回归支持向量机的改进序列最小优化学习算法*

张浩然+, 韩正之

(上海交通大学 自动化系,上海 200030)

An Improved Sequential Minimal Optimization Learning Algorithm for Regression Support Vector Machine

ZHANG Hao-Ran+, HAN Zheng-Zhi

(Department of Automation, Shanghai Jiaotong University, Shanghai 200030, China)

+ Corresponding author: Phn: 86-21-62932083 ext 810, E-mail: gczhr@sjtu.edu.cn

http://www.sjtu.edu.cn

Received 2002-11-04; Accepted 2003-03-04

Zhang HR, Han ZZ. An improved sequential minimal optimization learning algorithm for regression support vector machine. *Journal of Software*, 2003,14(12): 2006~2013.

http://www.jos.org.cn/1000-9825/14/2006.htm

Abstract: Support vector machine (SVM) is a learning technique based on the structural risk minimization principle, and it is also a class of regression method with a good generalization ability. This paper presents an improved SMO (sequential minimal optimization) algorithm to train the regression SVM, which gives an the analytical solution to the QP problem of size two. A new working set selection method and a stopping condition are developed. The simulation results show that the improved SMO algorithm is significantly faster and more precise than the original SMO one.

Key words: support vector machine; kernel method; regression; SMO (sequential minimal optimization)

摘 要: 支持向量机(support vector machine,简称SVM)是一种基于结构风险最小化原理的学习技术,也是一种新的具有很好泛化性能的回归方法,提出了实现回归支持向量机的一种改进的SMO(sequential minimal optimization)算法,给出了两变量子优化问题的解析解,设计了新的工作集选择方法和停止条件,仿真实例说明,所提出的SMO算法比原始SMO算法具有更快的运算速度.

关键词: 支持向量机;核方法;回归;序列最小优化

中图法分类号: TP18 **文献标识码**: A

自从 Vapnik[1]等人提出支持向量机(support vector machine,简称 SVM)理论以来,由于其出色的泛化能力,已在模式识别和函数估计方面取得了越来越多的进展[2~4].但是,由于 SVM 的训练方法归结为求解一个受约束的二次型规划问题,对于较多训练样本数据的情况而言,训练支持向量机的计算量是比较大的,需要很大的内存和很长的运行时间,因此限制了它的实际应用.为了简化支持向量机的训练,降低其计算复杂度,Platt[5]提出了序列最小优化(sequential minimal optimization,简称 SMO)方法,其主要思想是把一个大的优化问题分解成一系列

* **第一作者简介**: 张浩然(1972－),男,安徽灵璧人,博士生,主要研究领域为神经网络,支持向量机,智能信号处理.

只含有两个变量的优化问题.由于SMO算法的出色性能,Smola和Scholkopf[6,7]提出一种训练回归SVM的SMO算法,这个算法是Platt算法的类比扩展.S.K. Shevade[8] 和Gary William Flake[9]对Smola的算法进行了一些改进.本文提出另一种改进的 SMO 算法,这种算法的特点是充分利用支持向量机的几何意义,以便提高算法的快速性.

1 回归SVM简介

对于回归问题,给定训练样本$x_i \in R^n$, $y_i \in R, i=1,...,l$,采用式(1)拟合样本集:

$$f(x)=wT\phi(x)+b. \tag{1}$$

这里,$\phi(\cdot)$是指由输入空间到特征空间的非线性映射,$f(\cdot)$在特征空间中表示为一个线性函数.回归SVM的目的是用函数 $f(\cdot)$去拟合数据样本,同时保证能得到很好的泛化能力.为了增加回归的鲁棒性,Vapnik[1]提出了ε-不灵敏损失函数,其特点是忽略小于ε的拟合误差.综合考虑拟合误差和函数复杂度,回归SVM可以表示为如下的约束优化问题:

$$\begin{aligned}&\min\left[\frac{1}{2}w^T w+\frac{1}{2}C\sum_{i=1}^{l}(\zeta_i+\zeta_i^*)\right]\\ &\text{s.t.}\quad \begin{aligned}&y_i-w^T\phi(x_i)-b\le\varepsilon+\zeta_i\\ &w^T\phi(x_i)+b-y_i\le\varepsilon+\zeta_i^*\\ &\zeta_i,\zeta_i^*\ge 0\end{aligned}\end{aligned} \tag{2}$$

