

【文章编号】 1004-1540(2010)02-0162-05

一种光滑支持向量回归机新函数的研究

沈进东

(中国计量学院 理学院, 浙江 杭州 310018)

【摘要】 为支持向量回归机提供了一个新的光滑函数,即运用三次样条函数和复合函数的方法,得到一种新的光滑支持向量回归机——三次样条光滑支持向量回归机(TSSSVR). 对该支持向量回归机的光滑函数进行了逼近性能和收敛性的分析,并说明该光滑函数比以往的光滑函数具有更高的逼近精度和收敛速度.

【关键词】 支持向量回归机; ϵ -不敏感损失函数; 光滑函数; 三次样条函数

【中图分类号】 TP311

【文献标识码】 A

Research on a new function for smooth support vector regression

SHEN Jin-dong

(College of Sciences, China Jiliang University, Hangzhou 310018, China)

Abstract: A new method in which the three-order spline function was used to smooth the model of support vector regression (SVR) was presented. A third-order spline smooth support vector regression (TSSSVR) was obtained. Moreover, by analyzing the function approximation and convergence capability, it is shown TSSSVR has higher precision than other previous smoothing functions.

Key words: support vector regression (SVR); ϵ -insensitive loss function; smoothing function; three-order spline function

光滑函数在一些重要的数学规划中得到成功应用^[1-6],特别是在分类问题和回归问题中获得良好的效果^[3,5-8]. 对于分类问题,光滑函数用于支持向量机的研究较为深入. 而回归问题一般可用于解决函数的逼近问题,与以往的函数逼近方法^[7,8]相比,支持向量机有其自身的优点. 对于回归问题,2005年 Lee 等人按照用光滑技术解决分类问题的基本思路,采用 p_ϵ^2 -函数作为光滑函数,逼近 ϵ -不敏感损失函数的平方,对原回归模型中

的目标函数作光滑处理,提出 ϵ -不敏感的光滑支持向量回归机模型(ϵ -insensitive smooth support vector regression, ϵ -SSVR)^[9],结果表明, ϵ -SSVR 的回归性能和效率比传统的 LIBSVM 和 SVM^{light} 方法^[7]好. 2008年,熊金志等将分类问题中的多项式光滑技术运用到支持向量回归机模型中,得到了一类多项式光滑支持向量回归机模型 (polynomial smooth support vector regression, PSSVR)^[10],所得结果显示,此方法比 ϵ -SSVR 方法好.

【收稿日期】 2010-03-11

【作者简介】 沈进东(1981-),男,安徽六安人,博士,主要研究方向为智能计算.

本文采用三次样条函数和复合函数的方法,提出一种新的支持向量回归机模型. 首先将求 ϵ -不敏感损失函数平方的光滑函数问题转化为求正号函数的光滑逼近问题;其次求它的复合函数,得到支持向量回归机的一类新光滑函数. 此外,我们还证明该类光滑函数同以往的光滑函数相比,具有更好的逼近性能.

1 三次样条光滑支持向量回归机

文献[7]中,支持向量回归机问题用如下的无约束最优化问题来描述:

$$\min_{(w,b) \in R^{n+1}} \frac{1}{2}(w^T w + b^2) + \frac{C}{2} \sum_{i=1}^m |A_i w + b - y_i|_\epsilon^2, \tag{1}$$

其中 $|x|_\epsilon = \max\{0, |x| - \epsilon\}$, 是不敏感损失函数, 具有如下性质:

$$\begin{aligned} |x|_\epsilon &= \max\{0, |x| - \epsilon\} = \\ &= \max\{0, x - \epsilon\} + \max\{0, -x - \epsilon\} = \\ &= (x - \epsilon)_+ + (-x - \epsilon)_+. \end{aligned}$$

同时注意到, 对所有的 $x \in R$ 和 $\epsilon > 0$ 有 $(x - \epsilon)_+ \cdot (-x - \epsilon)_+ = 0$, 那么

$$|x|_\epsilon^2 = (x - \epsilon)_+^2 + (-x - \epsilon)_+^2. \tag{2}$$

非约束最优化模型(1) 就是我们所要讨论的对象, 其优化目标函数是强凸非光滑的, 这主要是由其中的 ϵ -不敏感损失函数的平方造成的. 显然 $|x|_\epsilon^2$ 是一个对称函数, 在 $(-\epsilon, 0)$ 和 $(\epsilon, 0)$ 两点上不光滑, 如图 1 所示.

