文章编号: 1001-0920 (2013) 12-1817-05

基于特征值分解的最大间隔支持向量回归机

陈伟杰, 邵元海, 叶娅芬

(浙江工业大学 之江学院，杭州 310024)

摘 要: 广义特征值中心支持向量回归机 (GEPSVR) 是一种有效的核回归算法, 但其在求解优化问题时易导致奇异性问题. 为此, 提出一种基于特征值分解的支持向量回归机, 简称 IGEPSVR. 与 GEPSVR 相比, IGEPSVR 的主要优势有: 结合最大间隔准则和 GEPSVR 几何思想给出了新的距离度量准则; 在优化模型中引入 Tikhonov 正则项, 克服了可能产生的奇异性问题; IGEPSVR 仅需求解两个标准特征值, 降低了计算复杂度. 实验结果表明, 较 GEPSVR 算法, IGEPSVR 不仅提高了学习能力, 而且缩短了训练时间.

关键词: 支持向量回归机；广义特征值中心支持向量机；非平行不敏感函数；特征值分解

中图分类号: TP391 **文献标志码:** A

Maximum margin eigenvalue proximal support vector regressor

CHEN Wei-jie, SHAO Yuan-hai, YE Ya-fen

(Zhijiang College，Zhejiang University of Technology，Hangzhou 310024，China．Correspondent：CHEN Wei-jie, E-mail：wjcper2008@126.com)

Abstract: The generalized eigenvalue proximal support vector regressor(GEPSVR) is an effective kernel-based regression algorithm. However, the generalized eigenvalue problems may be ill-conditioned in the GEPSVR. Therefore, a maximum margin eigenvalue proximal support vector regressor(IGEPSVR) is proposed. The main advantages are as following by defining the distances between the insensitive functions and data points, a novel optimization model is proposed according to the maximum margin criterion and GEPSVR; the possible ill-conditioned problem is overcome by introducing the meaningful Tikhonov regularization terms; the generalized eigenvalue decomposition is replaced by the standard eigenvalue decomposition, leading to simpler optimization problems. Experimental results on a series of datasets show that IGEPSVR is superior to GEPSVR in both generalization and training speed.

Key words: support vector regression；generalized eigenvalue proximal support vector regressor；nonparallel insensitive functions；eigenvalue decomposition

0 引 言

支持向量回归机 (SVR) 自Vapnik[1]提出以来, 便成为机器学习领域备受关注的核方法之一. 与其他学习方法相比, SVR 具有良好的泛化能力, 在社会各个领域都得到了广泛应用[2], 其中以 ε-SVR 模型最为典型. ε-SVR 的核心思想[1]是: 通过求解一个二次规划问题 (QPP)来构造一个回归函数 $f(x)$; 极小化两个平行边界不敏感函数 $f(x)-\varepsilon$ 与 $f(x)+\varepsilon$ 所形成的 ε-不敏感损失带区域的带宽, 使得训练样本尽可能落在该区域内; 同时, 引入正则项实现结构风险极小化原则, 使得 $f(x)$ 尽可能平滑.

为了降低 ε-SVR 的计算开销, 受双子支持向量机 (TWSVM)[3-6]的启发, Peng[7]提出了双子支持向量回归机 (TSVR). TSVR 通过分别构造两个非平行 ε-不敏感边界函数, 将回归问题转换为两个小规模的 SVM 型 QPPs, 将时间开销缩减为 ε-SVR 的 1/4. 随后, Shao 等[8]提出了ε-TSVR, 通过引入 Tikhonov 正则项, 克服了 TSVR 的 Hessian 矩阵可能存在的奇异性问题. 此外, 基于广义特征值中心支持向量机 (GEPSVM)[9]思想, Khemchandani 等[10]通过求解两个广义特征值问题构造一对最优非平行ε-不敏感边界函数, 并提出了 GEPSVR 算法. 文献 [10] 的实验结果表明, GEPSVR 在一系列数据集上具有良好的学习性能, 但与 GEPSVM 类似, GEPSVR 在求解广义特征值问题时

收稿日期: 2012-08-30；**修回日期:** 2013-03-19.

基金项目: 国家自然科学基金项目(11201426, 11071252, 61203133)；浙江省自然科学基金项目(LQ12A01020, LQ13F030010)；浙江省教育厅科研基金项目(Y201225179, Y201225256).

