

基于粒子群优化的 VB-LSSVM 算法 研究辛烷值预测建模^{*}

李 炜¹, 石连生¹, 梁成龙²

(1 兰州理工大学电气工程与信息工程学院 兰州 730050)

(2 兰州石化分公司油品储运厂 兰州 730060)

摘 要: 针对现有红外线分析仪表无法实现阶段在线检测车用汽油调合中, MMT 抗爆剂对辛烷值的影响问题, 考虑到样本数据较少的因素, 提出一种基于粒子群优化算法的矢量基最小二乘支持向量机方法, 首先以粒子群优化的方法来选取最小二乘支持向量机的模型参数, 然后用矢量基判据选择支持向量, 使最小二乘支持向量机的解具有稀疏性。该方法不但克服了常用的交叉验证法的耗时与盲目性问题, 发挥了最小二乘支持向量机的小样本学习和计算简单的特点, 而且提高了最小二乘支持向量机模型的泛化能力, 将其应用于汽油调合系统中研究法辛烷值的预测, 仿真结果表明, 该方法是可行且有效的。

关键词: 汽油调合; 辛烷值; 粒子群优化; 矢量基; 最小二乘支持向量机

中图分类号: TP273 **文献标识码:** A **国家标准学科分类代码:** 510 80

Forecasting model of research octane number based on PSO-VB-LSSVM

Li Wei, Shi Liansheng, Liang Chenglong

(1 Department of Electrical and Information Engineering, Lanzhou University of Technology, Lanzhou 730050, China)

(2 Factory of Oil Deposit and Transportation, Petrochemical Company of Lanzhou, Lanzhou 730060, China)

Abstract: Octane number can be improved by adding MMT when gasoline is blended. But the octane number can not be obtained by infrared analyzer. Considering the samples are few, a model is proposed in this paper, which integrates PSO (particle swarm optimization), VB (vector base) and LS-SVM (least square support vector machine). Firstly, the model parameters are selected using PSO algorithm, then the support vectors are selected using vector base learning, which makes the support vectors sparse. The method not only overcomes the time-consuming and blind problems that commonly used cross-validation method has, but also makes use of the small sample learning ability of LS-SVM and features simple calculation. The proposed method was applied to the prediction of research octane number (RON) in a system of gasoline blending. The obtained results demonstrate that the model based on PSO-VB-LSSVM can achieve good accuracy.

Key words: gasoline blending; octane number; particle swarm optimization; vector base; least square support vector machine

1 引言

辛烷值是车用汽油最重要的品质指标, 因此, 汽油辛

烷值的检测在石油炼制工业中占有十分重要的地位。高辛烷值汽油不仅可适应较高的压缩比, 提高发动机的热效率, 还可以改善汽车排放性能, 减少尾气对环境的污染。

采用抗爆剂是提高车用汽油辛烷值的重要手段。为

了弥补汽油无铅化造成的辛烷值降低,人们研发了不少新型抗爆剂,包括以 MTBE 为代表的含氧化合物、丙二酸酯、TKC 助剂、碳酸二甲酯、甲基环戊二烯三羰基锰 MMT^[3] 等。它们的结构不同,抗爆机理和效能也不同,而以 MMT 的抗爆效能为最好,几乎与烷基铅相当且添加量少,经济效益好,具有广阔的应用前景。但是随之带来的问题是现存的红外线分析仪表无法分析加入 MMT 抗爆剂对成品油研究辛烷值的影响,因而只能靠人的经验在确保产品较高的质量指标的前提下保守生产,从而造成一定的 MMT 抗爆剂浪费。

本文针对车用汽油一次调合之后,通过加入 MMT 抗爆剂提高辛烷值,而现有的红外线分析仪表无法实现阶段在线分析,考虑到生产中数据样本较少的因素,在提出了粒子群优化的矢量基最小二乘支持向量机 (PSO-VB-LSSVM) 方法的基础上,将其用于对成品油辛烷值检测建模。以粒子群优化的方法来选择最小二乘支持向量机的模型参数,既克服了常用的交叉验证法的耗时与盲目性问题,又发挥了最小二乘支持向量机的小样本学习能力;采用矢量基判断准则选择支持向量,使最小二乘支持向量机的解具有稀疏性,从而提高了最小二乘支持向量机模型的泛化能力和预测的鲁棒性。仿真结果表明, PSO-VB-LSSVM 模型较 PSO-LSSVM 模型和 LS-SVM 模型有更强的泛化能力和预测精度。

