

DOI:10.19853/j.zgjsps.1000-4602.2025.09.008

# 基于BP神经网络的给水厂混凝剂投加量预测

王 坤<sup>1,2</sup>, 孙新洋<sup>1,2</sup>, 黄显怀<sup>1,2</sup>, 唐玉朝<sup>1,2</sup>, 伍昌年<sup>1,2</sup>, 尹翠琴<sup>3</sup>,  
张良霄<sup>3</sup>

(1. 安徽建筑大学 环境与能源工程学院, 安徽 合肥 230601; 2. 安徽建筑大学 环境污染控制与资源化安徽省重点实验室, 安徽 合肥 230601; 3. 合肥水务集团有限公司, 安徽 合肥 230011)

**摘 要:** 选取合肥市某给水厂运行数据,使用相关系数法进行筛选后,选取流量、浊度、耗氧量、pH和水温作为预测模型的输入参数,利用多层变分模态分解(VMD)算法捕捉数据信息,通过遗传算法(GA)优化BP神经网络权重和偏置,建立基于VMD-GA-BP神经网络的给水厂混凝剂投加量预测模型。该模型评价指标平均绝对误差(MAE)、均方根误差(RMSE)、平均绝对百分比误差(MAPE)均低于单层VMD-GA-BP模型、输入参数未经多层VMD分解的GA-BP模型和未利用GA优化的VMD-BP模型,相比单层VMD-GA-BP模型,MAE下降37.26%、RMSE下降36.19%、MAPE下降2.44%;与GA-BP模型相比,MAE下降27.03%、RMSE下降23.94%、MAPE下降1.43%;与VMD-BP模型相比,MAE下降40.99%、RMSE下降41.47%、MAPE下降2.83%。结果表明,多层VMD算法与GA的参与提高了模型预测的准确性和稳定性,模型能有效拟合混凝剂投加量变化趋势。

**关键词:** 混凝剂投加量; 变分模态分解(VMD); 遗传算法(GA); BP神经网络

**中图分类号:** TU991 **文献标识码:** A **文章编号:** 1000-4602(2025)09-0053-06

## Prediction of Coagulant Dosage in Water Supply Plants Using Back Propagation Neural Network

WANG Kun<sup>1,2</sup>, SUN Xin-yang<sup>1,2</sup>, HUANG Xian-huai<sup>1,2</sup>, TANG Yu-chao<sup>1,2</sup>,  
WU Chang-nian<sup>1,2</sup>, YIN Cui-qin<sup>3</sup>, ZHANG Liang-xiao<sup>3</sup>

(1. School of Environment and Energy Engineering, Anhui Jianzhu University, Hefei 230601, China; 2. Anhui Provincial Key Laboratory of Environmental Pollution Control and Resource Reuse, Anhui Jianzhu University, Hefei 230601, China; 3. Hefei Water Group Co. Ltd., Hefei 230011, China)

**Abstract:** The operational data of a water supply plant in Hefei were selected and screened using correlation coefficient method. Flow rate, turbidity, oxygen consumption, pH, and water temperature were identified as input parameters for the prediction model. Data characteristics were extracted using multi-layer variational mode decomposition (VMD) algorithm, while genetic algorithm (GA) was employed to

基金项目: 国家自然科学基金资助项目(52370001); 安徽省高校自然科学基金项目(2024AH050229); 国家重点研发计划项目(2023YFC3205705); 安徽省协同创新项目(GXXT-2023-046); 安徽省自然科学基金资助项目(2208085US20); 安徽省新时代育人质量工程项目(2022zyxwxalk121)

