一种基于密度加权的最小二乘支持向量机稀疏化算法

司刚全，曹晖，张彦斌，贾立新
（西安交通大学电气工程学院，710049，西安）

摘要：针对最小二乘支持向量机失去标准支持向量机稀疏特性的问题，提出了一种基于密度加权的稀疏化算法. 首先计算样本的密度信息，对样本估计误差进行密度加权获得该样本对模型的可能贡献度；然后选取具有最大可能贡献度的样本作为支持向量，同时对支持向量样本邻域内的其他样本密度信息进行削减，从而避免相似样本被选中为支持向量；再选择剩余样本中具有最大可能贡献度的样本添加到支持向量集中，直到模型性能满足要求. 仿真和实际应用表明，与 Suykens 提出的标准稀疏化算法相比，所提出的算法能有效剔除冗余支持向量，具有更好的稀疏性和鲁棒性.

关键词：最小二乘支持向量机；密度加权；稀疏化；磨机负荷
中图分类号：TP18 **文献标志码**：A **文章编号**：0253-987X(2009)10-0011-05

Density Weighted Pruning Method for Sparse Least Squares Support Vector Machines

SI Gangquan, CAO Hui, ZHANG Yanbin, JIA Lixin
(School of Electrical Engineering, Xi′an Jiaotong University, Xi′an 710049, China)

Abstract: For lack of sparseness characteristic in support vector machines (SVM) by least squares solution (LSSVM), a density weighted pruning algorithm to improve the sparseness of the LSSVM regression model is investigated. The estimation error of training sample is weighted by corresponding density value and the potential contribution of training sample is obtained. Then the sample with the greatest value in the sorted spectrum of the potential contribution is selected as the support vector. The density and the potential contribution of training sample are updated and the potential contribution of the neighborhood sample of support vectors is significantly reduced. Thus the rechoosing of similar sample as support vector is properly avoided. More support vectors in the training sample set are iteratively selected, until the user-defined performance is achieved, thus the sparse LSSVM model is obtained. The simulation and practical applications indicate that the proposed method performs more effectively than Suykens standard sparse method for removing the redundant support vector with better sparseness and robustness.

Keywords: least square support vector machine; density weighted; sparse; mill load

最小二乘支持向量机(LSSVM)是标准支持向量机的一种扩展，由于损失函数采用等式约束使得求解难度大大降低，从而得到广泛应用. 但是，由于 $\alpha_k = \gamma e_k$ 的条件，使得 LSSVM 失去了标准 SVM 的稀疏特性[1]，所有训练样本都成为支持向量，导致模型变得非常“臃肿”，因此对 LSSVM 模型进行稀疏化成为其应用过程中的一个关键步骤. 最常用的稀

收稿日期：2009-04-02. 作者简介：司刚全(1980—)，男，博士生；贾立新(联系人)，男，副教授. 基金项目：国家高技术研究发展计划资助项目(2006AA04Z180)；教育部高等学校博士学科点专项科研基金资助项目(20070698059).

疏化方法是 Suykens 提出的去除部分具有较小|α|样本的方法[1]，文献[2]采用去掉具有最小引入误差的样本来获得更好的稀疏性能. 以上两种方法都是基于单一样本误差来确定样本对模型贡献度，这就导致具有相似性的样本不能进一步稀疏化，最终结果模型中可能包含冗余支持向量. 文献[3-5]采用获得稀疏训练样本子集的方式，得到稀疏 LSSVM 模型. 由于稀疏训练子集的获取并未考虑到样本对模型的贡献度，因此获得的训练子集并非最佳候选集.

本文提出一种基于密度加权的最小二乘支持向量机稀疏化方法(DW-LSSVM)，综合了样本对模型的贡献度以及样本分布信息，从样本集整体出发，保留具有最大可能贡献度的样本作为支持向量，从而有效避免了支持向量中的冗余样本，得到稀疏性更好的模型.