作者简介: 陈伟杰(1985−), 男, 讲师, 博士, 从事智能计算、数据挖掘、生物信息学的研究；邵元海(1983−), 男, 讲师, 博士, 从事优化理论、模式识别、人工智能的研究.

易导致奇异性问题, 从而影响了算法的泛化能力.

为解决上述GEPSVR存在的问题, 本文结合最大间隔准则(MMC)[11], 提出了一个基于特征值分解的最大间隔支持向量回归机, 简称为IGEPSVR. 较GEPSVR, 该方法的主要优势有: 1) 根据MMC准则, 定义了ε-不敏感边界函数与训练样本集之间的距离关系, 并依据GEPSVR的几何思想, 给出了IGEPSVR优化模型; 2) 在优化模型中引入Tikhonov正则项, 克服了可能产生的奇异性问题; 3) 将最优ε-不敏感边界函数的求解问题转换为标准特征值问题, 提高了算法的学习效率, 并从理论上保障了优化问题的最优解是特征值问题最小特征值所对应的特征向量. 最后通过实验验证了IGEPSVR算法的有效性.

1 广义特征值中心支持向量机(GEPSVR)

考虑n维实空间R^n下的回归问题, 记训练样本集$T=(A,Y)=\{(A_i,y_i),i=1,2,\cdots,m\}$. 其中: m为样本个数, 行向量$A_i\in R^n$为第i个训练样本, $y_i\in R$为样本点A_i所对应的回归值.

GEPSVR[10]核心思想是针对训练样本集T, 通过寻找如下一对边界不敏感函数:

$$f_1(x)=w_1^{\mathrm T}x+b_1=0,$$
$$f_2(x)=w_2^{\mathrm T}x+b_2=0, \tag{1}$$

来构造一个ε-不敏感损失带区域, 使得样本点尽量落在该区域内. 其中: $f_1(x)$和$f_2(x)$分别为下界和上界非平行不敏感函数; 权向量$w_1,w_2\in R^n$; 偏移量$b_1,b_2\in R$. GEPSVR通过求解如下优化问题来确定$f_1(x)$和$f_2(x)$:

$$\min_{(w_1,b_1)\neq 0}\frac{\|Aw_1+eb_1-(Y-e\varepsilon)\|^2+\delta\left\|\begin{bmatrix}w_1\\ b_1\\ -1\end{bmatrix}\right\|^2}{\|Aw_1+eb_1-(Y+e\varepsilon)\|^2}, \tag{2}$$

$$\min_{(w_2,b_2)\neq 0}\frac{\|Aw_2+eb_2-(Y+e\varepsilon)\|^2+\delta\left\|\begin{bmatrix}w_2\\ b_2\\ -1\end{bmatrix}\right\|^2}{\|Aw_2+eb_2-(Y-e\varepsilon)\|^2}. \tag{3}$$

其中: $\|\cdot\|$表示L_2范数, $\varepsilon>0$为损失带区域宽度, $\delta>0$为正则化因子, e为分量全为1的列向量. 为求解式(2)的最优解(w_1,b_1), 首先将其转化为Rayleigh商[12]形式, 即

$$\min_{z_1\neq 0}\frac{z_1^{\mathrm T}R_1z_1}{z_1^{\mathrm T}S_1z_1}. \tag{4}$$

其中: $z_1\in R^{(n+2)\times 1}$; R_1和S_1为$(n+2)\times(n+2)$阶对称矩阵. 定义如下:

$$z_1=[w_1^{\mathrm T}\quad b_1\quad -1]^{\mathrm T}, \tag{5a}$$
$$R_1=[A\quad e\quad (Y-e\varepsilon)]^{\mathrm T}[A\quad e\quad (Y-e\varepsilon)]+\delta I, \tag{5b}$$
$$S_1=[A\quad e\quad (Y+e\varepsilon)]^{\mathrm T}[A\quad e\quad (Y+e\varepsilon)]. \tag{5c}$$

然后利用Rayleigh商的性质[12], 进一步将优化问题(4)转化为求解如下广义特征值问题:

$$R_1z_1=S_1z_1,\ z_1\neq 0. \tag{6}$$

因此, 最初始的GEPSVR优化问题(2)的最优解(w_1, b_1)可由广义特征值问题(6)的最小特征值$\eta_{\min}$所对应的特征向量$z_{\min}$获得. 由于式(5a)所定义的向量z_1的第$(n+2)$个分量必须为-1, 在求得$z_{\min}$后, 需使用$-z_{\min}^{(n+2)}$分量去标准化$z_{\min}$, 即

$$z_1=z_{\min}/-z_{\min}^{(n+2)}. \tag{7}$$

得到z_1后, 可由w_1和b_1构建下界不敏感回归函数$f_1(x)=w_1^{\mathrm T}x+b_1$. 同理可求得上界不敏感回归函数$f_2(x)$. 最后, 取$f_1(x)$和$f_2(x)$的均值作为目标回归函数$f(x)$.

