第 14 卷 第 6 期
2009 年 6 月

中国图象图形学报
Journal of Image and Graphics

Vol. 14, No. 6
Jun., 2009

基于线性无关度的稀疏最小二乘支持向量回归机

赵永平 孙健国
(南京航空航天大学能源与动力学院,南京 210016)

摘 要 基于线性无关度提出了一种在高维特征空间中选择近似基的方法,并采用不完全抛弃法,充分利用非支持向量中的信息来建立稀疏最小二乘支持向量回归机的数学模型。另外,采用递推法加快了其模型的建立。该模型在保持预测精度基本不变的情况下,使支持向量的数目大大减少。最后,通过 3 个 UCI 数据集验证了该模型的有效性。

关键词 最小二乘支持向量回归机 线性无关 近似基 不完全抛弃

中图法分类号:TP181 **文献标识码**:A **文章编号**:1006-8961(2009)06-1136-05

Sparse Least Squares Support Vector Regression Machine Based on the Scale of Linear Independency

ZHAO Yong-ping, SUN Jian-guo
(*College of Energy and Power Engineering, Nanjing University of Aeronautics & Astronautics, Nanjing* 210016)

Abstract A novel method which selects the approximate bases of high dimensional feature space based on the scale of linear independency is proposed; and after combining the presented method with the partial reduction strategy, SLS-SVRM (Sparse Least Squares Support Vector Regression Machine) is built. In addition, the recursive trick is used to accelerate the establishment of SLS-SVRM. SLS-SVRM obviously decreases the number of support vector without loss of the predicted accuracy. Finally, three UCI (university of California at irvine) datasets confirm the effectiveness of the proposed model.

Keywords least squares support vector regression machine, linear independency, approximate bases, partial reduction

1 引 言

支持向量机[1]是一种有力的解决模式识别和非线性函数估计的机器学习方法[2-3]。基于不等式约束的原始支持向量机的解具有稀疏性的显著特点,并且与人工神经网络相比具有较强的泛化能力,成为目前机器学习的热点。通过用等式约束代替不等式约束,Suykens 等人提出了最小二乘支持向量机[4],并对其进行了深入的研究[5]。最小二乘支持向量机利用解线性方程组取代了原支持向量机的二次优化问题,降低了复杂度,加快了训练速度,目前已提出了一些求解最小二乘支持向量机的高效方法,如共轭梯度法[6]及其改进算法[7]、序贯最小优化法[8]和降阶法[9]。这些算法在一定程度上减少了计算量,加快了训练速度。但是,这些方法求解得到的最小二乘支持向量机的解都是非稀疏性的[10],这就在某种程度上限制了最小二乘支持向量机的应用。一个数学模型要有较强的泛化能力,一般通过两个途径来实现[11]:(1)引入正则化参数,像岭回归;(2)降低模型的复杂度,减少回归模型的输入项。由于最小二乘支持向量机的解缺乏稀疏性,这就导致了其泛化能力不强,并且加长了预测时间。

为了得到稀疏最小二乘向量机,人们提出了一

基金项目:国家自然科学基金项目(50576033)
收稿日期:2007-11-05;**改回日期**:2008-03-27
第一作者简介:赵永平(1982~),男。南京航空航天大学能源与动力学院博士研究生。主要研究方向为机器学习、控制理论与控制方法。E-mail:y. p. zhao@ nuaa. edu. cn

些剪枝算法[12]及其改进算法[13-17]。除此之外,人们还提出了一些选择高维特征空间近似基的算法,Suykens 等人首先基于最大化二次熵提出了固定窗法[18],Chen 等人[19]提出了基于矢量基学习的稀疏最小二乘支持向量机,还有基于相关性[20]在特征空间选择近似基的方法。这些方法的共性都是通过选择高维特征空间的近似基来实现稀疏最小二乘支持向量机的。

在相关性选择特征空间近似基的基础上,本文提出了一种改进的选择高维特征空间近似基的方法,该法根据线性无关度来选择近似基,并且采用不完全抛弃的策略充分利用非基中的信息来建立稀疏最小二乘支持向量回归机。为了避免加入新样本需要重新训练所有样本的问题,本文提出了动态建模的递推算法。最后用 UCI(university of california of irrine)数据集验证了该方法的有效性。

