自适应误差惩罚支撑向量回归机

陈晓峰[①]　王士同[①]　曹苏群[①②]
[①](江南大学信息学院　无锡　214122)
[②](淮阴工学院机械系　淮安　223001)

摘　要：该文提出一种支撑向量回归机 AEPSVR。它首先用 ε-SVR 求得一个近似的支撑向量回归函数，在此基础上，引入一种新自适应误差惩罚函数，通过迭代，得到鲁棒的支撑向量回归机。该方法因以 ε-SVR 为基础，故可以应用各种求解 SVR 的优化算法。实验表明，该支撑向量回归机 AEPSVR 能显著地降低离群点的影响，具有良好的泛化性能。

关键词：支撑向量回归；离群点；自适应误差惩罚
中图分类号：TP181　**文献标识码**：A　**文章编号**：1009-5896(2008)02-0367-04

SVR with Adaptive Error Penalization

Chen Xiao-feng[①]　Wang Shi-tong[①]　Cao Su-qun[①②]
[①](*School of Information*, *Southern Yangtz University*, *Wuxi* 214122, *China*)
[②](*Department of Mechanical Engineering*, *Huaiyin Institute of Technology*, *Huaian* 223001, *China*)

Abstract: A novel support vector regression method AEPSVR is proposed in this paper. First, an approximate regression function is obtained using ε-SVR method, and then a new adaptive error penalization function is introduced to enhance the robust performance of SVR such that a robust support vector regression is derived. Because the proposed AEPSVR here is based on ε-SVR, so various optimization methods for SVR can be used. Experimental results show that the proposed AEPSVR can reduce the affect of outliers, and have the very good generalization capability.

Key words: Support Vector Regression (SVR); Outlier; Adaptive Error Penalization (AEP)

1　引言

支撑向量回归机(SVR)是支撑向量方法在函数拟合与回归估计的应用[1, 2]。支撑向量回归机在处理含有离群点数据时，拟合性能要受到离群点的影响。离群点与其他训练样本相比对应较大的训练误差，导致训练结果与没有离群点时相比会有很大变化。减小离群点的影响对构造鲁棒的支撑向量回归机有很重要的意义。

在支撑向量分类问题中，根据距离[3]，最小化“留一法”交叉验证的训练误差上界[4]，hinge 损失函数[5]，非线性的惩罚函数[6]等减小离群点影响。在支撑向量回归问题中，文献[7]的方法是，首先用常规 SVR 得到一个近似的支撑向量回归函数，然后对目标函数加权，通过迭代得到优化结果。文献[8]的方法是用加权平方误差损失函数代替常规支撑向量回归方法中的 $\varepsilon-$不敏感函数。文献[9]的方法是首先用常规 SVR 得到一个有支撑向量的初始径向基函数网，再用鲁棒损失函数调节网络参数。文献[10]研究了一种分为三段的代价函数，用 ε 不敏感函数计算低频变量的噪声代价，用平方误差函数计算样本检测误差代价，用线性函数限制离群点的影响。文献[11,12]提出基于证据框架的鲁棒支撑向量参数优选方法。

本文提出的 AEPSVR 支撑向量回归机算法，主要是通过引入一种新的代价函数消减离群点的影响。

2　自适应误差惩罚支撑向量回归机(AEPSVR)

2.1　ε-SVR

假定训练样本集为 $\{(x_1,y_1),\cdots,(x_l,y_l)\}\subset R^N\times R$，$x_i$ 为输入变量的值，y_i 为对应的输出值，N 为输入变量的维数，l 为训练样本数。在 ε-SVR 中，估计函数 $f(x)$，使得对于所有训练样本，有 $|f(x_i)-y_i|\le\varepsilon(i=1,\cdots,l)$，且 $f(x)$ 要尽可能平滑。假设 $f(x)$ 为线性函数(可以通过核方法扩展到非线性情况)，即 $f(x)=<w,x>+b$，其中 $w\in R^N$，$b\in R$，$<\cdot,\cdot>$ 代表内积。求函数 $f(x)$ 即可表述为如下求极小值问题：

$$\text{minimize}\ \ \frac{1}{2}\|w\|^2+C\sum_{i=1}^{l}(\xi_i+\xi_i^*)$$

2006-07-20 收到，2007-01-31 改回
2004 年教育部优秀人才支持计划(NCET-04-0496)，模式识别国家重点实验室开放课题，南京大学软件新技术国家重点实验室开放课题，教育部重点科学研究项目(105087)和国防应用基础研究基金项目(A1420061266)资助课题

$$\text{s.t.}\quad \begin{cases} y_i - \langle w, x_i \rangle - b \le \varepsilon + \xi_i \\ \langle w, x_i \rangle + b - y_i \le \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \ge 0 \end{cases}$$