通信作者: 王坤 E-mail: wangkun@ahjzu.edu.cn

optimize the weights and biases of the back propagation (BP) neural network. Consequently, a prediction model for coagulant dosage in a water supply plant based on VMD-GA-BP neural network was successfully established. The mean absolute error (MAE), root mean square error (RMSE), and mean absolute percentage error (MAPE) of the model evaluation indices were significantly lower than those of the single-layer VMD-GA-BP model, the GA-BP model without multivariate VMD decomposition of input parameters, and the VMD-BP model without GA optimization. Specifically, compared to the single-layer VMD-GA-BP model, the MAE, RMSE and MAPE decreased by 37.26%, 36.19%, and 2.44% respectively. Compared with the GA-BP model, the MAE, RMSE and MAPE was reduced by 27.03%, 23.94%, and 1.43% respectively. Compared with the VMD-BP model, the MAE, RMSE and MAPE was reduced by 40.99%, 41.47%, and 2.83% respectively. The integration of multi-layer VMD algorithm and GA enhances both the accuracy and stability of model predictions, allowing the model to effectively capture the changing trend of coagulant dosage.

**Key words:** coagulant dosage; variational mode decomposition (VMD); genetic algorithm (GA); back propagation (BP) neural network

投加混凝剂可去除原水中悬浮固体颗粒及藻类,其投加量直接影响给水厂出水水质和运行成本。目前,给水厂通常依靠烧杯试验来确定混凝剂投加量,无法根据水质变化精准投加。因此,在确保水质达标的前提下,寻求混凝剂最佳投加量是当前给水行业的重要研究方向。

韩梅等<sup>[1]</sup>在预测混凝剂投加量时,将预测模型与水厂智慧管控平台相融合,实现了混凝投药的自动化控制。詹咏等<sup>[2]</sup>根据给水厂原水水质等情况,建立了反映助凝剂投加量、混凝剂投加量等多指标的数学模型。庾婧艺等<sup>[3]</sup>将粒子群优化的径向基函数(RBF)神经网络应用于混凝剂投加量预测,建立的模型具有良好的泛化能力。白桦等<sup>[4]</sup>引用神经网络预测理论,构建了混凝剂投加量预测模型,并以水厂运行数据进行验证,体现了模型的自学习性和适应性。Oladipupo等<sup>[5]</sup>依据进水水质的影响因素,建立线性回归与机理模型,有效提高了投药精度。刘旺等<sup>[6]</sup>采用了遗传算法(GA)优化的反向传播(BP)神经网络预测水厂混凝剂投加量,平均绝对误差为3.78 mg/L,模型具有一定的参考价值。但这些应用均未考虑水质指标波动对混凝剂投加量预测产生的影响,如不进行适当处理,水质波动会导致预测模型精度下降。

笔者以合肥市某给水厂作为研究对象,采集原水日流量、浊度、耗氧量、pH、水温、沉淀池出水浊度和混凝剂投加量等数据,利用变分模态分解(VMD)

算法进行数据预处理,并通过遗传算法优化BP神经网络参数,建立VMD-GA-BP预测模型,旨在确保水质达标的同时实现混凝剂精准投加,为给水厂优化运行提供新思路。

## 1 给水厂数据处理与分析

### 1.1 原水水质指标的重要性

在给水厂实际生产中,原水各项水质指标对混凝过程有不同程度的影响,若将所有影响因素作为输入层的影响因子,对混凝剂投加量影响较小的因子可能会干扰模型的精度与泛化性,因此在收集到给水厂2023年10月—2024年3月的运行数据后,对7种常规水质参数进行相关性分析。

由于各项水质指标有不同的计量单位,在量纲和数据尺度上存在差异,需将不同范围、不同单位的数据进行统一以便于比较<sup>[7]</sup>。在分析前使用最大-最小归一化来处理原始数据,见式(1)。

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (1)$$

式中: $x$ 为水质参数原始值; $x_{\min}$ 为原始水质参数的最小值; $x_{\max}$ 为原始水质参数的最大值; $x'$ 为将水质参数原始值映射为区间 $[0,1]$ 的值。

归一化后的数据需要进行皮尔逊(Pearson)相关性系数分析,该方法是一种能够准确评价两个数据序列之间相关程度的统计学方法。Pearson系数绝对值越大,代表参与计算的两个指标的相关程度越强,这在数据相关性分析领域应用极为广泛。计