这里,$C>0$是函数复杂度和损失误差的一个平衡量.优化问题(2)的Lagrange函数为

$$L=\frac{1}{2}w^T w+C\sum_{i=1}^{l}(\zeta_i+\zeta_i^*)-\sum_{i=1}^{l}\alpha_i(w^T\phi(x_i)+b-y_i+\varepsilon+\zeta_i)-\sum_{i=1}^{l}\alpha_i^*(y_i-w^T\phi(x_i)-b+\varepsilon+\zeta_i^*)-\sum_{i=1}^{l}(\upsilon_i\zeta_i+\upsilon_i^*\zeta_i^*).$$

KKT优化条件要求Lagrange函数相对于变量w,b,ζ_i,ζ_i^*的偏导数为0.

$$\frac{\partial L}{\partial w}=0\Rightarrow w=\sum_{i=1}^{l}(\alpha_i-\alpha_i^*)\phi(x_i),$$

$$\frac{\partial L}{\partial b}=0\Rightarrow \sum_{i=1}^{l}(\alpha_i-\alpha_i^*)=0,$$

$$\frac{\partial L}{\partial \zeta_i}=0\Rightarrow \upsilon_i=C-\alpha_i,$$

$$\frac{\partial L}{\partial \zeta_i^*}=0\Rightarrow \upsilon_i^*=C-\alpha_i^*.$$

把上述KKT条件代入优化问题(2)的目标函数中,根据对偶原理和核函数技术,可以得到式(2)的对偶优化问题:

$$\begin{aligned}&\min\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}(\alpha_i-\alpha_i^*)(\alpha_j-\alpha_j^*)k(x_i,x_j)+\varepsilon\sum_{i=1}^{l}(\alpha_i+\alpha_i^*)-\sum_{i=1}^{l}y_i(\alpha_i-\alpha_i^*)\\ &\text{s.t.}\quad \sum_{i=1}^{l}(\alpha_i-\alpha_i^*)=0\\ &\qquad 0\le\alpha_i,\alpha_i^*\le C,\quad i=1,...,l\end{aligned}, \tag{3}$$

这里,$k(x_i,x_j)=\phi(x_i)^T\phi(x_j)$是核函数.

SVM的输出为

$$f(x)=\sum_{i=1}^{l}(\alpha_i-\alpha_i^*)k(x_i,x)+b. \tag{4}$$

2 改进的SMO算法

为了推导方便,作一些变量替换,令

$$\begin{aligned}&\overline{\alpha}_i=\alpha_i-\alpha_i^*,\\&\left|\overline{\alpha}_i\right|=\alpha_i+\alpha_i^*,\\&\overline{\alpha}_0=b.\end{aligned}$$

则式(3)、式(4)分别变为

$$\begin{aligned}&\min\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\overline{\alpha}_i\overline{\alpha}_j k_{ij}+\varepsilon\sum_{i=1}^{l}\left|\overline{\alpha}_i\right|-\sum_{i=1}^{l}\overline{\alpha}_i y_i\\&\text{s.t.}\quad\sum_{i=1}^{l}\overline{\alpha}_i=0\qquad,\\&\quad-C\le\overline{\alpha}_i\le C,\quad i=1,\dots,l\end{aligned}\tag{5}$$

$$f(x,\lambda)=\sum_{i=1}^{l}\overline{\alpha}_i k(x_i,x)+\overline{\alpha}_0,\tag{6}$$

核函数$k(x_i,x_j)$简写为k_{ij}.

