

doi:10.16018/j.cnki.cn32-1650/n.202104003

基于 PSO 优化 LSSVM 的出水 BOD₅ 预测建模

崔心惠¹, 李文萱¹, 张祝威²

(1. 滁州职业技术学院 电气工程学院, 安徽 滁州 239000;
2. 南京电研电力自动化股份有限公司, 江苏 南京 210000)

摘要:针对城市污水处理中有机物污染度指标 BOD₅ 缺少运行状态信息难以做到实时检测, 提出了一种基于粒子群算法(PSO)优化最小二乘法支持向量机(LSSVM)的出水 BOD₅ 预测控制策略。在保证出水水质参数稳定达标条件下, 提取城市污水处理过程中输入输出参数数据, 通过 LSSVM 对被控对象出水 BOD₅ 进行建模, 同时利用 PSO 对 LSSVM 模型进行参数寻优, 获得最佳正则化参数 γ 和核函数参数 σ 。仿真结果说明, 该模型提高了对出水 BOD₅ 值的预测精度并具有良好的泛化能力, 达到了实时性的效果。

关键词: 粒子群; LSSVM; 出水 BOD₅; 建模

中图分类号: TP23 **文献标志码:** A **文章编号:** 1671-5322(2021)04-0011-06

城市生活污水主要由有机污染物和无机污染物组成。生化需氧量 BOD 是水体中好氧微生物进行生化反应所需消耗的溶解氧量, 是间接判断水体有机污染物含量的一个重要指标^[1], 其数值大小反映了污水处理的效果, 是污水处理过程中的关键水质参数之一。通常在 20℃ 下以 5 d 作为测定 BOD 的标准时间, 计算微生物氧化过程所消耗的溶解氧量, 称为五日生化需氧量 BOD₅, 可用于评价污水有机污染水平, 选择水质处理标准。

通常 BOD₅ 通过分别测定实验前和在培养瓶中放置 5 d 后污水的 DO 浓度来计算。此方法不仅有一定的周期性还会导致二次污染, 难以实现对参数动态变化的实时检测。BOD₅ 作为评价污水水质的重要指标, 能否对其进行实时检测已成为提高治污质量的关键^[2]。因此, 设计一个能够合理有效预测 BOD₅ 值的特征模型, 对确保控制过程水质达标, 提高系统运行品质具有一定的意义。

随着科学技术的不断发展, 控制领域预测模

型的研究方法日趋多样。针对污水处理关键水质参数的预测, 何金成等^[3]在红外光谱技术下能够快速有效地对污水中的 COD、BOD 和 pH 等指标展开检测, 并用 PLS 回归建模。王凯等^[4]采用最小二乘法曲线拟合分析得到不同水体中 COD 与 BOD 之间存在一定的线性相关性, 通过比较计算值与理论值进行合理性误差分析。Golabi 等^[5]通过小波变换分解输入量, 基于特征选择算法对生化需氧量值进行发展预测并通过对比得出最佳模型。邱禹^[6]基于深度学习堆叠自编码对污泥体积指数进行软测量预测。本文利用工艺参数与水质影响因素建立机器学习预测模型, 开发污水处理控制系统。尽管神经网络逼近系统被不断提出并广泛应用于非线性系统建模, 但其存在易陷入局部极小值问题^[7]。而最小二乘支持向量机(LSSVM)是基于统计学理论, 采用经验风险最小化原则, 分化能力较强, 因此本文根据污水处理过程的输入输出参数值, 利用 LSSVM 建立污水处理过程的 BOD₅ 控制模型, 并通过粒子群算法来优化 LSSVM 的正则化参数 γ 和核函数宽度系数

收稿日期: 2021-03-02

基金项目: 2020 年度安徽省高校优秀青年人才支持计划重点项目(gxyqZD2020068); 滁州职业技术学院校级自然科学研究项目(YJY-2020-25)。

作者简介: 崔心惠(1994—), 女, 安徽滁州人, 助教, 硕士, 主要研究方向为污水处理系统建模和智能控制。

σ ,从而达到提高模型预测精度的要求。

1 LSSVM 回归算法

LSSVM 采用平方项作为优化指标,引入等式约束,将其损失函数由传统的二次规划函数替换为最小二乘法函数以便于解决分类和函数估计问题,使得最优化问题变成求解线性方程,降低算法复杂程度的同时提高运算效率^[8-10]。

