计算机研究与发展
Journal of Computer Research and Development
DOI:10.7544/issn1000-1239.2015.20131444
52(3):561-568,2015

光滑 CHKS 孪生支持向量回归机

黄华娟[1,2,3]　丁世飞[1,2]　史忠植[2]
[1](中国矿业大学计算机科学与技术学院　江苏徐州　221116)
[2](中国科学院计算技术研究所智能信息处理重点实验室　北京　100190)
[3](广西民族大学信息科学与工程学院　南宁　530006)
(hhj-025@163.com)

Smooth CHKS Twin Support Vector Regression

Huang Huajuan[1,2,3], Ding Shifei[1,2], and Shi Zhongzhi[2]
[1](*School of Computer Science and Technology, China University of Mining and Technology, Xuzhou, Jiangsu* 221116)
[2](*Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, Beijing* 100190)
[3](*College of Information Science and Engineering, Guangxi University for Nationalities, Nanning* 530006)

Abstract　Twin support vector regression (TSVR) was proposed recently as a novel regressor that tries to find a pair of nonparallel planes, i.e., ε insensitive down- and up- bounds, by solving two related SVM-type problems. However, it may incur suboptimal solution since its objective function is positive semi-definite and it is lack of complexity control. In order to address this shortcoming, smooth twin support vector regression (STSVR) is introduced using sigmoid function as smoothing technique to convert the original problems into unconstrained minimization, which can improve the training speed. However, its accuracy needs to be improved. In this paper, aiming at the low approximation ability of sigmoid function of STSVR, using CHKS (chen-harker-kanzow-smale) function which has better approximation ability as the smooth function, a new version of smooth TSVR called smooth CHKS twin support vector regression (SCTSVR) model is proposed. In SCTSVR, CHKS function is used to approximate the non-differential term of twin support vector regression. Then Newton-Armijo algorithm is used to solve the corresponding model. We have proved that SCTSVR is not only strictly convex, but also can meet the arbitrary order smooth performance. Meanwhile, the experimental results on several artificial and benchmark datasets show that SCTSVR has better regression performance than STSVR.

Key words　twin support vector regression; chen-harker-kanzow-smale (CHKS); smooth function; Newton-Armijo algorithm; strictly convex

摘　要　针对目前光滑孪生支持向量回归机(smooth twin support vector regression, STSVR)中采用的 Sigmoid 光滑函数逼近精度不高,从而导致算法泛化能力不够理想的问题,引入一种具有更强逼近能力的光滑(chen-harker-kanzow-smale, CHKS)函数,采用 CHKS 函数逼近孪生支持向量回归机的不可微项,并用 Newton-Armijo 算法求解相应的模型,提出了光滑 CHKS 孪生支持向量回归机(smooth CHKS twin support vector regression, SCTSVR). 不仅从理论上证明了 SCTSVR 具有严格凸,能满足任意阶光滑和全局收敛的性能,而且在人工数据集和 UCI 数据集上的实验表明了 SCTSVR 比 STSVR 具有更好的回归性能.

收稿日期:2013-11-06;**修回日期**:2014-08-25
基金项目:国家"九七三"重点基础研究发展计划基金项目(2013CB329502);国家自然科学基金项目(61379101)
通信作者:丁世飞(dingsf@cumt.edu.cn)

关键词 孪生支持向量回归机;CHKS 函数;光滑函数;Newton-Armijo 算法;严格凸

中图法分类号 TP18

支持向量机(support vector machine, SVM)是一种基于统计学习理论的机器学习方法[1-3]. 和人工神经网络相比[4],SVM 成功解决了高维问题和局部极小值问题,因此具有更好的泛化能力.

然而,SVM 的训练时间非常高,达到 $O(l^3)$,其中,l 是整个训练集的样本数. 为了提高 SVM 的训练速度,学者们提出了多种改进算法,最经典的有选块算法[5](chunking algorithm)和序列最小优化算法[6](sequential minimal optimization, SMO)等. 这些改进方法虽然在一定程度上提高了 SVM 的学习性能,但其算法实现比较复杂. 近年来,学者们开始研究基于标准 SVM 的变形算法[7-8]. 特别是 2007 年 Jayadeva 等人[9]提出了孪生支持向量机(twin support vector machines, TWSVM). TWSVM 是要得到 2 个非平行的分类超平面,使得每一个超平面靠近其中的一类样本,而远离另一类样本. TWSVM 在形式上类似于标准 SVM,只是标准 SVM 解决的是 1 个大的二次规划问题(quadratic programming problem, QPP),而 TWSVM 需要解决的是 2 个小的 QPP,这个策略使得 TWSVM 的训练速度大大提高. 鉴于其明显的分类优势,TWSVM 已被应用于说话人识别[10]、医学检测[11]等领域.

2010 年基于 TWSVM 的思想,Peng 等人[12]提出了孪生支持向量回归机 (twin support vector regression, TSVR). TSVR 产生一对超平面,分别确定目标回归函数的 ε 不敏感上、下界. TSVR 仅需要求解一对较小规模的 QPP,每个 QPP 所含约束条件的数目仅为传统 SVR 的一半,并且 TSVR 的对偶问题中没有等式约束,这使得 TSVR 的训练速度大为提高. 然而,TSVR 是在对偶空间中求解 2 个 QPP,这种求解方法对于样本数目较大的问题将受到时间和内存的制约. 2010 年 Chen 等人[13]采用 Sigmoid 积分函数,对原空间 TSVR 的目标函数作光滑处理,提出了光滑孪生支持向量回归机(smooth twin support vector regression, STSVR),并采用快速算法进行求解,其实验结果表明回归效果比 TSVR 好. Sigmoid 函数的积分函数虽然具有任意阶光滑,但其逼近精度较差,使得 STSVR 模型的效果有待进一步提高. 为了进一步改善 STSVR 的回归效果,寻找性能更好的光滑函数是研究光滑孪生支持向量回归机的一个关键问题. 文献[14]将 CHKS 函数引入到支持向量回归机中,提出了光滑 CHKS 支持向量回归机. 该文实验结果表明 CHKS 作为光滑函数比 Sigmoid 的逼近精度高. 基于此,在本文中,采用 CHKS 函数作为光滑函数来逼近 TSVR 的目标函数,最后采用 Newton-Armijo 方法求解得到的新模型,从而提出光滑 CHKS 孪生支持向量回归机(smooth CHKS twin support vector regression, SCTSVR). 实验结果表明,与 STSVR 相比,SCTSVR 可通过批处理训练来提高训练速度,节约学习过程的存储空间,在训练速度和逼近精度上均有明显提高.

