

文章编号: 2095-4980(2013)04-0641-05

基于 GA 优化 LSSVM 高炉铁水含硅量预报

王龙辉^a, 高嵩^b, 屈星^{b*}

(南华大学 a.经济管理学院; b.电气工程学院, 湖南 衡阳 421001)

摘要: 铁水含硅量是表征高炉生铁质量的重要指标, 也是反映高炉内部热状态的重要参数。为了提高铁水含硅量测量精确度, 保证高炉顺行, 提出一种基于最小二乘支持向量机(LSSVM)的铁水含硅量预报模型, 采用遗传算法(GA)确定模型参数的优化组合, 以改善模型性能。将某钢管厂高炉的实际运行数据经过预处理后作为模型的训练和测试样本, 进行模型预报实验, 并与神经网络模型和时间序列分析模型的预报性能进行了比较。基于 GA 优化参数的 LSSVM 模型对铁水含硅量预报的最大相对误差为 5.8%, 相关系数为 0.926 375, 预报精确度比直接 LSSVM 模型提高了 2.1%, 比前向神经网络模型提高了 4.3%。

关键词: 高炉; 铁水含硅量; 预报; 最小二乘支持向量机; 遗传算法

中图分类号: TN911.6

文献标识码: A

doi: 10.11805/TKYDA201304.0641

Prediction of blast furnace hot metal silicon content based on LSSVM optimized by GA

WANG Long-hui^a, GAO Song^b, QU Xing^b

(a.College of Economics and Management; b.College of Electrical Engineering, University of South China, Hengyang Hunan 421001, China)

Abstract: Hot metal silicon content has long been used as one of the most important indices to represent the hot metal quality and the thermal state of a blast furnace. In order to improve the measurement precision and attain stable operation of the blast furnace, a novel model for predicting silicon content by using Least Square Support Vector Machine(LSSVM) is presented. It adopts Genetic Algorithm(GA) to determine the optimum parameter set and therefore improves the model performance. By training and testing the operational data from blast furnace at a steel tube plant, the experimental results indicate that the proposed model can predict silicon content in hot metal with a maximum relative error of 5.8 % and correlation coefficient of 0.926 375, whose accuracy can be improved by 2.1% and 4.3% than that of the direct LSSVM and the feed-forward network with the same data set, respectively.

Key words: blast furnace; hot metal silicon content; prediction; Least Square Support Vector Machine; Genetic Algorithm

铁水含硅量是衡量高炉生产的生铁质量的重要指标, 也是反映高炉内部热状态的重要参数。准确测量和严格控制铁水含硅量对高炉稳定运行极其重要^[1]。由于生铁生产过程的复杂性和高炉内部的极端状态, 不能在线直接测量铁水含硅量, 因此建立相应的模型对铁水含硅量进行预报, 使高炉操作人员实时掌握铁水含硅量变化情况, 采取相应的操作以保证高炉顺行^[2]是十分重要的。

对高炉铁水含硅量预报建模问题已经进行了大量的研究。早期常用解析方法建立铁水含硅量预报模型^[2-4]。解析模型对于深入理解高炉炼铁过程动态特性发挥了较大的作用, 也能为高炉操作提供一定的参考, 但作用有限^[2]; 同时解析模型依赖许多参数值, 而这些参数值很难准确获得; 高炉内部高达 1 500°C 的极端状态也妨碍了模型所需数据的获取; 其他不确定因素, 如原料成分的变化, 高炉内部随机波动的影响, 模型输入参数相互间的影响等都会影响这些模型的预报精确度。另一种常用的铁水含硅量预报方法是时间序列分析法^[2,5]。这一方法利用高炉铁

收稿日期: 2013-02-28; 修回日期: 2013-04-17

基金项目: 湖南省科技厅科研资助项目(2012FJ4332)

*通信作者: 屈星 sukkqu@163.com

水含硅量的历史数据建立铁水含硅量预报模型, 与传统的解析模型相比, 预报精确度有所提高; 不足之处在于该分析法只利用了高炉铁水含硅量的历史数据, 因此所建立的模型不能反映很多与铁水含硅量相关的其他参数与铁水含硅量之间的复杂关系。