1 最小二乘支持向量机函数估计

LSSVM 采用如下函数对未知函数进行估计[1]

$$y(\boldsymbol{x}) = \boldsymbol{w}^{\mathrm{T}}\boldsymbol{\phi}(\boldsymbol{x}) + b \tag{1}$$

式中：非线性函数 $\boldsymbol{\phi}(\cdot)$：$\mathbf{R}^n \rightarrow \mathbf{R}^{n_h}$ 将输入空间映射为高维的特征空间. 给定一个含有 N 个样本的训练样本数据集$\{\boldsymbol{x}_k, y_k\}_{k=1}^N$，其中输入为 m 维向量 $\boldsymbol{x}_k \in \mathbf{R}^n$，输出为 $y_k \in \mathbf{R}$，且所有训练样本已全部归一化到一个超立方体中. LSSVM 定义如下优化问题

$$\min_{w,b,e} J(\boldsymbol{w}, e) = \frac{1}{2}\boldsymbol{w}^{\mathrm{T}}\boldsymbol{w} + \gamma\frac{1}{2}\sum_{k=1}^{N} e_k^2 \tag{2}$$

满足以下等式约束

$$y_k = \boldsymbol{w}^{\mathrm{T}}\boldsymbol{\phi}(\boldsymbol{x}_k) + b + e_k \quad k = 1, \cdots, N \tag{3}$$

式中：e_k 为误差变量；γ 为正则化因子.

为求解该优化问题，定义如下拉格朗日函数

$$\boldsymbol{L}(\boldsymbol{w}, b, e; \boldsymbol{\alpha}) = J(\boldsymbol{w}, e) - \sum_{k=1}^{N}\alpha_k\{\boldsymbol{w}^{\mathrm{T}}\boldsymbol{\phi}(\boldsymbol{x}_k) + b + e_k - y_k\} \tag{4}$$

式中：$\alpha_k \in \mathbf{R}$ 为拉格朗日乘子. 分别求出 $\boldsymbol{L}(\boldsymbol{w}, b, e, \boldsymbol{\alpha})$ 对 $\boldsymbol{w}$、b、e、$\boldsymbol{\alpha}$ 的偏微分，可以得到式(2)的最优条件如下

$$\left.\begin{aligned}
&\frac{\partial \boldsymbol{L}}{\partial \boldsymbol{w}} = 0 \rightarrow \boldsymbol{w} = \sum_{k=1}^{N}\alpha_k\boldsymbol{\phi}(x_k)\\
&\frac{\partial \boldsymbol{L}}{\partial b} = 0 \rightarrow \sum_{k=1}^{N}\alpha_k = 0\\
&\frac{\partial \boldsymbol{L}}{\partial e_k} = 0 \rightarrow \alpha_k = \gamma e_k\\
&\frac{\partial \boldsymbol{L}}{\partial \alpha_k} = 0 \rightarrow \boldsymbol{w}^{\mathrm{T}}\boldsymbol{\phi}(\boldsymbol{x}_k) + b + e_k - y_k = 0\\
&k = 1, \cdots, N
\end{aligned}\right\} \tag{5}$$

由式(5)可知，由于 $\alpha_k = \gamma e_k$，而函数估计结果一般 $e_k \neq 0$，因此 $\alpha_k \neq 0$，LSSVM 也就失去了标准 SVM 的稀疏特性. 用 α_k 和 b 来表示式(5)中的 e_k 和 $\boldsymbol{w}$，可得到线性系统如下

$$\begin{bmatrix} 0 & \mathbf{1}^{\mathrm{T}} \\ 1 & \boldsymbol{\Omega} + \boldsymbol{\gamma}^{-1}\boldsymbol{I} \end{bmatrix}\begin{bmatrix} b \\ \boldsymbol{\alpha} \end{bmatrix} = \begin{bmatrix} 0 \\ \boldsymbol{y} \end{bmatrix} \tag{6}$$