2 最小二乘支持向量回归估计

支持向量机 (SVM) 是由 Vapnik 最初在统计学习理论的基础上建立起来的一种非常有效的机器学习方法。它是近几年发展起来的一种人工智能技术,最初用于模式识别,目前在信号处理、函数逼近及图像处理等领域也得到了应用。然而对于标准的 SVM 算法,随着样本数据个数的增多,其所求解的二次规划问题显得越复杂,计算速度越慢;同时,抗噪声能力较差。解决此类问题的一个有效方法就是 LS-SVM 算法。

设给定 l 个样本数据 $\{x_j, y_j\}$, 其中 $j=1, 2, \dots, l$, $x_j \in R^n$ 为 m 维样本输入; $y_j \in R$ 为样本输出。LS-SVM 利用非线性映射, $\varphi(\cdot): X \rightarrow F$, 将训练样本集非线性地映射到一个高维特征空间,从而使非线性函数估计问题转化为高维特征空间中的线性函数估计问题,设回归函数 $y = \langle w \circ \varphi(x) \rangle + b$ 。

在最小二乘支持向量回归估计中,回归问题变为:

$$\begin{cases} \min \frac{1}{2} w^T w + \frac{1}{2} \sum_{j=1}^l \xi_j \\ s.t. \quad y_j = w^T \cdot \varphi(x_j) + b + \xi_j, \quad j=1, 2, \dots, l \end{cases} \quad (1)$$

式中: $\xi \in R^l$ 为误差向量, φ 为正规化参数。引入拉格朗日乘子 $a \in R^{k \times 1}$, 式 (1) 可化为:

$$\min J_{SSVR} = \frac{1}{2} w^T w + \frac{1}{2} \sum_{j=1}^l \xi_j - \sum_{j=1}^l a_j (w^T \cdot \varphi(x_j) + b + \xi_j - y_j) \quad (2)$$

由 KKT 最优条件, 式 (2) 的解为:

$$\begin{bmatrix} b \\ a \end{bmatrix} = \begin{bmatrix} 0 & \tilde{I}^T \\ \tilde{I} & ZZ^T + \frac{1}{c} I \end{bmatrix}^{-1} \times \begin{bmatrix} 0 \\ Y \end{bmatrix} \quad (3)$$

式中: $Z = [\varphi(x_1), \dots, \varphi(x_l)]^T$, $\tilde{a} = [a_1, \dots, a_l]^T$, $Y =$

$[y_1, \dots, y_l]^T$, \tilde{I} 为 $l \times l$ 单位列向量。非线性函数拟合的 ZZ^T 内积运算, 可用满足 Mercer 条件的核函数 $K(x_i, x_j)$ 替代。常见的核函数有线性函数、多项式函数、径向基函数、多层感知器函数。本文采用径向基函数: $K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{\sigma^2}\right)$ 。

由于 $w = \sum_{j=1}^l a_j \varphi(x_j)$, 因此非线性回归估计函数的表达式为:

$$y = \sum_{j=1}^l a_j K(x, x_j) + b \quad (4)$$

3 基于粒子群算法优化的矢量基最小二乘支持向量机估计

LS-SVM 模型属于黑盒子一类的模型, 在建模中, 只需关心对象的输入与输出, 而不必关心对象的具体结构, 输入与输出的影射关系由式 (4) 来完成。

在进行 LS-SVM 建模时, 有两类参数是非常重要的参数, 即正规化参数 c 和核参数 σ 。如何寻找最佳正规化参数和核参数的问题是支持向量机研究领域的一个公开问题。常用的参数选取方法是交叉验证法, 然而这种方法不仅耗时, 而且有一定的盲目性。因此本文提出采用粒子群优化的方法来选取 LS-SVM 的正规化参数 c 和核参数 σ 。

在 LS-SVM 的建模中, 由于每个样本对估计器都有贡献, 从而使 LS-SVM 失去了支持向量解的稀疏性优点, 当样本量增大时, 则模型的维数也随之增大。冗余信息的噪声则会被全部拟合到模型参数里, 削弱了模型的鲁棒性。基于此本文采用矢量基学习的方法判别支持向量, 使最小二乘支持向量机的解具有稀疏性。

3.1 粒子群算法基本原理

粒子群优化 PSO (Particle swarm optimization) 算法是一类新兴的基于群智能的随机优化算法, 简单易于实现且具有更强的全局优化能力。