2　最小二乘支持向量回归机

通过把支持向量机的不等式约束改成等式约束,Suykens 等给出的最小二乘支持向量回归机的数学模型为:

$$\min_{\boldsymbol{w},\boldsymbol{e}} J(\boldsymbol{w},\boldsymbol{e}) = \frac{1}{2}\boldsymbol{w}^{\mathrm{T}}\boldsymbol{w} + \frac{C}{2}\sum_{i=1}^{N} e_i^2 \tag{1}$$

$$s.t.\quad d_i = \boldsymbol{w}^{\mathrm{T}}\varphi(\boldsymbol{x}_i) + b + e_i,\ i = 1,\cdots,N$$

式中,$d_i \in \mathbf{R}^1$ 代表系统的真实值,$\boldsymbol{x}_i \in \mathbf{R}^n$ 是系统的输入变量,e_i 代表系统预测值与真实值之间的误差,b 代表偏差量,$C \in \mathbf{R}^+$ 是正则化参数,$\varphi(\cdot)$ 为输入空间到特征空间的非线性映射,$\boldsymbol{w}$ 代表特征空间中超平面的法向量。为了解这个约束优化问题,建立 Lagrange 函数:

$$L(\boldsymbol{w},b,\boldsymbol{e};\boldsymbol{\alpha}) = J(\boldsymbol{w},\boldsymbol{e}) - \sum_{i=1}^{N}\alpha_i(\boldsymbol{w}^{\mathrm{T}}\varphi(\boldsymbol{x}_i) + b + e_i - d_i) \tag{2}$$

式中,$\boldsymbol{\alpha} = [\alpha_1,\alpha_2,\cdots,\alpha_N]^{\mathrm{T}}$,$\boldsymbol{e} = [e_1,e_2,\cdots,e_N]^{\mathrm{T}}$。由 KKT(Karush-Kuhn-Tucker)条件可知:

$$\begin{cases} \dfrac{\partial L}{\partial \boldsymbol{w}} = 0 \rightarrow \boldsymbol{w} = \sum_{i=1}^{N}\alpha_i\varphi(\boldsymbol{x}_i) \\ \dfrac{\partial L}{\partial b} = 0 \rightarrow \sum_{i=1}^{N}\alpha_i = 0 \\ \dfrac{\partial L}{\partial e_i} = 0 \rightarrow \alpha_i = Ce_i \\ \dfrac{\partial L}{\partial \alpha_i} = 0 \rightarrow \boldsymbol{w}^{\mathrm{T}}\boldsymbol{\varphi}(\boldsymbol{x}_i) + b + e_i - d_i = 0 \end{cases} \tag{3}$$

从方程组(3)中消去 e_i 和 $\boldsymbol{w}$ 后可以得到:

$$\begin{bmatrix} 0 & \boldsymbol{1}^{\mathrm{T}} \\ \boldsymbol{1} & \boldsymbol{\Omega} \end{bmatrix}\begin{bmatrix} b \\ \boldsymbol{\alpha} \end{bmatrix} = \begin{bmatrix} 0 \\ \boldsymbol{d} \end{bmatrix} \tag{4}$$

式中,$\boldsymbol{d} = [d_1,d_2,\cdots,d_N]^{\mathrm{T}}$,$\boldsymbol{1} = [1,1,\cdots,1]^{\mathrm{T}}$,$\boldsymbol{\Omega} = \boldsymbol{K} + C^{-1}\boldsymbol{I}$,$\boldsymbol{I}$ 为 N 维单位矩阵,$\boldsymbol{K}$ 是 $N \times N$ 维的核矩阵,$\boldsymbol{K}_{ij} = k(\boldsymbol{x}_i,\boldsymbol{x}_j) = (\varphi(\boldsymbol{x}_i))^{\mathrm{T}}\varphi(\boldsymbol{x}_j)$,$k(\boldsymbol{x}_i,\boldsymbol{x}_j)$ 是满足 Mercer 条件的核函数,其中常用的核函数有 RBF(radial basis function)核,多项式核和 Sigmoid 核等。于是可以得到最小二乘支持向量回归机的数学模型:

$$y = f(\boldsymbol{x}) = \sum_{i=1}^{N}\alpha_i k(\boldsymbol{x},\boldsymbol{x}_i) + b \tag{5}$$