神经网络具有对复杂非线性关系建模的能力, 作为铁水含硅量建模的有力工具, 吸引了众多关注^[6-9]。A. Nurkkala 等人提出了一种高炉铁水含硅量前向神经网络预报模型^[7], 该模型采用非线性变量误差反传学习算法修改连接权值, 获得了比线性时间序列方法好得多的预报结果。H. Singh^[8]等人曾分别采用反向传播(Back Propagation, BP)网络、动态学习率网络、功能链路网络和模糊神经网络 4 种不同的神经网络模型, 使用同一组输入变量进行铁水含硅量预报。研究表明: 与传统的预报模型相比, 这 4 种神经网络模型均提高了铁水含硅量预报能力。

神经网络已成为高炉铁水含硅量预报的重要工具, 但是如何确定神经网络的结构以及相应神经网络的输入仍然是充满挑战性的问题。在文献[1]和[7]中, 提出用剪枝算法来确定神经网络的输入变量及其时延, 以及神经网络的连接关系。文献[9]提出一种优化 BP 神经网络高炉铁水含硅量预报模型, 采用遗传算法确定模型的输入变量, 优化神经网络的结构。

虽然神经网络在铁水含硅量预报建模中已经并将继续发挥重要的作用, 但它们也存在一些不足, 例如容易陷入局部极小而得不到最优解, 神经网络的结构也会影响预报质量。V. N. Vapnik^[10]提出的基于结构风险最小化原理的支持向量机(Support Vector Machine, SVM)具有很强的逼近任意非线性关系的能力和泛化能力, 已被广泛用于函数逼近、数据估计与预报等回归问题^[11]。在各种 SVM 算法中, 最小二乘支持向量机(LSSVM)^[12]用最小二乘线性系统将二次规划优化问题转化为线性方程组求解, 降低了 SVM 计算的复杂性。本文采用 LSSVM 来建立铁水含硅量预报模型, 并采用遗传算法(GA)优化模型的参数组合以提高预报精确度。

1 预报模型

1.1 LSSVM模型

对给定的 N 个训练样本 $\{(x_i, y_i): i=1, 2, \dots, N\}$, LSSVM 用下式函数对待测参数进行预报^[12]:

$$y_i = \varphi(x_i) + b (i=1, 2, \dots, N) \quad (1)$$

式中: x_i 为输入; y_i 为输出; $\varphi(x_i)$ 为核空间的映射函数; b 为偏差量。将输入数据映射到高维空间, 参数预报问题可以描述为求解下述优化问题:

$$\min(J, e) = \frac{1}{2} \omega^T \omega + \frac{C}{2} \sum_{i=1}^N e_i^2 \quad (2)$$

$$s.t. y_i = -\omega^T \varphi(x_i) + b + e_i (i=1, 2, \dots, N) \quad (3)$$

式中: ω 为权向量; $C > 0$ 为惩罚因子; $J = \omega^T \omega$; e_i 为预报误差, $e_i = y_i - \hat{y}_i$, \hat{y}_i 为预报模型输出。

由式(2)和式(3), 导出最小值问题的拉格朗日函数为:

$$L(\omega, b, e, \alpha) = \frac{1}{2} \omega^T \omega + \frac{C}{2} \sum_{i=1}^N e_i^2 - \sum_{i=1}^N \alpha_i [\omega^T \varphi(x_i) + b + e_i - y_i] \quad (4)$$

式中 α_i 为拉格朗日乘子。对式中 ω, b, e, α 求偏导且令其等于 0, 消去 ω 和 e , 得到线性方程组:

$$\begin{bmatrix} 0 & 1 & \cdots & 1 \\ 1 & k(x_1)(x_1) + r^{-1} & \cdots & k(x_1)(x_N) \\ & & \ddots & \\ 1 & k(x_N)(x_1) & \cdots & k(x_N)(x_N) + r^{-1} \end{bmatrix} \begin{bmatrix} b \\ \alpha_1 \\ \vdots \\ \alpha_N \end{bmatrix} = \begin{bmatrix} 0 \\ y_1 \\ \vdots \\ y_N \end{bmatrix} \quad (5)$$