在 PSO 中, 每个粒子都是解空间 (n 维) 的一点, 并且都具有一个速度 (n 维的矢量), 不同粒子具有对应于与目标函数相关的个体适应度。每个粒子根据自身的飞行经验和群体的飞行经验来调整自己的飞行轨迹, 向最优优点靠拢。对于某个粒子 i 其位置表示为 $q_i = (q_{i1}, q_{i2}, \dots, q_{in})$, 飞行速度表示为 $v_i = (v_{i1}, v_{i2}, \dots, v_{in})$, 经历过的最好位置记为 $p_i = (p_{i1}, p_{i2}, \dots, p_{in})$, 能获取的群体的最好位置记为 $p_g = (p_{g1}, p_{g2}, \dots, p_{gn})$ 。在每一步中, 粒子根据以下公式更新自己的速度和位置:

$$v_{id}(t+1) = Wv_{id}(t) + \zeta_1 r_1 (p_{id}(t) - q_{id}(t)) + \zeta_2 r_2 (p_{gd}(t) - q_{id}(t)) \quad (5)$$

$$q_{id}(t+1) = q_{id}(t) + v_{id}(t+1) \quad (6)$$

式中: W 称为惯性权重; $d=1, 2, \dots, n$ 为当前进化迭代次数; $i=1, 2, \dots, S$ S 为种群规模; r_1 和 r_2 为分布于 $[0, 1]$ 之间的随机数; ζ_1 和 ζ_2 为两个正常数, 称为加速因子。

PSO 中, 惯性权重 W 的选择很关键, 较大的 W 有较好的全局搜索能力, 而较小的 W 有较强的局部搜索能力。因此, 随着迭代次数的增加, 惯性权重应不断减少, 从而使得微粒群算法在初期具有较强的全局收敛能力, 而晚期具有较强的局部收敛能力。

本文取 W 随 t 的增加从 0.9 线性递减至 0.4 其中 W 与 t 的关系表达式为:

$$W(t) = 0.9 - \frac{t}{T_{\max}} \times 0.5 \quad (7)$$

T_{\max} 为最大迭代次数。当加速因子 ζ_1 较大时, 会使粒子过多地在局部范围徘徊, 而当加速因子 ζ_2 较大时, 会促使粒子过早收敛到局部最小值, 本文为了平衡随机因素的作用, 参考文献 [10], 将 ζ_1 和 ζ_2 均取为 2。

3.2 矢量基及其选择判据

定义 1 设存在一组矢量 v_k ($k=1, \dots, m$; $m \leq n$) 其高维空间的映射为 $\phi(v_k)$, 这些矢量的线性组合为:

$$\omega = \alpha (\phi^T(v_1), \dots, \phi^T(v_m)) \quad (8)$$

式中: α 为系数矢量, 如果 ω 线性无关, 则称矢量 v_k 为基矢量 (base vector BV)。

定义 2 由基矢量 v_k ($k=1, \dots, m$) 组成的集合, 称为基矢量集 (base vector set BVS)。

定义 3 在最小二乘支持向量建模中, 由支持向量基矢量集张成的子空间称为该模型的矢量空间。

下面分析如何确定 BVS 的 m 个 BV。假设 BVS 已经有两个 BV $\phi(v_1)$ 、 $\phi(v_2)$, 分别对应图 1 的 A、B 两点。这两个矢量张成一个矢量空间, 该空间可表示为 $\Gamma = (\lambda_1, \lambda_2) (\phi(v_1), \phi(v_2))^T$, 假设另外两个新矢量 (分别为图 1 的 C 和 D), 与 Γ 的夹角分别为 θ_1 、 θ_2 , 如果角度很小如 θ_1 , 则认为矢量 C 可由 A、B 近似表示, 而 D 的 θ_2 角较大, 表明用 A、B 近似线性表示 D 的误差较大, 因而

将 D 引入到 BVS 中。通过此种方法, 就可确定 BVS 以及矢量空间, 从而在这个矢量空间里, 最终得到期望的模型参数 λ 的维数, 即模型的阶数。通过新样本矢量与 BVS 张成的矢量空间的角度 θ 来确定其成为 BV 的条件, 可采用如下选择判据。