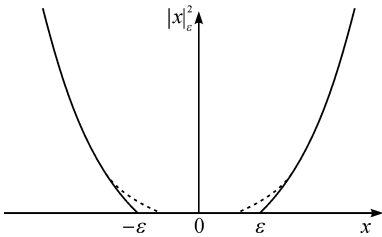


图 1 $|x|_\epsilon^2$ 的光滑函数

Figure 1 Smoothing function for $|x|_\epsilon^2$

这一特性导致很多好的算法不能用于求解问题(1), 这是因为很多求解优化问题的算法都要求目标函数满足一定的光滑性, 至少是一阶光滑的. 因此, 我们便试图采用光滑技术将不敏感损失函数 $|x|_\epsilon^2$ 加以处理, 使目标函数(1) 变为光滑.

在 ϵ -SSVR^[9] 中, Lee 采用 Sigmoid 函数的积分函数作为光滑函数, 来代替不光滑的 ϵ -不敏感损失函数的平方, 解决原回归模型不光滑的问题, 在 PSSVR^[11] 中, 熊金志等提出一类多项式光滑函数来逼近 ϵ -不敏感损失函数的平方. 而本文将采用如下的三次样条函数作为光滑函数:

$$T(x, k) = \begin{cases} 0, & x < -\frac{1}{k} \\ \frac{k^2}{6}x^3 + \frac{k}{2}x^2 + \frac{1}{2}x + \frac{1}{6k}, & -\frac{1}{k} \leq x < 0 \\ -\frac{k^2}{6}x^3 + \frac{k}{2}x^2 + \frac{1}{2}x + \frac{1}{6k}, & 0 \leq x < \frac{1}{k} \\ x, & x > \frac{1}{k} \end{cases} \tag{3}$$

也即直接以如下的光滑逼近函数代替式(2) 中的 ϵ -不敏感损失函数的平方 $|x|_\epsilon^2$:

$$T_\epsilon^2(x, k) = (T(x - \epsilon, k))^2 + (T(-x - \epsilon, k))^2. \tag{4}$$

从而得到一个三次样条光滑支持向量回归机模型 (Third-order spline smooth support vector regression, TSSSVR):

$$\begin{aligned} \min_{(w,b) \in R^{n+1}} \phi_{\epsilon,k}(w, b) &:= \min_{(w,b) \in R^{n+1}} \frac{1}{2}(w^T w + b^2) + \\ &+ \frac{C}{2} \sum_{i=1}^m T_\epsilon^2(A_i w + b - y_i, k) = \\ &= \min_{(w,b) \in R^{n+1}} \frac{1}{2}(w^T w + b^2) + \\ &+ \frac{C}{2} \mathbf{1}^T T_\epsilon^2(A \mathbf{w} + \mathbf{1}b - \mathbf{y}, k), \end{aligned} \tag{5}$$

其中 $T_\epsilon^2(A \mathbf{w} + \mathbf{1}b - \mathbf{y}, k) \in R^m$ 是一个向量, 其第 i 个元素为

$$T_\epsilon^2(A \mathbf{w} + \mathbf{1}b - \mathbf{y}, k)_i = T_\epsilon^2(A_i w + b - y_i, k),$$

$\mathbf{1}$ 是元素为 1 的列向量, $k > 0$ 是光滑因子.

2 光滑函数的性能分析

这一节, 我们将比较三次样条光滑回归机 (TSSSVR) 与以往其它光滑回归机的逼近性能和光滑性能.

引理 1^[9] 对于 p_ϵ^2 -函数:

$$p_\epsilon^2(x, k) = p(x - \epsilon, k)^2 + p(-x - \epsilon, k)^2,$$

其中

$$p(x, k) = x + \frac{1}{k} \ln(1 + e^{-kx}), k > 0.$$

e 是自然对数的底, 则

- 1) p_{ϵ}^2 - 函数关于 x 具有任意阶光滑性;
- 2) $p_{\epsilon}^2(x, k) \geq |x|_{\epsilon}^2$;
- 3) 任意给定的 $\rho > 0$, 当 $|x| < \rho$ 时,

$$p_{\epsilon}^2(x, k) - |x|_{\epsilon}^2 \leq 2\left(\frac{\log 2}{k}\right)^2 + \frac{2\rho}{k} \log 2.$$