由式(5)解出 α_i 和 b , 最后得到所求数据的 LSSVM 预报模型:

$$y(x) = \sum_{i=1}^N \alpha_i K(x \cdot x_i) + b \quad (6)$$

式中函数 $K(x_i, x_j) = \varphi(x_i)^T \varphi(x_j)$ 是满足 mercer 条件的核函数。本文核函数取径向基函数, 函数形式为:

$$k(x_i, x_j) = \exp(-\|x - x_i\|^2 / 2\sigma^2) \quad (7)$$

式中 σ 为径向基函数的宽度。

1.2 模型参数优化

惩罚因子 C 和核函数的参数 σ 是影响 LSSVM 模型性能的主要因素, 对这 2 个参数的选取目前尚无系统的方

法,多凭经验确定,难以实现参数的优化组合。遗传算法是一种随机优化算法,具有极强的全局搜索能力,遗传算法从在解空间中随机产生初始种群开始搜索,通过反复迭代进化,能够在解空间快速寻求全局优化解。本文采用遗传算法搜索 LSSVM 模型的参数 C 和 σ^2 的优化组合。算法中采用二进制编码表示参数 C 和 σ^2 ,位串长度为 19,确定初始种群后,选择测试集数据的均方误差 MSE 的倒数作为遗传算法的适应度函数:

$$f(X) = \frac{1}{MSE} = \frac{1}{mse(T-\hat{T})} = \frac{1}{\sum_{i=1}^n (\hat{t}_i - t_i)^2} \tag{8}$$

式中: $\hat{T} = \{\hat{t}_1, \hat{t}_2, \dots, \hat{t}_n\}$ 为测试样本的预报值; $T = \{t_1, t_2, \dots, t_n\}$ 为测试样本的实际值; n 为测试样本的样本数目。

初始种群确定后,经过一次次迭代进化,当满足迭代终止条件时,即可获得参数 C 和 σ^2 的优化组合。

2 实验仿真

2.1 数据集和数据预处理

实验所用数据来自某钢管公司炼铁厂高炉数据库中某年 4 月 14 日至 5 月 24 日的运行记录(其中剔除了 5 月 10 日至 11 日检修的 16 h 的数据),这些数据中包含了 35 个可用于铁水含硅量预报的参数。预报模型输入参数的选择对预报结果影响很大,如果选择的输入参数太少,一些与铁水含硅量关联度高的参数未被选择,会降低铁水含硅量预报的准确率;如果选择的参数中含有与铁水含硅量关联度不高的参数,会大大增加计算工作量,延长计算时间,从而导致收敛时间过长,可能还会降低预报命中率。根据对高炉工艺流程的分析以及操作人员的经验,在 35 个参数中选择了与铁水含硅量关联度高的 16 个参数作为模型的输入参数,这些参数包括:小时料速、富氧含量、风量、风温、压差、热风压力、顶压、顶温、透气性指数、铁水温度、每炉出铁中的含硫量、碱度 R_2 、铁量差、煤气利用率、综合负荷和前一炉的铁水含硅量。

高炉数据库中的数据由于测量误差或系统故障等原因,常常存在不精确、不完整和重复的数据,不能直接用于铁水含硅量预报,因此要对样本数据进行预处理。本文主要采取下述处理方法:如果某个参数存在缺失值,采用其之前时刻的已知值代替;对严重超出测量范围的异常数据,采用 K-近邻线性回归算法对该时刻点的参数值进行估计,用估计值代替严重超出范围的异常数据;对所有输入输出数据都进行归一化处理,使所有输入输出参数的取值均在 $[0,1]$ 区间。经过预处理后得到 479 组生产数据,按顺序取前面 379 组数据作为高炉铁水含硅量预报模型的训练样本,后面的 100 组数据作为测试样本。

2.2 实验结果

实验中首先用遗传算法对模型参数进行优化,用 g 表示式(7)中的 $1/\sigma^2$,得到 LSSVM 的优化参数值,如图 1 和图 2 所示。图 1 和图 2 分别为 GA 优化 LSSVM 铁水含硅量预报模型参数 C 和 σ^2 的等高线图和 3-D 图。然后用经过预处理的样本数据对模型进行训练和测试,用均方误差 MSE 和相关系数 R 来评估模型的预报性能。图 3(a)

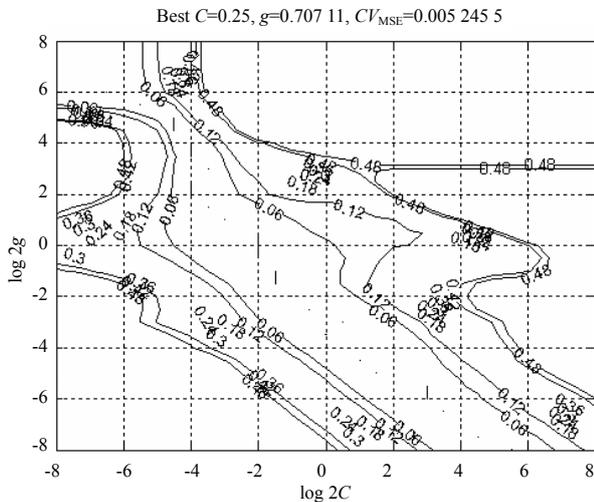


Fig.1 Contour map for optimization of LSSVM parameters with GA
图 1 GA 优化 LSSVM 模型参数等高线图