给定样本集合 $U\{(x_i, y_i)\}$, 其中, $i = 1, 2, \dots, n$, $x_i \in \mathbb{R}^n, y_i \in \mathbb{R}$, 利用函数 $\phi(\cdot)$ 将输入量通过非线性方式映射到高维特征空间 H , 此时样本集为:

$$f(x_i) = \omega \cdot \phi(x_i) + b \quad (1)$$

式中, ω 是权值, b 是偏置常数, x_i 为污水处理中影响 BOD_5 预测的因素, $f(x_i)$ 为待预测值。

LSSVM 求解的优化函数:

$$\min_{\omega, b, e} J(\omega, b, e) = \frac{1}{2} \omega^T \omega + \frac{1}{2} \gamma \sum_{i=1}^n e_i^2 \quad (2)$$

式中, γ 为正则化参数且大于 0, $e_i \in \mathbb{R}$ 为误差变量。

样本集 U 中所有的样本都需满足下述约束条件:

$$y_i = \omega \cdot \phi(x_i) + b + e_i, i = 1, 2, \dots, n \quad (3)$$

用 L 表示对 J 构造的拉格朗日函数, 其中为 λ 拉格朗日函数乘子:

$$L = (\omega, b, e, \lambda) = J(\omega, b, e) - \sum_{i=1}^n \lambda_i [\omega \cdot \phi(x_i) + b + e_i - y_i] \quad (4)$$

根据 KKT 最优条件, 对式(4)中的 ω, b, e, λ 求偏导, 可得:

$$\begin{cases} \frac{\partial L}{\partial \omega} = 0 \rightarrow \omega = \sum_{i=1}^n \lambda_i \phi(x_i) \\ \frac{\partial L}{\partial b} = 0 \rightarrow \sum_{i=1}^n \lambda_i = 0 \\ \frac{\partial L}{\partial e} = 0 \rightarrow \lambda_i = \gamma e_i \\ \frac{\partial L}{\partial \lambda} = 0 \rightarrow \omega \cdot \phi(x_i) + b + e_i - y_i = 0 \end{cases} \quad (5)$$

将式(5)中的 ω 和 e 两个变量约去, 得 b 和 λ 的最优解为:

$$\begin{bmatrix} b \\ \lambda \end{bmatrix} = \begin{bmatrix} 0 & Z^T \\ Z & K + \gamma^{-1} I \end{bmatrix} \begin{bmatrix} 0 \\ y \end{bmatrix} \quad (6)$$

式中, $\lambda = [\lambda_1, \lambda_2, \dots, \lambda_n]^T$, $Z = [1, 1, \dots, 1]^T$, $y = [y_1, y_2, \dots, y_n]^T$, 即训练集待预测值的集合, I 为单位矩阵。由 mercer 条件可知, K 为核函数矩

阵, 则 LSSVM 预测模型为:

$$f(x) = \sum_{i=1}^n \lambda_i K(x_i, x_j) + b \quad (7)$$

本文选取适用性广泛的 RBF 核函数作为 LSSVM 的核函数进行调用:

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \quad (8)$$

其中, σ 为核函数的核宽度系数。

2 基于 PSO 算法优化 LSSVM 参数

2.1 粒子群算法(PSO)

粒子群算法由 Eberhart 和 Kennedy 于 1995 年提出, 是一种模仿自然界鸟群觅食行为的拟态化求解方法^[11], 可形象地将优化问题演变为搜索空间中的鸟, 即“粒子”。初始化形成全局范围内的粒子群, 群体内部通过信息共享和竞争不断迭代寻找最优粒子, 即最优解。其原理如下:

每个粒子的位置对应所求的最优解向量, 设置一个 m 维空间的 q 个粒子种群, 群体第 i 个粒子的位置为 $x_i = (x_{i1}, x_{i2}, \dots, x_{im})$, 移动速度 $v_i = (v_{i1}, v_{i2}, \dots, v_{im})$, 用 $P_{best, i} = (P_{i1}, P_{i2}, \dots, P_{im})$ 表示个体粒子寻找的最优位置, 此时整个群体的全局最优位置为 $G_{best} = (P_{g1}, P_{g2}, \dots, P_{gm})$ 。每个粒子通过给定的适应度函数不断更新 P_{best} 和 G_{best} , 直至找到最优解或达到迭代次数。其中, 每个粒子可通过公式(9)和公式(10)更新自身的速度和位置^[12]。

