

基于 IBA—LSSVM 的光合细菌发酵软测量

朱湘临¹, 陈 威¹, 丁煜涵¹, 王 博¹, 朱 莉¹, 姜哲宇², 宋 彦¹

(1. 江苏大学 电气信息工程学院, 江苏 镇江 212013; 2. 无锡太湖水务有限公司, 江苏 无锡 214000)

摘要: 针对光合细菌 (PSB) 发酵过程活菌浓度难以在线检测, 离线测量又存在很大延时及易染菌的问题, 提出一种基于改进蝙蝠—最小二乘支持向量机 (IBA—LSSVM) 的软测量建模方法; 先对 BA 的速度更新公式进行改进, 且将差分进化算法 (DE) 的变异机制引入 BA, 增加了种群多样性, 进而提升了 BA 算法的全局及局部搜索能力, 然后构建了活菌浓度的 IBA—LSSVM 软测量模型, 并与 BA—LSSVM 软测量模型进行对比; 仿真结果表明, 改进的模型相较于 BA—LSSVM 模型有着更好的学习能力和预测性能, 测量误差为 0.135 8, 可为光合细菌发酵过程的优化控制提供准确有效的指导, 有一定的实际应用价值。

关键词: 光合细菌; 改进蝙蝠算法; 最小二乘支持向量机; 软测量模型

Soft Sensor of Photosynthetic Bacteria Fermentation Based on IBA—LSSVM

Zhu Xianglin¹, Chen Wei¹, Ding Yuhan¹, Wang Bo, Zhu Li¹, Jiang Zheyu², Song Yan¹

(1. College of Electrical and Information Engineering, Jiangsu University, Zhenjiang 212013, China;

2. Wuxi Taihu Water Service limited company, Wuxi 214000, China)

Abstract: In view of the living cell concentration is difficult to measure on line in photosynthetic bacteria (PSB) fermentation process and the off-line measurement is accompanied by large time-delay error and easy to stain bacteria, a soft sensor model based on IBA—LSSVM was proposed. The velocity equation of the BA algorithm was improved and the variance mechanism of DE algorithm was introduced in the BA algorithm. Thus, the diversity of the population can be increased and the global and local searching ability of the BA algorithm can be enhanced. Furthermore, the IBA—LSSVM soft sensor model was established for the living cell concentration and compared with BA—LSSVM soft sensor model. Finally, the simulation results showed that the improved model was better learning ability and prediction performance than BA—LSSVM, the measurement error is 0.1358. The improved model could provide accurate guidance for the photosynthetic bacteria fermentation control optimization. This model has certain practical value.

Keywords: photosynthetic bacteria; improved BA algorithm; least squares support vector machine; soft sensing model

0 引言

光合细菌 (PSB) 是具有光能合成体系的原核生物, 在新能源开发、环境保护、养殖、医药等方面有广泛应用^[1-5]。然而其发酵过程中菌液红色会不断加深, 越来越浑浊, 是一个高度时变性、强耦合复杂过程。由此导致直接反应发酵品质的关键参量活菌浓度难以在线测量, 进而影响发酵过程的优化控制。目前主要采用的是人工离线检测的手段, 但存在着较大的测量延时及误差, 且易染菌, 会降低发酵品质。软测量技术^[6-8]是解决此问题的有效手段。

最小二乘支持向量机 (LSSVM)^[9-12]是一种基于统计学习理论的机器学习方法, 具有优秀的小样本数据学习能力和预测能力, 很适合样本数据较少的发酵过程^[13-15]。实践表明, LSSVM 的核参数与惩罚参数的取值对模型的泛化能力与精度起关键作用, 不恰当的选择可能会使 LSSVM 预测模型容易出现过度拟合和泛化能力不佳的问题。针对这个问题, 大量学者针对 LSSVM 模型参数的选择研究了各种优化算法。文献 [16] 将混沌优化算法与 LSSVM 进行有机结

合并用于发酵过程中, 得到误差较小的发酵预测曲线; 文献 [17] 采用粒子群算法优化 LSSVM 的模型参数, 提高了模型的函数逼近能力, 取得了良好的预测效果; 文献 [18] 以遗传算法完成对 LSSVM 的模型参数选择, 能够有效地跟踪因变量和预测变量之间复杂的非线性关系, 显示出很好的预测与泛化能力。