式中：$\boldsymbol{y} = [y_1, \cdots, y_N]^{\mathrm{T}}$；$\mathbf{1} = [1, \cdots, 1]^{\mathrm{T}}$；$\boldsymbol{\alpha} = [\alpha_1, \cdots, \alpha_N]^{\mathrm{T}}$；$\boldsymbol{\Omega}$ 是一个方阵，其第 k 行 d 列的元素为

$$\Omega_{kd} = \boldsymbol{\phi}(\boldsymbol{x}_k)^{\mathrm{T}}\boldsymbol{\phi}(\boldsymbol{x}_d) = K(\boldsymbol{x}_k, \boldsymbol{x}_d) \quad k, d = 1, \cdots, N \tag{7}$$

式中：$K(\boldsymbol{x}_k, \boldsymbol{x}_d)$ 是核函数，它可以是满足 Mercer 条件的任意对称函数.

最后得到的用于函数估计的 LSSVM 模型为

$$y(x) = \sum_{k=1}^{N}\alpha_k K(\boldsymbol{x}, \boldsymbol{x}_k) + b \tag{8}$$

式中：α_k、b 是式(6)的解.

2 基于密度加权的 LSSVM 稀疏化算法

2.1 算法描述

定义样本$\{\boldsymbol{x}_k, y_k\}$处的密度指标为

$$D_k = \sum_{j=1}^{N}\exp\left[-\frac{\|\{\boldsymbol{x}_k, y_k\} - \{\boldsymbol{x}_j, y_j\}\|^2}{(\gamma_{\mathrm{a}}/2)^2}\right] \tag{9}$$

式中：γ_{a} 为一正数，表示其邻居半径. 显然，样本$\{\boldsymbol{x}_k, y_k\}$和$\{\boldsymbol{x}_j, y_j\}$之间距离越近，其对密度指标 D_k 的贡献越大，如果一个样本有多个相邻的其他样本点，则该样本具有较高的密度指标值. γ_{a} 定义了样本$\{\boldsymbol{x}_k, y_k\}$的一个邻域，$\gamma_{\mathrm{a}}$ 以外的样本点对该样本的密度指标影响较小. 对样本$\{\boldsymbol{x}_k, y_k\}$的误差 e_k 进行密度加权，得到其对模型的可能贡献度指标

$$P_k^{\bigstar} = D_k e_k \tag{10}$$

经过密度加权，$P_k^{\bigstar}$ 既包含了样本$\{\boldsymbol{x}_k, y_k\}$对模型的误差贡献信息，又包含了一部分训练样本集的样本分布信息，也即综合了样本$\{\boldsymbol{x}_k, y_k\}$及其邻域内其他样本对模型的误差贡献. 对|$P_k^{\bigstar}$|进行排序，显然具有最大|$P^{\bigstar}$|的样本对模型的贡献度最大. 选择对应样本为最大可能支持向量，设其为$\{\boldsymbol{x}_{\mathrm{s1}}, y_{\mathrm{s1}}\}$，对应的密度指标为 D_{s1}，重新计算每个样本点的密度指标

$$D_k = D_k - D_{\mathrm{s1}}\exp\left[-\frac{\|\{\boldsymbol{x}_k, y_k\} - \{\boldsymbol{x}_{\mathrm{s1}}, y_{\mathrm{s1}}\}\|^2}{(\gamma_{\mathrm{b}}/2)^2}\right] \tag{11}$$

式中：$\gamma_{\mathrm{b}} = \lambda\gamma_{\mathrm{a}}$($\lambda$ 为正数)为最大可能支持向量$\{\boldsymbol{x}_{\mathrm{s1}}, y_{\mathrm{s1}}\}$的邻域范围. 从式(11)中可以看出，由于$\{\boldsymbol{x}_{\mathrm{s1}}, y_{\mathrm{s1}}\}$已经被选中为支持向量，重新计算其密度指标

时，结果变为 0，从而避免了在后续过程中被再次选中. 在样本$\{\boldsymbol{x}_{sl}, y_{sl}\}$邻域内的其他样本的 D_k 得以大幅度降低，被选中为支持向量的可能性也被减少，从而有效减少了最终估计模型的冗余支持向量.

