

文章编号: 1001-0920(2011)05-0707-05

基于智能融合策略的钴离子浓度预测模型

晏密英, 桂卫华, 阳春华

(中南大学 信息科学与工程学院, 长沙 410083)

摘要: 针对湿法炼锌净化过程中钴离子浓度难以准确检测的问题, 提出一种基于智能融合策略的钴离子浓度组合预测模型. 首先从提高模型预测精度的角度出发, 考虑不同核函数对预测性能的影响, 分别建立两个在线支持向量回归子模型, 并采用改进粒子群优化算法进行子模型参数寻优; 然后通过熵值法智能融合策略建立组合预测模型. 仿真实验表明, 组合模型具有良好的预测性能, 预测效果能满足硫酸锌溶液净化过程中对钴离子浓度值的误差要求.

关键词: 钴离子浓度; 在线支持向量回归; 改进粒子群优化算法; 熵值法; 组合预测模型

中图分类号: TP273

文献标识码: A

Prediction model of cobalt ion concentration based on intelligent fusion strategy

YAN Mi-ying, GUI Wei-hua, YANG Chun-hua

(School of Information Science and Engineering, Central South University, Changsha 410083, China.

Correspondent: YAN Mi-ying, E-mail: ymy137@163.com)

Abstract: To deal with the problem of detecting the cobalt ion concentration in the purification process of zinc hydrometallurgy, a cobalt ion concentration combination prediction model based on intelligent fusion strategy is proposed. In order to improve the prediction accuracy, two online support vector sub-models are established by taking into account the impact of different kernel functions on the prediction performance. And an improved particle swarm optimization algorithm is applied to the optimization of parameters in these models. Then the combination prediction model is established through the intelligent fusion strategy of entropy method. Simulation results show that, the combination model has good prediction performance which satisfies the error request of the cobalt ion concentration in the zinc sulfate solution purification process.

Key words: cobalt ion concentration; online support vector regression; improved particle swarm optimization algorithm; entropy method; combination prediction model

1 引言

锌作为一种重要的有色金属原材料, 在国民经济中有着广泛的应用. 湿法炼锌目前已成为锌冶炼的主导发展方向, 其过程主要由焙烧、浸出、净化、电解和熔铸5个工序组成^[1]. 其中, 净化过程是湿法炼锌中的一个重要环节, 其主要目的是除去中性上清液中的各种杂质离子, 为电解提供合格的新液. 在中性上清液中, 主要存在钴、镉、镍、铜等离子, 钴离子对电解的影响很大, 不仅影响电解的效率, 而且超标严重时还可能导致电解烧板, 影响产品的质量. 因此, 中性上清液中钴离子浓度是影响电解效率与产品质量的一个关键参数.

目前国内多数湿法炼锌企业自动化水平不高, 在

净化阶段, 对钴离子浓度的检测基本都是采用人工化验方法, 检测的滞后使得操作工不能及时获得钴离子浓度的变化情况, 失去了对过程控制的最佳时间. 操作工通常依据经验采用过量添加锌粉的方法来保证生产出合格的新液, 但这样却增加了企业的生产成本. 因此, 实现净化过程中钴离子浓度的实时在线检测, 对后续优化控制锌粉添加量、实现节能降耗的目标具有重要的意义.

为解决复杂工业过程中的预测建模问题, 众多学者进行了大量的研究工作^[2-4]. 这些工作通过深入的机理分析建立工业过程混合机理模型, 为后续优化控制提供了坚实的基础, 对于稳定生产、提高产品产量与质量发挥了重要作用. 然而, 由于净化过程反应机

收稿日期: 2010-01-28; 修回日期: 2010-04-12.

基金项目: 国家自然科学基金重点项目(60634020).

作者简介: 晏密英(1972-), 女, 副教授, 博士生, 从事工业过程预测与优化控制的研究; 桂卫华(1950-), 男, 教授, 博士生导师, 从事复杂工业过程建模与优化、大系统理论与应用等研究.

理复杂,导致净化过程的机理建模极其困难.随着生产企业基础自动化水平的大幅提升,净化过程中积累了大量工业运行数据,为净化过程的钴离子浓度预测模型的建立提供了有利条件.

针对湿法炼锌净化过程中钴离子浓度的预测建模问题,文献[5]提出了一种基于神经网络补偿灰色预测误差的钴离子浓度预测模型,并通过仿真和生产实际验证了该方法的有效性.受湿法炼锌净化过程时变特性的影响,该模型的预测精度会随时间变差,必须每过一定时间利用新采集的数据对模型参数重新离线训练,以使模型具有一定的自适应能力.但这种自适应能力是被动的,不能实时调整.