引理 2^[12] 对于 $p_{1\epsilon}^2$ - 函数:

$$p_{1\epsilon}^2(x, k) = p_1(x - \epsilon, k)^2 + p_1(-x - \epsilon, k)^2,$$

其中

$$p_1(x, k) = \begin{cases} x, & x \geq \frac{1}{k} \\ \frac{k}{4}x^2 + \frac{1}{2}x + \frac{1}{4k}, & -\frac{1}{k} \leq x < \frac{1}{k}, k > 0 \\ 0, & x \leq -\frac{1}{k} \end{cases}$$

则

- 1) $p_{1\epsilon}^2(x, k)$ 具有一阶光滑性;
- 2) $p_{1\epsilon}^2(x, k) \geq |x|_{\epsilon}^2$;
- 3) $p_{1\epsilon}^2(x, k) - |x|_{\epsilon}^2 \leq 0.2/k^2$.

引理 3^[12] 对于 $p_{2\epsilon}^2$ - 函数:

$$p_{2\epsilon}^2(x, k) = p_2(x - \epsilon, k)^2 + p_2(-x - \epsilon, k)^2, \text{ 其中}$$

$$p_2(x, k) = \begin{cases} x, & x \geq \frac{1}{k} \\ -\frac{1}{16k}(kx + 1)^3(kx - 3), & -\frac{1}{k} \leq x < \frac{1}{k}, k > 0 \\ 0, & x \leq -\frac{1}{k} \end{cases}$$

则

- 1) $p_{2\epsilon}^2(x, k)$ 具有二阶光滑性;
- 2) $p_{2\epsilon}^2(x, k) \geq |x|_{\epsilon}^2$;
- 3) $p_{2\epsilon}^2(x, k) - |x|_{\epsilon}^2 \leq 0.08779/k^2$.

在讨论三次样条光滑回归机的性能之前, 先介绍下面由文献[3]给出的结论:

引理 4^[3] 假设 $x \in \mathbf{R}$, x_+ 是正号函数, 那么对于式(3)所定义的函数 $T(x, k)$ 有如下性质:

- 1) $T(x, k)$ 具有二阶光滑性;
- 2) $T(x, k) \geq x_+$;
- 3) $0 < T(x, k)^2 - x_+^2 \leq \frac{1}{24k^2}$.

由此给出下面关于 T_{ϵ}^2 - 函数的性能分析.

定理 1 T_{ϵ}^2 - 函数即 $T_{\epsilon}^2(x, k)$ 由式(4)给出, 则

- 1) $T_{\epsilon}^2(x, k)$ 具有二阶光滑性;
- 2) $T_{\epsilon}^2(x, k) \geq |x|_{\epsilon}^2$;
- 3) $T_{\epsilon}^2(x, k) - |x|_{\epsilon}^2 \leq 0.06944/k^2$.

证明 1) 对 $T_{\epsilon}^2(x, k)$ 求关于 x 一阶和二阶导, 分别可得

$$\begin{aligned} (T_{\epsilon}^2(x, k))' &= 2T(x - \epsilon, k)T'(x - \epsilon, k) - \\ &\quad 2T(-x - \epsilon, k)T'(-x - \epsilon, k), \\ (T_{\epsilon}^2(x, k))'' &= 2T'(x - \epsilon, k)^2 + 2T(x - \epsilon, k) \\ &\quad T''(x - \epsilon, k) + 2T'(-x - \epsilon, k)^2 + \\ &\quad 2T(-x - \epsilon, k)T''(-x - \epsilon, k). \end{aligned}$$

由引理 4 知 $T(x, k)$ 二阶光滑, 即 $T''(x, k)$ 存在且连续, 由上式知 $(T_{\epsilon}^2(x, k))''$ 也存在且连续, 所以 $(T_{\epsilon}^2(x, k))''$ 也二阶光滑.

2) 由引理 4 知, $T(x, k) \geq x_+ \geq 0$, 因(2)式, 故有

$$T_{\epsilon}^2(x, k) - |x|_{\epsilon}^2 = (T(x - \epsilon, k)^2 - (x - \epsilon)_+^2) + (T(-x - \epsilon, k)^2 - (-x - \epsilon)_+^2) \geq 0.$$

3) 分三种情况加以讨论.