利用选出的支持向量训练 LSSVM 获得估计模型，然后计算出对应于该模型下的各样本的误差，与更新后的 D_k 相乘获得新的可能贡献度指标，从而可以选出剩余训练样本中的最大可能支持向量. 在后续样本密度指标更新过程中，当 $D_k<0$ 时，则将其设置为 0，相应的该样本的最大可能贡献度指标也变为 0，从而排除了此样本点被选中为支持向量的可能性. 从剩余的样本数据中继续采用上面类似的方法寻找支持向量，直到模型的性能指标满足需要.

在稀疏化过程中，采用通用的均方差指标来评价 LSSVM 模型的性能

$$\sigma_{\mathrm{MSE}}=\frac{1}{N-1}\sum_{k=1}^{N}(y_k-f(\boldsymbol{x}_k))^2 \tag{12}$$

式中：y_k 是样本数据集中第 k 个样本的输出；$f(\boldsymbol{x}_k)$ 为对应于第 k 个样本的 LSSVM 模型估计值.

2.2　算法步骤

输入：训练样本集 $S_{\mathrm{T}}=\{\boldsymbol{x}_k, y_k\}_{k=1}^{N}$，空支持向量样本集 $S_{\mathrm{S}}=\varnothing$，模型性能指标下降阈值 $\xi^{S_{\mathrm{T}}}$，每个稀疏化循环过程中增加的支持向量个数 M，计算样本密度指标过程中的邻居半径 γ_{a} 和邻域范围因子 λ.

输出：支持向量样本集 S_{S}，稀疏化后的 LSSVM 模型.

(1)采用 S_{T} 训练 LSSVM 模型，获得 S_{T} 中每个样本误差 e_k 和模型初始性能指标 $P_0^{S_{\mathrm{T}}}$；

(2)根据式(9)，计算 S_{T} 中每个样本的初始密度指标 D_k；

(3)根据式(10)，计算每个样本的可能贡献度指标 $P_k^{\bigstar}$；

(4)对可能贡献度指标$|P_k^{\bigstar}|$进行排序；

(5)从排序后的$|P_k^{\bigstar}|$谱中选择具有最大$|P_k^{\bigstar}|$的样本作为最大支持向量，添加到 S_{S} 中；

(6)根据式(11)，更新 S_{T} 中每个样本的密度指标 D_k，转向步骤(3)，直到本循环中选出的支持向量达到 M 个；

(7)第 $l(l=1,2,\cdots)$ 个循环，采用 S_{S} 训练 LSSVM，用当前训练后的 LSSVM 模型计算性能指标 $P_l^{S_{\mathrm{T}}}$ 和 S_{T} 中每个样本误差 e_k，判断模型性能是否满足 $P_0^{S_{\mathrm{T}}}\geqslant\xi^{S_{\mathrm{T}}}P_l^{S_{\mathrm{T}}}$，不满足则转步骤(3)，增选 M 个支持向量进入 S_{S}，满足则进入下一步；

(8)算法结束，S_{S} 为支持向量样本集，最终的 LSSVM 模型为经过稀疏化后的模型.

2.3　参数选择

本文提出的稀疏化算法需要选择两类参数：一类为 LSSVM 模型的超参数(γ,δ)，通常采用交叉验证法获得；另外一类是稀疏化参数$(\gamma_{\mathrm{a}},\lambda,M)$. γ_{a} 的选择与样本密度指标有着密切关系，选择过小意味着每个样本都是独立样本，算法将退化为基于单个样本误差对模型贡献度的方法，选择过大则样本之间的密度指标差别较小，不利于筛选具有最大贡献度的样本，通常情况下，γ_{a} 在 0.05～0.20 之间选择. λ 与 γ_{a} 共同决定支持向量的邻域影响范围，通常情况下 λ 在 0.8～1.2 之间选择. 当设定的模型性能指标阈值较高时，λ 应选择较小值，避免由于邻域范围太大导致大多数剩余样本的密度指标变为 0 而无可供选择的支持向量；反之可选择较大值，加快稀疏化速度. M 典型取值为样本数的 1%，M 取值较小可获得更好的稀疏性能，但大样本集则导致稀疏化时间变长，因此应平衡稀疏性能和时间加以选择.