算原水各项水质指标与混凝剂投加量之间的皮尔逊系数,结果如图1所示[\*\*表示皮尔逊系数置信度(双侧) $\leq 0.01$ ,相关性显著]。可以看出,混凝剂投加量与流量、浊度、耗氧量、pH、水温相关性显著。给水厂混凝剂投加量预测模型选取这5个指标作为BP神经网络输入值的维度,混凝剂投加量作为输出值的维度。

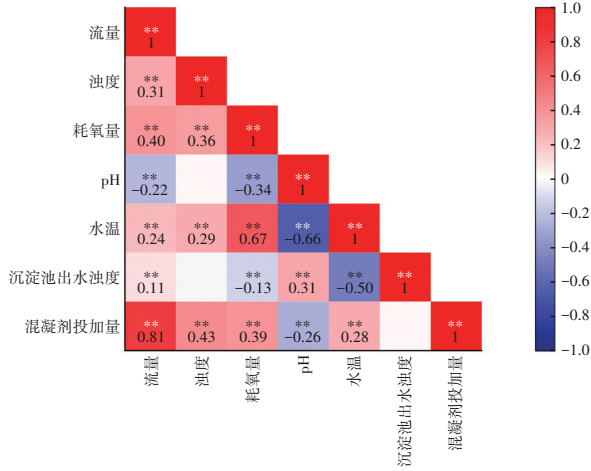


图1 Pearson相关性热力图

Fig.1 Pearson correlation heatmap

## 1.2 基于VMD算法的数据多层分解

VMD算法是一种自适应、完全非递归的分解算

$$L(\{u_k\}, \{\omega_k\}, \lambda) = \alpha \sum_{k=1}^K \left\| \partial_t \left[ \left( \delta(t) + \frac{j}{\pi t} \right) u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 + \left\| f(t) - \sum_{k=1}^K u_k(t) \right\|_2^2 + \left\langle \lambda(t), f(t) - \sum_{k=1}^K u_k(t) \right\rangle \quad (4)$$

利用交替方向乘子方法对上式中的  $u_k$ 、 $\omega_k$ 、 $\lambda$  进行更新,见式(5)~(7)。

$$\hat{u}_k^{n+1}(\omega) = \frac{\hat{f}(\omega) - \sum_{i \neq k} \hat{u}_i(\omega) + \frac{\hat{\lambda}^n(\omega)}{2}}{1 + 2\alpha(\omega - \omega_k^n)^2} \quad (5)$$

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega |\hat{u}_k^{n+1}(\omega)|^2 d\omega}{\int_0^\infty |\hat{u}_k^{n+1}(\omega)|^2 d\omega} \quad (6)$$

$$\hat{\lambda}^{n+1}(\omega) = \hat{\lambda}^n(\omega) + \tau \left[ \hat{f}(\omega) - \sum_{k=1}^K |\hat{u}_k^{n+1}(\omega)| \right] \quad (7)$$

定义迭代停止条件见式(8)。

$$\sum_{k=1}^K \frac{\|\hat{u}_k^{n+1} - \hat{u}_k^n\|_2^2}{\|\hat{u}_k^n\|_2^2} < e \quad (8)$$

式中: $e$ 为收敛误差。

本研究采用多层VMD分解,共三层分解。针对

法。本研究对给水厂混凝剂投加量预测模型的输入指标进行了多层VMD分解,进一步分解低频分量并保留高频分量。低频分量分解提取更加细致的趋势信息,可有效去除低频噪声,提高信号的平滑度和稳定性。同时保留高频分量中的细节信息,使得模型能够捕捉原始信号中的短期波动和小尺度特征。通过对输入指标进行多层VMD分解,不仅能优化特征提取和信号处理过程,而且在后续模型训练中,有利于提高模型的预测准确性和泛化能力,从而实现对混凝剂投加量的精准预测。