(i) 当 $-\epsilon \leq x \leq \epsilon$ 时, $|x|_{\epsilon} = 0$. 因 $T(x, k)^2$ 是单调增函数, 所以

$$T_{\epsilon}^2(x, k) - |x|_{\epsilon}^2 = T(x - \epsilon, k)^2 + T(-x - \epsilon, k)^2,$$

又因此时 $x - \epsilon \leq 0$ 和 $-x - \epsilon \leq 0$, 有

$$T_{\epsilon}^2(x, k) - |x|_{\epsilon}^2 \leq 2T(0, k)^2 = 2\left(\frac{1}{6k}\right)^2 = 0.05556/k^2 < 0.06944/k^2.$$

(ii) 当 $x > \epsilon$ 时, $-x + \epsilon < 0$, 显然有 $-x - \epsilon < -x + \epsilon < 0$, 因此 $(-x - \epsilon)_+ = 0$. 由引理 4 知

$$\begin{cases} T(-x - \epsilon, k) \leq T(0, k), \\ T(x - \epsilon, k)^2 - (x - \epsilon)_+^2 \leq \frac{1}{24k^2}, \end{cases}$$

所以

$$\begin{aligned} T_{\epsilon}^2(x, k) - |x|_{\epsilon}^2 &= (T(x - \epsilon, k)^2 - \\ &\quad (x - \epsilon)_+^2) + T(-x - \epsilon, k)^2 \leq \\ &\quad \frac{1}{24k^2} + T(0, k)^2 = \frac{5}{72k^2} \approx 0.06944/k^2 \end{aligned}$$

(iii) 当 $x < -\epsilon$ 时, 证明类似(ii), 这里省略.

所以, 对 $x \in \mathbf{R}$, 有 $T_{\epsilon}^2(x, k) - |x|_{\epsilon}^2 \leq 0.06944/k^2$. 证毕.

3 TSSSVR 的收敛性分析

本小节中,将对三次样条支持向量回归机模型 TSSSVR(式(5))进行收敛性分析.可以证明,此模型的最优解在 k 趋于正无穷时能逼近原来的模型(式(1))的最优解.

定理 2 设 $A \in \mathbf{R}^{m \times n}, b \in \mathbf{R}^{m \times 1}$. 实函数 $h_\epsilon(x): \mathbf{R}^n \rightarrow \mathbf{R}, g_\epsilon(x, k): \mathbf{R}^n \times N \rightarrow \mathbf{R}$ 定义如下:

$$h_\epsilon(x) = \frac{1}{2} \sum_{i=1}^m |(\mathbf{A}_i x - b)|_\epsilon^2 + \frac{1}{2} \|x\|_2^2,$$

$$g_\epsilon(x, k) = \frac{1}{2} \sum_{i=1}^m T_\epsilon^2(\mathbf{A}_i x - b, k) + \frac{1}{2} \|x\|_2^2,$$

其中 $T_\epsilon^2(x, k)$ 由式(4)定义,则如下结论成立:

- 1) $h_\epsilon(x), g_\epsilon(x, k)$ 是强凸函数;
- 2) 优化问题 $\min_x h_\epsilon(x)$ 存在唯一解,记为 \bar{x} ;

优化问题 $\min_x g_\epsilon(x, k)$ 存在唯一解,记为 \bar{x}^k ;

- 3) 对任意 $k \geq 1$,有

$$\|\bar{x}^k - \bar{x}\|_2^2 \leq \frac{5m}{144k^2}$$

成立;

- 4) $\lim_{k \rightarrow \infty} \bar{x}^k = \bar{x}$.

证明:

- 1) 易知 $\sum_{i=1}^m \|(\mathbf{A}_i x - b)\|_\epsilon^2 = \|\mathbf{A}x - b\|_\epsilon^2$,

则由 $\|\cdot\|_2^2$ 的强凸性可得 $h_\epsilon(x), g_\epsilon(x, k)$ 是强凸函数.

2) 为了证明优化问题 $\min_x h_\epsilon(x)$ 和 $\min_x g_\epsilon(x, k)$ 的解的存在性,注意到 $T_\epsilon^2(x, k) \geq |x|_\epsilon^2$,则对任意 $v \geq 0$,水平集 $L_v(g_\epsilon(x, k))$ 和 $L_v(h_\epsilon(x))$ 满足:

$$L_v(g_\epsilon(x, k)) \subseteq L_v(h_\epsilon(x)) \subseteq \{x \mid \|x\|_2^2 \leq 2v\}.$$

因此, $L_v(g_\epsilon(x, k))$ 和 $L_v(h_\epsilon(x))$ 是 \mathbf{R}^n 中的紧子集,从而 $\min_x h_\epsilon(x)$ 和 $\min_x g_\epsilon(x, k)$ 的解的存在性得证;又由 1) 知道,目标函数具有强凸性,从而解具有唯一性,从而结论得证.