2 基于最大间隔准则的IGEPSVR

2.1 IGEPSVR的优化准则

类似于GEPSVR, IGEPSVR的核心思想是: 通过构造一对不敏感回归函数$f_1(x)$和$f_2(x)$, 使得下界函数$f_1(x)$尽可能地靠近$(Y-e\varepsilon)$且尽可能地远离$(Y+e\varepsilon)$, 上界函数$f_2(x)$尽可能地靠近$(Y+e\varepsilon)$且尽可能地远离$(Y-e\varepsilon)$. 为此, 对距离作如下定义:

1) $f_1(x)$到$(Y-e\varepsilon)$之间的距离

$$D_{(f_1,-\varepsilon)}=\frac{\|Aw_1+eb_1-(Y-e\varepsilon)\|^2}{\|w_1\|^2+b_1^2}; \tag{8a}$$

2) $f_1(x)$到$(Y+e\varepsilon)$之间的距离

$$D_{(f_1,+\varepsilon)}=\frac{\|Aw_1+eb_1-(Y+e\varepsilon)\|^2}{\|w_1\|^2+b_1^2}; \tag{8b}$$

3) $f_2(x)$到$(Y-e\varepsilon)$之间的距离

$$D_{(f_2,-\varepsilon)}=\frac{\|Aw_2+eb_2-(Y-e\varepsilon)\|^2}{\|w_2\|^2+b_2^2}; \tag{8c}$$

4) $f_2(x)$到$(Y+e\varepsilon)$之间的距离

$$D_{(f_2,+\varepsilon)}=\frac{\|Aw_2+eb_2-(Y+e\varepsilon)\|^2}{\|w_2\|^2+b_2^2}. \tag{8d}$$

为获得最优边界函数$f_1(x)$和$f_2(x)$, 依据MMC准则[11]得到如下优化问题:

$$\min_{(w_1,b_1)\neq 0}D_{(f_1,-\varepsilon)}-\nu D_{(f_1,+\varepsilon)}, \tag{9}$$

$$\min_{(w_2,b_2)\neq 0}D_{(f_2,+\varepsilon)}-\nu D_{(f_2,-\varepsilon)}. \tag{10}$$

其中参数$\nu>0$为权值因子, 用于权衡目标函数中两项间的权重.

2.2 线性IGEPSVR模型

下面给出线性IGEPSVR模型及其求解方法. 定义

$$M=[A\quad e\quad (Y-e\varepsilon)]^{\mathrm{T}}[A\quad e\quad (Y-e\varepsilon)],$$

$$H=[A\quad e\quad (Y+e\varepsilon)]^{\mathrm{T}}[A\quad e\quad (Y+e\varepsilon)],$$

且

$$z_1=\begin{bmatrix} w_1\\ b_1\\ -1\end{bmatrix},\ z_2=\begin{bmatrix} w_2\\ b_2\\ -1\end{bmatrix}.$$

同时, 令 $\|z_1\|^2=\alpha_1, \|z_2\|^2=\alpha_2$, 则可将优化问题 (9) 和 (10) 转换为

$$\begin{aligned}&\min_{z_1,\alpha_1}\frac{1}{\alpha_1}z_1^{\mathrm{T}}Mz_1-\nu\frac{1}{\alpha_1}z_1^{\mathrm{T}}Hz_1,\\ &\text{s.t. } \|z_1\|^2=\alpha_1,\ \alpha_1>0;\end{aligned}\tag{11}$$

$$\begin{aligned}&\min_{z_2,\alpha_2}\frac{1}{\alpha_2}z_2^{\mathrm{T}}Hz_2-\nu\frac{1}{\alpha_2}z_2^{\mathrm{T}}Mz_2,\\ &\text{s.t. } \|z_2\|^2=\alpha_2,\ \alpha_2>0.\end{aligned}\tag{12}$$