2.1 子优化问题的解析解

SMO 算法是把一个大的 QP 问题转化成一系列的具有两个变量的 QP 问题,参照文献[9]的一些方法,假设待求的两个优化变量为 $\overline{\alpha}_a,\overline{\alpha}_b$,把式(5)的目标函数表示成变量 $\overline{\alpha}_a,\overline{\alpha}_b$ 的函数,而把其他 Lagrange 乘子视为常数,可得:

$$L(\overline{\alpha}_a,\overline{\alpha}_b)=\frac{1}{2}\overline{\alpha}_a^2k_{aa}+\frac{1}{2}\overline{\alpha}_b^2k_{bb}+\overline{\alpha}_a\overline{\alpha}_b k_{ab}+\varepsilon\left|\overline{\alpha}_a\right|+\varepsilon\left|\overline{\alpha}_b\right|-\overline{\alpha}_a y_a-\overline{\alpha}_b y_b+\overline{\alpha}_a v_a^*+\overline{\alpha}_b v_b^*+L',\tag{7}$$

其中 $v_i^*=\sum_{j=1,j\ne a,b}^{l}\overline{\alpha}_j^*k_{ij}=f_i^*-\overline{\alpha}_a^*k_{ai}-\overline{\alpha}_b^*k_{bj}-\overline{\alpha}_0$,L'是式(5)的目标函数中不含有 $\overline{\alpha}_a,\overline{\alpha}_b$ 的项,即相对于 $\overline{\alpha}_a,\overline{\alpha}_b$ 来说是不变的.具有上标*的量是指前一次迭代的值.由优化问题的等式约束,可以假设:

$$s^*=\overline{\alpha}_a^*+\overline{\alpha}_b^*=\overline{\alpha}_a+\overline{\alpha}_b,\tag{8}$$

由式(8)可以得到:

$$\overline{\alpha}_a=s^*-\overline{\alpha}_b,\tag{9}$$

把式(9)代入式(7),有

$$\begin{aligned}L(\overline{\alpha}_b)=&\frac{1}{2}(s^*-\overline{\alpha}_b)^2k_{aa}+\frac{1}{2}\overline{\alpha}_b^2k_{bb}+(s^*-\overline{\alpha}_b)\overline{\alpha}_b k_{ab}+\varepsilon\left|s^*-\overline{\alpha}_b\right|+\varepsilon\left|\overline{\alpha}_b\right|-\\&(s^*-\overline{\alpha}_b)y_a-\overline{\alpha}_b y_b+(s^*-\overline{\alpha}_b)v_a^*+\overline{\alpha}_b v_b^*+L'.\end{aligned}\tag{10}$$

为了求式(10)的极小值,必须求偏微分 $\frac{\partial L(\overline{\alpha}_b)}{\partial\overline{\alpha}_b}$,然而由于式中含有绝对值项,式(10)不是严格可微的,但是可以定义:

$$\frac{\mathrm{d}|x|}{\mathrm{d}x}=\mathrm{sgn}(x),$$

$$\mathrm{sgn}(x)=\begin{cases}1 & x>0\\ \text{待定} & x=0\\ -1 & x<0\end{cases}.$$

这种定义在数学上是允许的,这样可求得:

$$\frac{\partial L(\overline{\alpha}_b)}{\partial\overline{\alpha}_b}=\varepsilon(\mathrm{sgn}(\overline{\alpha}_b)-\mathrm{sgn}(s^*-\overline{\alpha}_b))+y_a-y_b+(\overline{\alpha}_b-s^*)k_{aa}+\overline{\alpha}_b k_{bb}+(s^*-2\overline{\alpha}_b k_{ab})-v_a^*+v_b^*.\tag{11}$$