1 光滑 CHKS 孪生支持向量回归机

1.1 SCTSVR 原理及其性质

类似于 TWSVM 的思想,TSVR 将在训练数据点两侧产生一对不平行的函数,分别确定回归函数的 ε 不敏感上、下界.

对于线性情况,TSVR 通过训练数据的 ε_1 不敏感下界

$$f_1(x) = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{w}_1 + b_1 \tag{1}$$

与 ε_2 不敏感上界

$$f_2(x) = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{w}_2 + b_2 \tag{2}$$

确定最终的回归函数,而这对函数的确定可以通过求解下面的一对 QPP:

$$\min \quad \frac{1}{2}\|\boldsymbol{Y} - \boldsymbol{e}\boldsymbol{\varepsilon}_1 - (\boldsymbol{A}\boldsymbol{w}_1 + \boldsymbol{e}b_1)\|^2 + C_1\boldsymbol{e}^{\mathrm{T}}\boldsymbol{\xi},$$

$$\text{s.t.} \quad \boldsymbol{Y} - (\boldsymbol{A}\boldsymbol{w}_1 + \boldsymbol{e}b_1) \geqslant \boldsymbol{e}\boldsymbol{\varepsilon}_1 - \boldsymbol{\xi}, \boldsymbol{\xi} \geqslant 0; \tag{3}$$

$$\min \quad \frac{1}{2}\|\boldsymbol{Y} + \boldsymbol{e}\boldsymbol{\varepsilon}_2 - (\boldsymbol{A}\boldsymbol{w}_2 + \boldsymbol{e}b_2)\|^2 + C_2\boldsymbol{e}^{\mathrm{T}}\boldsymbol{\eta},$$

$$\text{s.t.} \quad (\boldsymbol{A}\boldsymbol{w}_2 + \boldsymbol{e}b_2) - \boldsymbol{Y} \geqslant \boldsymbol{e}\boldsymbol{\varepsilon}_2 - \boldsymbol{\eta}, \boldsymbol{\eta} \geqslant 0. \tag{4}$$

其中,$C_1, C_2 > 0$,$\varepsilon_1, \varepsilon_2 > 0$ 为常数,$\boldsymbol{\xi}, \boldsymbol{\eta}$ 为松弛变量,$\boldsymbol{e}$ 为 $l \times 1$ 维的单位列向量.

在求解式(3)和式(4)时,一般都在对偶空间中进行求解. 当训练样本数目较多时这种求解方法将使算法效率降低. 为了提高 TSVR 的效率,本文直接在原空间求解式(3)和式(4).

由 KKT 条件,式(3)和式(4)的松弛变量 $\boldsymbol{\xi}$ 和 $\boldsymbol{\eta}$ 可以变为

$$\boldsymbol{\xi} = \max(0, \boldsymbol{e}\boldsymbol{\varepsilon}_1 - \boldsymbol{Y} + \boldsymbol{A}\boldsymbol{w}_1 + \boldsymbol{e}b_1), \tag{5}$$

$$\boldsymbol{\eta} = \max(0, \boldsymbol{e}\boldsymbol{\varepsilon}_2 + \boldsymbol{Y} - (\boldsymbol{A}\boldsymbol{w}_2 + \boldsymbol{e}b_2)). \tag{6}$$

将式(5)和式(6)分别带入式(3)和式(4),则可以得到无约束的 TSVR 模型:

$$\min \quad \frac{1}{2}\|\boldsymbol{Y}-\boldsymbol{e}\varepsilon_1-(\boldsymbol{A}\boldsymbol{w}_1+\boldsymbol{e}b_1)\|^2+C_1\boldsymbol{e}^{\mathrm{T}}\max(0,\boldsymbol{e}\varepsilon_1-\boldsymbol{Y}+\boldsymbol{A}\boldsymbol{w}_1+\boldsymbol{e}b_1), \tag{7}$$

$$\min \quad \frac{1}{2}\|\boldsymbol{Y}+\boldsymbol{e}\varepsilon_2-(\boldsymbol{A}\boldsymbol{w}_2+\boldsymbol{e}b_2)\|^2+C_2\boldsymbol{e}^{\mathrm{T}}\max(0,\boldsymbol{e}\varepsilon_2+\boldsymbol{Y}-(\boldsymbol{A}\boldsymbol{w}_2+\boldsymbol{e}b_2)). \tag{8}$$

令:

$$\phi_1(\boldsymbol{w}_1,b_1)=\max(0,\boldsymbol{e}\varepsilon_1-\boldsymbol{Y}+\boldsymbol{A}\boldsymbol{w}_1+\boldsymbol{e}b_1)=(u_1)_+, \tag{9}$$

$$\phi_2(\boldsymbol{w}_2,b_2)=\max(0,\boldsymbol{e}\varepsilon_2+\boldsymbol{Y}-(\boldsymbol{A}\boldsymbol{w}_2+\boldsymbol{e}b_2))=(u_2)_+, \tag{10}$$

其中,$(u_1)_+$和$(u_2)_+$为正号函数.

则式(7)和式(8)可以写成:

$$\min \frac{1}{2}\|\boldsymbol{Y}-\boldsymbol{e}\varepsilon_1-(\boldsymbol{A}\boldsymbol{w}_1+\boldsymbol{e}b_1)\|^2+C_1\boldsymbol{e}^{\mathrm{T}}(u_1)_+, \tag{11}$$

$$\min \frac{1}{2}\|\boldsymbol{Y}+\boldsymbol{e}\varepsilon_2-(\boldsymbol{A}\boldsymbol{w}_2+\boldsymbol{e}b_2)\|^2+C_2\boldsymbol{e}^{\mathrm{T}}(u_2)_+. \tag{12}$$

定理 1. 无约束的孪生支持向量回归机模型由式(11)和式(12)给出,该模型连续但不光滑.