与 ε-TSVR[8] 类似, 在原始问题中引入 Tikhonov 正则项[13] $\delta\|z_1\|^2$ 和 $\delta\|z_2\|^2$, 可得到如下的优化问题:

$$\begin{aligned}&\min_{z_1,\alpha_1} f_{(1,\mathrm{obj})}(z_1)=\delta\|z_1\|^2+z_1^{\mathrm{T}}Mz_1-\nu z_1^{\mathrm{T}}Hz_1,\\ &\text{s.t. } \|z_1\|^2=\alpha_1,\ \alpha_1>0;\end{aligned}\tag{13}$$

$$\begin{aligned}&\min_{z_2,\alpha_2} f_{(2,\mathrm{obj})}(z_2)=\delta\|z_2\|^2+z_2^{\mathrm{T}}Hz_2-\nu z_2^{\mathrm{T}}Mz_2,\\ &\text{s.t. } \|z_2\|^2=\alpha_2,\ \alpha_2>0.\end{aligned}\tag{14}$$

其中参数 $\delta>0$ 为调节因子.

下面给出用于构造下边界函数 $f_1(x)$ 的优化问题 (13) 的几何解释. 目标函数 (13) 的第 1 项 $\|z_1\|^2$ 控制 $f_1(x)$ 模型的复杂程度, 同时极小化该项意味着 $f_1(x)$ 越光滑越好; 第 2 项则是优化 $f_1(x_i)=w_1^{\mathrm{T}}x_i+b_1$ 与 $(y_i-\varepsilon)$ 之间的距离平方和, 使其越小越好; 极小化最后一项表明 $f_1(x_i)=w_1^{\mathrm{T}}x_i+b_1$ 与 $(y_i+\varepsilon)$ 之间的距离平方和越大越好. 式 (14) 的几何意义与式 (13) 类似.

引入 Lagrange 乘子 λ_1 和 λ_2, 对式 (13) 构造如下 Lagrange 函数:

$$\begin{aligned}L(z_1,\alpha_1)=\,&z_1^{\mathrm{T}}Mz_1+\delta\|z_1\|^2-\nu z_1^{\mathrm{T}}Hz_1-\\ &\lambda_1(\|z_1\|^2-\alpha_1)-\lambda_2\alpha_1.\end{aligned}\tag{15}$$

令 $\nabla_{z_1}L=0$, 并整理, 得

$$2(M+\delta I)z_1-2\nu Hz_1-2\lambda_1z_1=0,\tag{16}$$

其中 I 为 $(n+2)\times(n+2)$ 阶单位阵.

显然, 式 (16) 是个 Rayleigh 问题[12], 其最优解可通过求解如下的标准特征值问题得到:

$$((M+\delta I)-\nu H)z_1=\lambda_1z_1.\tag{17}$$

类似地, 可将优化问题 (14) 转换为如下的特征值问题:

$$((H+\delta I)-\nu M)z_2=\lambda_2z_2.\tag{18}$$

由上述分析, 易得到以下结论.

定理 1　最优化问题 (13) 和 (14) 的最优解 (w_1, b_1) 和 (w_2,b_2) 分别为矩阵 $(M+\delta I)-\nu H$ 和 $(H+\delta I)-\nu M$ 的最小特征值所对应的特征向量.

证明　对于优化问题 (13), 将 $\|z_1\|^2=z_1^{\mathrm{T}}Iz_1$ 代入其目标函数 $f_{(1,\mathrm{obj})}(z_1)$, 得

$$\begin{aligned}f_{(1,\mathrm{obj})}(z_1)=\,&\delta\|z_1\|^2+z_1^{\mathrm{T}}Mz_1-\nu z_1^{\mathrm{T}}Hz_1=\\ &\delta z_1^{\mathrm{T}}Iz_1+z_1^{\mathrm{T}}Mz_1-\nu z_1^{\mathrm{T}}Hz_1=\\ &z_1^{\mathrm{T}}((M+\delta I)-\nu H)z_1.\end{aligned}\tag{19}$$

设矩阵 $(M+\delta I)-\nu H$ 的特征值为 $\lambda_1\leqslant\lambda_2\leqslant\cdots\leqslant\lambda_{(n+2)}$, 其对应的标准正交特征向量为 $p_1,p_2,\cdots,p_{(n+2)}$. 由文献 [12] 可知, 式 (13) 的任意可行解为其特征向量的线性组合, 即 $z_1=\sum_{i=1}^{n+2}\beta_ip_i$, 其中 $\beta=(\beta_1,\cdots,\beta_{(n+2)})\geqslant 0$. 因此, 结合式 (19), 问题 (13) 可写为

$$\begin{aligned}&\min_{\beta} f_{(1,\mathrm{obj})}(\beta)=\sum_{i=1}^{n+2}\beta_i^2\lambda_i,\\ &\text{s.t. } \sum_{i=1}^{n+2}\beta_i^2=\alpha_1.\end{aligned}\tag{20}$$

显然, $\beta^*=(\sqrt{\alpha_1},0,\cdots,0)$ 满足 $\sum_{i=1}^{n+2}\beta_i^{*2}=\alpha_1$. 因此, β^* 是式 (20) 的一个可行解.