$$v_{im}^{t+1} = wv_{im}^t + c_1 \text{rand}_1(P_{im}^t - x_{im}^t) + c_2 \text{rand}_2(P_{gm}^t - x_{im}^t) \quad (9)$$

$$x_{im}^{t+1} = x_{im}^t + v_{im}^{t+1} \quad (10)$$

式中, w 是惯性权重因子; c_1, c_2 为学习因子, rand 是 $[0, 1]$ 取值的随机函数, t 为迭代次数。

w 采用非线性二次递减的惯性权重设置, 用来调节局部和全局之间的粒子平衡^[13], 其取值由适应度值来确定, 即:

$$\begin{cases} w_i^t = w_{\max} - \left(\frac{t}{t_{\max}}\right)^2 + (w_{\max} - w_{\min}), \\ f(x_i^{t+1}) - f(x_i^t) < 0 \\ w_i^t = 0, f(x_i^{t+1}) - f(x_i^t) \geq 0 \end{cases} \quad (11)$$

式中, w_{\max}, w_{\min} 分别为最大、最小惯性权重因子, t_{\max} 为最大迭代次数, $f(x_i^t)$ 为第 i 个粒子第 t 次迭代过程对应的适应度值。

2.2 基于PSO算法的LSSVM参数寻优

进行城市污水处理水质回归预测时,LSSVM模型中的正则化参数 γ 和核函数关键参数 σ 决定了其学习能力和泛化能力,需结合经验以及实际情况手动调整确定,选取不当会导致取值并非全局最优,且会大大降低模型精度^[14]。参数过大或过小对预测结果的影响如表1所示。 γ 和 σ 选取的重要性不言而喻。

表1 粒子群优化参数对LSSVM影响
Table 1 Influence of particle swarm optimization parameters on LSSVM

优化参数	参数趋势	模型预测影响
正则化参数 γ	过大	欠学习
	过小	过学习
核函数宽度系数 σ	过大	欠学习
	过小	过学习

采用PSO算法将问题合理转化为选择区间内参数的最优适应搜索,达到自动寻优效果^[15]。通过LSSVM模型建立城市污水处理输出水质参数BOD₅与其影响因素之间的非线性关系,进行BOD₅的预测,可提高模型预测精度和收敛速度,并实现对LSSVM的性能优化。构建基于PSO-LSSVM的出水BOD₅预测模型的具体步骤如下,模型处理流程图如图1所示。

(1)收集污水处理过程中采集的关键数据,

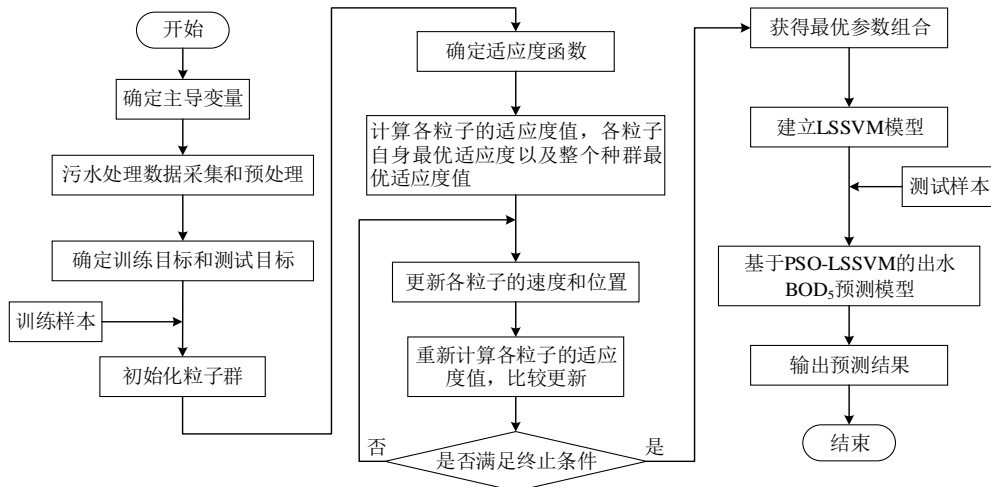


图1 PSO-LSSVM模型处理流程图

Fig 1 Flow chart of PSO-LSSVM model processing

3 PSO-LSSVM预测模型应用与分析

3.1 出水BOD₅预测模型结构

城市污水处理工艺单元环环相扣,内部生化反应过程复杂,输入输出变量存在耦合关系且影

响大小不一,因此及时获取重要过程变量BOD₅值尤为重要^[16]。本文在多变量系统中建立PSO-LSSVM的污水处理出水BOD₅预测模型,综合污水处理工艺,选取3种对输出影响较大且可测的辅助变量,确定以下样本作为训练、验证以及预