由式(3)中的 $\alpha_i = Ce_i$ 可知,在最小二乘支持向量回归机中几乎每一个训练数据都是支持向量,丧失了支持向量机解的稀疏性优点。

3　特征空间近似基的选择

由式(3)中的 $\boldsymbol{w} = \sum_{i=1}^{N}\alpha_i\varphi(\boldsymbol{x}_i)$ 可知,$\boldsymbol{w}$ 是输入向量在特征空间中的线性组合。如果能够获得输入向量在特征空间中的最大线性无关组,那么就可以在一定程度上实现最小二乘支持向量回归机解的稀疏性。

3.1　输入数据在特征空间中的最大线性无关组

通过一个非线性映射 $\varphi(\cdot)$,训练数据集 $\{(\boldsymbol{x}_i, d_i)\}_{i=1}^{N}$ 在特征空间中的像集为 $\{(\varphi(\boldsymbol{x}_i),d_i)\}_{i=1}^{N}$。由矩阵论的知识可知,若 $\{\varphi(\boldsymbol{x}_i)\}_{i=1}^{l}$ 线性相关,则至少存在一个 $\varphi(\boldsymbol{x}_\sigma) = \sum_{i=1,i\neq\sigma}^{l}\mu_i\varphi(\boldsymbol{x}_i)$,其中 $\mu_i \in \mathbf{R}$。虽然 $\varphi(\cdot)$ 无法显式表达出来,但 $k(\boldsymbol{x}_\sigma,\boldsymbol{x}_\sigma) = \sum_{i=1,i\neq\sigma}^{l}\sum_{j=1,j\neq\sigma}^{l}\mu_i\mu_j k(\boldsymbol{x}_i,\boldsymbol{x}_j)$。因此,由以下步骤可以得到 $\{\varphi(\boldsymbol{x}_i)\}_{i=1}^{N}$ 的一个最大线性无关组:

(1) 初始化集合 $\boldsymbol{P} = \varnothing$,集合 $\boldsymbol{Q} = \{1,\cdots,N\}$;

(2) 从集合 $\boldsymbol{Q}$ 中随机取出一个数据 j,若 $\boldsymbol{P} = \varnothing$,则把 j 放到 $\boldsymbol{P}$ 中,否则计算

$$\min_{\boldsymbol{\mu}} g(\boldsymbol{\mu}) = (\varphi(\boldsymbol{x}_j) - \sum_{i\in\boldsymbol{P}}\mu_i\varphi(\boldsymbol{x}_i))^{\mathrm{T}}(\varphi(\boldsymbol{x}_j) - \sum_{i\in\boldsymbol{P}}\mu_i\varphi(\boldsymbol{x}_i)) \tag{6}$$

(3) 若式(6)等于零,则抛弃数据 j,否则把 j 放到集合 $\boldsymbol{P}$ 中;

(4) 若 $\boldsymbol{Q} = \varnothing$ 则终止,否则转到步骤(2)。

通过以上步骤可以得到 $\{\varphi(\boldsymbol{x}_i)\}_{i=1}^N$ 的一个最大线性无关组,其中式(6)等于零,说明 $\varphi(\boldsymbol{x}_j)$ 可以由 $\{\varphi(\boldsymbol{x}_i)\mid i\in\boldsymbol{P}\}$ 线性表示,否则说明 $\varphi(\boldsymbol{x}_j)$ 不可以由 $\{\varphi(\boldsymbol{x}_i)\mid i\in\boldsymbol{P}\}$ 线性表示,则 $\{\varphi(\boldsymbol{x}_j),\varphi(\boldsymbol{x}_{i\in\boldsymbol{P}})\}$ 线性无关,这就是一般基于线性相关性选择高维特征空间近似基的方法。式(6)的求解过程如下,令 $\frac{\partial g}{\partial\boldsymbol{\mu}}=0$ 可得 $\boldsymbol{\mu}=\boldsymbol{K}_{PP}^{-1}\boldsymbol{K}_{Pj}$,其中 $\boldsymbol{K}_{PP}=\boldsymbol{K}(\boldsymbol{x}_{i\in\boldsymbol{P}},\boldsymbol{x}_{j\in\boldsymbol{P}})$,$\boldsymbol{K}_{Pj}=\boldsymbol{K}(\boldsymbol{x}_{i\in\boldsymbol{P}},\boldsymbol{x}_j)$,则