下面针对式 (20), 证明对于任意其他可行解 β 都有 $f_{(1,\mathrm{obj})}(\beta)\geqslant f_{(1,\mathrm{obj})}(\beta^*)$. 如果 β 是可行解, 则有

$$\begin{aligned}f_{(1,\mathrm{obj})}(\beta)=\sum_{i=1}^{n+2}\beta_i^2\lambda_i\geqslant\sum_{i=1}^{n+2}\beta_i^2\lambda_1=\\ \alpha_1\lambda_1=\sum_{i=1}^{n+2}\beta_i^{*2}\lambda_i=f_{(1,\mathrm{obj})}(\beta^*).\end{aligned}\tag{21}$$

由 Rayleigh 定理[12] 可知, 当且仅当 λ 是式 (19) 的最小特征时, 优化问题 (13) 的目标函数 $f_{(1,\mathrm{obj})}(z_1)$ 取得全局极小值. 此时, 最小特征值 λ_1 对应的特征向量 $z_1=\sum_{i=1}^{n+2}\beta_i^*p_i=\sqrt{\alpha_1}p_1$ 是 $f_1(x)$ 的最优解. 同理可得, 式 (14) 的最优解 (w_2,b_2) 为矩阵 $(H+\delta I)-\nu M$ 的最小特征值所对应的特征向量. □

2.3　非线性 IGEPSVR 模型

为了将 IGEPSVR 推广到非线性情况, 本文应用非线性核映射 $\varphi(\cdot)$ 将原空间中的样本集 A 映射到高维特征空间 $\mathcal{H}$ 中. 依据 Mercer 定理[14], 可使用核函数 $K(x,y)=\varphi(x)\varphi(y)$ (如 RBF 核函数 $K(x,x')=\mathrm{e}^{-\mu\|x-x'\|^2}$) 代替样本点在特征空间 $\mathcal{H}$ 中的内积运算, 以降低核运算带来的计算复杂度. 在特征空间 $\mathcal{H}$ 中, 考虑如下一对边界不敏感函数:

$$f_1(x)=K(x^{\mathrm{T}},A^{\mathrm{T}})u_1+b_1=0,\tag{22a}$$

$$f_2(x)=K(x^{\mathrm{T}},A^{\mathrm{T}})u_2+b_2=0.\tag{22b}$$

类似于线性情况, 根据最大最小间隔准则, 可将

非线性 IGEPSVR 优化问题描述为

$$\min_{(u_1,b_1)\neq 0} \frac{\|K(A,A^{\mathrm{T}})u_1+eb_1-(Y-e\varepsilon)\|^2}{u_1^{\mathrm{T}}K(A,A^{\mathrm{T}})u_1+b_1^2}-\nu\frac{\|K(A,A^{\mathrm{T}})u_1+eb_1-(Y+e\varepsilon)\|^2}{u_1^{\mathrm{T}}K(A,A^{\mathrm{T}})u_1+b_1^2}, \quad (23)$$

$$\min_{(u_2,b_2)\neq 0} \frac{\|K(A,A^{\mathrm{T}})u_2+eb_2-(Y+e\varepsilon)\|^2}{u_2^{\mathrm{T}}K(A,A^{\mathrm{T}})u_2+b_2^2}-\nu\frac{\|K(A,A^{\mathrm{T}})u_2+eb_2-(Y-e\varepsilon)\|^2}{u_2^{\mathrm{T}}K(A,A^{\mathrm{T}})u_2+b_2^2}, \quad (24)$$