在VMD算法中,关键在于建立和解决变分问题,建立的变分约束问题见式(2)、(3)。

$$\min \left\{ \sum_{k=1}^K \left\| \partial_t \left[ \left( \delta(t) + \frac{j}{\pi t} \right) u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\} \quad (2)$$

$$\text{s. t.} \quad \sum_{k=1}^K u_k = f \quad (3)$$

式中: $u_k$ 为第 $k$ 个模态分量; $\omega_k$ 为对应分量的中心频率; $f$ 为原始信号; $\delta(t)$ 为狄拉克函数;s. t. 表示约束条件,确保所有分解后的模态分量之和等于原始信号; $j$ 为虚数单位; $t$ 为时间。

引入拉格朗日乘子 $\lambda$ 和二次惩罚因子 $\alpha$ ,转化为无约束的变分问题,见式(4)。

输入信号,第一层分解首先产生两组特征分量,即低频特征分量和高频特征分量。第二层分解采用相同的方法,对前一层的低频分量进行进一步的分解,得到低、高频两种分量。第三层再将前一层的低频分量进行分解,得到低、高频两种分量。保留第一层和第二层的高频分量,以及第三层的两个特征分量。当这四个分量排列顺序不同时,模型会有不同的预测结果,经反复试验,排序方式为:第三层的低频分量、第三层的高频分量、第一层的高频分量、第二层的高频分量,这种排序方式有助于模型更好地捕捉和利用数据信息,以提高模型预测的精度。VMD各参数设置如下:惩罚参数 $\alpha=180$ ;模态数 $K=2$ ;初始化方式initial=1(均匀初始化模态频率中心);收敛判据 $r=10^{-7}$ 。