$$g_{\min}=\boldsymbol{K}_{jj}-\boldsymbol{K}_{Pj}^{\mathrm{T}}\boldsymbol{K}_{PP}^{-1}\boldsymbol{K}_{Pj}\tag{7}$$

3.2 基于线性无关度选择特征空间中的近似基

若只按 $g_{\min}$ 是否等于零来选择特征空间中的近似基的话,就有可能导致如下两个问题:(1)根本无法实现最小二乘支持向量回归机解的稀疏性,主要是因为特征空间是一个高维的空间,输入数据只要稍微有一点不同或者受噪声的干扰反映在特征空间中就是线性无关,也就是说这种选择近似基的方法鲁棒性很差;(2)由式(7)可知,当 $g_{\min}$ 很小时,在下一步计算 $\boldsymbol{K}_{PP}^{-1}$ 时数值很不稳定,可以导致 $g_{\min}<0$ 不正常情况的出现,主要是因为此时 $\boldsymbol{K}_{PP}$ 接近奇异。以上两种情况在做仿真试验的过程中都出现过,故提出了下面基于线性无关度选择特征空间近似基的方法。此方法的算法步骤如下:

(1) 初始化集合 $\boldsymbol{P}=\varnothing$,集合 $\boldsymbol{Q}=\{1,\cdots,N\}$,阈值 ε(一个小的正数);

(2) 从集合 $\boldsymbol{Q}$ 中随机取出一个数据 j,若 $\boldsymbol{P}=\varnothing$,则把 j 放到 $\boldsymbol{P}$ 中,$\boldsymbol{K}_{PP}^{-1}=\frac{1}{k(\boldsymbol{x}_j,\boldsymbol{x}_j)}$;否则计算式(7);

(3) 若 $g_{\min}<\varepsilon$,则抛弃数据 j,否则由 Sherman-Morrison 定理[21]可知

$$\boldsymbol{K}_{(P+j)(P+j)}=\begin{bmatrix}\boldsymbol{K}_{PP}^{-1} & \boldsymbol{0}\\ \boldsymbol{0} & 0\end{bmatrix}+\begin{bmatrix}\boldsymbol{K}_{PP}^{-1}\boldsymbol{K}_{Pj}\\ -1\end{bmatrix}r[\boldsymbol{K}_{Pj}^{\mathrm{T}}\boldsymbol{K}_{PP}^{-1}\quad -1]\tag{8}$$

其中 $r=\frac{1}{\boldsymbol{K}_{jj}-\boldsymbol{K}_{Pj}^{\mathrm{T}}\boldsymbol{K}_{PP}^{-1}\boldsymbol{K}_{Pj}}$,并把 j 放到集合 $\boldsymbol{P}$ 中,$\boldsymbol{K}_{PP}=\boldsymbol{K}_{(P\cup j)(P\cup j)}$;

(4) 若 $\boldsymbol{Q}=\varnothing$ 则终止,否则转到步骤(2)。

以上选择特征空间近似基的方法避免了用最大线性无关组作为特征空间近似基带来的问题,同时可以实现 $\boldsymbol{K}_{PP}^{-1}$ 的迭代求解,减少了计算量,增加了数值的稳定性。