3) 为了证明 $\min_x g_\epsilon(x, k)$ 的最优解在 k 趋于正无穷时收敛于 $\min_x h_\epsilon(x)$ 的最优解,由一阶最优性条件和 $h_\epsilon(x), g_\epsilon(x, k)$ 的强凸性,有

$$h_\epsilon(\bar{x}^k) - h_\epsilon(\bar{x}) \geq \nabla h_\epsilon(\bar{x})(\bar{x}^k - \bar{x}) +$$

$$\frac{1}{2} \|\bar{x}^k - \bar{x}\|_2^2 = \frac{1}{2} \|\bar{x}^k - \bar{x}\|_2^2,$$

$$g_\epsilon(\bar{x}, k) - g_\epsilon(\bar{x}^k, k) \geq \nabla g_\epsilon(\bar{x}, k)(\bar{x} - \bar{x}^k) +$$

$$\frac{1}{2} \|\bar{x} - \bar{x}^k\|_2^2 = \frac{1}{2} \|\bar{x} - \bar{x}^k\|_2^2,$$

将上面两式相加,并注意到 $T_\epsilon^2(x, k) \geq |x|_\epsilon^2$,从而 $g_\epsilon(x, k) \geq h_\epsilon(x) \geq 0$,因此有

$$\|\bar{x}^k - \bar{x}\|_2^2 \leq (g_\epsilon(\bar{x}, k) - h_\epsilon(\bar{x})) -$$

$$(g_\epsilon(\bar{x}^k, k) - h_\epsilon(\bar{x}^k)) \leq$$

$$(g_\epsilon(\bar{x}, k) - h_\epsilon(\bar{x})) =$$

$$\frac{1}{2} \sum_{i=1}^m T_\epsilon^2(\mathbf{A}_i \bar{x} - b, k) -$$

$$\frac{1}{2} \sum_{i=1}^m |(\mathbf{A}_i \bar{x} - b)|_\epsilon^2$$

应用定理 1,得到 $\|\bar{x}^k - \bar{x}\|_2^2 \leq \frac{5m}{144k^2}$,从而

结论得证.

- 4) 由 3) 的结论易知

$$\lim_{k \rightarrow \infty} \|\bar{x}^k - \bar{x}\|_2^2 \leq \lim_{k \rightarrow \infty} \frac{5m}{144k^2} = 0,$$

因此, $\lim_{k \rightarrow \infty} \bar{x}^k = \bar{x}$.

证毕.

4 与其它方法比较

4.1 逼近精度的比较

四种光滑函数均能逼近 ϵ -不敏感损失函数的平方,但是它们的逼近精度却有所不同,由图 2 及表 1 可见本文所提的 TSSSVR 方法效果最好,精度最高.

表 1 四类光滑函数的逼近精度和收敛速度

Table 1 Comparison of approximation precision and convergence rate among four smoothing functions

模型	光滑函数表达式	误差上界	收敛上界
ϵ -SSVR	$x + \frac{1}{k} \log(1 + \epsilon^{-kx})$	1.385 4 / k^2	0.692 7m / k^2
1PSSVR	$\frac{k}{4}x^2 + \frac{1}{2}x + \frac{1}{4k}$	0.2 / k^2	0.1m / k^2
2PSSVR	$-\frac{1}{16k}(kx+1)^3(kx-3)$	0.087 79 / k^2	0.043 895m / k^2

$$\text{TSSSVR} \begin{cases} \frac{k^2}{6}x^3 + \frac{k}{2}x^2 + \frac{1}{2}x + \frac{1}{6k}, \\ \quad -\frac{1}{k} \leq x < 0 \\ -\frac{k^2}{6}x^3 + \frac{k}{2}x^2 + \frac{1}{2}x + \frac{1}{6k}, \\ \quad 0 \leq x < \frac{1}{k} \end{cases} \begin{matrix} 0.069 44 & 0.034 72m \\ /k^2 & /k^2 \end{matrix}$$