其中正常量C为惩罚因子，协调函数$f(x)$的平滑性与训练精度。$\xi_i^{(*)}$(表示ξ_i或ξ_i^*)由下述ε-不敏感损失函数计算：

$$|\xi|_\varepsilon = \begin{cases} 0, & |\xi| \le \varepsilon \\ |\xi| - \varepsilon, & \text{其他} \end{cases}$$

上述问题可以通过 Lagrange 求极值方法转为如下的 Wolfe 对偶的二次规划问题：

$$\text{minimize}\begin{cases} -\dfrac{1}{2}\sum_{i,j=1}^{l}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) \langle x_i, x_j \rangle \\ -\varepsilon\sum_{i=1}^{l}(\alpha_i + \alpha_i^*) + \sum_{i=1}^{l} y_i(\alpha_i - \alpha_i^*) \end{cases}$$

且$\sum_{i=1}^{l}(\alpha_i - \alpha_i^*) = 0$且$\alpha_i, \alpha_i^* \in [0, C]$。

待求量w, b可表示为

$$w = \sum_{i=1}^{l}(\alpha_i - \alpha_i^*) x_i \tag{1}$$

$$\begin{aligned} &\max\{-\varepsilon + y_i - \langle w, x_i \rangle \mid \alpha_i < C \text{或} \alpha_i^* > 0\} \le b \\ &\le \min\{-\varepsilon + y_i - \langle w, x_i \rangle \mid \alpha_i > 0 \text{或} \alpha_i^* < C\} \end{aligned} \tag{2}$$

2.2 AEPSVR

构建鲁棒支撑向量回归机AEPSVR的关键在于构造合适的代价函数。为了减小离群点对回归的影响，代价函数必须是有界的。此外，代价函数应当具有合适的结构，以方便求解二次规划对偶问题，使得正则化参数能根据训练误差$\xi_i^{(*)}(i = 1, \cdots, l)$自动地进行调整，本文称这种性质为自适应性。本文引入一种新的代价函数$\text{erf}(\xi, \sigma) = \dfrac{2}{\sqrt{\pi}\sigma}\int_0^{\xi} e^{-x^2/\sigma^2} \text{d}x$(简称为erf函数)来构建鲁棒支撑向量回归机AEPSVR。erf函数是有界函数。如图1所示，erf函数是奇函数，当$\lim_{\xi \to +\infty} \text{erf}(\xi, \sigma) = 1$，离群点对应的$\xi$值比较大，但无论$\xi$多大，erf函数极限是1，因此，用它做代价函数可以大大降低离群点的影响。参数σ调节erf函数的收敛速度，当σ较小时，代价函数收敛的速度更快。

以ε-SVR为基础，重新建立的规划问题如下(下面的公式推导假定为$f(x)$为线性，可以通过核方法扩展到非线性情况)：

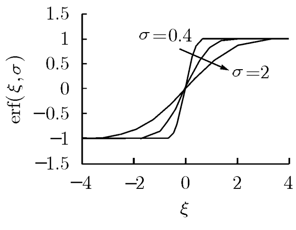

图1 erf函数

$$\text{minimize}\quad \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}(\text{erf}(\xi_i, \sigma) + \text{erf}(\xi_i^*, \sigma))$$

$$\text{s.t.}\quad \begin{cases} y_i - \langle w, x_i \rangle - b \le \varepsilon + \xi_i \\ \langle w, x_i \rangle + b - y_i \le \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \ge 0 \end{cases}$$

其中正的常量C为惩罚因子，$\xi_i^{(*)}$由下述ε-不敏感损失函数计算

$$|\xi|_\varepsilon = \begin{cases} 0, & |\xi| \le \varepsilon \\ |\xi| - \varepsilon, & \text{其他} \end{cases}$$

自适应惩罚函数定义为

$$\text{erf}(\xi_i^{(*)}, \sigma) = \frac{2}{\sqrt{\pi}\sigma}\int_0^{\xi_i^{(*)}} e^{-\frac{x^2}{\sigma^2}} \text{d}x \tag{3}$$

