 $\alpha''_1(\alpha''_1=\alpha'_1\times(1-l_1))$ 进行剪枝,得到舍去的样本个数 r_1。

5)给定一固定值 M,得到添加到剪枝后样本集中的样本增量 $M_1(M_1=M+r_1)$,组成新的样本集 N_2。

6)$N_1=N_2$,返回到第二步,直到所有样本都进行了训练,得到最后分类器 Γ。

3.2 仿真试验

本试验采用了一个具有 170 个随机产生的两类二维样本,在工程数学软件 Matlab 中做仿真试验。实验环境为 Window XP 操作系统,CPU 主频 1.79 GHz,内存为 384 M,Matlab 版本 7.0.1。

采用 Matlab 中 LS_SVMlab1.5[9] 工具箱,结合本文提出的算法,取 RBF 核函数,$\sigma^2=0.2$,$\gamma=10$,$M=20$。得到如下仿真结果。

图 1(a)是从给定训练样本集 N 中取大小为 50 的初始样本集 N_1,应用最小二乘支持向量机算法对其训练的结果,图 1(b)是根据本文算法剪枝后的训练结果,$l_1=0.132$,$\alpha''_1=0.503\ 4$。

图 2(a)是二次迭代后训练得到的结果,图 2(b)是通过对训练样本剪枝后得到的训练结果,$l_1=0.128$,$\alpha''_1=0.553\ 6$。

图 3(a)为给定样本集 N 应用最小二乘支持向量机算法对其训练的结果,图 3(b)是本文算法最终训练得到的结果,从两图对比的结果可以看出,本文提出的算法对样本集 N 剪枝后所得分类超平面的性能指标没有明显下降。