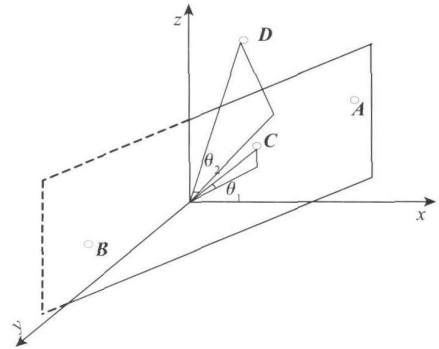


图 1 平面矢量空间确定 BV 的示意图

Fig. 1 Plane vector space defines BV

设 BVS 的映射 $\phi(v_k)$ ($k=1, \dots, m$) 张成的解空间为 Γ , $\hat{\Gamma}_1 = (\phi(x_{t_1}), \phi(x_{t_{t+1}}))$, $\hat{\Gamma}_{t+1} = (\phi(x_{t_{t+1}}), \phi(v_k))$ ($k=1, \dots, m$), $\tilde{\Gamma}_m = (\phi(v_k), \phi(v_g))$ ($k=1, \dots, m$)。则新增加的 $\phi(x_{t_1})$ 与 Γ 的夹角为:

$$\theta = \arcsin \Omega \quad (9)$$

$$\text{式中: } \Omega = \sqrt{\frac{\hat{\delta}_{1,b}}{N \hat{\Gamma}_1}}, \quad \hat{\delta}_{1,b} = \hat{\Gamma}_1 - \tilde{\Gamma}_{t+1}^T \tilde{\Gamma}_m^{-1} \tilde{\Gamma}_{t+1}.$$

如果 $\hat{\delta}_{1,b} \geq \hat{\Gamma}_1 (\sin \theta)^2$, 则说明这个新样本不可以由 BV 近似表示, 该样本应引入到 BVS 反之则不引入。由此可知, 运用 θ 角的大小可以任意调整 BV 的数量, 这样就可以减少 λ 的维数, 达到削减支持向量的目的, 使最小二乘支持向量机的解具有稀疏性。

3.3 基于 PSO-VB-LSSVM 的估计模型

基于 PSO-VB-LSSVM 的估计模型, 即首先采用 PSO 优化方法, 对 LS-SVM 的正规化参数 ρ 和核参数 σ 进行优化选择, 使模型的估计值与期望值的逼近程度达到预期的目标, 然后采用矢量基判据来选择支持向量, 使最小二乘支持向量机的解具有稀疏性。其算法步骤如下:

1) PSO 与 LS-SVM 的初始化。初始化加速因子 ζ_1 、 ζ_2 和惯性权重 W 及最大进化迭代次数 T_{\max} ; 并把正规化参数 ρ 和核参数 σ 映射为一群粒子, 初始化粒子的位置与速度, 将每个粒子的初始位置设为初始最好位置, 再根据适应度函数计算每个粒子的适应值, 将粒子中最小的适应值所对应的初始位置设为初始群体最好位置。这里, 每个粒子的适应度函数定义为: $F_i = \frac{1}{l} \sum_{j=1}^l (y_{ij} - \hat{y}_{ij})^2$, 其中, y_{ij} 为第 i 个粒子的 LS-SVM 训练输出值, $j=1, 2, \dots, l$ 。

..., \hat{y}_j 为 LS-SVM 期望输出值。

2) 按式 (5) 和 (6) 更新粒子的速度和位置, 生成新一代的种群。

3) 根据设定的适应度函数, 评价每个粒子的适应值。粒子的适应值越小, 则粒子的位置越好。

4) 检查终止条件, 若满足, 则将全局最优粒子映射为正规化参数 c 和核参数 σ , 以此优化结果作为 LS-SVM 的模型参数。否则转入 2。终止条件为寻优达到最大进化代数 T_{\max} 或者评价值小于给定精度。

5) 初始化值 θ , 将第 1 个训练样本作为第 1 个基矢量, 从第 2 个样本开始, 根据引理 1 的结论判断其是否为基矢量, 直到从全部的训练样本中选出所有基矢量为止, 将其组成基矢量集。

6) 用这些基矢量作为新的训练样本, 采用已经确定的正规化参数 c 和核参数 σ 作为 LS-SVM 的模型参数进行建模。

7) 用建立好的 LS-SVM 模型进行预测。

4 工业实例仿真

汽油调合的生产工艺是将炼油厂常减压装置、催化裂化装置、烷基化装置、MTBE 装置、连续重整装置、芳烃抽提装置等生产出的直馏汽油、催化汽油、车用异辛烷、MTBE 车用生成油、非芳烃等半成品按照工艺卡片要求的调合比例, 添加一定比例的添加剂, 在调合罐中通过调合喷嘴循环或经过管道调合均匀后, 生产出达到质量要求的合格产品。