蝙蝠算法 (BA) 是由杨新社根据蝙蝠的生物特性于 2010 年提出的一种新型智能算法^[19], 相较于粒子群算法、遗传算法等传统智能算法, BA 算法具有模型简单、搜索能力强、收敛速度快等特点, 有效性和准确性方面也有明显的提高。目前已在电机运行监测^[20]、数据挖掘^[21]、风速预测^[22]、水力发电^[23]等领域得到广泛应用。但 BA 算法仍有着易陷入局部极小的问题^[24-25], 基于此, 本文对 BA 算法的速度更新公式进行改进, 并将差分进化算法 (DE) 的变异机制引入 BA, 提升了 BA 算法的全局及局部搜索能力。然后将改进的蝙蝠算法 (Improve Bat Algorithm, IBA) 对 LSSVM 的模型参数进行组合寻优, 建立光合细菌发酵过程活菌浓度的 IBA—LSSVM 模型。仿真结果表明, 该软测量模型有不错的泛化能力和预测精度。

1 算法描述

1.1 最小二乘支持向量机

最小二乘支持向量机是 Suykens^[26] 在支持向量机

收稿日期: 2018-11-29; 修回日期: 2018-12-15。

基金项目: 镇江市重点研发计划 (SH2017002)。

作者简介: 朱湘临 (1963-), 男, 江苏镇江人, 硕士, 研究员, 研究生导师, 主要从事仪表检测、过程控制方向的研究。

(SVM) 的基础上提出的新型支持向量机, 用于解决模型分解和函数估计问题。其建模原理如下:

设有 l 个训练样本 $\{(x_i, y_i) \mid i = 1, 2, \dots, l\}$, 其中样本本为 n 维向量, $x_i \in R^n$ 为样本输入, $y_i \in R^n$ 为样本输出, 对样本数据逼近的 LSSVM 可表示为:

$$\min_{\omega, b, \xi} J(\omega, \xi) = \frac{1}{2} \omega^T \omega + \frac{g}{2} \sum_{i=1}^n \xi_i^2 \quad (1)$$

$$\text{s.t. } y_i = \omega^T \varphi(x_i) + b + \xi_i, (i = 1, 2, \dots, l) \quad (2)$$

式中, ω 为权矢量; $g \in R^+$ 是惩罚参数; ξ_i 是误差变量; b 为偏差量; $\varphi(\cdot)$ 是一个非线性映射, 能把 x_i 从输入空间映射到高维 (甚至无限维) 的特征空间。

用拉格朗日法对上述问题进行优化:

$$L(\omega, b, \xi, \alpha) = \frac{1}{2} \omega^T \omega + \frac{g}{2} \sum_{i=1}^l \xi_i^2 - \sum_{i=1}^l \alpha_i (\omega^T \varphi(x_i) + b + \xi_i - y_i) \quad (3)$$

式中, α_i 是拉格朗日乘子。根据 KKT 条件将求解的优化问题最终可转换为求解线性方程:

$$\begin{bmatrix} 0 & 1_l^T \\ 1_l & K + g^{-1} I_l \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix} \quad (4)$$

式中, $y = [y_1, y_2, \dots, y_l]^T$; $1_l = [1, \dots, 1]^T$; I_l 为 l 阶单位矩阵; $\alpha = [\alpha_1, \dots, \alpha_l]^T$; K 为满足 Mercer 条件的核函数矩阵, $K = K(x_i, x_j) = \varphi^T(x_i) \varphi(x_j), i, j = 1, \dots, l$ 。

本文选用径向基函数作为核函数:

$$K = K(x, x_i) = e^{-\frac{\|x-x_i\|^2}{\sigma^2}} \quad (5)$$

式中, σ 为核函数宽度。

最终得到 LSSVM 的函数估计为:

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

1.2 蝙蝠算法

蝙蝠算法 (BA) 是基于蝙蝠的捕食机理而模拟出的一种非线性全局随机搜索算法, 算法中包含三个要素: 脉冲搜索频率、脉冲音强及发射脉冲的频度。假设蝙蝠种群数目为 N , 第 i 只蝙蝠在 t 和 $t-1$ 时刻的速度与位置分别表示为 $v_i^t, v_i^{t-1}, z_i^t, z_i^{t-1}, z^*$ 表示当前搜索过程中全局最佳位置, 蝙蝠 i 在全局搜索过程中的状态更新如下:

$$f_i = f_{\min} \times (f_{\max} - f_{\min}) \theta \quad (9)$$

$$v_i^t = v_i^{t-1} + (z_i^t - z^*) f_i \quad (10)$$

$$z_i^t = z_i^{t-1} + v_i^t \quad (11)$$

式中, $\theta \in [0, 1]$ 是均匀分布的随机数; $f_i \in [f_{\min}, f_{\max}]$ 是第 i 只蝙蝠的搜索脉冲频率。