3　实验结果及分析

3.1　sinc 函数仿真结果

为了验证所提出的基于密度加权的 LSSVM 稀疏化算法性能，与文献[1]相同，采用 sinc 函数产生 200 个训练样本

$$y=\frac{10\sin(x)}{x}+\varepsilon \tag{13}$$

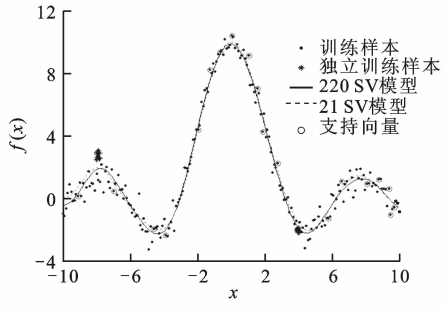
式中：输入 $x\in[-10,10]$，等间隔采样；ε 是均值为 0、方差为 0.5 的高斯噪声. 为了检验所提出的模型对相似样本的处理能力，人为增加了 20 个独立的样本点，其中在 $x\in[-8,-7.8]$范围内的 10 个样本点加入了均值为 1.5、方差为 0.2 的高斯噪声，用于模拟多个相似的具有较大偏差的样本情况. 在 $x\in[3.9,4.0]$范围内的 10 个样本点加入了均值为 −0.2、方差为 0.1 的高斯噪声，用于模拟多个相似的具有较小偏差的样本情况. 所有独立样本点在图 1 中均以“*”加以标识.

为比较本文所提出的稀疏化算法性能，同时采用 Suykens 提出的标准稀疏化方法[1]进行处理，LSSVM 模型均采用 RBF 核函数，超参数(γ,δ)采用十折交叉验证法获得，分别为 $\gamma=5.954\,1$ 和 $\delta^2=0.099\,1$，在稀疏化过程中，超参数不做调整. 本文算法中选择 $\gamma_{\mathrm{a}}=0.05$，$\lambda=1.0$，$M=1$，$\xi^{S_{\mathrm{T}}}=0.85$. 两种算法的稀疏化结果如图 1 所示，其中 $f(x)$ 为 LSS-

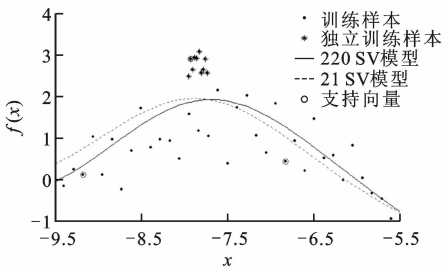
VM 模型估计的结果. 从图中可以看出，采用本文提出的方法，在满足预先设定的性能指标情况下，模型仅保留 21 个支持向量，而采用 Suykens 的标准方法的模型在 21 个支持向量时，估计结果已经严重偏离真实值，不能满足函数估计需要. 未进行稀疏化的 LSSVM 模型(220SV)在训练样本集上的 σ_{MSE} 为 0.278 9，采用两种方法稀疏化后的 σ_{MSE} 分别为 0.324 4 和 1.868 5，即稀疏化后的模型性能分别下降为初始性能的 85.97%和 14.93%. 从图 1b 和图 1e 可以看出，在左侧区域具有较大偏差的独立样本点，经过本文提出的算法稀疏化后，仅有 1 个样本点被保留为支持向量，且该样本点并不是偏离真实值最大的点，即非估计误差最大的样本点，而是具有最大可能贡献度的样本点. 在采用 Suykens 的标准方法稀疏化后的模型中，有 5 个样本点被保留为支持向量，且全部为偏离真实值(估计误差)最大的点. 同样从图 1c 和图 1f 可以看出，在右侧区域具有较小偏差的独立样本点，在采用 Suykens 方法稀疏化后，由于各样本点具有较小的偏差，因此在稀疏化进程的前期就已经全部被丢弃，经过本文提出的算法稀疏化后，有 1 个样本点被保留下来，从而使得模型在 $x\in[2,10]$ 区域内保持了良好的估计性能. 由此可以看出，本文提出的方法能够很好地综合每个样本点的误差贡献度以及样本整体分布信息，从而保留具有最大贡献度的样本作为支持向量. 与 Suykens 提出的方法相比，本文方法能够有效避免冗余支持向量的出现，同时保留了对模型具有关键作用的隐性支持向量，在稀疏化性能和模型估计性能上也有进一步的提高.