由定理 1 可知,式(11)和式(12)的第 2 项不光滑,无法用快速的梯度方法进行求解,因为使用这些优化方法的前提条件要求目标函数是一阶或二阶可微的. 为此,本文采用 CHKS 函数 $\phi(u,\varepsilon)=\frac{u+\sqrt{u^2+4\varepsilon^2}}{2}$来逼近 $\phi(u)=\max\{0,u\}$,其中 ε 是足够小的参数. 令 $f_1=\boldsymbol{e}\varepsilon_1-\boldsymbol{Y}+\boldsymbol{A}\boldsymbol{w}_1+\boldsymbol{e}b_1$,$f_2=\boldsymbol{e}\varepsilon_2+\boldsymbol{Y}-(\boldsymbol{A}\boldsymbol{w}_2+\boldsymbol{e}b_2)$,则可得 $\phi_1(\boldsymbol{w}_1,b_1)$ 和 $\phi_2(\boldsymbol{w}_2,b_2)$ 的光滑函数分别为

$$\phi_1(\boldsymbol{w}_1,b_1,\varepsilon)=\frac{f_1+\sqrt{f_1+4\varepsilon^2}}{2}, \tag{13}$$

$$\phi_2(\boldsymbol{w}_2,b_2,\varepsilon)=\frac{f_2+\sqrt{f_2+4\varepsilon^2}}{2}. \tag{14}$$

采用式(13)和式(14)作为光滑函数可得到光滑 CHKS 孪生支持向量回归机的模型为

$$\min \Phi_1(\boldsymbol{w}_1,b_1,\varepsilon)=\frac{1}{2}\|\boldsymbol{Y}-\boldsymbol{e}\varepsilon_1-(\boldsymbol{A}\boldsymbol{w}_1+\boldsymbol{e}b_1)\|^2+C_1\boldsymbol{e}^{\mathrm{T}}\phi_1(\boldsymbol{w}_1,b_1,\varepsilon), \tag{15}$$

$$\min \Phi_2(\boldsymbol{w}_2,b_2,\varepsilon)=\frac{1}{2}\|\boldsymbol{Y}+\boldsymbol{e}\varepsilon_2-(\boldsymbol{A}\boldsymbol{w}_2+\boldsymbol{e}b_2)\|^2+C_2\boldsymbol{e}^{\mathrm{T}}\phi_2(\boldsymbol{w}_2,b_2,\varepsilon). \tag{16}$$

定理 2. 光滑 CHKS 孪生支持向量回归机由式(15)和式(16)给出,则:

1) 光滑 CHKS 孪生支持向量回归机满足任意阶光滑,即对任意的 $\boldsymbol{w}_1\in\mathbb{R}^n,\boldsymbol{w}_2\in\mathbb{R}^n,b_1\in\mathbb{R},b_2\in\mathbb{R}$,$\Phi_1(\boldsymbol{w}_1,b_1,\varepsilon)$和 $\Phi_2(\boldsymbol{w}_2,b_2,\varepsilon)$分别关于 $\boldsymbol{w}_1,b_1$ 和 $\boldsymbol{w}_2,b_2$ 任意阶光滑;

2) 对任意的 $\boldsymbol{w}_1\in\mathbb{R}^n,\boldsymbol{w}_2\in\mathbb{R}^n,b_1\in\mathbb{R},b_2\in\mathbb{R}$,$\Phi_1(\boldsymbol{w}_1,b_1,\varepsilon)$和 $\Phi_2(\boldsymbol{w}_2,b_2,\varepsilon)$是单调递增函数;

3) 对任意的 $\boldsymbol{w}_1\in\mathbb{R}^n,\boldsymbol{w}_2\in\mathbb{R}^n,b_1\in\mathbb{R},b_2\in\mathbb{R}$,$\varepsilon>0$,有:

$$\Phi_1(\boldsymbol{w}_1,b_1)\leqslant\Phi_1(\boldsymbol{w}_1,b_1,\varepsilon)\leqslant\Phi_1(\boldsymbol{w}_1,b_1)+\varepsilon,$$

$$\Phi_2(\boldsymbol{w}_2,b_2)\leqslant\Phi_2(\boldsymbol{w}_2,b_2,\varepsilon)\leqslant\Phi_2(\boldsymbol{w}_2,b_2)+\varepsilon;$$

4) 对任意的 $\varepsilon>0$,光滑 CHKS 孪生支持向量回归机是连续可微且是严格凸的.

证明.

1) 容易证明 $\phi_1(\boldsymbol{w}_1,b_1,\varepsilon)$和 $\phi_2(\boldsymbol{w}_2,b_2,\varepsilon)$是任意阶光滑的,因此结论成立,证明略.

2) 在此我们只证明 $\Phi_1(\boldsymbol{w}_1,b_1,\varepsilon)$是单调递增的,$\Phi_2(\boldsymbol{w}_2,b_2,\varepsilon)$同理可证.

对任意的 $\boldsymbol{x}\in\mathbb{R}^n,\varepsilon>0$,有:

$$\frac{\partial\phi_1(x,\varepsilon)}{\partial\varepsilon}=\frac{2\varepsilon}{\sqrt{x^2+4\varepsilon^2}}>0,$$

$$\frac{\partial\phi_1(x,\varepsilon)}{\partial x}=\frac{1}{2}\left(1+\frac{x}{\sqrt{x^2+4\varepsilon^2}}\right)=\frac{\sqrt{x^2+4\varepsilon^2}+x}{2\sqrt{x^2+4\varepsilon^2}}>0,$$

因此函数 $\phi_1(x,\varepsilon)$关于 ε 是单调递增函数,即对任意的 $0<\varepsilon_1<\varepsilon_2$ 有 $\phi_1(x,\varepsilon_1)<\phi_1(x,\varepsilon_2)$,结论显然是成立的.

3) 由于:

$$\phi_1(x,\varepsilon)-\max\{0,x\}=\frac{(\sqrt{x^2+4\varepsilon^2}+x)}{2}-\frac{x+\sqrt{x^2}}{2}=\frac{\sqrt{x^2+4\varepsilon^2}-\sqrt{x^2}}{2},$$

因此对任意的 $\boldsymbol{w}_1\in\mathbb{R}^n,b_1\in\mathbb{R},\varepsilon>0$,有 $0\leqslant\phi_1(\boldsymbol{w}_1,b_1,\varepsilon)-\phi_1(\boldsymbol{w}_1,b_1)\leqslant\varepsilon$.

由 $\Phi_1(\boldsymbol{w}_1,b_1,\varepsilon)$的定义易得:$\Phi_1(\boldsymbol{w}_1,b_1)\leqslant\Phi_1(\boldsymbol{w}_1,b_1,\varepsilon)\leqslant\Phi_1(\boldsymbol{w}_1,b_1)+\varepsilon$.

而 $\Phi_2(\boldsymbol{w}_2,b_2)\leqslant\Phi_2(\boldsymbol{w}_2,b_2,\varepsilon)\leqslant\Phi_2(\boldsymbol{w}_2,b_2)+\varepsilon$ 同理可证.