蝙蝠算法的局部搜索是通过随机扰动实现的, 随机在当前最优解集中选择一个解 z_r , 并在其附近产生一个局部新解 z_{new} , 如式 (12) 所示:

$$z_{new} = z_r + \mu A^t \quad (12)$$

式中, A^t 为当前蝙蝠种群的平均脉冲音强; μ 是 $[-1, 1]$ 上均匀分布的随机数。

蝙蝠 i 的搜索脉冲频度和脉冲音强更新公式为:

$$A_i(t+1) = \beta A_i(t) \quad (13)$$

$$R_i(t+1) = R_0 [1 - \exp(-\gamma t)] \quad (14)$$

式中, R_0 表示蝙蝠群体的最大脉冲频度; γ 表示频度增加系数; β 表示脉冲音强衰减系数; $A_i(t)$ 表示 t 时刻的脉冲音强; $A_i(t+1)$ 表示 $t+1$ 时刻的脉冲音强; $R_i(t+1)$ 表示 $t+1$ 时刻的脉冲频度。

1.3 改进的蝙蝠算法

由于蝙蝠算法存在易陷入局部极小的缺陷, 会导致算法的优化性能下降, 本文引入了一种自适应方法^[27], 对蝙蝠算法的速度 v_i^t 的更新公式进行优化改进如下:

$$v_i^t = qv_i^{t-1} + \lambda_1(z_i^t - z^*) f_i + \lambda_2(z_i^t - z_r) f_i \quad (15)$$

$$\lambda_2 = \frac{M_{\max} - M}{M_{\max}} \tau \quad (16)$$

$$\lambda_1 = 1 - \lambda_2 \quad (17)$$

$$q = \frac{(q_{\max} - q_{\min})(M_{\max} - M)}{M_{\max}} + q_{\min} \quad (18)$$

式中, M 是当前迭代次数, M_{\max} 是最大迭代次数, τ 是 $[0, 1]$ 的 1 个实数, 随着迭代次数的增大, λ_2 线性递减, 而 λ_1 则从 $1 - \tau$ 线性递增到 1, q_{\max} 和 q_{\min} 别为惯性权重的最大和最小值。

该自适应方法使得算法在迭代早期具有较强的全局搜索能力, 跳出局部极小, 而在后期以当前全局最优解为主导进行局部搜索。针对 BA 算法因缺乏变异机制而导致局部搜索能力不强的缺陷, 将 DE 算法中的变异机制引入 BA 算法中, 以丰富种群多样性, 强化算法局部搜索能力。本文采用标准 DE/rand/1/bin^[27] 进行变异操作, 公式如下:

$$\eta = e^{1 - M_{\max} / (M_{\max} - M + 1)} \quad (19)$$

$$F = F_0 \times 2^\eta \quad (20)$$

$$z_{new} = z_{r1} + F(z_{r2} - z_{r3}) \quad (21)$$

式中, η 为变异算子; F 为缩放因子, F_0 为缩放因子的初始值; z_{r1}, z_{r2}, z_{r3} 是从当前局部最优解集中随机选出的 3 个解。

综上所述, IBA 步骤如下:

Step 1: 初始化种群数目 N 、最大迭代次数 M_{\max} 或搜索精度 ϵ 、搜索脉冲频率范围 $[f_{\min}, f_{\max}]$ 、最大脉冲音强 A 、音强衰减系数 β 、最大脉冲频度 R_0 、频度增加系数 γ 、 τ 、权重范围 $[q_{\min}, q_{\max}]$ 、初始缩放因子 F_0 , 初始化种群速度和位置并计算每只蝙蝠的适应度, 找到当前全局最优解 z^* 。

Step 2: 依据式 (9)、式 (11)、式 (15)、式 (16)、式 (17) 及式 (18) 产生新一代解 z_i^t , 并计算适应度 $f(z_i^t)$ 。