用 Lagrange 求极值方法求解上述问题，则有

$$\begin{aligned} L_p(w, b, \xi_i) = &\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}(\text{erf}(\xi_i, \sigma) + \text{erf}(\xi_i^*, \sigma)) \\ &- \sum_{i=1}^{l}\alpha_i(\varepsilon + \xi_i - y_i + \langle w, x_i \rangle + b) \\ &- \sum_{i=1}^{l}\alpha_i^*(\varepsilon + \xi_i^* + y_i - \langle w, x_i \rangle - b) \\ &- \sum_{i=1}^{l}\mu_i\xi_i - \sum_{i=1}^{l}\mu_i^*\xi_i^* \end{aligned} \tag{4}$$

且$\alpha_i^{(*)} \ge 0$，$\mu_i^{(*)} \ge 0$ $(i = 1, \cdots, l)$。

根据 Kuhn-Tucker 条件，将上式转为为对偶问题，即为

$$\begin{aligned} \text{minimize} &-\frac{1}{2}\sum_{i,j=1}^{l}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) \langle x_i, x_j \rangle \\ &- \varepsilon\sum_{i=1}^{l}(\alpha_i + \alpha_i^*) + \sum_{i=1}^{l} y_i(\alpha_i - \alpha_i^*) \\ &+ C\sum_{i=1}^{l}(\xi_i \text{erf}'(\xi_i, \sigma) - \text{erf}(\xi_i, \sigma)) \\ &+ C\sum_{i=1}^{l}(\xi_i^* \text{erf}'(\xi_i^*, \sigma) - \text{erf}(\xi_i^*, \sigma)) \end{aligned} \tag{5}$$

且$0 \le \alpha_i^{(*)} \le C \cdot \text{erf}'(\xi_i^{(*)}, \sigma)$，$\xi_i^{(*)} \ge 0$ $(i = 1, \cdots, l)$。

w, b可表示为

$$w = \sum_{i=1}^{l}(\alpha_i - \alpha_i^*) x_i \tag{6}$$

$$\begin{aligned} &\max\{-\varepsilon + y_i - \langle w, x_i \rangle \mid \alpha_i < C \text{或} \alpha_i^* > 0\} \le b \\ &\le \min\{-\varepsilon + y_i - \langle w, x_i \rangle \mid \alpha_i > 0 \text{或} \alpha_i^* < C\} \end{aligned} \tag{7}$$

$$\text{erf}'(\xi_i^{(*)}, \sigma) = \frac{1}{\sqrt{\pi}\sigma} e^{-\frac{(\xi_i^{(*)})^2}{s^2}} \tag{8}$$

由此，AEPSVR转化为求解二次规划问题。正则化参数C_i $(i = 1, \cdots, l)$由$\text{erf}'(\xi, \sigma)$根据训练误差ξ自适应调整。如图2所示，$\text{erf}'(\xi, \sigma)$随着误差ξ变大而迅速减小，离群点对应的训练误差ξ远比非离群点大，因此，离群点对应的正则化参数C_i迅速降低，离群点在训练中的影响也随之降低。

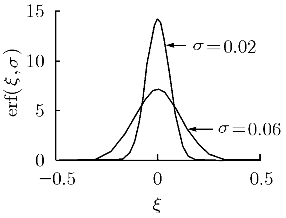

图 2　$\mathrm{erf}'(\xi,\sigma)$ 函数

自适应误差惩罚支撑向量回归机 AEPSVR 的实现步骤如下：

(1)用常规的 ε-SVR 训练出初始的支撑向量回归机；

(2)选择 erf 函数的参数 σ 和 σ 下降系数参数 $s(s>1)$；

(3)计算 $\sigma=\sigma/s$，若更新后的 $\sigma\geq \mathrm{sigmalb}$，则转到第(4)步，否则，转到第(7)步；

(4)计算 $\xi_i^{(*)}$，并根据计算对应得 $C_i=C\cdot \mathrm{erf}'(\xi_i^{(*)},\sigma)$；

(5)根据更新后的 C_i，训练新的 ε-SVR；

(6)返回第(3)步；

(7)停止运算，返回 w 和 b 的值。

3　实验结果及其分析

实验研究 sinc 函数和多元函数回归拟合问题。样本分别混入 3~8 个不同幅度的离群点，它们是 $[5\quad 15]\cup[-15\quad -5]$ 区间的随机数。参数取为 C=100，σ=100，s=5，$\mathrm{sigmalb}=0.1$，$\varepsilon=0.005$。在训练结果的图示中，“*”代表的是离群点，“.”代表测试样本点，实线为根据训练结果绘制的拟合曲线。