由 Vapnik 创立的支持向量机(SVM)^[6]建立在统计学习理论和结构风险最小的基础上,在理论上充分保证了模型的泛化能力,目前已广泛应用于复杂工业过程预测建模等领域^[7-8].在考虑模型动态特性的基础上,文献[9]提出了一种基于在线支持向量回归的钴离子浓度预测模型,仿真实验表明,该模型具有较好的泛化能力和较小的时间复杂度,能满足不同工况下对预测模型的要求.

本文以文献[9]中的在线支持向量回归模型为基础,从提高模型预测精度的角度出发,通过考虑不同核函数对预测性能的影响,分别建立两个在线支持向量回归子模型,并采用改进的粒子群优化算法(PSO)进行子模型的参数寻优;其次通过熵值法的智能融合策略进行子模型的集成;最后对所建立的组合预测模型进行验证,以说明方法的有效性.

2 湿法炼锌净化过程机理分析

某锌冶炼厂的湿法炼锌过程是将硫化锌精矿等原料经过焙烧得到氧化锌等产物,将其用稀硫酸浸出,经固液分离后,再对含有多种杂质的硫酸锌中性上清液净化,除去各种杂质.最后,对新液进行电解,使锌从溶液中沉积出来,电锌再熔铸成锌锭.

其中硫酸锌溶液净化过程,一般采用锌粉铈盐3段连续净化工艺,如图1所示.经过3段连续净化后,硫酸锌溶液杂质含量会降低到生产标准之内,再直接送往电解工序.其中第II段净化除钴过程是整个硫酸锌溶液净化过程中一个关键的环节,钴离子含量成为第II段净化过程中一个重要的技术指标.

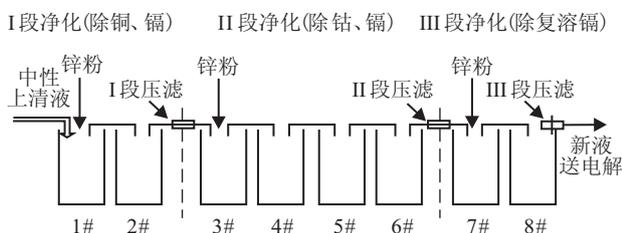


图1 铈盐锌粉3段净化工艺流程图

来自浸出过程的含有硫酸锌的中性上清液进入净化I段,主要除去杂质铅和铜离子.在I段净化过程,离子浓度波动大,工况不稳定.为了保证后续过程的正常生产,通常锌粉量都是设定成最大值来添加的.经过I段净化后,溶液中的杂质离子浓度一般都能达到相对稳定的状态.

I段出口净液经过压滤和加热后进入II段净化过程,添加锌粉和催化剂铈盐,沉淀除去钴离子,并继续沉淀铅离子.该过程中溶液的离子浓度相对较低,有利于控制锌粉的下料量,操作工可以根据II段入口和出口的每小时化验的钴和铅离子浓度来不断调整锌粉的下料量.操作工为保证正常生产,往往过量添加锌粉.

II段出口净液经过压滤后进入III段净化过程,在溶液的自然温度下,添加少量锌粉除去复溶的杂质铅离子.实际上,III段净化过程溶液中的钴和铅离子浓度都已经很低,基本达到了生产要求.III段出口新液经过压滤和冷却后直接送电解工序.

根据对湿法炼锌净化生产过程的调研,影响II段出口离子浓度并可实时检测的主要因素变量有:净化反应槽的入口溶液温度 x_{i1} ;净化反应槽的入口溶液流量 x_{i2} ;II段净化铈盐流量 x_{i3} ;II段净化过程中的锌粉添加量,间接表现为和锌粉下料装置相连的变频器的调整频率 x_{i4} ;II段入口溶液的钴离子浓度 x_{i5} 和铅离子浓度 x_{i6} .其中, x_{i1}, \dots, x_{i6} 表示在*i*时刻相关过程参数的检测值,并且II段净化过程中的铈盐流量 x_{i3} 是与入口溶液流量 x_{i2} 成连锁比例控制的.选取以上变量在现场每小时测得的实际数据,作为II段净化过程钴离子浓度预测模型的输入变量进行建模研究.