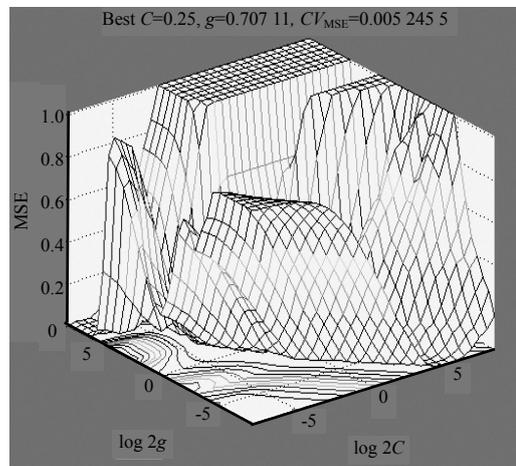


Fig.2 3-D chart for optimization of LSSVM parameters with GA
图 2 GA 优化 LSSVM 模型参数 3-D 图

给出了 GA 优化 LSSVM 模型对高炉铁水含硅量的预报值与实际值的比较情况,从图中可以看出预报值与实际值接近。图 3(b)是 GA 优化 LSSVM 模型的预报误差图。作为比较,图 4(a)给出了参数未经优化的直接 LSSVM 模型对高炉铁水含硅量的预报输出,图 4(b)是直接 LSSVM 模型的预报误差图。图 5(a)给出了采用相同输入参数和样本数据的基于 BP 神经网络的铁水含硅量模型的预报结果,图 5(b)为 BP 神经网络模型的预报误差图。所用神经网络都含有一个有 15 个节点的隐含层。

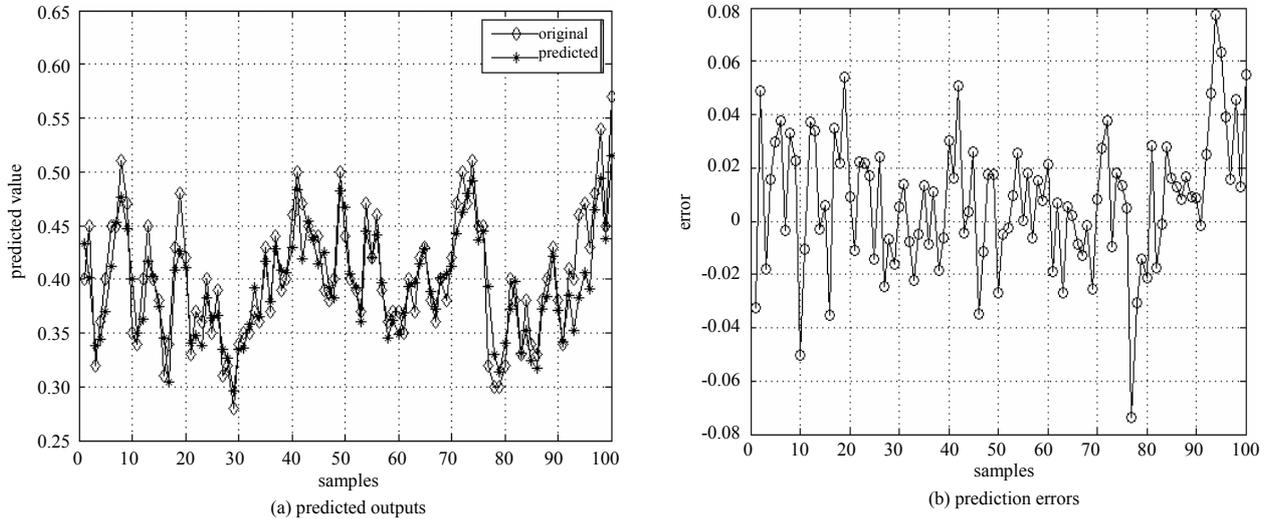


Fig.3 Prediction results of LSSVM model optimized by GA
图 3 GA 优化 LSSVM 模型铁水含硅量预报结果