4 稀疏最小二乘支持向量回归机

按第 3 节中的方法可以进行特征空间中近似基的选择,从而实现稀疏最小二乘支持向量回归机,若只用 P 中的向量来进行最小二乘支持向量回归机的建模,而完全抛弃其他的向量,这种做法有一定的不妥之处,因为其他向量虽然为非支持向量,但往往也包含有用的信息,这对最小二乘支持向量回归机的建模是有益的。同时在基于以上选择特征空间近似基的最小二乘支持向量回归机的建模中,每增加一个新样本需要进行一次矩阵求逆,模型的维数越高,运算量越大,为此提出了下面用递推法进行稀疏最小二乘支持向量回归机建模的过程:

(1) 初始化集合 $\boldsymbol{P}=\varnothing$,集合 $\boldsymbol{Q}=\{1,\cdots,N\}$,阈值 ε(一个小的正数);

(2) 当按照前面所述的线性无关度法从 $\boldsymbol{Q}$ 中取出第一个基时,把它放入 $\boldsymbol{P}$ 中,计算 $\bar{\boldsymbol{K}}_{OP}=\begin{bmatrix}0 & \boldsymbol{1}\\ \boldsymbol{1} & \tilde{\boldsymbol{K}}_{OP}\end{bmatrix}$,$\boldsymbol{\Omega}_{PP}=(\bar{\boldsymbol{K}}_{OP}^{\mathrm{T}}\bar{\boldsymbol{K}}_{OP})^{-1}$,其中 $\boldsymbol{O}=\{1,2,\cdots,N-1,N\}$,$\boldsymbol{1}=[1,\cdots,1]_{N\times1}^{\mathrm{T}}$,$\tilde{\boldsymbol{K}}_{OP}=\boldsymbol{K}(\boldsymbol{x}_{i\in\boldsymbol{O}},\boldsymbol{x}_{j\in\boldsymbol{P}})+\frac{\delta_{ij}}{C}$,$\delta_{ij}=\begin{cases}1 & i=j\\ 0 & i\neq j\end{cases}$;

(3) 按照线性无关法继续进行选基,当 $\boldsymbol{Q}$ 中的 j 被选中取出时,计算 $\boldsymbol{\Omega}_{Pj}=\bar{\boldsymbol{K}}_{OP}^{\mathrm{T}}\begin{bmatrix}1\\ \tilde{K}_{Oj}\end{bmatrix}$,由 Sherman-Morrison 定理可知 $\boldsymbol{\Omega}_{(P\cup j)(P\cup j)}=\begin{bmatrix}\boldsymbol{\Omega}_{PP} & \boldsymbol{0}\\ \boldsymbol{0} & 0\end{bmatrix}+\begin{bmatrix}\boldsymbol{\Omega}_{PP}\boldsymbol{\Omega}_{Pj}\\ -1\end{bmatrix}s[\boldsymbol{\Omega}_{Pj}^{\mathrm{T}}\boldsymbol{\Omega}_{PP}]$,其中 $s=\frac{1}{k[(\boldsymbol{x}_j,\boldsymbol{x}_j)-\boldsymbol{\Omega}_{Pj}^{\mathrm{T}}\boldsymbol{\Omega}_{PP}\boldsymbol{\Omega}_{Pj}]}$,令 $\boldsymbol{\Omega}_{PP}=\boldsymbol{\Omega}_{(P\cup j)(P\cup j)}$,并把 j 放入 $\boldsymbol{P}$ 中,然后计算 $\bar{\boldsymbol{K}}_{OP}=\left[\bar{\boldsymbol{K}}_{OP}\quad\begin{pmatrix}1\\ \tilde{\boldsymbol{K}}_{Oj}\end{pmatrix}\right]$;

(4)若 $\boldsymbol{Q}=\varnothing$,则终止建模,否则转到步骤(3)。

5 数值仿真

为了说明算法的效果,在 Matlab6.5,CPU Celeron(R) 1.70GHz,256M 内存环境下用 UCI 数据集[22]进行测试。数据集的描述如下:

(1) Boston housing data set:此数据集有 506 个

样本,13 个属性,用 400 个样本进行训练,106 个样本进行测试。

(2) Stock data set:此数据集有 950 个样本,9 个属性,用 600 个样本进行训练,350 个样本进行测试。

(3) Auto-MPG data set:此数据集有 392 个样本,7 个属性,用 350 个样本进行训练,42 个样本进行测试。

在仿真的过程当中,选的核函数为 RBF 核函数 $k(\boldsymbol{x}_i,\boldsymbol{x}_j) = \exp\left(-\frac{\|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2}{2\gamma^2}\right)$,核参数的选取采用留一交叉验证法[23-24]。定义一个检验模型性能的指标为 $\mathrm{RMSE} = \sqrt{\frac{1}{N_{\text{test}}}\sum_{i=1}^{N_{\text{test}}}(y_i - \hat{y}_i)^2}$,其中 N_{test} 代表测试数据的规模,y_i 代表真实值,$\hat{y}_i$ 代表模型的预测值。测试的结果如表 1 - 3 所示:

表 1　Boston housing 数据集测试结果

Tab.1　Testing results of boston housing dataset

	预测精度	训练时间(s)	预测时间(s)	支持向量的数目
标准	2.64	11.66	5.25	400
矢量基	2.74	14.89	4.17	318
固定窗	2.68	69.06	4.52	390
无关度	2.78	12.98	1.53	116

表 2　Stock 数据集测试结果

Tab.2　Testing results of stock dataset

	预测精度	训练时间(s)	预测时间(s)	支持向量的数目
标准	0.75	25.49	23.31	600
矢量基	0.80	25.42	16.39	410
固定窗	0.79	438.77	16.11	400
无关度	0.79	23.08	6.31	155

表 3　Auto-MPG 数据集测试结果

Tab.3　Testing results of Auto-MPG dataset

	预测精度	训练时间(s)	预测时间(s)	支持向量的数目
标准	2.19	8.05	1.69	350
矢量基	2.28	4.38	0.67	138
固定窗	2.27	548.20	1.05	210
无关度	2.22	1.94	0.11	21

表中的“标准”代表非稀疏最小二乘支持向量回归机,“矢量基”代表文献[19]中基于矢量基学习的稀疏最小二乘支持向量回归机,“固定窗”代表文献[18]中的固定窗法,“无关度”代表本文所提出的方法。

由表可知,在预测精度(RMSE)基本不变的情况下,固定窗法需要最长的训练时间,选择的支持向量也比较多。和固定窗法相比较,矢量基法需要的训练时间大大缩短,但选择支持向量的数目并没有得到明显的改善,因此,预测时间也就不会有大的改观。本文提出的方法,和矢量基法相比,训练时间不仅没有增加,选择的支持向量的数目大大减少,因而也就很大程度上缩短了预测时间。造成这样结果的主要原因是因为本文提出的方法选择的基具有较强的鲁棒性,并且采用不完全抛弃的策略充分利用了非支持向量中的信息,以及用递推法减少了稀疏最小二乘支持向量回归机建模的时间。

6　结　论

本文提出了一种根据线性无关度选择特征空间近似基的方法,依此选择的基不仅线性无关,而且具有较强的鲁棒性。同时,采用不完全抛弃的策略,充分利用非支持向量中的信息建立稀疏最小二乘支持向量回归机模型。另外,采用递推法减少了建模的时间,这就使得本文提出的方法,不仅训练时间没有增加,而且支持向量数目大大减少,因而也就缩短了预测时间。最后用 UCI 数据集验证了本文提出方法有效性。

参考文献(References)

1　Vapnik V N. The Nature of Statistical Learning Theory[M]. New York: Springer-Verlag, 1995, 85-124.

2　Übeyli E D. Multiclass support vector machines for diagnosis of erythemato-squamous diseases[J]. Expert Systems with Applications, 2008, **35**(4): 1733-1740.

3　Wang W J, Men C Q, Lu W Z. Online prediction model based on support vector machine[J]. Neurocomputing, 2008, **71**(4-6): 550-558.

4　Suykens J A K, Vandewalle J. Least squares support vector machine classifiers[J]. Neural Processing Letters, 1999, **9**(3): 293-300.

5　Suykens J A K, Van Gestel T, De Brabanter J, *et al*. Least Squares Support Vector Machines[M]. Singapore: World Scientific, 2002.