前100组数据作为训练样本 x_i ,50组数据作为模型准确度的测试样本 x_i' ,对数据进行预处理。
(2)初始化粒子群各参数:空间维数 m 取2;给定种群规模 q ,最大迭代次数 t_{max} ,学习因子 c_1 、 c_2 ,惯性权重因子 w_{max} 、 w_{min} ,设置 (γ, σ) 的取值范围。随机产生初代粒子群。

(3)通过各代 (γ, σ) 组合对LSSVM模型参数进行训练,选用均方根误差(RMSE)作为评价粒子适应度的函数,计算各代适应度值,计算公式为:

$$\delta_{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - y'_i)^2} \quad (12)$$

式中, n 为样本数量, y_i 为样本实际值, y'_i 为样本预测值。

(4)分别对比各粒子当前和历史最优位置适应度值 $f(x_i)$ 和 $f(P_{best,i})$,若 $f(x_i) < f(P_{best,i})$,则更新 $P_{best,i} = x_i$;再对比粒子最优位置的适应度值 $f(x_i)$ 和整个粒子群的最优位置适应度值 $f(G_{best})$,若 $f(x_i) < f(G_{best})$,则更新 $G_{best} = x_i$ 。

(5)通过不断更新粒子的速度和位置来产生新一代粒子,同时更新 $P_{best,i}$ 和 G_{best} 。

(6)若计算次数达到最大迭代次数或满足的设定适应度值,则终止算法、输出最优解,否则返回步骤(4)。

测的输入输出,预测模型如图 2 所示。

样本输入:进水 SS, $U(a)$; 进水 NH_3-N , $U(b)$; DO, $U(c)$ 。

样本输出: $BOD_5, y(t)$ 。

PSO-LSSVM 模型的当前输出值: $y_p(t)$ 。

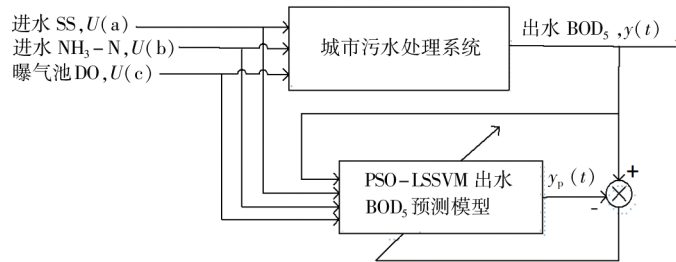


图 2 城市污水处理过程出水 BOD_5 预测模型

Fig 2 BOD_5 prediction model of effluent from municipal wastewater treatment process

城市污水处理是一个过程繁琐冗长且各因素关联程度高的非线性动态系统,当前状态下的出水 BOD_5 不仅与进水水质参数有关,还与前一状态下的 BOD_5 有一定的联系,因此污水厂的进出水口易随时间变化呈现出一定的时序性和规律性。基于污水处理 PSO-LSSVM 预测模型和公式(6)可构建如下非线性映射关系:

$$y_p(t+1) = f[y(t), U(t)] \quad (13)$$

则出水 BOD_5 预测控制模型关系如下:

$$\begin{cases} y_p(t+1) = y_{BOD_5}(t+1) \\ y(t) = y_{BOD_5}(t) \\ U_t = [U_a(t), U_b(t), U_c(t)]^T \end{cases} \quad (14)$$

3.2 仿真实验及结果

进行 PSO-LSSVM 出水 BOD_5 预测前,先初始化模型参数,各参数设置如下:种群规模 $q=30$,最大迭代数 $t_{max}=100$,学习因子 $c_1=1.5, c_2=1.7$,正则化参数 γ 的调整区间为 $[0.1, 100]$,核函数宽度系数 σ 调整区间为 $[0.01, 1000]$ 。由于原始数据存在较大的量纲差异,现从污水处理厂提供的数据中选取 150 组有效数据进行归一化处理,前 100 组数据作为 PSO-LSSVM 模型训练学习的样本集,后 50 组数据作为测试集进行预测。出水 BOD_5 预测模型适应度曲线如图 3 所示。