4) 由定理 1 可知,对任意的 $\varepsilon>0$,$\Phi_1(\boldsymbol{w}_1,b_1,\varepsilon)$显然是连续可微的. 下面证明它是严格凸函数. 由式(13)和式(15)可得:

$$\frac{\partial\Phi_1(\boldsymbol{w}_1,b_1,\varepsilon)}{\partial\boldsymbol{w}_1}=-\boldsymbol{A}(\boldsymbol{Y}-\boldsymbol{e}\varepsilon_1-(\boldsymbol{A}\boldsymbol{w}_1+\boldsymbol{e}b_1))+\frac{C_1\boldsymbol{e}^{\mathrm{T}}}{2}\boldsymbol{A}\left(1+\frac{f_1}{\sqrt{f_1+4\varepsilon^2}}\right), \tag{17}$$

$$\frac{\partial \Phi_1(\boldsymbol{w}_1,b_1,\varepsilon)}{\partial b_1} = -\boldsymbol{e}(\boldsymbol{Y}-\boldsymbol{e}\boldsymbol{\varepsilon}_1-(\boldsymbol{A}\boldsymbol{w}_1+\boldsymbol{e}b_1))+\frac{C_1\boldsymbol{e}^{\mathrm{T}}}{2}\boldsymbol{A}\left(1+\frac{f_1}{\sqrt{f_1+4\varepsilon^2}}\right). \tag{18}$$

于是有:

$$\nabla^2\Phi_1(\boldsymbol{w}_1,b_1,\varepsilon) = \begin{pmatrix}\boldsymbol{A}^2+\lambda(\boldsymbol{w}_1,b_1,\varepsilon) & \boldsymbol{A}\boldsymbol{e}+\lambda(\boldsymbol{w}_1,b_1,\varepsilon)\\ \boldsymbol{A}\boldsymbol{e}+\lambda(\boldsymbol{w}_1,b_1,\varepsilon) & \boldsymbol{e}^2+\lambda(\boldsymbol{w}_1,b_1,\varepsilon)\end{pmatrix}, \tag{19}$$

其中:

$$\lambda(\boldsymbol{w}_1,b_1,\varepsilon) = \frac{2C_1\boldsymbol{e}\boldsymbol{\varepsilon}^2}{(\sqrt{f_1^2+4\varepsilon^2})^3}.$$

任意的 $\boldsymbol{\xi}^{*\mathrm{T}}=(\boldsymbol{\xi}^{\mathrm{T}},\xi_0)\in\mathbb{R}^{n+1}$ 且 $\boldsymbol{\xi}^*\neq 0,\boldsymbol{\xi}\in\mathbb{R}^n$, 由于 $\lambda(\boldsymbol{w}_1,b_1,\varepsilon)>0$, 因此有:

$$\boldsymbol{\xi}^{*\mathrm{T}}\nabla^2\Phi_1(\boldsymbol{w}_1,b_1,\varepsilon)\boldsymbol{\xi}^* = \|\boldsymbol{\xi}\|^2(\boldsymbol{A}+\boldsymbol{e})^2+\|\boldsymbol{\xi}\|^2\lambda(\boldsymbol{w}_1,b_1,\varepsilon)>0.$$

因此, 对任意的 $\varepsilon>0$, $\Phi_1(\boldsymbol{w}_1,b_1,\varepsilon)$ 是严格凸函数, 同理可证 $\Phi_2(\boldsymbol{w}_2,b_2,\varepsilon)$ 也是凸函数. 所以, 对任意的 $\varepsilon>0$, 光滑 CHKS 孪生支持向量回归机连续可微且严格凸. 证毕.

1.2 SCTSVR 的 Newton-Armijo 算法

对二阶可微的无约束规划问题, 可以采用 Newton-Armijo 算法来求解. SCTSVR 的目标函数是任意阶光滑的, 因此可以使用具有快速收敛能力的 Newton-Armijo 方法进行求解, 而且 SCTSVR 的目标函数是严格凸的, 使用 Newton-Armijo 方法求解可以全局收敛, 即可以得到唯一的极小点. 基于 Newton-Armijo 方法的 SCTSVR 算法的具体步骤如算法 1 所示.

算法 1. SCTSVR 算法.

输入: 给定初始点 $(\boldsymbol{w}^0,\gamma^0)\in\mathbb{R}^{n+1}$, η, 令迭代步骤 $i=0$;

输出: 目标函数的最优值.

① 计算 $\boldsymbol{\Phi}^i=\Phi(\boldsymbol{w}^i,b^i;\varepsilon)$, $g^i=\nabla\Phi(\boldsymbol{w}^i,b^i;\varepsilon)$;

② 如果 $\|g^i\|\leqslant\eta$, 则选取 $(\boldsymbol{w}^*,b^*)=(\boldsymbol{w}^i,b^i)$, 算法停止; 否则由方程 $\nabla^2\Phi(\boldsymbol{w}^i,b^i;\varepsilon)d^i=-g^i$, 计算下降方向 d^i;

③ 取 $\delta\in\left(0,\frac{1}{2}\right)$, $\lambda_i=\max\left\{1,\frac{1}{2},\frac{1}{4},\cdots\right\}$, 使得:

$$\Phi(\boldsymbol{w}^i,b^i;\varepsilon)-\Phi((\boldsymbol{w}^i,b^i)+\lambda_i d^i;\varepsilon)\geqslant-\delta\lambda_i g^i d^i$$

成立, 令 $(\boldsymbol{w}^{i+1},b^{i+1})=(\boldsymbol{w}^i,b^i)+\lambda_i d^i$.

④ 令 $i\leftarrow i+1$, 转步骤②.

引理 1. 设 $(\boldsymbol{w}_1^k,b_1^k)$ 和 $(\boldsymbol{w}_2^k,b_2^k)$ 是目标函数式(11)和式(12)的极小点, $(\boldsymbol{w}_1^*,b_1^*)$ 和 $(\boldsymbol{w}_2^*,b_2^*)$ 是式(7)和式(8)的极小点, 则当光滑因子 ε 趋于无穷小时, 有:

$$0\leqslant\Phi_1(\boldsymbol{w}_1^k,b_1^k,\varepsilon)-\Phi_1(\boldsymbol{w}_1^*,b_1^*)\leqslant C\varepsilon, \tag{20}$$

$$0\leqslant\Phi_2(\boldsymbol{w}_2^k,b_2^k,\varepsilon)-\Phi_2(\boldsymbol{w}_2^*,b_2^*)\leqslant C\varepsilon, \tag{21}$$

其中, C 为常数.