由 $\frac{\partial L(\overline{\alpha}_b)}{\partial\overline{\alpha}_b}=0$ 可得:

$$\begin{aligned}\overline{\alpha}_b(k_{bb}+k_{aa}-2k_{ab}) &= y_b-y_a+\varepsilon(\mathrm{sgn}(s^*-\overline{\alpha}_b)-\mathrm{sgn}(\overline{\alpha}_b))+s^*(k_{aa}-k_{ab})+v_a^*-v_b^* \\ &= y_b-y_a+f_a^*-f_b^*+\varepsilon(\mathrm{sgn}(s^*-\overline{\alpha}_b)-\mathrm{sgn}(\overline{\alpha}_b))+\overline{\alpha}_a^*k_{aa}-\overline{\alpha}_a^*k_{ab}+\overline{\alpha}_b^*k_{aa}-\overline{\alpha}_b^*k_{ab}- \\ &\quad \overline{\alpha}_a^*k_{aa}-\overline{\alpha}_b^*k_{ab}-\overline{\alpha}_0^*+\overline{\alpha}_a^*k_{ab}+\overline{\alpha}_b^*k_{bb}+\overline{\alpha}_0^* \\ &= y_b-y_a+f_a^*-f_b^*+\varepsilon(\mathrm{sgn}(s^*-\overline{\alpha}_b)-\mathrm{sgn}(\overline{\alpha}_b))+\overline{\alpha}_b^*(k_{aa}+k_{bb}-2k_{ab}).\end{aligned} \tag{12}$$

由式(12)可以写出变量 $\overline{\alpha}_b$ 的递推计算公式,如下:

$$\overline{\alpha}_b=\overline{\alpha}_b^*+\eta[y_b-y_a+f_a^*-f_b^*+\varepsilon(\mathrm{sgn}(s^*-\overline{\alpha}_b)-\mathrm{sgn}(\overline{\alpha}_b))], \tag{13}$$

其中 $\eta=\dfrac{1}{k_{aa}+k_{bb}-2k_{ab}}$.

由于两个符号函数的存在,式(13)是递归的,必须求其解.如果 SVM 的核函数满足 Mercer 条件,就可以保证下式成立:

$$\eta=\frac{1}{k_{aa}+k_{bb}-2k_{ab}}\ge 0.$$

这样由式(11)可知偏导数 $\dfrac{\partial L(\overline{\alpha}_b)}{\partial\overline{\alpha}_b}$ 是 $\overline{\alpha}_b$ 的增函数,即随 $\overline{\alpha}_b$ 的增加而增加,而且,如果 $s^*\ne 0$,偏导数的变化曲线是具有两次突变的分段直线,如图 1 所示.

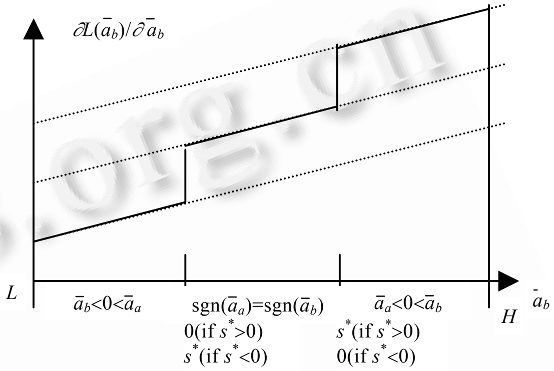

Fig.1 The curve of $\partial L(\overline{a}_b)/\partial\overline{a}_b$

图1 随变化曲线图

图 1 显示了对偶目标函数的偏导数 $\dfrac{\partial L(\overline{\alpha}_b)}{\partial\overline{\alpha}_b}$ 的变化情况.考虑以上因素我们来求方程(13)的解,这个方程存在 5 个可能的解,3 个解对应于 $(\mathrm{sgn}(s^*-\overline{\alpha}_b)-\mathrm{sgn}(\overline{\alpha}_b))$ 的值,取−2,0,2,另两个解是 $\overline{\alpha}_b$ 等于 0 和 S^*.