## 2 VMD-GA-BP模型建立

### 2.1 BP神经网络模型与GA优化

利用Pearson相关系数法筛选后的数据,在经多



层VMD算法处理后,采用GA优化BP模型,搭建VMD-GA-BP神经网络模型。模型框架如图2所示,其运算流程如图3所示。

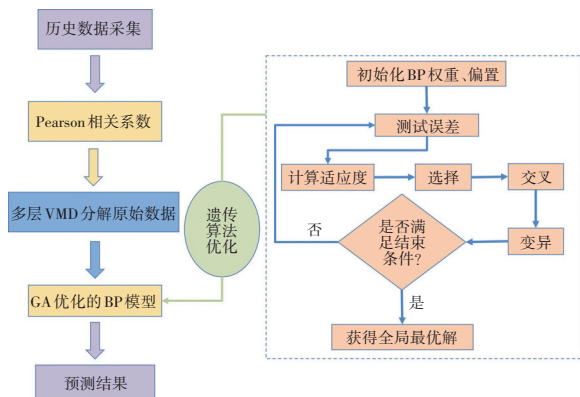


图2 VMD-GA-BP模型框架示意

Fig.2 Schematic diagram of VMD-GA-BP model framework

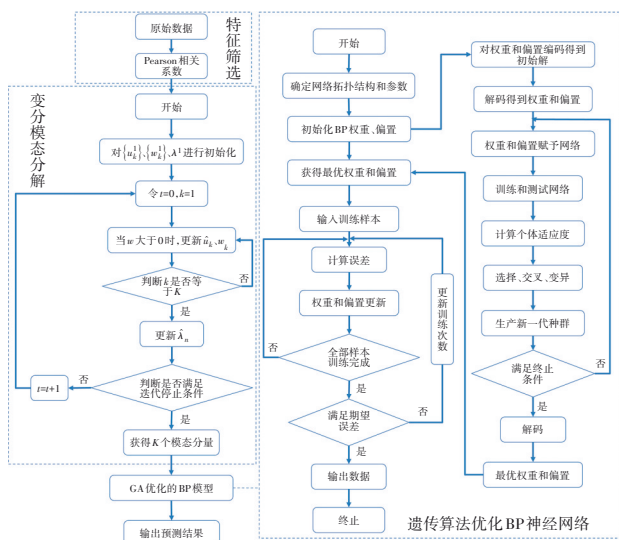


图3 VMD-GA-BP模型运算流程

Fig.3 Computational process of VMD-GA-BP model

BP神经网络是目前最常用的一种神经网络模型,其由输入层、隐含层和输出层构成,每一层根据实际需要,其节点数各不同<sup>[8]</sup>。模型样本使用VMD分解后随机排列的183组数据,其中30%作为测试样本,不参与模型训练。测试集对于模型来说是全新的数据,通过测试集在模型上的表现,可以评估模型的泛化能力,确保模型不会出现过拟合。在隐含层层数选择上,设置隐含层层数为一层,输入数据在经VMD分解以后,输入层的神经元数量为20个,隐含层神经元数量经调试以后设置为6个,输出层神经元数量为1个。在模型训练过程中学习速率为

0.01,训练次数最大值为1000次,期望误差为0.0001,神经网络训练函数选用Levenberg-Marquardt算法。

GA是一种基于自然选择和遗传机制的优化算法,适用于解决复杂的优化和搜索问题。通过模拟生物进化过程中的选择、交叉和变异等机制,GA能够不断优化解的集合,从而在多峰值搜索空间中找到全局最优解。使用GA对BP神经网络中输入层到隐含层的权重、隐含层到输出层的权重以及隐含层和输出层的偏置参数进行了优化,其具体步骤如下:

#### ① 实数编码和初始化种群

通过实数编码将BP神经网络中需要进行优化的权重和偏置参数编码成一个实数向量(染色体),编码长度见式(9)。

$$m = h \times i + h \times j' + j' + h \quad (9)$$

式中: $m$ 为染色体的编码长度; $i$ 为输入层节点数; $j'$ 为输出层节点数; $h$ 为隐含层节点数。

在遗传算法中,种群是由多个个体(也称为解)组成的集合,在算法开始时,需要初始化一个种群。本研究初始种群规模设置为60,初始化参数范围设置为 $(-3, 3)$ ,随机生成一组初始解。

#### ② 适应度函数和最大进化代数

适应度函数在GA优化BP神经网络权重和偏置参数的过程中起着至关重要的作用。它用来评估每个候选解,即神经网络的权重和偏置组合的性能,从而在搜索空间中找到最优的权重和偏置参数<sup>[9]</sup>。

在本研究中,以训练集和测试集均方误差(MSE)的平均值作为适应度函数。首先计算训练集和测试集的均方误差,然后对每个均方误差取平方根,最后再取两个值的平均值。这种方法可以平衡模型在训练集和测试集上的性能,避免模型出现过拟合或者欠拟合现象。适应度函数见式(10)。

$$\text{error} = \frac{\sqrt{\text{MSE}_{\text{train}}} + \sqrt{\text{MSE}_{\text{test}}}}{2} \quad (10)$$

式中: $\text{MSE}_{\text{train}}$ 为训练集的均方误差; $\text{MSE}_{\text{test}}$ 为测试集的均方误差。

最大进化代数是GA中的一个重要参数,它限制了算法运行过程中进行的最大迭代次数。为确保GA在合理的时间内对BP神经网络的权重和偏置参数进行充分优化,找到高质量的解,经多次调试,最大进化代数设置为90。

### ③ 选择、交叉和变异

在每一代中,GA通过选择、交叉和变异等操作生成新一代种群。选择操作采用余数随机采样从当前种群中挑选出适应度较高的个体,以生成下一代种群;交叉操作在这些被选中的个体之间进行信息交换以产生新的解;而变异操作则在种群中引入一些随机性以保持种群的多样性。交叉概率设为0.9,变异概率设为0.1。这些操作使得种群在每一代中都朝着最终的优化目标进化。

## 2.2 模型评价指标

为定量评价模型的预测精度,采用4种常用的模型评价指标,分别为平均绝对误差(MAE)、均方根误差(RMSE)、平均绝对百分比误差(MAPE)和决定系数 $R^2$ 。MAE、RMSE和MAPE从预测值与真实值之间的误差大小角度来评价模型预测的精度,其值越小说明预测效果越好; $R^2$ 是用于衡量回归模型拟合效果的指标,其值越大说明预测效果越好。4项指标分别按式(11)~(14)计算。