证明. 由定理 2 可知:

$$\Phi_1(\boldsymbol{w}_1^k,b_1^k)\leqslant\Phi_1(\boldsymbol{w}_1^k,b_1^k,\varepsilon)\leqslant\Phi_1(\boldsymbol{w}_1^k,b_1^k)+C\varepsilon, \tag{22}$$

$$\Phi_2(\boldsymbol{w}_2^k,b_2^k)\leqslant\Phi_2(\boldsymbol{w}_2^k,b_2^k,\varepsilon)\leqslant\Phi_2(\boldsymbol{w}_2^k,b_2^k)+C\varepsilon, \tag{23}$$

再由 $(\boldsymbol{w}_1^k,b_1^k)$ 和 $(\boldsymbol{w}_1^*,b_1^*)$ 分别是规划问题(式(11)和式(7))的极小点, $(\boldsymbol{w}_2^k,b_2^k)$ 和 $(\boldsymbol{w}_2^*,b_2^*)$ 分别是规划问题(式(12)和式(8))的极小点, 可证结论成立. 证毕.

定理 3. 光滑 CHKS 孪生支持向量回归机的解是全局收敛的, 即设 $(\boldsymbol{w}_1^k,b_1^k)$ 和 $(\boldsymbol{w}_2^k,b_2^k)$ 是目标函数式(11)和式(12)的极小点, 则存在式(7)和式(8)的极小点 $(\boldsymbol{w}_1^*,b_1^*)$ 和 $(\boldsymbol{w}_2^*,b_2^*)$, 使得:

$$\lim_{k\to\infty}(\boldsymbol{w}_1^k,b_1^k)=(\boldsymbol{w}_1^*,b_1^*), \tag{24}$$

$$\lim_{k\to\infty}(\boldsymbol{w}_2^k,b_2^k)=(\boldsymbol{w}_2^*,b_2^*). \tag{25}$$

证明. 由于 $\Phi_1(\boldsymbol{w}_1,b_1,\varepsilon)$ 和 $\Phi_2(\boldsymbol{w}_2,b_2,\varepsilon)$ 是严格凸函数, 可得式(11)和式(12)的极小值点 $(\boldsymbol{w}_1^k,b_1^k)$ 和 $(\boldsymbol{w}_2^k,b_2^k)$ 唯一. 由定理 2 可知:

$$0\leqslant\Phi_1(\boldsymbol{w}_1^k,b_1^k,\varepsilon)-\Phi_1(\boldsymbol{w}_1^*,b_1^*)\leqslant\varepsilon,$$

$$0\leqslant\Phi_2(\boldsymbol{w}_2^k,b_2^k,\varepsilon)-\Phi_2(\boldsymbol{w}_2^*,b_2^*)\leqslant\varepsilon,$$

因此点列 $\{(\boldsymbol{w}_1^k,b_1^k)\}_{k=1}^{+\infty}$ 和 $\{(\boldsymbol{w}_2^k,b_2^k)\}_{k=1}^{+\infty}$ 收敛, 再由 $\Phi_1(\boldsymbol{w}_1,b_1)$ 和 $\Phi_2(\boldsymbol{w}_2,b_2)$ 的连续性可证. 证毕.

1.3 非线性 CSTSVR

对于非线性情况, 我们利用含有核函数的非线性回归估计函数代替线性回归估计函数, 把线性 SCTSVR 推广到非线性 SCTSVR, 类似线性情况, 我们得到非线性回归估计函数:

$$f_1(x) = K(\boldsymbol{x}^{\mathrm{T}},\boldsymbol{A}^{\mathrm{T}})\boldsymbol{w}_1+b_1, \tag{26}$$

$$f_2(x) = K(\boldsymbol{x}^{\mathrm{T}},\boldsymbol{A}^{\mathrm{T}})\boldsymbol{w}_2+b_2, \tag{27}$$

确定最终的回归函数, 而这对函数可以通过求解下面的一对 QP 问题得到:

$$\min\ \frac{1}{2}\|\boldsymbol{Y}-\boldsymbol{e}\boldsymbol{\varepsilon}_1-(K(\boldsymbol{A},\boldsymbol{A}^{\mathrm{T}})\boldsymbol{w}_1+\boldsymbol{e}b_1)\|^2+C_1\boldsymbol{e}^{\mathrm{T}}\boldsymbol{\xi}$$

$$\text{s.t. } \boldsymbol{Y}-(K(\boldsymbol{A},\boldsymbol{A}^{\mathrm{T}})\boldsymbol{w}_1+\boldsymbol{e}b_1)\geqslant\boldsymbol{e}\boldsymbol{\varepsilon}_1-\boldsymbol{\xi},\boldsymbol{\xi}\geqslant 0, \tag{28}$$

$$\min\ \frac{1}{2}\|\boldsymbol{Y}+\boldsymbol{e}\boldsymbol{\varepsilon}_2-(K(\boldsymbol{A},\boldsymbol{A}^{\mathrm{T}})\boldsymbol{w}_2+\boldsymbol{e}b_2)\|^2+C_2\boldsymbol{e}^{\mathrm{T}}\boldsymbol{\eta}$$

$$\text{s.t. } (K(\boldsymbol{A},\boldsymbol{A}^{\mathrm{T}})\boldsymbol{w}_2+\boldsymbol{e}b_2)-\boldsymbol{Y}\geqslant\boldsymbol{e}\boldsymbol{\varepsilon}_2-\boldsymbol{\eta},\boldsymbol{\eta}\geqslant 0. \tag{29}$$

其中,$C_1,C_2>0,\varepsilon_1,\varepsilon_2>0$ 为常数,ξ,η 为松弛变量,e 为 $l\times 1$ 维的单位列向量.式(28)和式(29)可转化为如下无约束问题:

$$\min \frac{1}{2}\| \boldsymbol{Y}-\boldsymbol{e\varepsilon}_1-(K(\boldsymbol{A},\boldsymbol{A}^{\mathrm{T}})\boldsymbol{w}_1+\boldsymbol{e}b_1)\|^2+C_1\boldsymbol{e}^{\mathrm{T}}\max(0,\boldsymbol{e\varepsilon}_1-\boldsymbol{Y}+K(\boldsymbol{A},\boldsymbol{A}^{\mathrm{T}})\boldsymbol{w}_1+\boldsymbol{e}b_1), \tag{30}$$

$$\min \frac{1}{2}\| \boldsymbol{Y}+\boldsymbol{e\varepsilon}_2-(K(\boldsymbol{A},\boldsymbol{A}^{\mathrm{T}})\boldsymbol{w}_2+\boldsymbol{e}b_2)\|^2+C_2\boldsymbol{e}^{\mathrm{T}}\max(0,\boldsymbol{e\varepsilon}_2+\boldsymbol{Y}-(K(\boldsymbol{A},\boldsymbol{A}^{\mathrm{T}})\boldsymbol{w}_2+\boldsymbol{e}b_2)). \tag{31}$$

引入光滑 CHKS 函数:

$$\phi_{11}(\boldsymbol{w}_1,b_1,\varepsilon)=\frac{h_1+\sqrt{h_1+4\varepsilon^2}}{2},$$

$$\phi_{12}(\boldsymbol{w}_2,b_2,\varepsilon)=\frac{h_2+\sqrt{h_2+4\varepsilon^2}}{2},$$