3 基于不同核函数的改进在线SVM模型

SVM的基本思想是:对于给定的样本数据集,在高维特征空间中构造一个回归函数,其中的权系数向量和阈值可通过求解凸二次规划问题得到.针对优化问题中的目标函数,构造拉格朗日函数,把原始的凸二次规划问题转化为对偶问题,并引入核函数 $K(x_i, x_j)$,避免使用非线性映射.最终可得SVM回归函数为

$$f(x) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) K(x_i, x) + b. \quad (1)$$

其中: $\alpha_i \geq 0, \alpha_i^* \geq 0$ 为对偶问题的解, b 为阈值, n 为样本数据的组数.

根据KKT条件公式中边界函数 $h(x_i)$ 取值的不同,可将全部训练数据分为误差向量集、边界支持向量集和剩余样本向量集3个不同子集合.在线支持向量回归算法的目的就是把新样本添加到其中的一个

集合内,同时把冗余样本从其中删除掉,与此同时保持其他所有样本不变。

在文献[9]中,当数据集增加一个新样本时,递增算法就会更新原支持向量回归函数模型。其主要思想是:在有限的步数内,改变对应于新样本的系数偏差,使其满足KKT条件,同时保证训练样本集合中的现有样本在每一步更新过程中依然满足KKT条件。为保持一定的边界支持向量样本规模,在增加一个新边界支持向量样本的同时必然会删除一个旧边界支持向量样本,删除旧样本的前提是它不会对支持向量回归模型造成任何影响。经过增加新样本和删除旧样本的算法步骤后,便可得到一组更新后的支持向量回归模型参数。

当一个新增样本被划分为边界支持向量,并对原回归模型的性质产生影响时,必须更新矩阵 K_s 。同时由于硫酸锌溶液净化生产过程中存在大量的数据,将会出现复杂的矩阵运算。文献[9]中设计了一种分块矩阵的方法对更新矩阵 K_s 进行迭代更新,降低了计算过程的时间复杂性。

该在线支持向量回归预测模型中,核函数的类型和参数、正则化参数 C 和不敏感系数 ε 是影响支持向量机建模效果的主要因素。

将低维样本空间映射到高维特征空间是SVM建模的关键。核函数的类型和参数隐式确定了样本空间到特征空间的映射。其中,核函数的类型决定空间映射的方式,在核函数类型确定的情况下,随着核参数的变化,特征空间的复杂度以一定规律变化。特征空间复杂度过小时,模型精度将变差;反之,模型泛化能力将变差。目前采用较多的核函数主要有多项式核函数和高斯径向基核函数两种,其表达式分别为

$$K(x_i, x) = (x * x_i + 1)^d, \quad (2)$$

$$K(x_i, x) = \exp \left\{ -\frac{\|x - x_i\|^2}{\sigma^2} \right\}. \quad (3)$$

其中:多项式核函数属于全局核函数,具有全局特性,允许相距很远的点都可以对核函数的值有影响;高斯径向基核函数属于局部核函数,具有局部性,只允许相距很近的数据点对核函数的值有影响。

鉴于湿法炼锌净化过程本身的强非线性特点,本文分别采用多项式核函数和高斯径向基核函数建立在线SVM预测子模型。多项式核函数参数主要由 d 决定,当 $d=1$ 时退化为一个线性函数,随着 d 值的增大多项式的空间复杂度成指数增长,当 d 很大时,核函数的值会趋近于无穷,这给数据的处理带来很大的困难;高斯径向基核函数只含有一个参数 σ ,参数 σ 控制着核函数的作用半径,是有界的局部函数。

不敏感参数 ε 和正则化参数 C 以不同的方式控制着模型的训练精度和结构复杂度。不敏感参数 ε 控制不敏感带的宽度,影响支持向量机的稀疏性,一定程度上反映了对模型误差的容忍度。减小 ε 值,将增加支持向量个数,导致目标函数中训练误差的比重增大,使得模型训练误差减小但泛化能力降低;反之,将减小支持向量个数,导致模型训练误差增大但泛化能力提高。正则化参数 C 则直接控制目标函数中的模型复杂度和训练误差的权重比例。增加 C 值,将增加对超出 ε 不敏感带的预报误差的惩罚力度,从而增加目标函数中训练误差的比重,使得模型训练误差减小但泛化能力降低;反之,将导致模型训练误差增大但泛化能力提高。

为了获得最佳预测性能的在线SVM子模型,需要分别求得两个在线SVM子模型中的 d, ε, C 和 σ, ε, C 参数值。鉴于粒子群进化算法具有隐含的并行性和强大全局搜索能力,本文采用文献[10]中提出的改进PSO算法进行在线SVM子模型参数优化。