根据工艺分析, 在一次调合后, 影响成品油研究法辛烷值的因素有空白油研究法辛烷值、汽油抗爆剂的添加比例、机泵的循环压力、循环时油品的温度以及循环时间。而在研究汽油抗爆剂对油品研究法辛烷值提升的属性时, 可以忽略机泵的循环压力、温度以及时间对其的影响。因此这里把空白油研究法辛烷值、汽油抗爆剂的添加比例作为输入参数, 把成品油的研究法辛烷值作为输出参数。这里针对 93 车用汽油经过催化汽油、生成油、MTBE 和异辛烷调合以后, 再加入 MMI 抗爆剂对研究法辛烷值的影响问题, 采用文中提出的方法对成品油研究法辛烷值建模, 模型可以描述为:

$$Y(k) = f(O(k), M(k)) \quad (10)$$

式中: $Y(k)$ 表示成品油研究法辛烷值; $O(k)$ 表示空白油研究法辛烷值; $M(k)$ 表示 MMI 抗爆剂加入比例, mg/L 。

本次实验中共采集 104 组样本数据, 经过预处理后, 剩余 92 组数据, 用其中的 62 组进行训练, 其余 30 组用于模型测试, 最小二乘支持向量机的核函数采用径向基核函数。在 MATLAB 仿真环境下, 粒子群规模取 20, 解空间为 2 维, 分别表示正规化参数 c 和径向基核宽度 σ 。

的取值范围为 $[0.1, 150]$, σ 的取值范围为 $[0.1, 10]$, 对应 (c, σ) 的速度最大值向量为 $(0.5, 0.1)$; 初始惯性权值 W 取 0.9, 最大迭代次数 T_{\max} 取 200, 加速因子 c_1 和 c_2 同时取 2, $(\sin \theta)^2$ 取 0.2。

通过交叉验证法, 得到 LS-SVM 的正规化参数 $c = 25.943$, 径向基核宽度 $\sigma = 4.3291$ 。通过 PSO 优化 LS-SVM 的模型, 可得到优化的正规化参数 $c = 92.3149$, 径向基核宽度 $\sigma = 0.1527$ 。再采用矢量基判据通过学习获得了 8 个支持向量, 进而通过 LS-SVM 算法得到成品油研究法辛烷值的预测模型。将此模型应用于成品油研究法辛烷值的在线估计。图 2、图 3、图 4 分别给出了 PSO-VB-LSSVM、PSO-LSSVM 及 LS-SVM 等估计模型的成品油研究法辛烷值的估计结果。表 1 为 3 种不同模型 RMSE 和 MAXE 的比较。其中 RMSE 表示均方差根, MAXE 表示最大绝对误差, 计算表达式为:

$$\text{RMSE} = \left(\frac{1}{n} \sum_{j=1}^n (z_j - \hat{y}_j)^2 \right)^{1/2} \quad (11)$$