其中:

$$h_1=\boldsymbol{e\varepsilon}_1-\boldsymbol{Y}+K(\boldsymbol{A},\boldsymbol{A}^{\mathrm{T}})\boldsymbol{w}_1+\boldsymbol{e}b_1,$$

$$h_2=\boldsymbol{e\varepsilon}_2+\boldsymbol{Y}-(K(\boldsymbol{A},\boldsymbol{A}^{\mathrm{T}})\boldsymbol{w}_2+\boldsymbol{e}b_2).$$

可得非线性 SCTSVR 模型:

$$\min \Phi_1(\boldsymbol{w}_1,b_1,\varepsilon)=\frac{1}{2}\| \boldsymbol{Y}-\boldsymbol{e\varepsilon}_1-(K(\boldsymbol{A},\boldsymbol{A}^{\mathrm{T}})\boldsymbol{w}_1+\boldsymbol{e}b_1)\|^2+C_1\boldsymbol{e}^{\mathrm{T}}\phi_{11}(\boldsymbol{w}_1,b_1,\varepsilon), \tag{32}$$

$$\min \Phi_2(\boldsymbol{w}_2,b_2,\varepsilon)=\frac{1}{2}\| \boldsymbol{Y}+\boldsymbol{e\varepsilon}_2-(K(\boldsymbol{A},\boldsymbol{A}^{\mathrm{T}})\boldsymbol{w}_2+\boldsymbol{e}b_2)\|^2+C_2\boldsymbol{e}^{\mathrm{T}}\phi_{12}(\boldsymbol{w}_2,b_2,\varepsilon). \tag{33}$$

定理 1~3 也适用于非线性 SCTSVR,即规划问题(式(32)和式(33))是严凸的,任意阶可微的,因此可用 Newton-Armijo 算法求解该模型.

2 实验与分析

为了测试本文提出的 SCTSVR 算法的性能,我们将对人工数据集和 UCI 数据集进行测试,并和 SVR,TSVR,STSVR 的测试结果进行比较.以上 4 种方法都是在 Intel® Core™2 Duo CPU E4500,2 GB内存和 MATLAB7.11.0 的环境中运行的.和很多其他的机器学习方法一样,这些方法的学习性能对参数的选择非常敏感.在本文中,对于非线性情况,仅仅考虑高斯核函数:

$$K(x,x_i)=\exp\left(-\frac{\| x-x_i\|^2}{2\sigma^2}\right),$$

所有算法参数的选取是从实验数据集中随机抽取 10%作为子集上采用 10 次交叉测试所得结果的最优值.

2.1 人工数据集

$\sin c$ 函数经常被用来测试各种机器学习方法的回归性能,其表达式为

$$y=\sin c(x)=\frac{\sin \pi x}{\pi x},x\in[-4,4].$$

为了更有效地测试本文的算法,在训练样本中,我们分别加入不同的零平均值高斯噪声和均匀分布噪声.下面我们在 $\sin c$ 函数中加入 4 种不同的噪声,加入噪声后的 $\sin c$ 函数分别为

类型 1.

$$y_i=\frac{\sin \pi x_i}{\pi x_i}+\xi_i,x\sim U[-4,4],\quad \xi_i\sim N(0,0.1^2);$$

类型 2.

$$y_i=\frac{\sin \pi x_i}{\pi x_i}+\xi_i,x\sim U[-4,4],\quad \xi_i\sim N(0,0.2^2);$$

类型 3.

$$y_i=\frac{\sin \pi x_i}{\pi x_i}+\xi_i,x\sim U[-4,4],\quad \xi_i\sim U[-0.1,0.1];$$

类型 4.

$$y_i=\frac{\sin \pi x_i}{\pi x_i}+\xi_i,x\sim U[-4,4],\quad \xi_i\sim U[-0.2,0.2].$$

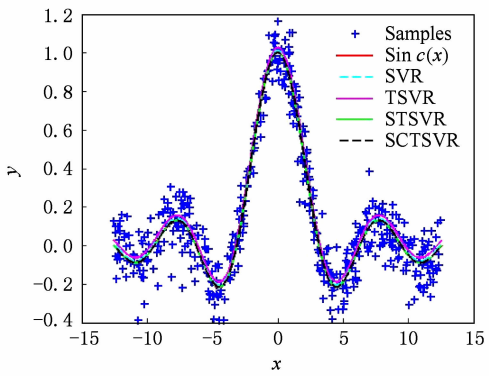
其中,$U[a,b]$表示的是在$[a,b]$上的均匀随机变量,$N(c,d^2)$表示的是均值为 c,方差为 d^2 的高斯随机变量.

为了提高比较结果的可靠性,用 Matlab7.11 工具箱对每种噪声分别产生 10 组噪声样本,每组噪声样本包括 500 个训练样本和 500 个测试样本.表 1 是 SVR,TSVR,STSVR,SCTSVR 分别运行 10 次的平均结果.图 1 是 SVR,TSVR,STSVR,SCTSVR 分别对 4 种带不同噪声的 sinc 函数运行一次的结果.从表 1 可以看出,SCTSVR 和 STSVR 能得到比较小的 SSE,SSE/SST 和较大的 SSR/SST,这说明 SCTSVR 的逼近精度比 STSVR 的逼近精度好.此外,SCTSVR 的 CUP 时间少于 STSVR 的时间.这些实验结果表明,和 STSVR 相比,SCTSVR 具有更好的回归效果.从图 1 可以看出,SCTSVR 的拟合能力比较好.

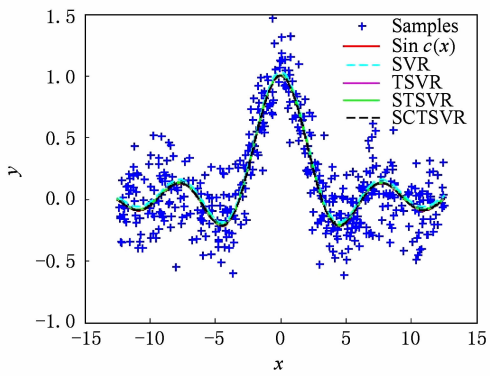
Table 1 Comparison of Four Algorithms on sin c Function with Different Types of Noises