由图 3 可知,当迭代次数达到 20 次时,出水 BOD_5 适应度曲线基本处于稳定状态,适应度值小。此时获得 BOD_5 预测模型的正则化参数 $\gamma=7.261$,核函数宽度系数 $\sigma=13.679$ 。根据上述参数用 PSO-LSSVM 预测模型分别对训练样本和测试样本进行回归仿真和误差实验,结果如图 4、图 5 所示。

由图 4、图 5 可知,模型训练的预测值与实验

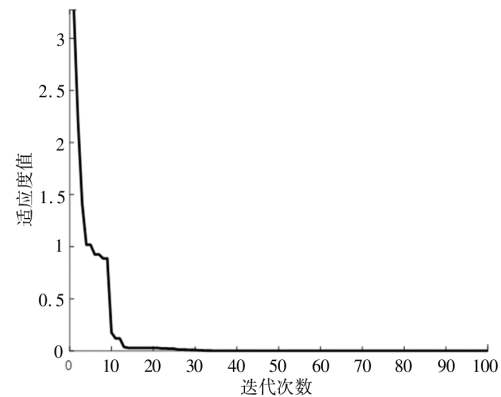


图 3 出水 BOD_5 预测模型适应度曲线

Fig 3 BOD_5 fitness curve of effluent prediction model

室离线分析值对比具有较高的拟合精度,误差基本在 0.6 以内,且绝大部分在 0.2~0.4 范围之间,说明该模型具有良好的回归拟合效果,能够满足城市污水处理过程的实际需求。

3.3 对比分析预测结果

为了验证 PSO-LSSVM 出水 BOD_5 预测模型不仅具有良好的学习能力,还具有较高的预测精度及优良的泛化能力,本文选取测试样本中的 30 组数据进行预测,分别建立未经优化的 LSSVM 和 BP 神经网络模型,出水 BOD_5 实际值和预测值的对比结果如图 6 所示。

由图 6 可以看出,BP 神经网络模型存在学习效率不高、收敛速度慢的特点,与其他两种模型相比仿真误差较大;LSSVM 模型与 BP 神经网络模型相比预测误差降低了近 20%,但与经过 PSO 算法优化的 PSO-LSSVM 模型相比,预测精度仍然无法满足要求。通过对比分析可知,PSO-LSSVM 对出水 BOD_5 的预测值更加切合实际运行情况,能够更加精准地跟踪实际出水 BOD_5 值,控制效果比其他两种模型更好,提高了预测精度及