Step 3: 生成随机数 $rand1$, 若 $rand1 > R_i$, 则随机在当前最优解集中选择一个解, 通过式 (12) 产生一个局部新解 z_{new} , 并计算其适应度 $f(z_{new})$, 若 $f(z_{new}) < f(z_i^t)$ 则以该位置替代当前蝙蝠个体 i 的位置; 若 $rand1 \leq R_i$, 则从当前局部最优解集中随机选出 3 个解 z_{r1}, z_{r2}, z_{r3} , 通过式 (19) ~ 式 (21) 产生新个体 z_{new} , 并计算其适应度 $f(z_{new})$, 若有 $f(z_{new}) < f(z_i^t)$ 则, 以该位置替代当前蝙蝠个体 i 的位置。

Step 4: 生成随机数 $rand2$, 若 $rand2 > A_i$ 且更新位置后的蝙蝠 i 优于当前全局最佳位置, 则接受此新位置为当前全

局最佳位置, 并根据式 (13)、式 (14) 更新脉冲频度 R_i 和 响度 A_i 。

Step 5: 当满足搜索精度或达到最大迭代次数时, 则算法结束, 输出结果; 否则转回 Step 2 继续迭代。

$rand\ 1$ 与 $rand\ 2$ 是 $[0, 1]$ 上均匀分布的随机数。

2 基于 IBA-LSSVM 的光合细菌发酵过程活菌浓度软测量

2.1 选取辅助变量

在光合细菌发酵过程中, 影响到关键参量活菌浓度, 且可实时在线检测的潜在辅助变量有很多, 如: 光照强度 E 、空气流量 H 、发酵罐压力 p 、发酵液温度 T 、发酵液体积 V 、氨水流加速率 S 、葡萄糖流加速率 C 、电机搅拌速度 U 、发酵液酸碱度 pH 等环境参量。软测量模型的精准度离不开对可测辅助变量的合理选取^[28-29], 故采用一致关联度法^[30]获得各参量与活菌浓度的关联度, 选择关联度最高的前 5 个作为辅助变量。表 1 为各潜在辅助变量与关键参量活菌浓度的关联度。

表 1 环境变量和主导变量的关联度

外部变量	关联度
发酵罐温度 T	0.819
发酵罐压力 p	0.157
电机搅拌转速 U	0.231
发酵液体积 V	0.132
空气流量 H	0.817
光照强度 E	0.951
葡萄糖流加速率 C	0.875
发酵液酸碱度 pH	0.796
氨水流加速率 S	0.525

由表 1 可知, 光照强度 E 、发酵罐温度 T 、空气流量 H 、葡萄糖流加速率 C 、发酵液酸碱度 pH 与活菌浓度的相关性更密切, 因此选择这 5 个变量作为辅助变量。

2.2 IBA-LSSVM 的软测量建模

采用离线训练构建基于 IBA-LSSVM 的光合细菌发酵过程活菌浓度软测量模型, 在训练过程中采用 IBA 对 LSSVM 的惩罚参数 g 与核函数宽度 σ 进行组合寻优, 得到最佳的模型参数, 然后将训练好的软测量模型进行测试, 验证 IBA 的优化性能。整个软测量模型构建流程如图 1 所示。

算法具体步骤如下:

1) 采集样本数据, 划分为测试样本和训练样本两部分, 并进行预处理。

2) 设定参数 g 和 σ 的搜索范围, 初始化 IBA 算法的相关参数。

3) 随机产生蝙蝠种群, 每个蝙蝠位置由 g 和 σ 构成, 使用训练样本进行训练, 取训练样本集的均方差作为适应度值。

4) 依据适应度值最小法则, 使用 IBA 算法进行迭代搜索以及必要的越界处理, 保留适应度值最优的位置。

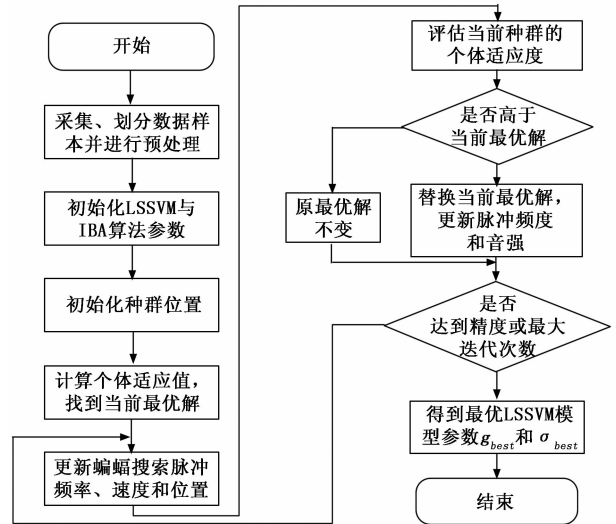


图 1 光合细菌发酵软测量模型构建流程图

5) 判断是否满足搜索精度或达到最大迭代次数, 若满足, 结束迭代, 转至 6); 否则返回 4) 循环进行迭代。

6) 选择全局适应度最佳的模型参数 g_{best} 和 σ_{best} 建立 LSSVM 软测量模型, 最终得到基于 IBA-LSSVM 的光合细菌发酵过程活菌浓度软测量模型。