其中 $\nu>0$ 为权值因子. 定义

$$z_1=\begin{bmatrix}u_1\\ b_1\\ -1\end{bmatrix},\ z_2=\begin{bmatrix}u_2\\ b_2\\ -1\end{bmatrix},$$

$$Q=[K(A,A^{\mathrm{T}})\quad e\quad (Y-e\varepsilon)]^{\mathrm{T}}[K(A,A^{\mathrm{T}})\quad e\quad (Y-e\varepsilon)],$$

$$P=[K(A,A^{\mathrm{T}})\quad e\quad (Y+e\varepsilon)]^{\mathrm{T}}[K(A,A^{\mathrm{T}})\quad e\quad (Y+e\varepsilon)].$$

同时, 引入 Tikhonov[13] 正则项, 则式 (23) 和 (24) 可转换为

$$\min_{z_1,\alpha_1}\delta\|z_1\|^2+z_1^{\mathrm{T}}Qz_1-\nu z_1^{\mathrm{T}}Pz_1,$$
$$\text{s.t. } \|z_1\|^2=\alpha_1,\ \alpha_1>0; \quad (25)$$

$$\min_{z_2,\alpha_2}\delta\|z_2\|^2+z_2^{\mathrm{T}}Pz_2-\nu z_2^{\mathrm{T}}Qz_2,$$
$$\text{s.t. } \|z_2\|^2=\alpha_2,\ \alpha_2>0. \quad (26)$$

非线性 IGEPSVR 在形式上与线性模型相似, 不同的是非线性模型将线性模型中的 A 替换为 $K(A, A^{\mathrm{T}})$.

类似于线性情况, 关于式 (25) 和 (26) 的最优解, 有如下定理.

定理 2 最优化问题 (25) 和 (26) 的最优解 (u_1, b_1) 和 (u_2,b_2) 分别为矩阵 $(Q+\delta I)-\nu P$ 和 $(P+\delta I)-\nu Q$ 的最小特征值所对应的特征向量.

定理 2 的证明类似于定理 1 的证明过程, 此略.

2.4 计算复杂度分析

综上可知, 线性 IGEPSVR 仅需求解两个标准的特征值问题, 其计算复杂度为 $\mathcal{O}(2n^2)$, 其中 n 为样本点的特征数, 而线性 GEPSVR 则需要求解两个广义特征值问题, 其计算复杂度为 $\mathcal{O}(2n^3)$. 对于非线性情况, IGEPSVR 的计算复杂度为 $\mathcal{O}(2m^2)$, GEPSVR 则为 $\mathcal{O}(2m^3)$, 其中 m 为样本的规模. 因此, 由理论分析可知, IGEPSVR 训练速度快比 GEPSVR 快一个数量级.