3.2 磨机负荷估计结果

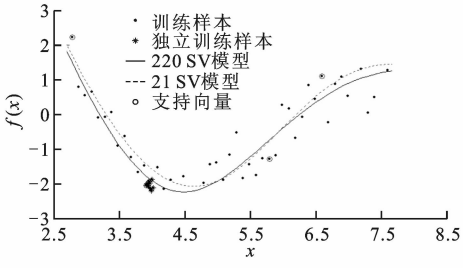
磨机负荷的准确检测是火电厂球磨机自动控制得以有效实施的关键，而在实际工业中由于环境恶劣、影响因素众多等原因一直未能得以解决[6]. 经过长期研究发现，磨机负荷可以通过噪声、振动、磨机



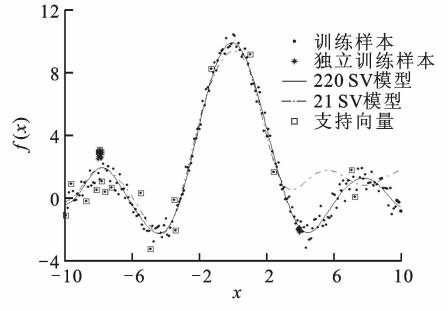
(a)本文方法稀疏化结果



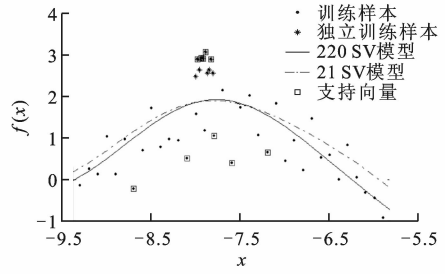
(b)a 图左侧区域局部放大



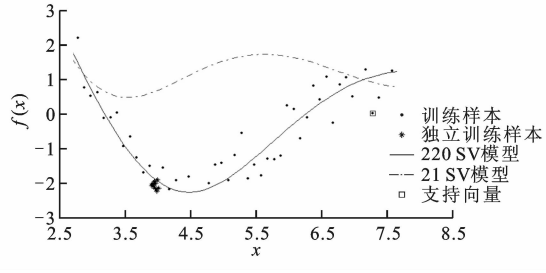
(c)a 图右侧区域局部放大



(d)Suykens 的标准方法稀疏化结果



(e)d 图左侧区域局部放大



(f)d 图右侧区域局部放大

图 1 LSSVM 模型稀疏化结果比较

电流、磨机出入口压差以及磨机出口温度等多个信号以及信号中包含的特征量进行综合获得[7]. 本文基于 DW-LSSVM 方法构建磨机负荷的软测量模型

$$y_{ml} = f(E_n, E_v, I, P_D, T) \tag{14}$$

式中：y_{ml}为磨机负荷；E_n、E_v、I、P_D 和 T 分别为磨机噪声特征值、磨机振动特征值、磨机电流、出入口差压和出口温度；$f(\cdot)$为经过稀疏化后的 LSSVM 估计模型.