2.3 实验与仿真

依据光合细菌发酵工艺进行分批发酵试验。发酵罐进行高温消毒后, 发酵过程温度控制在 $25 \sim 34^\circ\text{C}$, 电机搅拌转速为 $350 \sim 450 \text{ r} \cdot \text{min}^{-1}$, 罐压为 $0.04 \sim 0.06 \text{ Mpa}$, 通气量为 $0.2 \sim 0.6 \text{ V} \cdot (\text{V} \cdot \text{min})^{-1}$, pH 控制在 $6 \sim 8$, 光照强度控制在 $2000 \sim 5000 \text{ Lux}$, 葡萄糖流加速率控制在 $0.01 \sim 0.015 \text{ g} \cdot (\text{L} \cdot \text{h})^{-1}$ 发酵周期为 48 小时。由数字系统每隔 1 分钟采集发酵过程中光照强度、温度、发酵液 pH 值、葡萄糖流加速率, 由下位机传送到上位机, 形成数据库。在发酵正常工况下, 每 2 h 取样一次发酵液, 经离心、洗涤后, 于 105°C 烘干后得到菌体含量。汇集现场数据进行数据预处理, 一共采集 10 个发酵批次的数据, 从上述批次中取出 6 批数据 (包含 480 个样本) 作为训练样本, 2 批 (包含 120 个样本) 作为验证样本, 其余 2 批数据 (包含 480 个样本) 作为测试样本。

为了验证以上方法对光合细菌发酵过程软测量建模的可行性采用 IBA-LSSVM 软测量方法建立了光合细菌发酵过程活菌浓度的软测量模型, 选用标准 BA-LSSVM 与其做比较分析。模型参数设置为: 种群规模 $N = 50$, 最大迭代次数 $M_{\max} = 300$ 、搜索精度 $\epsilon = 0.05$ 、搜索脉冲频率范围 $[0, 10]$ 、最大脉冲音强 $A = 0.5$ 、音强衰减系数 $\beta = 0.95$ 、最大脉冲频度 $R_0 = 0.5$ 、频度增加系数 $\gamma = 0.9$ 、 $\tau = 0.6$ 、权重范围 $[0.2, 0.9]$ 、初始缩放因子 $F_0 = 1$, 参数 g 和 σ 的取值范围均设置为 $[0.01, 1000]$ 。两种优化模型的预测效果分别如图 2 和图 3 所示。

对比图 2 和图 3 可以发现, 图 3 中 IBA-LSSVM 模型的预测曲线比图 2 中 BA-LSSVM 模型的预测曲线更逼近

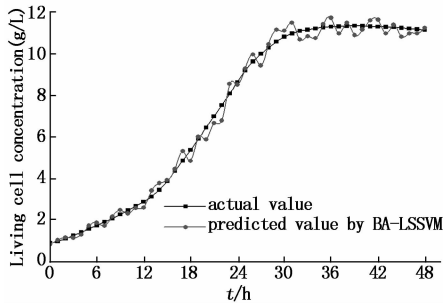


图 2 BA-LSSVM 预测效果图

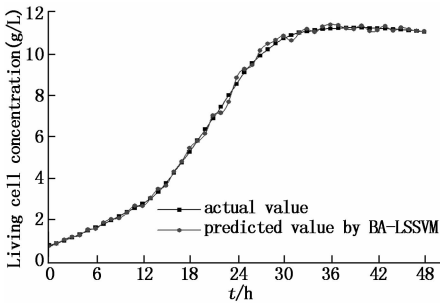


图 3 IBA-LSSVM 预测效果图

于离线化验值，即 IBA-LSSVM 模型的预测效果明显优于 BA-LSSVM 模型的预测效果。将两者对光合细菌活菌浓度的软测量误差在图 4 中体现出来，并计算两种模型的均方误差 (RMSE) 与最大相对误差 (MRE) 如表 2 所示，以便能更直观体现两种模型的预测性能。