基于在线SVM的钴离子浓度预测建模过程中,选取某锌冶炼厂的湿法炼锌净化过程的实际生产数据,采用 3σ 准则对异常数据进行处理,并对处理后的数据进行归一化变换,从而获得建模数据集 $\{x_i, y_i\} \in R^l \times R, i=1, 2, \dots, n, x_i$ 为 $l=6$ 维样本输入量(辅助变量), y_i 为样本输出值(主导变量)。数据集的大小为 $n=400$,其中280组数据用于模型训练,120组数据用于模型测试。

由于本文选取净化过程钴离子浓度预测模型的输入数据为6维,根据经验公式^[11],多项式的次数 d 一般为 $[2, 9]$ 。 σ 的范围在 $[0.1\sqrt{l}, \sqrt{l}]$, l 为输入数据的维数,则 $\sigma \in [0.2, 2.5]$ 。不敏感参数 ε 可由经验公式 $\varepsilon = 3\sigma_m \sqrt{\frac{\ln m}{m}}$ 得到, m 是训练集样本大小, σ_m 是噪声的标准离差, $\sigma_m = \frac{m^{0.2}k}{m^{0.2}k-1} \cdot \frac{1}{m} \sum_{i=1}^m (\hat{y}_i - y_i)^2$, k 的取值范围一般在 $[2, 6]$ 之间。结合净化过程钴离子浓度预测的实际情况,计得 $\varepsilon \in [0.01, 2]$ 。正则化参数 C 根据经验设置在 $C \in [10^{-1.5}, 10^{2.5}]$,则正则化参数 C 的范围为 $[0.05, 400]$ 。基于改进PSO算法的在线SVM预测子模型的训练过程如下:

Step 1: 改进PSO算法与在线SVM子模型的初始化。改进PSO算法的参数 $c_1 = c_2 = 2.5, \alpha = 0.9, a = 3.5, b = 0.35, \eta = 0.99, \xi \in [-5, 5]$,种群数为30,迭代次数为100。参数 d, ε, C 和 σ, ε, C 都采用实数编码,把 d, ε, C 和 σ, ε, C 值分别映射为两个种群的每个粒子。每个粒子的适应度函数定义为 $F_i = \frac{1}{m} \sum_{i=1}^m (\hat{y}_{ij} - y_i)$,其中 \hat{y}_{ij} 为第 j 个粒子的在现SVM

训练输出值, y_i 为对应数据集中的实际检测值, $i = m = 220$ 表示训练样本集的大小。

Step 2: 按照文献 [10] 中的速度和位置更新公式更新粒子的速度和位置, 生成新一代种群。

Step 3: 根据适应度函数计算每个粒子的适应度值, 该值越小, 则结果越好。

Step 4: 检查终止条件, 若满足, 则将全局最优粒子映射为对应的 d, ε, C 或 σ, ε, C 值, 以此结果作为在线 SVM 子模型的参数; 否则转入 Step 2。本文的终止条件为寻优达到最大进化代数或是适应度值小于给定的精度。

经过上述训练后得到最优参数值为 $d = 3, \varepsilon = 0.03, C = 17.82$ 和 $\sigma = 1.76, \varepsilon = 0.06, C = 26.18$ 。将此模型应用于净化过程钴离子浓度的在线预测, 图 2 给出了采用多项式核函数 SVM 预测模型和高斯径向基核函数 SVM 预测模型的预测结果。其中: 实线为实际检测值, 虚线为多项式 SVM 模型, 点划线为径向基 SVM 模型。表 1 为两个 SVM 子模型 MSE 和 MAPE 值的比较, 其中 MSE 表示均方误差, MAPE 表示平均绝对百分比误差, 计算表达式为

$$\text{MSE} = \sqrt{\sum_{i=1}^m (\hat{y}_i - y_i)^2 / m}, \quad (4)$$

$$\text{MAPE} = \frac{1}{m} \sum_{i=1}^m \left| \frac{\hat{y}_i - y_i}{y_i} \right|. \quad (5)$$