3 实验分析

为验证 IGEPSVR 的有效性, 在 7 个 UCI 回归数

表 1 4 种回归算法在 UCI 数据集上的性能对比

数据集	算法	NMSE	R^2	MAPE	时间/s	参数
Servo (167×4)	ε-SVR	0.252 8±0.155 2	0.929 3±0.107 9	0.361 2±0.154 3	0.032 8	$(2^{-3},0.3,2^{5})$
	TSVR	0.246 8±0.102 6	0.932 8±0.142 7	**0.344 5±0.130 5**	0.051 8	$(2^{-1},0.2,2^{-3})$
	GEPSVR	0.2766±0.1382	0.9114±0.1768	0.3734±0.1439	0.0842	$(2^{2},0.7,2^{-6})$
	IGEPSVR	**0.233 7±0.096 5**	**0.948 5±0.125 5**	0.352 4±0.122 1	**0.029 8**	$(2^{-4},2^{-1},0.4,2^{3})$
Auto Price (159×15)	ε-SVR	0.273 6±0.086 8	0.772 2±0.217 1	0.219 6±0.112 5	0.021 9	$(2^{2},0.5,2^{4})$
	TSVR	**0.260 8±0.072 3**	0.816 3±0.160 6	0.172 8±0.127 1	0.054 8	$(2^{-5},0.1,2^{0})$
	GEPSVR	0.281 8±0.097 2	0.798 6±0.128 1	0.184 4±0.019 1	0.098 3	$(2^{-4},0.3,2^{-1})$
	IGEPSVR	0.269 4±0.023 4	**0.831 5±0.246 1**	**0.164 8±0.095 6**	**0.020 2**	$(2^{-3},2^{0},0.2,2^{5})$
Machine CPU (209×7)	ε-SVR	0.1616±0.0326	0.9787±0.2579	0.2646±0.0925	**0.0382**	$(2^{2},0.9,2^{-4})$
	TSVR	**0.1498±0.0277**	0.9832±0.0948	0.2453±0.0795	0.1172	$(2^{-3},0.1,2^{2})$
	GEPSVR	0.157 6±0.024 1	0.982 0±0.077 4	0.253 0±0.015 7	0.096 6	$(2^{0},0.5,2^{-3})$
	IGEPSVR	0.151 8±0.019 9	**0.988 1±0.055 3**	**0.240 6±0.081 0**	0.040 7	$(2^{-2},2^{1},0.2,2^{1})$
Wisconsin BC (194×32)	ε-SVR	0.859 1±0.115 6	0.482 9±0.105 6	0.959 3±0.115 3	0.077 4	$(2^{2},0.3,2^{-6})$
	TSVR	0.840 6±0.116 8	**0.478 2±0.112 9**	**0.934 1±0.146 0**	0.087 1	$(2^{-4},0.3,2^{-1})$
	GEPSVR	0.871 5±0.124 5	0.503 7±0.119 4	0.981 6±0.137 4	0.114 5	$(2^{-3},0.6,2^{0})$
	IGEPSVR	**0.832 0±0.114 7**	0.480 2±0.118 5	0.935 2±0.123 3	**0.064 9**	$(2^{1},2^{-3},0.2,2^{3})$
Auto-Mpg (398×7)	ε-SVR	0.107 1±0.036 3	**0.988 2±0.142 9**	0.175 7±0.041 8	0.084 6	$(2^{5},0.3,2^{-2})$
	TSVR	0.112 1±0.030 5	0.975 8±0.166 5	0.184 4±0.035 1	0.094 1	$(2^{-1},0.6,2^{-4})$
	GEPSVR	0.128 4±0.042 5	0.941 9±0.146 9	0.206 1±0.041 2	0.138 4	$(2^{3},0.1,2^{-1})$
	IGEPSVR	**0.097 3±0.037 1**	0.981 6±0.137 0	**0.174 2±0.038**	**0.073 9**	$(2^{2},2^{5},0.2,2^{-3})$
Boston housing (506×13)	ε-SVR	0.124 7±0.061 7	0.914 7±0.132 5	0.572 2±0.025 8	0.951 3	$(2^{5},0.4,2^{-2})$
	TSVR	0.127 5±0.048 5	0.915 3±0.127 2	**0.567 1±0.031 5**	1.316 9	$(2^{-1},0.8,2^{-4})$
	GEPSVR	0.132 8±0.056 9	0.878 1±0.141 2	0.583 1±0.029 1	2.627 2	$(2^{3},0.2,2^{-1})$
	IGEPSVR	**0.121 6±0.046 1**	**0.923 3±0.120 9**	0.569 3±0.028 3	**0.589 4**	$(2^{1},2^{-2},0.5,2^{-3})$
Concrete CS ($1\,030\times 8$)	ε-SVR	0.1126±0.0274	**0.9735±0.0730**	0.3195±0.0574	3.8275	$(2^{4},0.7,2^{-6})$
	TSVR	0.1051±0.0273	0.9818±0.0785	0.3173±0.0528	4.3790	$(2^{-2},0.2,2^{4})$
	GEPSVR	0.117 9±0.032 4	0.962 9±0.083 1	0.325 9±0.067 3	6.582 4	$(2^{2},0.5,2^{-5})$
	IGEPSVR	**0.103 2±0.025 1**	0.984 1±0.073 4	**0.316 4±0.048 6**	**3.573 1**	$(2^{-1},2^{3},0.3,2^{-4})$

据集[7-8]上分别与 ε-SVR[1]、TSVR[7] 和 GEPSVR[10]进行比较. 所有实验均以 Matlab 7.0 为软件环境, 以 Intel P4 (2.9 GHz) 处理器、1 G 内存的 PC 为硬件平台. 在实验中, 使用 LIBSVM 工具包[15]实现 ε-SVR, 使用逐次超松弛技术 (SOR)[8]来求解 TSVR 的 QPP 问题, 使用 Matlab 的"eig"函数求解 GEPSVR 和 IGEPSVR 的特征值问题.