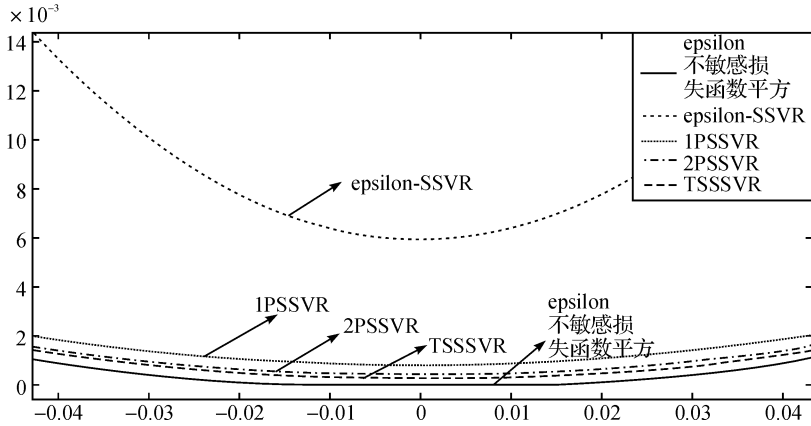


图2 $k=10, \epsilon=0.01$ 时, $p_{\epsilon}^2(x, k), p_{1\epsilon}^2(x, k), p_{2\epsilon}^2(x, k), T_{\epsilon}^2(x, k)$ 函数图像对比图

Figure 2 Comparison of function image among $p_{\epsilon}^2(x, k), p_{1\epsilon}^2(x, k), p_{2\epsilon}^2(x, k), T_{\epsilon}^2(x, k)$

4.2 收敛速度的比较

四种模型 (ϵ -SSVR, 1PSSVR, 2PSSVR 和 TSSSVR) 均是二次收敛的, 但是它们的收敛速度有快有慢. 由表 1 中的第四列可见, 本文的模型 TSSSVR 收敛速度最快, 同时相比 2PSSVR, 光滑函数的次数更低, 而且是奇数次, 解决了多项式函数作为光滑函数一直为偶数次的遗憾(表 1 第二列).

5 结 语

本文将三次样条函数作为光滑函数从而获得一种新的光滑支持向量回归机模型, 即三次样条支持向量回归机(TSSSVR). 与以往其它光滑函数所构造的光滑支持向量回归机相比, TSSSVR 无论是在逼近性能上还是收敛速度上, 都有较大的优势. 在未来, 如何获得更好的光滑函数仍然是一个值得进一步研究的课题.

【参 考 文 献】

[1] CHEN C H, MANGSARIAN O L. A class of smoothing functions for nonlinear and mixed complementarity problems[J]. Computational Optimization and Application, 1996, 5(2): 97-138.
 [2] 袁玉波, 严杰, 徐成贤. 多项式光滑的支撑向量机[J]. 计

算机学报, 2005, 28(1): 9-17.

[3] YUAN Y B, LI C Z. A new smooth support vector machine[C]// Lecture notes in artificial intelligence, 2005: 3801:594-598.
 [4] CHEN C, MANGASARIAN O L. Smoothing methods for convex inequalities and linear complementarity problems[J]. Mathematical Programming, 1995, 71: 51-69.
 [5] MANGASARIAN O L. Mathematical programming in neural networks[J]. ORSA Journal on Computing, 1993, 5(4): 349-360.
 [6] LEE Y J, MANGASARIAN O L. SSVM: A smooth support vectormachine for classification[J]. Computational Optimization and Applications, 2001, 22(1): 5-21.
 [7] 闫灿伟, 曹飞龙. m 依赖过程经验风险最小化算法的泛函性能[J]. 中国计量学院学报, 2009, 20(4): 357-361.
 [8] 陈晓, 谢庭藩. 人工神经网络的单调序列逼近[J]. 中国计量学院学报, 2009, 20(4): 354-356.
 [9] LEE Y J, HSIEH W F, HUANG C M. ϵ -SSVR: A smooth support vector machine for ϵ -insensitive regression[J]. IEEE Transactions on Knowledge and Data Engineering, 2005, 17(5): 5-22.
 [10] 后敏, 曹飞龙. 一类神经网络逼近可积函数[J]. 中国计量学院学报, 2007, 18(2): 155-158.
 [11] 熊金志, 胡金莲, 袁华强, 等. 一类光滑支持向量机新函数的研究[J]. 电子学报, 2007, 35(2): 366-370.
 [12] 熊金志, 胡金莲, 袁华强, 等. 支持向量回归机的光滑函数研究[J]. 模式识别与人工智能, 2008, 21(3): 273-279.