$$\text{MAXE} = \max_{j=1}^n |z_j - \hat{y}_j| \quad (12)$$

式中: z_j 表示第 j 个预测样本的输出, n 表示预测样本的个数。

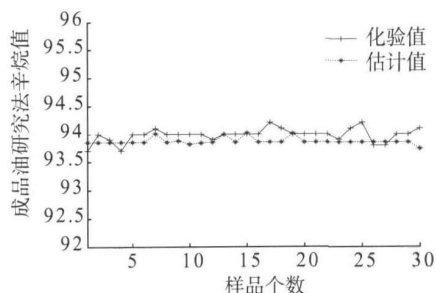


图 2 基于 PSO-VB-LSSVM 模型的成品油研究辛烷值的估计结果

Fig. 2 The estimated result of octane number based on PSO-VB-LSSVM

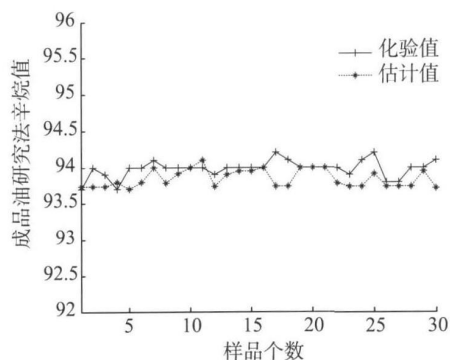


图 3 基于 PSO-LSSVM 模型的成品油研究辛烷值的估计结果

Fig. 3 The estimated result of octane number based on PSO-LSSVM

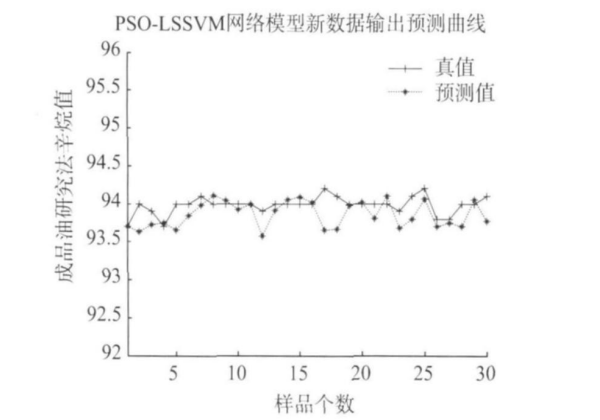


图 4 基于 LS-SVM模型的成品油研究辛烷值的估计结果
Fig 4 The estimated result of octane number based on LS-SVM

表 1 不同算法下的泛化结果比较 (应用实例)
Table 1 The generalized results under different algorithms

算法	RMSE(test)	MAXE(test)
PSO-VB-LSSVM	0.169 9	0.359 8
PSO-LSSVM	0.201 2	0.462 1
LS-SVM	0.214 6	0.550 5

由表 1和图 2、3、4的结果可以看出:

1)相对广义 LS-SVM建模方法, 基于粒子群优化的 LS-SVM具有更高的预测精度;

2)矢量基判据的引入, 使 LS-SVM的支持向量由 62个减少为 8个, 模型结构大为简化, LS-SVM的解具有了稀疏性, 因此有最佳的鲁棒性和辨识效果;

3)基于粒子群优化的 VB-LSSVM较其他两种方法具有更好的泛化能力和预测精度。

5 结 论

本文提出了基于粒子群优化的矢量基最小二乘支持向量机的建模方法, 并将其应用于成品油辛烷值的在线估计, 取得了较好的仿真估计结果。采用粒子群优化的方法对最小二乘支持向量机的参数进行选择, 避免了交叉验证法耗时与盲目的缺点; 采用矢量基判据准则选择支持向量, 使最小二乘支持向量机的解具有了稀疏性, 进而又提高了最小二乘支持向量机模型的泛化能力和预测的鲁棒性。因此三者的有机结合在非线性系统建模中有很大的应用潜力。

参考文献

[1] LIM X, LIN W M. The progress and advice to the production others in China. J. Natural Gas Chemical Indus. (C)1994-2019 China Academic Journal Electronic Publishing House. All rights reserved. <http://www.cnki.net>

1999 24(2): 53- 57

[2] FANG Y J, XIAO W D, LU W ZH. Methyl carbonate for the application of gasoline additive. J. Modern Chemical Industry 1998 (4): 20-22

[3] ZHU Y X. Cyclopentadiene manganese gasoline additive. J. Advances in fine Petrochemicals Abstracts 1999 13(4): 72-76

[4] VAPNIK V N. The nature of statistical learning theory. [M]. New York: Springer-Verlag 1995

[5] VAPNIK V N. The nature of statistical learning theory. [M]. 2nd ed. New York: Springer-Verlag 1999

[6] ZHANG X G. Introduction to statistical learning theory and support vector machines. J. Acta Automatic Sinica 2000 26(1): 34-42

[7] 单勇, 顾广华, 单海波. 基于支持向量机的彩色图像人脸检测. J. 仪器仪表学报, 2006 27(6): 772-773

SHAN Y, GU G H, SHAN H B. Face detection in color image based on support vector machines. J. Chinese Journal Scientific Instrument 2006 27(6): 772-773

[8] CHEN AI, J. SONG ZH 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

[9] ZHANG H M, WEI ZH N. A short-term load forecasting approach based on PSO support vector machine. J. Relay 2006 34(3): 28-31

[10] SHI Y H, EBERHARD R C. Parameter selection in particle swarm optimization. Q. Annual Conference on Evolutionary Programming San Diego 1998

作者简介



李炜, 教授, 主要从事工业过程先进控制、故障诊断与容错控制的理论及应用研究。
Li Wei is a professor and mainly engaged in the research of advanced industrial process control, fault diagnosis and fault tolerant control.

石连生, 于 2008年 6月获得兰州理工大学硕士学位, 研究方向为工业过程先进控制。
Email: shiliansheng2001@163.com

Shi Liansheng received master degree from Lanzhou University of Technology in June 2008 and main research direction is advanced industrial process control.

Email: shiliansheng2001@163.com