表 1 4 种算法在带不同类型噪声 sin c 函数下的比较结果

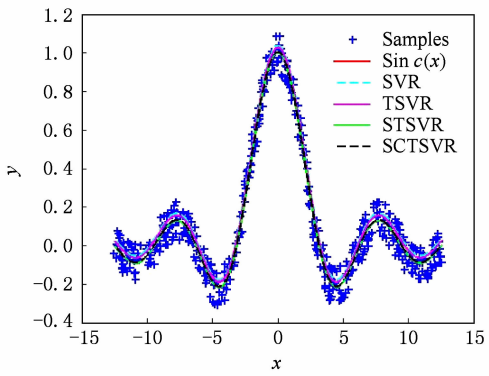
Function	Algorithm	SSE	SSE/SST	SSR/SST	CPU Running Time/s
Type 1	SVR	0.3327±0.1524	0.0061±0.0031	0.9528±0.0321	24.561
	TSVR	0.2554±0.0347	0.0055±0.0024	0.9715±0.0347	6.824
	STSVR	0.2551±0.1121	0.0053±0.0022	0.9731±0.1223	0.036
	SCTSVR	**0.2412±0.1105**	**0.0048±0.0011**	**0.9734±0.0068**	**0.021**
Type 2	SVR	1.1952±0.3462	0.0221±0.0038	0.9428±0.0221	35.678
	TSVR	0.9637±0.1473	0.0175±0.0034	0.9484±0.0352	9.031
	STSVR	0.9483±0.5042	0.0176±0.0052	0.9513±0.0133	0.0688
	SCTSVR	**0.8551±0.0332**	**0.0153±0.0011**	**0.9621±0.0431**	**0.0326**
Type 3	SVR	0.0856±0.0474	0.0016±0.0003	0.9989±0.0154	18.35
	TSVR	0.0693±0.0143	0.0013±0.0223	1.0023±0.0062	4.026
	STSVR	0.0743±0.0574	0.0012±0.0115	0.9878±0.0228	0.396
	SCTSVR	**0.0724±0.0216**	**0.0009±0.1503**	**1.0013±0.0036**	**0.175**
Type 4	SVR	0.5848±0.1356	0.0157±0.0025	0.9438±0.0123	19.24
	TSVR	0.5245±0.1361	0.0098±0.0012	0.9526±0.0512	2.741
	STSVR	0.5104±0.0467	0.0098±0.0041	0.9528±0.0345	0.189
	SCTSVR	**0.4921±0.0097**	**0.0096±0.1254**	**0.9727±0.0063**	**0.169**



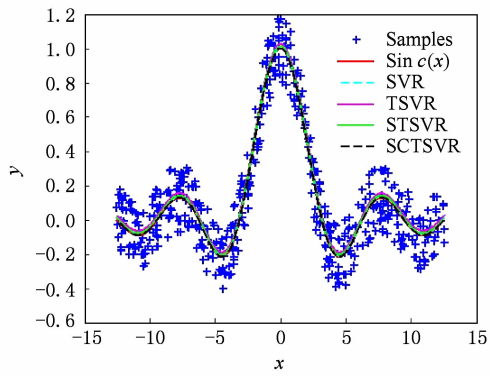
(a) Fitting predictions of four algorithms on sin c function with noises of type 1



(b) Fitting predictions of four algorithms on sin c function with noises of type 2



(c) Fitting predictions of four algorithms on sin c function with noises of type 3



(d) Fitting predictions of four algorithms on sin c function with noises of type 4

Fig. 1 Fitting predictions of four algorithms on sin c function with four types of noises.

图 1 4 种算法对 4 种带不同噪声的 sin c 函数的拟合结果

2.2 UCI 数据集

为了进一步检验本文算法的性能,我们将对文献[15]中的 10 个 UCI 数据集进行测试. 这 10 个 UCI 数据集分别是 Diabetes, Boston housing, Auto-

Mpg, Machine CPU, Servo, Concrete Cs, Auto price, Bodyfat, Triazines, Chwirut1. 这些数据集经常被用来测试机器学习算法的性能.

表 2 所示为 SVR, TSVR, STSVR, SCTSVR 分别对 10 个数据集进行 10 次测试的平均结果. 从表 2 可以看出,对于大部分的数据集,SCTSVR 获得的 SSE, SSE/SST 和 SSR/SST 比 STSVR 好. 这说明和 STSVR 相比,SCTSVR 的泛化能力有了一定的

Table 2 Comparison of Four Algorithms on UCI Dataset

表 2 4 种算法在 UCI 数据集上的比较结果

Dataset	Algorithm	SSE	SSE/SST	SSR/SST	CPU Running Time/s
Diabetes (43×3)	SVR	0.4507±0.0521	0.5192±0.5673	0.6014±0.0174	3.474
	TSVR	0.3688±0.1983	0.4165±0.5801	0.6108±0.0515	0.982
	STSVR	0.3512±0.1562	0.4129±0.3901	0.6294±0.1274	0.215
	SCTSVR	**0.3469±0.0024**	**0.4023±0.1019**	**0.6325±0.0246**	**0.112**
Boston Housing (506×14)	SVR	0.4056±0.0174	0.1287±0.0352	0.9056±0.1547	69.842
	TSVR	0.3974±0.2547	0.1235±0.0362	1.0257±0.1025	20.712
	STSVR	0.3722±0.1426	0.1208±0.0644	0.9928±0.1246	0.295 2
	SCTSVR	**0.3609±0.4526**	**0.1203±0.0037**	**1.0147±0.0356**	**0.099 5**
Auto-Mpg (392×8)	SVR	0.1247±0.1578	0.1141±0.0314	0.9875±0.0145	29.241
	TSVR	0.0867±0.0234	0.1125±0.0423	1.0108±0.0012	7.254
	STSVR	0.0843±0.2442	0.1063±0.0527	0.9964±0.0574	0.045
	SCTSVR	**0.0725±0.0169**	**0.1053±0.0402**	**1.0251±0.0213**	**0.032**
Machine CPU (209×9)	SVR	0.1024±0.1742	0.1049±0.0741	0.9678±0.0052	16.522
	TSVR	0.0854±0.5214	0.1112±0.0741	0.9801±0.0024	3.987
	STSVR	0.0835±0.5614	0.1019±0.0126	0.9860±0.0044	0.945
	SCTSVR	**0.0787±0.1453**	**0.1017±0.0753**	**0.9863±0.0046**	**0.506**
Servo (167×4)	SVR	0.2542±0.0547	0.1415±0.0089	0.9542±0.0085	12.549
	TSVR	0.2248±0.0145	0.1284±0.0021	0.9748±0.0012	3.821
	STSVR	0.2184±0.1201	0.1242±0.0078	0.9816±0.0046	0.698
	SCTSVR	**0.1874±0.0145**	**0.1057±0.0027**	**0.9945±0.0011**	**0.384**
Concrete Cs (1030×9)	SVR	0.1985±0.1457	0.1245±0.0052	0.9452±0.2454	215.254
	TSVR	0.1720±0.1701	0.1204±0.0041	0.9541±0.2103	51.362
	STSVR	0.1704±0.0124	0.1121±0.0078	0.9562±0.2452	11.252
	SCTSVR	**0.1517±0.0124**	**0.0898±0.0052**	**0.9702±0.1024**	**8.163**
Auto price (159×16)	SVR	0.2104±0.4556	0.1289±0.0056	0.9894±0.0145	45.26
	TSVR	0.1824±0.2014	0.1224±0.0004	1.0024±0.0412	10.85
	STSVR	0.1515±0.4178	0.1282±0.0034	0.9970±0.1212	4.287
	SCTSVR	**0.1387±0.4781**	**0.0940±0.0012**	**1.0007±0.0451**	**2.185**
Bodyfat (252×15)	SVR	0.0147±0.0142	0.0462±0.0451	0.9634±0.0854	36.242
	TSVR	0.0098±0.0412	0.0451±0.0322	0.9985±0.0542	8.923
	STSVR	0.0094±0.0589	0.0448±0.0742	0.9952±0.0784	1.055
	SCTSVR	**0.0089±0.0128**	**0.0426±0.0363**	**0.9978±0.0821**	**0.988**
Triazines (186×61)	SVR	0.2471±0.1472	0.3601±0.0522	0.9421±0.0854	8.524
	TSVR	0.2081±0.1041	0.3245±0.0152	0.9685±0.0717	2.129
	STSVR	0.2147±0.1240	0.3168±0.0451	0.9626±0.0742	0.974
	SCTSVR	**0.2015±0.1204**	**0.3018±0.0748**	**0.9701±0.0745**	**0.472**
Chwirut1 (214×2)	SVR	0.0145±0.0741	0.0784±0.0521	0.9821±0.0142	22.725
	TSVR	0.0138±0.0458	0.0727±0.0544	0.9954±0.0452	5.582
	STSVR	0.0042±0.0521	0.0641±0.0747	0.9945±0.0121	3.101
	SCTSVR	**0.0035±0.0146**	**0.0405±0.0575**	**0.9994±0.0524**	**1.012**