取 $(\mathrm{sgn}(\overline{\alpha}_a)-\mathrm{sgn}(\overline{\alpha}))$ 的值分别为−2,0,2,代入式(7)中,如果能使式(11)等于 0 则接受此新值,如果上面的取值失败,尝试取 $\overline{\alpha}_b$ 值为 0 或 S^*,利用 $\overline{\alpha}_b$ 值的正(负)扰动是否能使式(11)取正(负)值来判断取 0 还是 S^*.总结以上讨论,可以得到 $\overline{\alpha}_b$ 的更新律,见式(14),式中 $0^+,0^-,s^{*+},s^{*-}$ 分别表示 0 和 S^*的正负微小扰动.

$$\overline{\alpha}_b=\begin{cases}\overline{\alpha}_b^*+\eta[y_b-y_a+f_a^*-f_b^*+2\varepsilon] & \overline{\alpha}_a>0>\overline{\alpha}_b \\ \overline{\alpha}_b^*+\eta[y_b-y_a+f_a^*-f_b^*-2\varepsilon] & \overline{\alpha}_a<0<\overline{\alpha}_b \\ \overline{\alpha}_b^*+\eta[y_b-y_a+f_a^*-f_b^*] & \mathrm{sgn}(\overline{\alpha}_a)=\mathrm{sgn}(\overline{\alpha}_b) \\ 0 & \dfrac{\partial L(\overline{\alpha}_b)}{\partial\overline{\alpha}_b}(\overline{\alpha}_b=0^+)>0 \text{ and } \dfrac{\partial L(\overline{\alpha}_b)}{\partial\overline{\alpha}_b}(\overline{\alpha}_b=0^-)<0 \\ s^* & \dfrac{\partial L(\overline{\alpha}_b)}{\partial\overline{\alpha}_b}(\overline{\alpha}_b=s^{*+})>0 \text{ and } \dfrac{\partial L(\overline{\alpha}_b)}{\partial\overline{\alpha}_b}(\overline{\alpha}_b=s^{*-})<0\end{cases}. \tag{14}$$

考虑到 $\overline{\alpha}_b$ 所受的矩形约束,即 $-C\le\overline{\alpha}_b\le C$,我们定义:

$$L=\max(s^*-C,-C),$$
$$H=\min(C,s^*+C),$$

以上等式的高低边界限制可以确保优化变量 $\overline{\alpha}_a,\overline{\alpha}_b$ 的值在 $\pm C$ 之间.

根据KKT条件可求得常值偏差 $\overline{\alpha}_0$ 的更新,强迫使SVM的输出 $f_a=y_a,f_b=y_b$,有

$$\overline{\alpha}_0^a=\overline{\alpha}_0^{\mathrm{old}}+f_a^*-y_a+(\overline{\alpha}_a^{\mathrm{new}}-\overline{\alpha}_a^{\mathrm{old}})k_{aa}+(\overline{\alpha}_b^{\mathrm{new}}-\overline{\alpha}_b^{\mathrm{old}})k_{ab},$$
$$\overline{\alpha}_0^b=\overline{\alpha}_0^{\mathrm{old}}+f_b^*-y_b+(\overline{\alpha}_a^{\mathrm{new}}-\overline{\alpha}_a^{\mathrm{old}})k_{ab}+(\overline{\alpha}_b^{\mathrm{new}}-\overline{\alpha}_b^{\mathrm{old}})k_{bb}.$$

把以上两式平均,即可得到

$$\overline{\alpha}_0 = 0.5(\overline{\alpha}_0^a + \overline{\alpha}_0^b) .$$

2.2 选择待优化Lagrange乘子

在 SMO 算法中,如何选择子优化问题的工作集是非常重要的,它严重影响了算法的收敛性和快速性.事实上,即使我们不采用任何选择方法,只是按顺序抽取 $\overline{\alpha}_a,\overline{\alpha}_b$ 的所有组合进行优化,目标函数也会不断下降,直到任意一对 $\overline{\alpha}_a,\overline{\alpha}_b$ 都不能继续优化,目标函数就会收敛到极小值.但是这种情况下算法的收敛速度会很慢,为了提高算法的收敛速度,必须采用某种启发式方法来选择工作集.Platt 和 Smola 采用了一种两层循环的方法,外层循环遍历所有样本,找到训练集中违反 KKT 条件的样本,内层循环针对违反 KKT 条件的样本选择另一个样本与它配对优化(即指优化它们的 Lagrange 乘子),选择的依据是尽量使这样一对样本能取得最大优化步长.我们认为,这种选择优化变量的启发式策略存在一定的缺陷,就是第 1 个优化变量的选取过于随机.下面对这种优化变量的选择策略作进一步的改进,采用另一种启发式方法来选择待优化的 Lagrange 乘子 $\overline{\alpha}_a,\overline{\alpha}_b$.