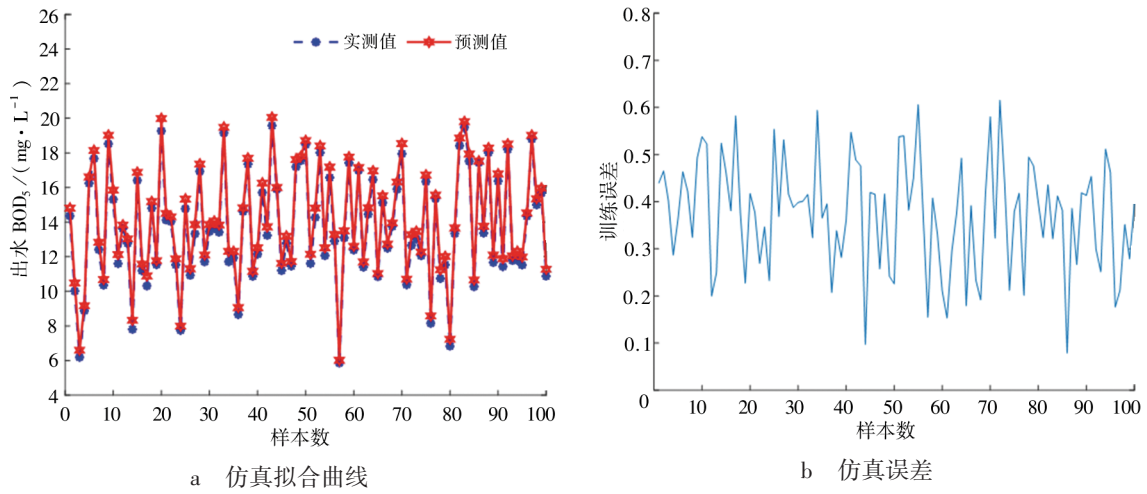


图4 BOD₅训练样本仿真结果
Fig 4 Training sample simulation results of BOD₅

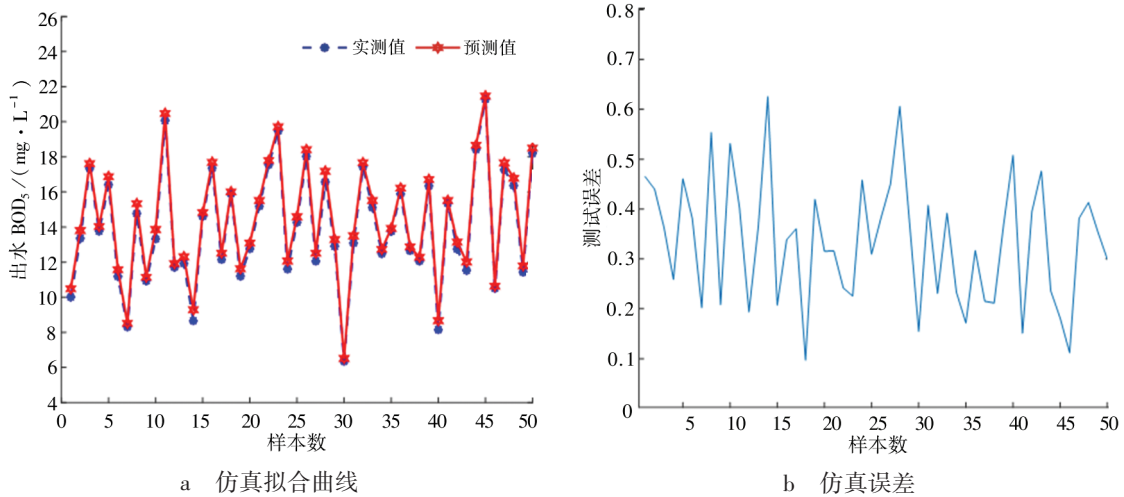


图5 BOD₅测试样本仿真结果
Fig 5 Test sample simulation results of BOD₅

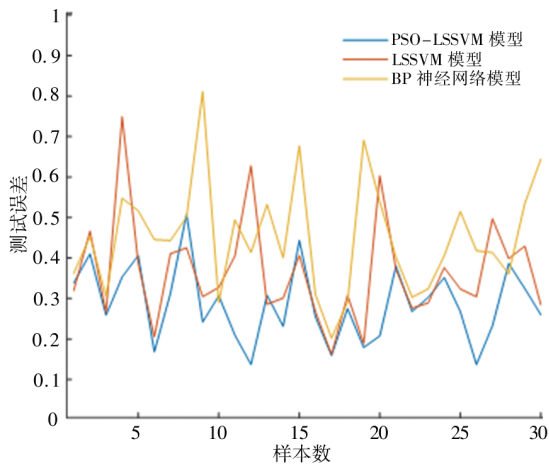


图6 3种模型的预测结果对比
Fig 6 Comparison of the prediction results of the three models

系统的稳定性。

4 结语

城市污水处理过程耦合关系强、滞后性大、非线性特点显著,关键水质参数难以实时在线检测。本文提出一种基于PSO-LSSVM的关键水质特征模型,实现了城市污水处理中出水BOD₅浓度在线预测。该模型利用PSO对LSSVM的正则化参数 γ 和核函数参数 σ 进行优化,建立关键水质参数与主要工艺变量之间的相关性,从而提升模型精度,并将其与LSSVM和BP神经网络预测模型进行对比。通过比较可知,本文提出的方法具有较好的模型预测性能和泛化性能,可更好地满足污水处理过程的需求,具有良好的应用前景。

参考文献:

- [1] GAWDZIK A, GAWDZIK J, GAWDZIK B, et al. Prediction of BOD₅ content of the inflow to the treatment plant using different methods of black box- the case study[J]. Desalination and Water Treatment, 2020, 195: 58-66.
- [2] 牛桂荣. 工业污水五日生化需氧量(BOD₅)测定方法的优化与探讨[J]. 广东化工, 2019, 46(17): 135-136.
- [3] 何金成, 杨祥龙, 王立人, 等. 基于近红外光谱法的废水 COD、BOD₅、pH 的快速测量[J]. 环境科学学报, 2007, 27(12): 2105-2108.
- [4] 王凯, 周怀兰, 刘年, 等. 最小二乘法曲线拟合分析 COD 与 BOD 的关系[J]. 能源与环境, 2018(1): 78-79.
- [5] GOLABI M R, FARZI S, KHODABAKHSHI F, et al. Biochemical oxygen demand prediction: development of hybrid wavelet-random forest and M5 model tree approach using feature selection algorithms[J]. Environmental Science and Pollution Research, 2020, 27: 34322-34336.
- [6] 邱禹. 面向污水处理的软测量建模研究及其应用[D]. 广州: 华南理工大学, 2018.
- [7] 李明河, 王健, 孟凡许. 基于支持向量机 α 阶逆系统方法的污水处理内模控制[J]. 控制工程, 2016, 23(2): 185-189.
- [8] 李琨, 韩莹, 黄海礁. 基于 IBH-LSSVM 的混沌时间序列预测及其在抽油井动液面短期预测中的应用[J]. 信息与控制, 2016, 45(2): 241-247, 256.
- [9] YANG Y F, QIN Y, JIA L M, et al. Traffic safety region estimation based on SFS-PCA-LSSVM: an application to highway crash risk evaluation[J]. International Journal of Software Engineering and Knowledge Engineering, 2016, 26(9/10): 1555-1570.
- [10] GAO S Z, LI T C, ZHANG Y M. Rolling bearing fault diagnosis of PSO-LSSVM based on CEEMD entropy fusion[J]. Transactions of the Canadian Society for Mechanical Engineering, 2020, 44(3): 405-418.
- [11] YU H H, CHEN Y Y, HASSAN S G, et al. Prediction of the temperature in a Chinese solar greenhouse based on LSSVM optimized by improved PSO[J]. Computers and Electronics in Agriculture, 2016, 122: 94-102.
- [12] 张扬, 朱志宇. 改进 PSO 算法优化 LSSVM 的柴油机排气预测[J]. 重庆理工大学学报(自然科学版), 2020, 34(1): 103-110.
- [13] 岳中文, 范皓宇, 马鑫民. 基于 PSO-SVM 的煤矿巷道爆破效果预测关键技术研究[J]. 爆破, 2019, 36(3): 31-36, 55.
- [14] 何邦乐, 黄勇, 叶颀, 等. 基于 PSO-LSSVM 的高压电力电缆接头温度预测[J]. 电力工程技术, 2019, 38(1): 31-35.
- [15] 郭宝龙, 王健, 闫允一, 等. 基于多目标 PSO 算法的 DSP 防护优化设计[J]. 计算机工程, 2018, 44(4): 74-80.
- [16] 宋翼颀. 污水处理机理建模与预测控制研究[D]. 株洲: 湖南工业大学, 2016.

Predictive Modeling of Effluent BOD₅ Based on PSO Optimized LSSVM

CUI Xinhui¹, LI Wenxuan¹, ZHANG Zhuwei²

(1.College of Electrical Engineering, Chuzhou Polytechnic, Chuzhou Anhui 239000, China ;
2.Nanjing Dianyan Electric Power Automation Co., Ltd., Nanjing Jiangsu 210000, China)

Abstract: In view of the fact that BOD₅ is lack of running state information in urban sewage treatment, and it is difficult to achieve real-time detection, a predictive control strategy of effluent BOD₅ based on particle swarm optimization (PSO) and least support vector machine (LSSVM) is proposed. Under the condition that the effluent quality parameters are stable and up to the standard, the input and output parameter data in the process of urban sewage treatment are extracted, the effluent BOD₅ of the controlled object is modeled by LSSVM, and the parameters of LSSVM model are optimized by PSO to obtain the optimal regularization parameters and kernel parameters. The simulation results show that the model improves the prediction accuracy of effluent BOD₅ value, and has good generalization ability, achieving the real-time effect.

Keywords: particle swarm optimization; LSSVM; effluent BOD₅; modeling

(责任编辑:熊璐璐)