提高. 并且,SCTSVR 的 CPU 运行时间比 STSVR 的要少一些,并远远少于 TSVR 的 CPU 运行时间,特别是对于样本数目较大的数据集来说这种优势更明显.

3 结束语

本文通过采用 CHKS 函数逼近 TSVR 的不可微项,并用 Newton-Armijo 算法求解相应的模型,提出了一种新的光滑孪生支持向量回归机——光滑 CHKS 孪生支持向量回归机(SCTSVR). 我们不仅从理论上证明了 SCTSVR 具有严凸性,能满足任意阶光滑的性能,而且在人工数据集和 UCI 数据集上的实验表明了 SCTSVR 比 STSVR 具有更好的回归性能. 鉴于 SCTSVR 的良好性能,把 SCTSVR 应用到其他领域是我们下一步的工作.

参 考 文 献

[1] Hua Xiaopeng, Ding Shifei. Locality preserving twin support vector machines [J]. Journal of Computer Research and Development, 2014, 51(3): 590-597 (in Chinese)
(花小朋, 丁世飞. 局部保持对支持向量机[J]. 计算机研究与发展, 2014, 51(3): 590-597)

[2] Ding Shifei, Huang Huajuan, Shi Zhongzhi. Weighted smooth CHKS twin support vector machines [J]. Journal of Software, 2013, 24(11): 2548-2557 (in Chinese)
(丁世飞, 黄华娟, 史忠植. 加权光滑 CHKS 孪生支持向量机[J]. 软件学报, 2013, 24(11): 2548-2557)

[3] Ding Shifei, Hua Xiaopeng. Recursive least squares projection twin support vector machines [J]. Neurocomputing, 2014, 130: 3-9

[4] Ding Shifei, Jia Hongjie, Chen Jinrong, et al. Granular neural networks [J]. Artificial Intelligence Review, 2014, 41(3): 373-384

[5] Cortes C, Vapnik V N. Support vector networks [J]. Machine Learning, 1995, 20: 273-297

[6] Platt J C. Using analytic QP and sparseness to speed training of support vector machines [C] //Advances in Neural Information Processing Systems 11. Cambridge, MA: MIT Press, 1999: 557-563

[7] Fung G, Mangasarian O L. Proximal support vector machine classifiers [C] //Proc of 2001 ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining. New York: ACM, 2001: 77-86

[8] Mangasarian O L, Wild W. Multi-surface proximal support vector machine classification via generalized eigenvalues [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2006, 28(1): 69-74

[9] Jayadeva, Khemchandni R, Chandra S. Twin support vector machines for pattern classification [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2007, 29(5): 905-910

[10] Cong Hanhan, Yang Chengfu, Pu Xiaorong. Efficient speaker recognition based on multi-class twin support vector machines and GMMs [C] //Proc of 2008 IEEE Conf on Robotics, Automation and Mechatronics. Piscataway, NJ: IEEE, 2008: 348-352

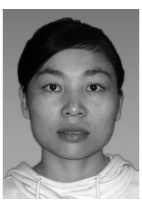
[11] Zhang Xinsheng, Gao Xinbo, Wang Ying. Twin support vector machine for MCs detection [J]. Journal of Electronics, 2009, 26(3): 318-325

[12] Peng Xinjun. TSVR: An efficient twin support vector machine for regression [J]. Neural Networks, 2010, 23: 365-372

[13] Chen Xiaobo, Yang Jian, Liang Jun. Smooth twin support vector regression [J]. Neural Computing & Applications, 2012, 21(3): 505-513

[14] Wu Qing. A novel smooth support vector regression based on CHKS function [J]. Applied Mechanics and Materials, 2011, 44: 3746-3751

[15] Zhong Ping, Xu Yitian, Zhao Yaohong. Training twin support vector regression via linear programming [J]. Neural Computing & Applications, 2012, 21(2): 399-407



Huang Huajuan, born in 1984. PhD candidate in applied computer technology from China University of Mining and Technology. Her current research interests include machine learning, data mining and pattern recognition.

Ding Shifei, born in 1963. Professor and PhD supervisor at China University of Mining and Technology. Senior member of China Computer Federation. His main research interests include pattern recognition, data mining, intelligent information processing, machine learning and granular computing, etc.

Shi Zhongzhi, born in 1941. Professor and PhD supervisor at the Institute of Computing Technology, Chinese Academy of Sciences. Fellow member of China Computer Federation. His main research interests include machine learning, data mining, pattern recognition and intelligence science, etc.