对于优化问题(5),当只考虑等式约束时,可以得到目标函数:

$$L = \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\overline{\alpha}_i\overline{\alpha}_j k_{ij} + \varepsilon\sum_{i=1}^{l}\left|\overline{\alpha}_i\right| - \sum_{i=1}^{l}\overline{\alpha}_i y_i + \lambda\sum_{i=1}^{l}\overline{\alpha}_i .$$

根据文献[10]可知,Lagrange乘子 λ 等于待求函数的常值参数 b ,即$\lambda=b$,因此有

$$L = \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\overline{\alpha}_i\overline{\alpha}_j k_{ij} + \varepsilon\sum_{i=1}^{l}\left|\overline{\alpha}_i\right| - \sum_{i=1}^{l}\overline{\alpha}_i y_i + b\sum_{i=1}^{l}\overline{\alpha}_i .$$

可求得目标函数对优化变量 $\overline{\alpha}_b$ 的梯度分量为

$$\begin{aligned}\frac{\partial L}{\partial \overline{\alpha}_b} &= \sum_{i=1}^{l}\overline{\alpha}_i k_{ib} + b - y_b + \varepsilon\,\mathrm{sgn}(\overline{\alpha}_b)\\ &= f(x_b) - y_b + \varepsilon\,\mathrm{sgn}(\overline{\alpha}_b)\\ &= E_b + \varepsilon\,\mathrm{sgn}(\overline{\alpha}_b)\ ,\end{aligned}$$

其中sgn(·)是符号函数.

对于一个目标函数来说,一个优化变量的梯度分量的绝对值越大,这个优化变量的变化对目标函数值的改变量就越大,因此我们就根据这一点来选择第 1 个优化变量,对于第 2 个优化变量的选择,像 Smola 所使用的方法一样,使优化变量本身获得最大的变化量.综上所述,选择两个待优化变量的策略如下:

① 对第 1 个优化变量的选择是算法的外部循环,首先遍历训练样本集中的非边界样本,找出样本中违反 KKT 条件的样本,若找不到则遍历整个训练集,找出样本中违反 KKT 条件的样本,如果还是不存在违反 KKT 条件的样本则退出循环,结束程序;从违反 KKT 条件的样本中找出使目标函数梯度绝对值最大的样本(即 $\mathrm{Max}(|E_b+\varepsilon\mathrm{sgn}(\overline{\alpha}_b)|)$)所对应的优化变量 $\overline{\alpha}_b$ 作为第 1 个优化变量.

② (A) 在确定第 1 个优化变量之后,进入算法的内部循环,再从违反 KKT 条件的非边界样本中找出使 $|E_a-E_b|$最大所对应的 Lagrange 乘子 $\overline{\alpha}_a$ 作为第 2 个优化变量进行优化;

(B) 若(A)步所选择的优化变量不能产生改进,则遍历非边界样本,从违反 KKT 条件的优化变量中随机找出 Lagrange 乘子 $\overline{\alpha}_a$ 进行优化;

(C) 若(B)也无法使优化变量产生改进,则遍历整个训练集,从违反 KKT 条件的优化变量中随机找出 Lagrange 乘子 $\overline{\alpha}_a$ 进行优化;

(D) 若(A),(B),(C)都不行,则算法忽略外部循环中所选择的第 1 个样本,继续外部循环.

这种选择策略的第①步是使所选取的优化变量能够使目标函数的下降最多,第②步是使优化变量本身的变化量大,进而使目标函数的变化量大.