对于非线性情况, 选择如下 RBF 核作为核函数:

$$K(x,x') = \mathrm{e}^{-\gamma\|x-x'\|^2},$$

其中 γ 为核参数. 参数选择对于各算法性能起着至关重要的作用, 因此, 对于不同数据集, 使用 10-折交叉验证方法为算法选择最优参数 (即随机将原始数据集分成 10 等份, 然后轮流选取其中的 9 份作为训练集, 剩余的 1 份作为预测集, 最后将 10 次结果求均值作为最终的结果). 各算法的选参范围是 $\{c,\delta,\nu,\gamma=2^i|i=-6,-5,\cdots,5,6\}$ 和 $\{\varepsilon=j|j=0.1,0.2,\cdots,0.9\}$. 此外, 实验将采用如下几个回归算法性能评价指标: 归一化平均平方误差 NMSE, 拟合优度 R^2, 平均绝对百分误差 MAPE 以及训练时间. 上述指标的详细定义可参见文献 [8].

表 1 给出了 4 个算法对上述 7 个数据集的 10-折交叉验证结果和具体选参, 其中黑体数字代表特定数据集下某个指标获得的最好结果. 由结果可知, 其中有 5 个数据集 IGEPSVR 获得 NMSE 最小, 4 个数据集 R^2 最小, 4 个数据集 MAPE 最小, 这表明了 IGEPSVR 算法的有效性. 此外, 由于 GEPSVR 存在奇异性问题, 算法性能受到较大影响. 另一方面, 相对于 GEPSVR, IGEPSVR 用求解标准特征值代替求解 GEPSVR 的广义特征值问题来获得最优解, 从而降低了训练时间, 提高了算法的学习效率. 例如数据集 Auto Price, 相对于 GEPSVR (训练时间为 0.098 3 s), IGEPSVR (训练时间为 0.020 2 s) 的训练速度提高了 4 倍多. 综上所述, IGEPSVR 具有较好的学习和泛化能力以及较高的学习效率.

4　结　论

针对 GEPSVR 存在求解广义特征值时易导致奇异性的问题, 本文提出了一个基于特征值分解的最大间隔支持向量回归机 (IGEPSVR). 该算法结合最大间隔准则, 引入 Tikhonov 正则项, 将优化问题转化为求解两个标准特征值问题, 并从理论上证明了优化问题的最优解是特征值问题最小特征值所对应的特征向量. 由多个 UCI 回归数据集的实验结果可知, 较 ε-SVR、TSVR 和 GEPSVR 算法, 本文提出的 IGEPSVR 不仅具有较好的泛化能力, 而且具有较快的训练速度.

参考文献(References)

[1] Vapnik V. Statistical learning theory[M]. New York: Wiley Press, 1998: 401-421.

[2] 王晓明, 王士同. 最小方差支撑向量回归[J]. 控制与决策, 2010, 25(4): 556-561.
(Wang X M, Wang S T. Minimum variance support vector regression[J]. Control and Decision, 2010, 25(4): 556-561.)

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

[4] Shao Y, Zhang C, Wang X. Improvements on twin support vector machines[J]. IEEE Trans on Neural Networks, 2011, 22(6): 962-968.

[5] Shao Y, Deng N, Yang Z. Least squares recursive projection twin support vector machine for classification[J]. Pattern Recognition, 2012, 45(6): 2299-2307.

[6] Shao Y, Deng N, Chen W. Improved generalized eigenvalue proximal support vector machine[J]. IEEE Signal Processing Letters, 2013, 20(3): 213-216.

[7] Peng X. TSVR: An efficient twin support vector machine for regression[J]. Neural Networks, 2010, 23(3): 365-372.

[8] Shao Y, Zhang C, Yang Z. An ε-twin support vector machine for regression[J]. Neural Computing & Applications, 2012, 23(1): 175-185.

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

[10] Khemchandani R, Karpatne A, Chandra S. Generalized eigenvalue proximal support vector regressor[J]. Expert Systems with Applications, 2011, 38(10): 13136-13142.

[11] Jiang L, Zhang K. Efficient and robust feature extraction by maximum margin criterion[C]. Advances in Neural Information Processing Systems. Cambridge: MIT Press, 2004: 97-104.

[12] Parlett B. The symmetric eigenvalue problem[M], Philadelphia: SIAM Press, 1998: 61-80.

[13] Tikhonov A, Arsen V. Solutions of ill-posed problems[M]. New York: Wiley Press, 1977: 20-29.

[14] Mercer J. Functions of positive and negative type and the connection with the theory of integal equations[J]. Philosophical Trans of the Royal Society of London, 1909, 209(1): 415-446.

[15] Chang C, Lin C. LIBSVM: A library for support vector machines[J]. ACM Trans on Intelligent Systems and Technology, 2011, 2(3): 1-27.