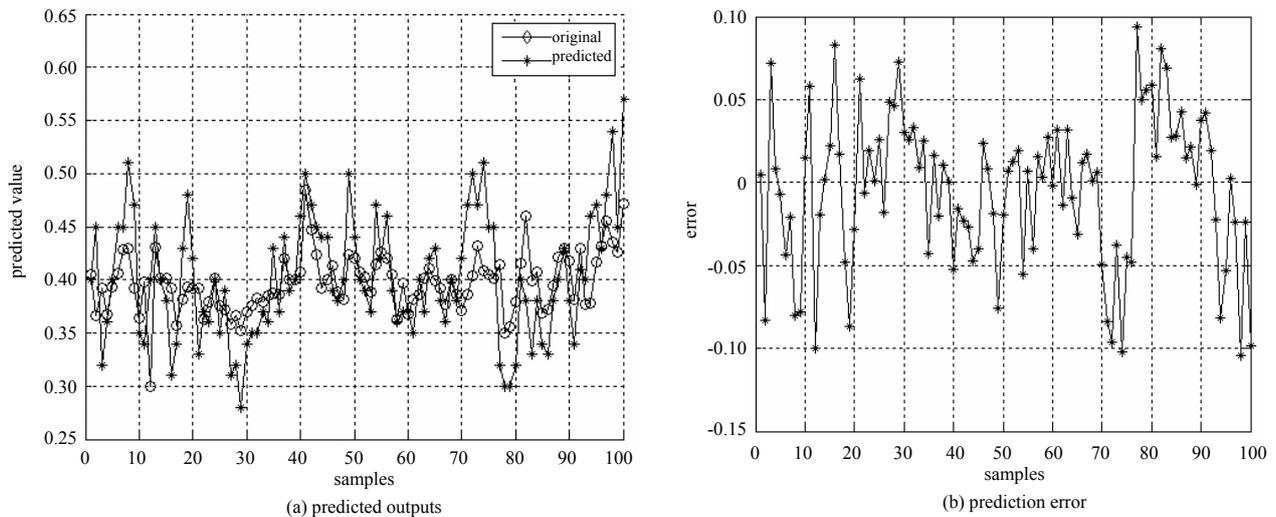


Fig.4 Prediction results of direct LSSVM model
图 4 直接 LSSVM 模型铁水含硅量预报结果

仿真得到基于 GA 优化参数的 LSSVM 模型的铁水含硅量预报最大相对误差为 5.8%, 均方误差 $MSE=0.000\ 248\ 049$, 相关系数 $R=0.926\ 375$; 基于直接 LSSVM 模型的铁水含硅量预报最大相对误差为 7.9%, 均方误差 $MSE=0.000\ 675$, 相关系数 $R=0.821\ 735$; 基于前向神经网络模型的最大相对预报误差为 10.1%, 均方误差 $MSE=0.000\ 964\ 7$, 相关系数 $R=0.796\ 375$ 。采用 GA 优化参数的 LSSVM 模型的预报精确度比直接 LSSVM 模型提高了 2.1%, 比前向神经网络模型高出 4.3%。

3 结论

利用 LSSVM 逼近高炉铁水含硅量和相关参数间的复杂非线性关系,对铁水含硅量进行预报。采用遗传算法对模型参数组合进行寻优,以高炉实际运行数据经过预处理后作为样本对模型进行训练和测试。实验结果表明,与直接 LSSVM 模型和神经网络模型相比,本文提出的预报模型具有更高的预报精确度,对优化高炉操作有重要的参考价值。