其中: \hat{y}_i 是净化过程钴离子浓度的预测值, y_i 是净化过程钴离子浓度的实际检测值, m 是测试数据的个数。

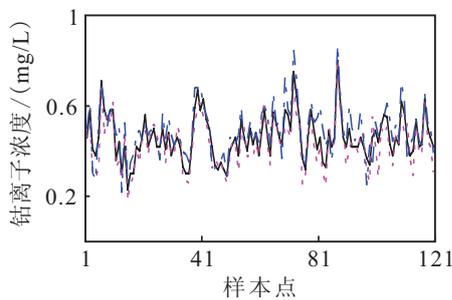


图 2 多项式模型和径向基模型的预测曲线

表 1 多项式模型和径向基模型的预测结果对比

评价指标	多项式模型	径向基模型
MSE	0.1269	0.1029
MAPE	0.1491	0.1181

4 基于熵值法的组合预测模型

为使预测结果尽可能接近钴离子浓度的实际值, 基于集成建模思想^[11]将上述两个单一 SVM 子模型进行加权集成, 建立净化过程钴离子浓度组合预测模

型; 采用熵值法^[12]确定最优加权系数, 综合两个子模型的输出作为最终的预测结果。组合预测模型的结构框图如图 3 所示。

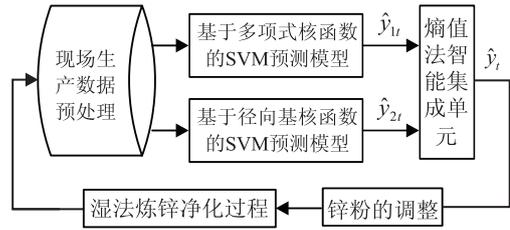


图 3 净化过程钴离子浓度组合预测模型

图 3 中, \hat{y}_{1t} 和 \hat{y}_{2t} 分别表示基于多项式核函数的 SVM 预测模型和基于高斯径向基核函数的 SVM 预测模型在第 t 时刻的钴离子浓度预测值, \hat{y}_t 是集成预测模型在第 t 时刻的钴离子浓度预测值。设 e_{it} 为第 i 个预测模型在第 t 时刻的预测相对误差, 按下式计算得到:

$$e_{it} = \begin{cases} |(y_t - \hat{y}_{it})/y_t|, & 0 \leq |(y_t - \hat{y}_{it})/y_t| < 1; \\ 1, & |(y_t - \hat{y}_{it})/y_t| \geq 1. \end{cases} \quad (6)$$

其中: y_t 表示钴离子浓度实际检测值; $i = 1, 2, t = 1, 2, \dots, m$ 。利用熵值法确定加权系数的步骤如下:

Step 1: 计算第 i 个单一预测模型在第 t 时刻的预测相对误差的比重

$$p_{it} = e_{it} / \sum_{t=1}^n e_{it}. \quad (7)$$

Step 2: 计算第 i 个单一预测模型预测相对误差的熵值

$$E_i = -\frac{1}{\ln n} \sum_{t=1}^n p_{it} \ln p_{it}. \quad (8)$$

Step 3: 计算第 i 个单一预测模型预测相对误差序列的变异程度系数

$$d_i = 1 - E_i. \quad (9)$$

Step 4: 计算各单一预测模型加权系数

$$w_i = \frac{1}{m-1} \left(1 - d_i / \sum_{i=1}^m d_i \right). \quad (10)$$

通过上述步骤计算得到权值 $w_1 = 0.468, w_2 = 0.532$, 按照 $\hat{y}_t = w_1 \hat{y}_{1t} + w_2 \hat{y}_{2t}$ 可以获得组合预测模型的钴离子浓度输出值。图 4 为组合预测模型的预测

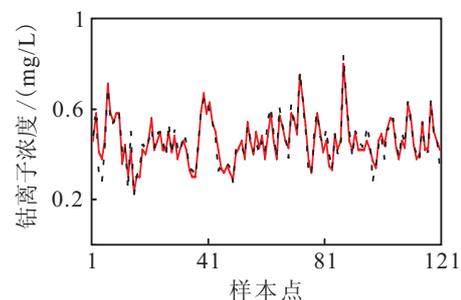


图 4 组合预测模型的预测曲线

结果, 通过计算组合得到预测模型的MSE和MAPE值分别为0.0595和0.0804.

由图2, 图4及组合预测模型的MSE和MAPE指标可知, 得到的组合预测模型具有较高的预测精度. 与基于多项式核函数的SVM预测模型和基于高斯径向基核函数的SVM预测模型相比, 组合预测模型的MSE指标分别提高了0.0674%和0.0434%, MAPE指标分别提高了0.0687%和0.0377%, 说明组合预测模型的预测精度和性能要高于单一子预测模型. 本文提出的组合预测模型具有优良的建模性能, 适合于净化过程钴离子浓度的预测, 预测效果能满足硫酸锌溶液净化过程中对钴离子浓度值的误差要求.