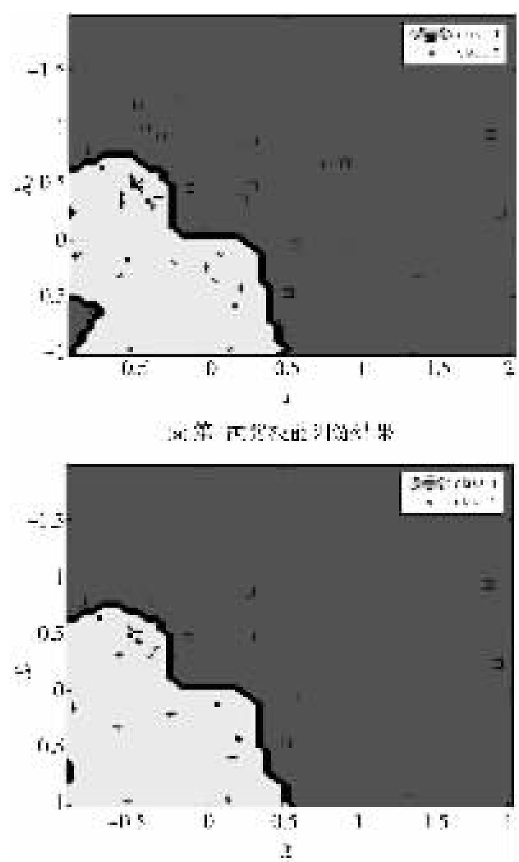

图 1　第 1 次训练结果

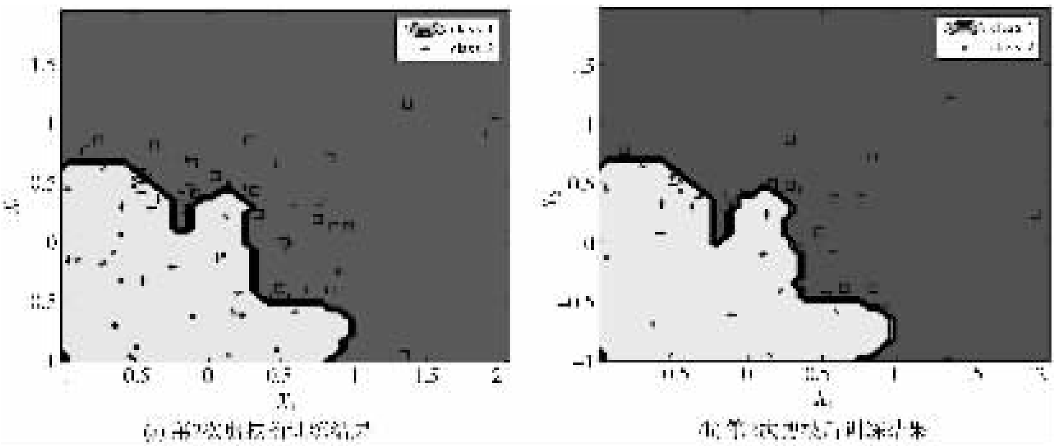

图2 第2次训练结果

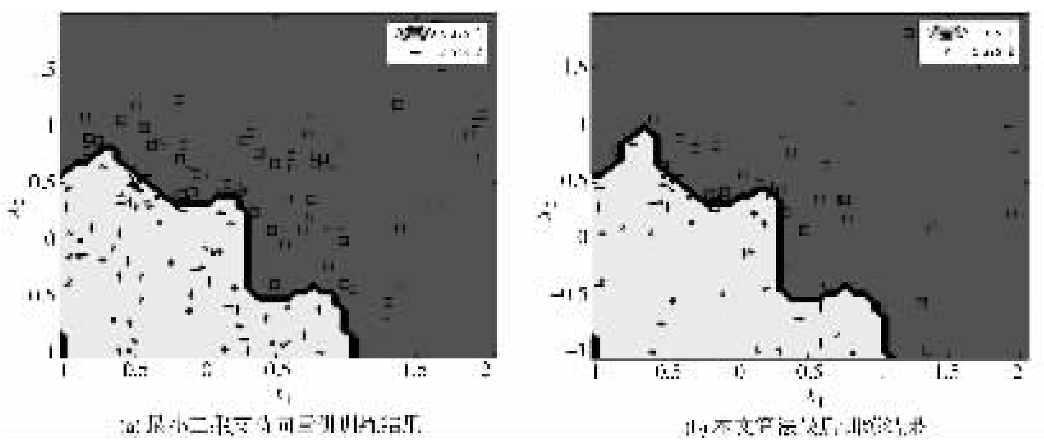

图3 最后训练结果

4 结语

最小二乘支持向量机相比传统的支持向量机,丧失了解的稀疏性,影响了二次学习的效率,针对此提出了一种改进的最小二乘支持向量机增量学习方法,采用自适应方法对样本进行剪枝,根据得到的分类器性能来设定剪枝阈值和样本增量的大小,提高了对大样本学习的速度。实验结果表明所得分类超平面性能指标没有明显下降。

参考文献:

[1] 张学工. 关于统计学习理论与支持向量机[J]. 自动化学报,2000,26(1):32-42.

[2] 曾绍华,魏延. 供应商评价的支持向量机模型及应用研究[J]. 重庆师范大学学报(自然科学版),2007,24(1):29-33.

[3] Suykens J A K,Vandewalle J. Least squares support vector machine classifiers[J]. Neural Processing Letters,1999,9(3):293-300.

[4] 周欣然,滕召胜,易钊. 构造稀疏最小二乘支持向量机的快速剪枝算法[J]. 电机与控制学报,2009,13(4):626-630.

[5] 许亮. 改进的模糊最小二乘支持向量机模型[J]. 计算机工程,2009,35(14):236-240.

[6] 萧嵘. 一种SVM增量学习算法-ISVM[J]. 软件学报,2001,12(12):1811-1824.

[7] Ratsaby J. Incremental learning with sample queries[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,1998,20(8):883-888.

[8] Suykens J A K,Lukas L,Vandwealle J. Sparse least squares support vecter machines for adaptive communication channel equalization[J]. International Journal of Applied Science and Engineering,2005,11(3):51-59.

An Improved Least Square Support Vector Machines Algorithm

WAN Hui[1], *WEI Yan*[2]

(1. Dept. of Science Research; 2. Information Sciences and Engineer College, Chongqing Normal University, Chongqing 400047, China)

Abstract: Least Square Support Vector Machine is expansion of the standard Support Vector Machine. It is the form under the Support Vector Machine in quadratic loss function. It uses equality constraints instead of inequality constraints. The solution process becomes a solution group of equality equation, and avoids time-consuming in quadratic programming problem solving, but the least square support vector machine loses the sparseness, which would influence the efficiency of re-learning. In view of the above questions, this paper presents an improved incremental least squares support vector machine learning methods, In the solution of the sparsity by using adaptive pruning method, according to the initial classification to set performance pruning threshold and increase the size of samples. If you get good performance classifier, the pruning threshold and the sample increment will be greater, Otherwise, pruning threshold and the sample increment will be smaller. The result is improved efficiency of least square support vector machine training, and will solve the problem of sparse. Finally, the simulation results show that the algorithm feasible.

Key words: least squares support vector machine; incremental learning; sparsity

(责任编辑 游中胜)