经过各种现场实验收集了 4 800 组样本，包含了磨机处于低负荷、正常负荷以及高负荷等工况，从中抽取了 1 200 组样本作为训练样本集，320 组作为测试样本集. 以模型性能下降阈值为 85%进行稀疏化，采用本文方法稀疏化后的模型支持向量为 408 个，采用 Suykens 方法稀疏化后，支持向量为 934 个. 由此可知，实际工业过程收集的训练样本中存在大量的相似样本，从而影响了模型稀疏化效果. 进一步比较发现，采用本文提出的方法，支持向量较好地分布于各种工况下，而 Suykens 方法获得的模型中的支持向量更多地分布于低负荷和高负荷工况. 其原因在于该两种工况下的样本具有较大的方差水平，这样就导致对正常工况下样本的估计性能下降，显然这在实际应用中不能满足自动控制的要求. 经过本文提出的稀疏化方法得到的最终模型在测试样本集上的估计结果及误差如图2和图3所示. 从图

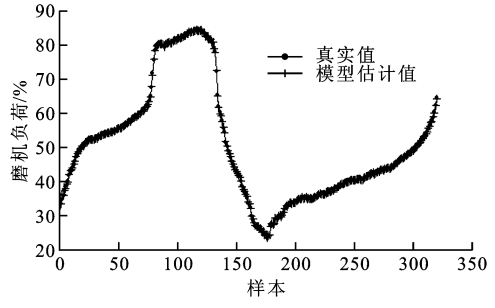


图 2　测试数据的估计输出

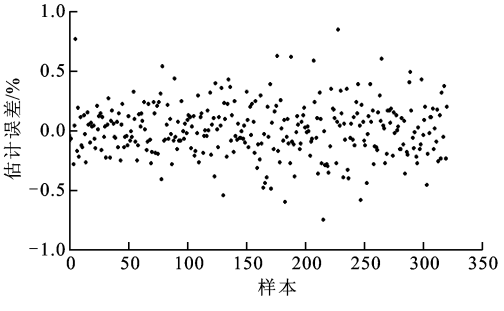


图 3　测试数据的估计误差

中可以看出，所建立的模型在各种工况下都具有较好的估计性能，鲁棒性得到加强，且由于支持向量的大幅度减少，使得模型复杂度大大降低，从而有利于在工业中实际应用.

4　结　论

本文针对实际训练样本集中存在大量相似样本的特点，提出了一种基于密度加权的最小二乘支持向量机稀疏化算法. 该方法充分考虑样本误差对模型的贡献以及样本分布信息，从而选择具有最大可能贡献度的样本作为支持向量，同时削减邻域内样本再次被选中的可能性. 通过函数仿真以及在工业实际中的应用表明，该方法能够显著提高模型的稀疏性，且模型中不再含有冗余支持向量，同时支持向量的分布更加合理，使得模型能够较好地适应工业实际需要，具有更好的鲁棒性.

参考文献：

[1] SUYKENS J A K, DE BRABANTER J, LUKAS L, et al. Weighted least squares support vector machines: robustness and sparse approximation [J]. Neurocomputing, 2002, 48(1/4): 85-105.

[2] DE KRUIF B J, DE VRIES T J A. Pruning error minimization in least squares support vector machines [J]. IEEE Transactions on Neural Networks, 2003, 14(3): 696-702.

[3] MILLER A J. Subset selection in regression [M]. 2nd ed. Boca Raton, USA: Chapman & Hall/CRC Press, 1990.

[4] CHEN S, COWAN C, GRANT P. Orthogonal least squares learning algorithm for radial basis function networks[J]. IEEE Transactions on Neural Networks, 1991, 2(2): 302-309.

[5] BAUDAT G, ANOUAR F. Kernel based methods and function approximation[C]// Proceedings of IJCNN'01. Washington DC, USA: IEEE Press, 2001: 1244-1249.

[6] KOLACZ J. Measurement system of the mill charge in grinding ball mill circuits[J]. Minerals Engineering, 1997, 10(12): 1329-1338.

[7] SI Gangquan, CAO Hui, ZHANG Yanbin, et al. Ball mill load measurement using self-adaptive feature extraction method and LS-SVM model[C]// Proceedings of ICIC 2008. Shanghai, China: Springer-Verlag Press, 2008: 275-284.

（编辑　杜秀杰）