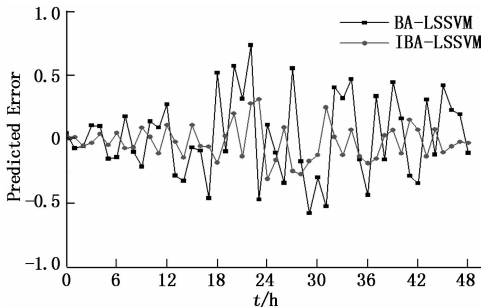


图 4 软测量误差

表 2 两种模型误差对比

模型	RMSE	MRE/%
BA-LSSVM	0.3174	9.7
IBA-LSSVM	0.1358	4.1

从图 4 和表 2 可以看出相较于 BA-LSSVM, IBA-LSSVM 的软测量效果更优异, MSE 仅为 4.1% 表明模型的预测精度高, RMRE 为 0.1358 表明模型的可行性佳。

3 结论

针对光合细菌发酵过程中活菌浓度难以在线测量的问题, 以 BA 算法为基础, 对 BA 的速度更新公式进行改进, 并引入 DE 的变异性, 使得改进的 IBA 算法有更好的局部

收敛能力及收敛精度, 将该方法确定 LSSVM 的模型参数, 建立了基于 IBA-LSSVM 的光合细菌发酵过程活菌浓度的软测量模型。仿真实验结果表明, 该软测量方法的学习能力、预测性能优于 BA-LSSVM, 为某些工业中存在难以在线测量的参量提供了一种可行的测量方法。

参考文献:

- [1] Hallenbeck P C, Liu Y. Recent advances in hydrogen production by photosynthetic bacteria [J]. International Journal of Hydrogen Energy, 2016, 41 (7): 4446 - 4454.
- [2] Qi X, Ren Y, Liang P, et al. New insights in photosynthetic microbial fuel cell using anoxygenic phototrophic bacteria [J]. Bioresource Technology, 2018, 258: 310 - 317.
- [3] Meng F, Yang A, Wang H, et al. One-step treatment and resource recovery of high-concentration non-toxic organic wastewater by photosynthetic bacteria [J]. Bioresource Technology, 2017, 251: 121 - 127.
- [4] 张健, 冯学愚, 刘小彬, 等. 光合菌发酵啤酒糟制鱼饲料工艺优化 [J]. 农业工程学报, 2011, 27 (8): 387 - 392.
- [5] Sasaki K, Watanabe M, Suda Y, et al. Applications of photosynthetic bacteria for medical fields [J]. Journal of Bioscience & Bioengineering, 2005, 100 (5): 481 - 488.
- [6] Zhu X, Ji X. Soft sensing of key state variables in fermentation process based on relevance vector machine with hybrid kernel function [J]. Sensors & Transducers, 2014, 173 (6): 237 - 243.
- [7] 刘国海, 周大为, 徐海霞, 等. 基于 SVM 的微生物发酵过程软测量建模研究 [J]. 仪器仪表学报, 2009, 30 (6): 1228 - 1232.
- [8] Wu J Q, Sun Y K, Huang Y H, et al. Soft sensor modeling based on GRNN for biological parameters of marine protease fermentation process [A]. Control Conference [C]. IEEE, 2014: 5102 - 5106.
- [9] Giorgi M G D, Campilongo S, Ficarella A, et al. Comparison between wind power prediction models based on wavelet decomposition with least-squares support vector machine (LS-SVM) and artificial neural network (ANN) [J]. Energies, 2014, 7 (8): 5251 - 5272.
- [10] 王强, 田学民. 基于 KPCA-LSSVM 的软测量建模方法 [J]. 化工学报, 2011, 62 (10): 2813 - 2817.
- [11] 钱晓山, 阳春华, 徐丽莎. 基于改进差分进化和最小二乘支持向量机的铝酸钠溶液浓度软测量 [J]. 化工学报, 2013, 64 (5): 1704 - 1709.
- [12] Li M, Chen W, Zhang T. A novel seizure diagnostic model based on kernel density estimation and least squares support vector machine [J]. Biomedical Signal Processing & Control, 2018, 41: 233 - 241.
- [13] Sun Y M, Du N, Sun Q Y, et al. Research and application of biological potency soft sensor modeling method in the industrial fed-batch chlortetracycline fermentation process [J]. Cluster Computing, 2018 (5): 1 - 12.