2.3 完整的SMO算法

总结以上各部分的讨论,完整的 SMO 算法可以表示如下:

(1) 初始化 $\overline{\alpha}_a^{\mathrm{old}},\overline{\alpha}_b^{\mathrm{old}}$;

(2) 利用第 2.2 节提供的方法选择优化变量 $\overline{\alpha}_a,\overline{\alpha}_b$;

(3) 利用式(14)计算 $\overline{\alpha}_b^{\text{new}}$;

(4) 如果 $\overline{\alpha}_b^{\text{new}} > H$,设定 $\overline{\alpha}_b^{\text{new}} = H$,如果 $\overline{\alpha}_b^{\text{new}} < L$,设定 $\overline{\alpha}_b^{\text{new}} = L$;

(5) $\overline{\alpha}_a^{\text{new}} = \overline{\alpha}_a^{\text{old}} + \overline{\alpha}_b^{\text{old}} - \overline{\alpha}_b^{\text{new}}$;

(6) $\overline{\alpha}_0^{\text{new}} = \frac{1}{2}(\overline{\alpha}_0^a + \overline{\alpha}_0^b)$;

(7) 如果对于所有的样本均满足 KKT 条件,go to (8),否则 go to (2);

(8) 输出 $\overline{\alpha}$,停止程序.

3 仿真实验

本节将采用一个具体的实例来验证我们所提出的算法,选择所需要拟合的函数为

$$y(x)=4\sin(2x)-x.$$

首先产生训练和测试样本集,产生一个随机数列 x,具有 200 个点,代入上式生成 200 个样本$(x(k),y(k))$,如图 2 所示.总样本集中的前 100 个作为训练样本,用来训练 SVM,后 100 个作为测试样本,用来检验 SVM 的学习性能和泛化能力.

选择SVM的有关设计参数,核函数选RBF函数,即 $k(x_i,x_j)=\exp\left(-\frac{\|x_i-x_j\|^2}{2\sigma^2}\right)$,并且设定: $\frac{1}{2\sigma^2}=0.4$,ε=0.01, C=50, τ=0.5.采用上文提到的SMO算法来训练SVM,同时也利用原始的SMO算法来训练SVM,以便于比较.得到如图3和图4所示的仿真结果.

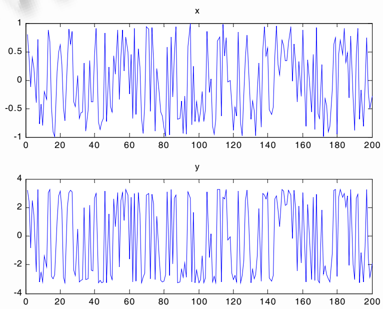

Fig.2 Samples set

图2 样本集

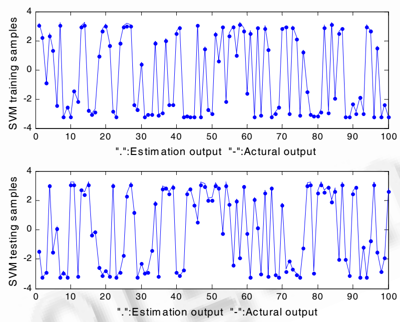

Fig.3 The simulation result of improved SMO algorithm

图3 改进的SMO仿真结果

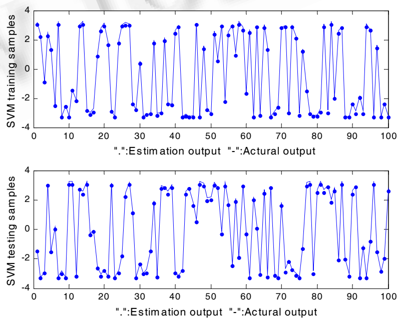

Fig.4 The simulation result of original SMO algorithm

图4 原始SMO算法仿真结果

图中,点(.)代表 SVM 的输出,实线(—)代表实际样本的输出.从仿真结果图 3 和图 4 可以看出,两个训练方法都能够使 SVM 具有良好的学习性能.原始 SMO 的运行时间约为 5.6s,训练误差为 0.11,测试误差为 0.12;我们提

出的 SMO 算法的运行时间为 1.7s,训练误差为 0.09,测试误差为 0.11.可以明显看出,我们提出的 SMO 算法比原始算法优越,不但运行时间短,而且拟合精度也好.