6　Suykens J A K, Lukas L, Dooren P V, *et al*. Least squares support vector machine classifiers: a large scale algorithm [A]. In:

Proceedings of the European Conference on Circuit Theory and Design [C], Stresa, Italy, 1999: 839-842.

7 Chu W, Ong C J, Keerthi S S. An improved conjugate gradient scheme to the solution of least squares SVM[J]. IEEE Transactions on Neural Networks, 2005, **16**(2): 498-501.

8 Keerthi S S, Shevade S K. SMO algorithm for least squares SVM [A]. In: Proceedings of the International Joint Conference on Neural Networks 2003[C], Portland, OR, USA, 2003: 2088-2093.

9 Chua K S. Efficient computations for large least squares support vector machine classifiers[J]. Pattern Recognition Letters, 2003, **24**(1-3): 75-80.

10 Suykens J A K, Brabanter J D, Lukas L, *et al.* Weighted least squares support vector machines: robustness and sparse approximation [J]. Neurocomputing, 2002, **48**(1-4): 85-105.

11 Gao J B, Shi D, Liu X M. Significant vector learning to construct sparse kernel regression models[J]. Neural Networks, 2007(7): 791-798.

12 Suykens J A K, Lukas L, Vandewalle J. Sparse approximation using least squares support vector machines[A]. In: Proceeding of IEEE International Symposium on Circuits and System[C], Geneva, Switz, 2000: 757-760.

13 Kruif B J, Vries J A. Pruning error minimization in least squares support vector machines[J]. IEEE Transactions on Neural Networks, 2003, **14**(3): 696-702.

14 Kuh A, Wilde P D. Comments on "pruning error minimization in least squares support vector machines"[J]. IEEE Transactions on Neural Networks, 2007, **18**(2): 606-609.

15 Hoegaerts L, Suykens J A K, Vandewalle J, *et al.* A comparison of pruning algorithms for sparse least squares support vector machines [A]. In: Proceeding of International Conference on Neural Information Processing 2004 [C], Calcutta, India, 2004: 1247-1253.

16 Zeng X Y, Chen X W. SMO-based pruning methods for sparse least squares support vector machines[J]. IEEE Transactions on Neural Networks,2005, **16**(6): 1541-1546.

17 Jiao L C, Bo L F, Wang L. Fast sparse approximation for least squares support vector machine[J]. IEEE Transactions on Neural Networks, 2007, **18**(3): 685-697.

18 Espinoza M, Suykens J A K, Moor B D. Least squares support vector machines and primal space estimation[A]. In: Proceeding of the 42nd IEEE Conference on Decision and Control[C], Maui, HI, USA, 2003: 3451-3456.

19 Chen A J, Song Z H, Li P. Modeling method of least squares support vector regression based on vector base learning[J]. Control Theory & Applications, 2007, **24**(1): 1-5. [陈爱军,宋执环,李平. 基于矢量基学习的最小二乘支持向量机建模[J]. 控制理论与应用, 2007,**24**(1):1-5.]

20 Gan L Z, Liu H K, Sun Y X. Sparse least squares support vector machine for function estimation [A]. In: Proceedings of the 3rd International Symposium on Neural Networks[C], Chengdu, China, 2006: 1016-1021.

21 Zhang X D. Matrix Analysis and Applications [M]. Beijing: Tsinghua Press, 2004. [张贤达. 矩阵分析与应用[M]. 北京:清华大学出版社,2004.]

22 Murphu P M, Aha D W. UCI Repository of Machine Learning Database[EB/OL]. 2007,available from: http://www.ics.uci.edu/~mlearn/MLRepository.html.

23 Zhao Y, Keong K C. Fast leave-one-out evaluation and improvement on inference for LS-SVMs [A]. In: Proceedings of the 17th International Conference on Pattern Recognition [C], Cambridge, UK, 2004: 494-497.

24 An S J, Liu W Q, Venkatesh S. Fast cross-validation algorithms for least squares support vector machine and kernel ridge regression[J]. Pattern Recognition, 2007, **40**(8): 2154-2162.