核函数对支撑向量回归的效果影响很大，在拟合非线性数据时，rbf 核与 erbf 核的性能比线性核与多项式核性能更好。核参数 σ 相同时，erbf 核函数的回归受离群点的影响更明显。在实验中选择 erbf 核，核函数形式如下：

$$K(x,x')=e^{-\frac{\|x-x'\|}{2s^2}} \tag{9}$$

加权鲁棒支撑向量回归方法(WRSVR)是张讲社等在文献[7]中提出的，它与 WLS-SVM[8]和 RSVR[9]相比也具有良好的拟合性能。用 WRSVR 与 AEPSVR 做对比实验，WRSVR 参数按照文献[7]中的建议取为 $p=1.5$，$\eta=0.6$。

用 Err (样本集训练误差平方和)衡量测试的拟合误差，计算方法如下：

$$\mathrm{Err}=\sum_{i=1}^{l}(y_i-\overline{y}_i)^2 \tag{10}$$

3.1 sinc 函数拟合

拟合 sinc 函数 $y=\sin(x)/x$，x 初始点为−5，训练样本的间距为 0.2。训练样本数量均为 51 个。测试样本数量均为 50 个，均匀的取在定义域上。实验结果如表 1 所示。

3.2 多元函数拟合

拟合 $y=0.1x_1^3+0.12x_2^2+0.3x_3^2+0.4x_3x_4$，其中 x_1，x_2，x_3 和 x_4 的初始点分别为−5，−6，1，−3，样本间距分别为 0.2，0.2，0.1，0.15，训练样本数量均为 51 个。测试样本数量为 50 个，均匀的取在定义域上。实验结果如表 2 所示。

3.3 离群点幅度对算法性能的影响

以拟合 sinc 函数为例，测试离群点幅度对算法的影响。训练样本、测试样本和训练参数设置同 3.1 节。用离群点绝对值的均值(E)和标准差(Std)为指标衡量离群点的幅度，其计算方法如下：

$$E=\frac{1}{l}\sum_{i=1}^{l}|y_i| \tag{11}$$

$$\mathrm{Std}=\sqrt{\frac{\sum_{i=1}^{l}(|y_i|-E)^2}{l-1}} \tag{12}$$

实验表明，离群点幅度从 E=11，Std=2.0 到 E=4.4，Std=0.8 时，AEPSVR 对幅度较大的离群点训练效果很稳定，基本上不受离群点影响，当离群点幅度降到 E=4.4，Std=0.8 以下时，训练误差和测试误差开始变大。导致这种结果的原因是训练参数 C 相对与离群点的幅度而言过大，适当地减小 C 即可避免出现这种情况。例如，在此例中，取 $C=8$，其他参数不变，当 E=1.1，Std=0.2 时，训练误差和测试误差才开始变大，当 E=0.5，Std=0.1 时，部分离群点已很难和正常样本区分开，但 AEPSVR 仍能有较好的训练效果。

3.4 实验结果分析

AEPSVR 对每个训练样本分别设置正则化参数 C_i，在训练中，逐渐减小离群点对应的正则化参数，由此减小离群点的影响。参数 C 具有两个作用：其一，调节正则化参数 C_i；其二，参数 C 的取值和离群点的幅度相关，如果离群点的幅度较大，C 可以取的比较大，反之 C 要取小。参数 σ，s 和

表 1　sinc 函数拟合

	Err					
	3 离群点		6 离群点		8 离群点	
	训练	测试	训练	测试	训练	测试
ε-SVR	12.1289	6.1181	24.1308	12.1294	32.1716	16.1700
WRSVR	0.0036	0.0039	0.0048	0.0045	0.0121	0.0108
AEPSVR	0.0028	0.0033	0.0030	0.0035	0.0032	0.0037

表 2 多元函数拟合

	Err					
	3 离群点		6 离群点		8 离群点	
	训练	测试	训练	测试	训练	测试
ε-SVR	772.4672	385.1052	799.9389	412.2536	835.2873	441.4451
WRSVR	6.1478	6.3419	10.6882	10.9776	18.3347	19.0509
AEPSVR	5.9537	5.4713	3.5335	3.8869	10.4421	9.8871

sigmalb 确定对离群点对应的正则化 C_i 约束效果。参数 s 调控 σ 的下降速度，一般取大于 1 的正数，较小的 σ 对应较小的 C_i，即离群点对应的正则化参数逐步渐小。