5 结 论

本文针对湿法炼锌II段净化过程出口钴离子浓度难以准确预测的问题, 根据净化过程中的大量生产数据, 通过现场调研确定影响钴离子浓度的主要影响因素, 对相关生产过程数据进行预处理后, 建立II段净化过程出口钴离子浓度的在线SVM子模型, 通过熵值法建立组合预测模型. 仿真分析说明组合模型具有良好的预测性能, 能及时为生产操作提供参考, 并为下一步指导锌粉添加量的优化控制奠定了基础.

参考文献(References)

- [1] 杨大锦, 朱华山, 陈加希. 湿法提锌工艺与技术[M]. 北京: 冶金工业出版社, 2006.
(Yang D J, Zhu H S, Chen J X. Zinc hydrometallurgy process and technology[M]. Beijing: Metallurgy Industry Press, 2006.)
- [2] 杨辉, 谭明皓, 柴天佑. 基于神经网络的多元稀土萃取组分含量的软测量[J]. 中国稀土学报, 2003, 21(4): 425-430.
(Yang H, Tan M H, Chai T Y. Neural networks based component content soft-sensor in countercurrent rare-earth extraction[J]. J of Chinese Rare Earth Society, 2003, 21(4): 425-430.)
- [3] 贾润达, 毛志忠, 常玉清, 等. 钴湿法冶炼萃取过程中的组分含量软测量[J]. 控制与决策, 2009, 24(4): 632-646.
(Jia R D, Mao Z Z, Chang Y Q, et al. Soft sensing of component content in cobalt hydrometallurgy extraction process[J]. Control and Decision, 2009, 24(4): 632-646.)
- [4] 王雅琳, 桂卫华, 阳春华, 等. 基于有限信息的铜吹炼动态过程智能集成建模[J]. 控制理论与应用, 2009, 26(8): 860-866.
(Wang Y L, Gui W H, Yang C H, et al. Intelligent integrated modeling for the dynamic copper-converting process based on limited data information[J]. Control Theory & Applications, 2009, 26(8): 860-866.)
- [5] 晏密英, 桂卫华, 王凌云. 基于神经网络补偿灰色预测误差的钴离子浓度预测研究[J]. 计算机与应用化学, 2008, 25(7): 805-808.
(Yan M Y, Gui W H, Wang L Y. Research of prediction in cobalt ions concentration based on a neural network compensating the error of grey forecast[J]. Computers and Applied Chemistry, 2008, 25(7): 805-808.)
- [6] Vapnik V. The nature of statistical learning[M]. New York: Springer, 1995.
- [7] 张英, 苏宏业, 褚健. 基于模糊最小二乘支持向量机的软测量建模[J]. 控制与决策, 2005, 20(6): 621-624.
(Zhang Y, Su H Y, Chu J. Soft sensor modeling based on fuzzy least squares support vector machines[J]. Control and Decision, 2005, 20(6): 621-624.)
- [8] 梁绍华, 郑立刚, 周昊, 等. 基于支持向量机与高斯分布估计的低NO_x排放[J]. 化工学报, 2009, 60(1): 223-229.
(Liang S H, Zheng L G, Zhou H, et al. Low NO_x emissions based on support vector machine and Gaussian estimation of distribution[J]. CIESC J, 2009, 60(1): 223-229.)
- [9] 王凌云, 桂卫华, 刘梅花, 等. 基于改进在线支持向量机回归的离子浓度预测模型[J]. 控制与决策, 2009, 24(4): 537-541.
(Wang L Y, Gui W H, Liu M H, et al. Prediction model of ion concentration based on improved online support vector regression[J]. Control and Decision, 2009, 24(4): 537-541.)
- [10] Chen X, Li Y M. A modified PSO structure resulting in high exploration ability with convergence guaranteed[J]. IEEE Trans on Systems, Man, and Cybernetics, 2007, 37(5): 1271-1289.
- [11] 王雅琳. 智能集成建模理论及其在有色冶炼过程优化控制中的应用研究[D]. 长沙: 中南大学信息科学与工程学院, 2001.
(Wang Y L. Research on theory of intelligent integrated modeling and its applications to optimization and control of nonferrous metallurgical process[D]. Changsha: School of Information Science and Engineering, Central South University, 2001.)
- [12] Fu X P, Zou M. Application of combination weighting method in contract risk's evaluation of third party logistics[J]. J of Southeast University, 2007, 23(S1): 128-132.