Table 1 The contrast of improved SMO and original SMO

表1 改进的SMO和原始SMO的比较

	Training error	Testing error	Running time (s)
Improved SMO algorithm	0.09	0.11	1.7
Original SMO algorithm	0.11	0.12	5.4

在改变不灵敏参数ε时,可得到学习机的ε参数和支持向量的个数的关系,如图 5 所示.

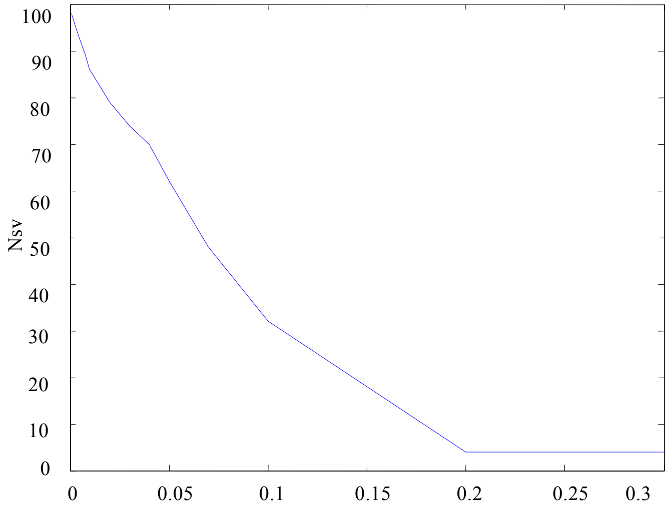

Fig.5 The number of support vector with parameter ε

图 5 支持向量数随ε变化的曲线

可以看出,随着不灵敏参数ε的增大,支持向量数递减,支持向量机解的稀疏性增加.

4 结 论

针对回归支持向量机的学习方法计算复杂度过高的难题,本文提出一种改进的 SMO 算法.仿真实验表明,它具有运算速度快、拟合精度高等优点,是较为实用的学习算法.未来的研究目标是设计出在线 SMO 学习算法,能用它来解决 SVM 的在线学习问题.

References:

[1] Vapnik VN. The Nature of Statistical Learning Theory. New York: Springer-Verlag, 1995.

[2] Cherkassky V, Mulier F. Learning from Data-Concepts, Theory and Methods. New York: John Wiley Sons, 1998.

[3] Joachims T. Text categorization with support vector machines: Learning with many relevant features. In: Proceedings of the European Conference on Machine Learning (ECML). Berlin: Springer-Verlag, 1998. 37~142.

[4] Weston GJ, Barnhill S. Gene selection for cancer classification using support vector machines. Machine Learning, 2002,46(1-3): 389~422.

[5] Platt JC. Fast training of support vector machines using sequential minimal optimization. In: Scholkopf B, Burges C, Smola A, eds. Advances in Kernel Methods: Support Vector Machines. Cambridge: MIT Press, 1998. 185~208.

[6] Smola AJ. Learning with kernels [Ph.D. Thesis]. University of Birlinghoven, 1998.

[7] Smola AJ, Scholkopf B. A tutorial on support vector regression. Technical Report, TR 1998-030. London: Royal Holloway College, 1998.

[8] Shevade SK, Keerthi SS, Bhattacharyya C. Improvements to SMO algorithm for SVM regression. IEEE Transactions on Neural Networks, 2000,11(5):1188~1194.

[9] Flake GW, Lawrence S. Efficient SVM regression training with SMO. Machine Learning Special Issue on SVMs, 2000,46(1~3): 271~290.

[10] Martin M. On-Line support vector machines for function approximation. Technical Report, LSI-02-11-R. Catalunya: Department of Software, Universitat Politecnica de Catalunya, 2002.