实验表明，AEPSVR 算法的误差在 3 种方法中是最小的。ε-SVR 算法对离群点非常敏感，随着离群点数量的增多，误差迅速增大。当离群点数量较少时，WRSVR 算法和 AEPSVR 算法的结果差不多，AEPSVR 性能稍好。如离群点距离正常样本较远时，WRSVR 和 AEPSVR 性能接近，都具有较好的抗离群点效果，WRSVR 的误差稍大。当离群点数量较多时，AEPSVR 算法性能明显好于 WRSVR 和 ε-SVR，达到了很高的拟合精度和泛化性能，而且随着离群点增加，AEPSVR 的误差只是缓慢增加，说明 AEPSVR 在离群点幅度较小或数量较多的情况下性能仍然稳定。在多元函数拟合实验中，离群点的幅度相对较小，AEPSVR 和 WRSVR 受离群点的影响比较大，因此误差比 sinc 函数拟合时要大一些，但精度仍然较好。

4 结束语

本文提出了一种自适应误差惩罚支撑向量回归机（AEPSVR），它改进支撑向量回归算法的鲁棒性，将离群点在训练中影响大幅度降低。由于该算法是以 ε-SVR 为基础，因此能将各种求解常规 SVR 的优化方法用于 AEPSVR 的求解。迭代方式的鲁棒支撑向量机的缺点是训练时间较长，进一步的研究工作将是提高它的泛化性能和求解速度。另外，研究一般代价函数形式及其对支撑向量回归算法的影响也很有必要。

参考文献

[1] Smola A J and Schölkopf B. A tutorial on support vector regression. *Statistics and Computing*, 2004, 14(3): 199–222.

[2] Schölkopf B, Smola A J and Williamson R C, *et al*.. New support vector algorithm. *Neural Computation*, 2000, 12(12): 1207–1245.

[3] Song Q, Hu W, and Xie W. Robust support vector machine with bullet hole image classification. *IEEE Trans. on Systems, Man and Cybernetics C*, 2002, 32(4): 440–448.

[4] Weston J and Herbrich R. Adaptive margin support vector machines. Neural Information Processing Systems (NIPS) Conference Workshop on Advance in Large Margin Classifiers, Breckenridge Colorado USA, 1998: 281–296.

[5] Xu Linli, Crammer K, and Schuurmans D. Robust support vector machine training via convex outlier ablation. In Proceedings of the 21st National Conference on Artificial Intelligence(AAAI-06), Boston USA, 2006: 536–546.

[6] Zhan Yiqiang and Shen Dinggang. An adaptive error penalization method for training an efficient and generalized SVM. *Pattern Recognition*, 2006, 39(3): 342–350.

[7] 张讲社，郭高. 加权稳健支撑向量回归方法. 计算机学报, 2005, 28(7): 1171–1177.
Zhang Jiang-she and Guo Gao. Reweighted robust support vector regression method. *Chinese Journal of Computer*, 2005, 28(7): 1171–1177.

[8] Suykens J A K, De Brahanter J, Lukas L, and Vandewalle J. Weighted least squares support vector machine: robustness and sparse approximation. *Neurocomputing*, 2002, 48(1-4): 85–105.

[9] Chuang C C, Su F F, Jeng J T, and Hsiao C C. Robust support regression networks for function approximation with outliers. *IEEE Trans. on Neural Networks*, 2002, 13(6): 1322–1330.

[10] Yong Zhan and HaoZhong Cheng. A robust support vector algorithm for harmonic and interharmonic analysis of electric power system. *Electric Power Systems Research*, 2005, 73(3): 393–400.

[11] Wang Shitong, Zhu Jiagang and Chung Fu-Lai, *et al*.. Experimental study on parameter choices in norm-r support vector regression machines with noisy input. *Soft Computing*, 2006, 10(3): 219–223.

[12] Wang Shitong, Zhu Jiagang and Chung Fu-Lai, *et al*.. Theoretically optimal parameter choices for support vector regression machines with noisy input. *Soft Computing*, 2005, 9(10): 732–741.

陈晓峰： 男，1977 年生，博士生，从事人工智能、模式识别的研究.

王士同： 男，1964 年生，教授，博士生导师，从事人工智能、机器学习等研究.

曹苏群： 男，1975 年生，讲师，博士生，从事人工智能、软件工程等研究.