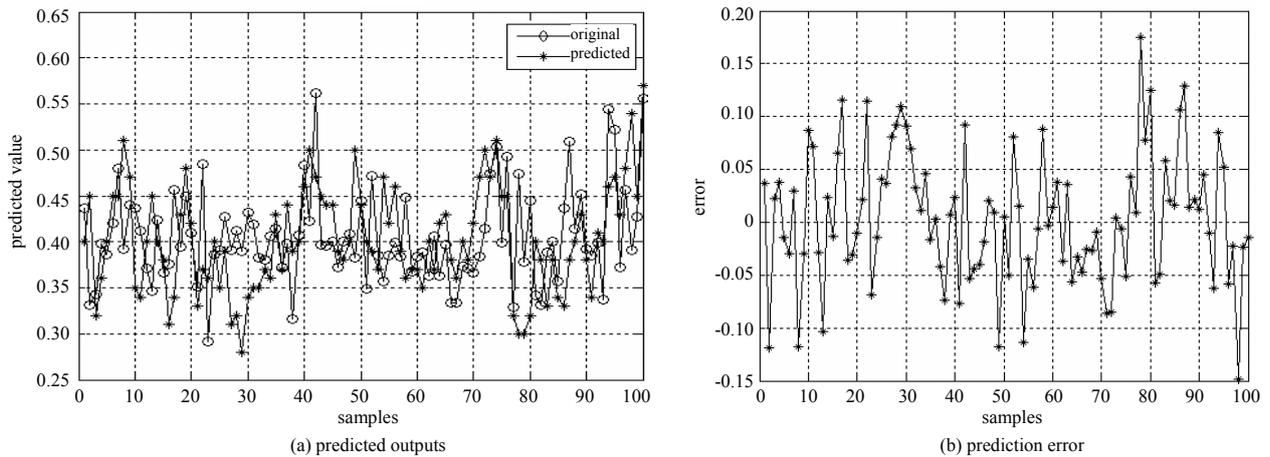


Fig.5 Prediction results of BP Artificial Neural Network model

图5 BP神经网络模型铁水含硅量预报结果

参考文献:

- [1] Saxén H, Pettersson F. Nonlinear prediction of the hot metal silicon content in the blast furnace[J]. ISIJ International, 2007,47(12):1732-1737.
- [2] Ueda S, Natsui S, Nogami H, et al. Recent progress and future perspective on mathematical modeling of Blast Furnace[J]. ISIJ International, 2010,50(7):914-923.
- [3] Nogami H, Chu M, Yagi J. Multi-dimensional transient mathematical simulator of blast furnace process based on multi-fluid and kinetic theories[J]. Computer and Chemical Engineering, 2005,29(11-12):2438-2448.
- [4] LIU Xiangguan, ZENG Jiusun, ZHAO Min. Mathematical model and its hybrid dynamic mechanism in intelligent control of ironmaking[J]. Journal of Iron and Steel Research, 2007,14(1):7-11.
- [5] Saxén H, Pettersson F. Evolving nonlinear time-series models of the hot metal silicon content in the blast furnace[J]. Materials and Manufacturing Processes, 2007,22(5):577-584.
- [6] 邱东, 仝彩霞, 祁晓钰, 等. 基于神经网络的高炉铁水硅含量预报模型的研究[J]. 冶金分析, 2009,29(2):49-52. (QIU Dong, TONG Caixia, QI Xiaoyu, et al. Study on prediction model of silicon content in blast furnace liquid iron based on neural network[J]. Metallurgical Analysis, 2009,29(2):49-52.)
- [7] Nurkkala A, Pettersson F, Saxén H. Nonlinear modeling method applied to prediction of hot metal silicon in the ironmaking blast furnace[J]. Ind. Eng. Chem. Res., 2011,50(15):9236-9248.
- [8] Singh H, Sridhar N V, Deo B. Artificial neural nets for prediction of silicon content of blast furnace hot metal[J]. Steel Research, 1996,67(12):521-527.
- [9] 张军红, 金永龙, 沈峰满, 等. 应用优化BP神经网络建立铁水硅含量的预测模型[J]. 钢铁研究学报, 2007,19(11):60-63. (ZHANG Junhong, JIN Yonglong, SHEN Fengman, et al. Prediction Model of Silicon Content in Hot Metal Using Optimized BP Network[J]. Journal of Iron and Steel Research, 2007,19(11):60-63.)
- [10] Vapnik V N. Statistical Learning Theory[M]. New York: Wiley-Interscience, 1998.
- [11] 尚伟, 梁玉英, 朱赛. 具有信任度输出的雷达信号SVM回归分析模型[J]. 信息与电子工程, 2012,10(2):192-195. (SHANG Wei, LIANG Yuying, ZHU Sai. Regression analysis model for radar signal based on SVM with credibility evaluation[J]. Information and Electronic Engineering, 2012,10(2):192-195.)
- [12] Suykens J A K, Vandewalle J. Recurrent least squares support vector machines[J]. Circuits and Systems I: IEEE Transactions on Fundamental Theory and Applications, 2000,47(1):1109-1114.

作者简介:



王龙辉(1962-), 男, 长沙市人, 学士, 讲师, 主要研究方向为数据处理. email: wlhqj@sina.com

高嵩(1958-), 男, 湖南省衡阳市人, 学士, 高级工程师, 主要研究方向为嵌入式技术.

屈星(1979-), 男, 湖南省邵阳市人, 硕士, 实验师, 主要研究方向为FPGA与集成电路设计.