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (11)$$

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

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\% \quad (13)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (14)$$

式中: $y_i$ 为真实值; $\hat{y}_i$ 为预测值; $\bar{y}$ 为真实值的平均值。

## 2.3 模型表现与分析

采用MATLAB软件进行神经网络模型的搭建和相应的遗传算法优化,模型VMD-GA-BP经训练后,训练集决定系数 $R^2$ 为0.88,MAE为94.05 kg, RMSE为146.95 kg,MAPE为4.01%;测试集决定系数 $R^2$ 为0.90,MAE为98.99 kg, RMSE为127.42 kg, MAPE为4.53%。模型训练集和测试集的 $R^2$ 都接近1,表明模型的拟合效果好,并且在测试集上的表现略优于训练集,说明模型没有过拟合,具有良好的泛化能力。模型VMD-GA-BP预测运行结果表现优秀,测试集的预测值与真实值平均绝对百分比误差小于5%,模拟精度较高。

同时,为验证GA对模型的优化效果及数据预处理阶段采用的多层VMD分解对模型的影响,采用相同的建模方式建立了单层VMD-GA-BP、VMD-BP及GA-BP神经网络模型,预测效果如图4所示。

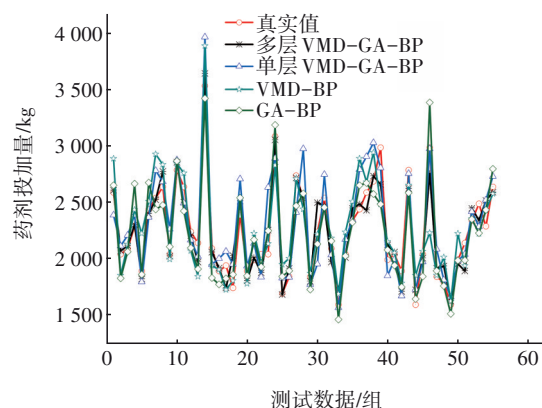


图4 各模型预测效果比较

Fig.4 Comparison of prediction performance among various models

由图4可以看出,单层VMD-GA-BP模型、未经GA优化过的VMD-BP模型和未使用VMD算法的GA-BP模型,这三种模型预测的总体误差较大,且波动明显,对实际值的拟合能力差。而在数据处理阶段采用多层VMD算法时,经GA优化过的多层VMD-GA-BP模型预测精度明显提高,各项指标都有所下降(见表1)。

表1 模型间运行精度对比

Tab.1 Comparison of operational accuracy among models

预测模型	MAE/kg	RMSE/kg	MAPE/%
多层 VMD-GA-BP	98.99	127.42	4.53
单层 VMD-GA-BP	157.78	199.70	6.97
VMD-BP	167.75	217.71	7.36
GA-BP	135.65	167.52	5.96

相比于单层VMD-GA-BP模型,多层VMD-GA-BP模型的MAE下降37.26%、RMSE下降36.19%、MAPE下降2.44%;相比于未对模型进行GA优化的VMD-BP模型,MAE下降40.99%、RMSE下降41.47%、MAPE下降2.83%;与未对输入数据进行VMD预处理的GA-BP模型相比,MAE下降27.03%、RMSE下降23.94%、MAPE下降1.43%。可见,GA与多层VMD算法对模型均起到了有效的优化作用,多层VMD-GA-BP神经网络模型的预测效果更加理想,稳定性和准确性更好,能很好地预

测给水厂混凝剂投加量。

### 3 结论

① 在对原水各项水质指标进行 Pearson 相关性分析的基础上,选取流量、浊度、耗氧量、pH 和水温作为 BP 神经网络的输入参数,并对数据进行多层 VMD 分解,有效提取出了信号中的重要趋势信息,去除了噪声,提高了信号平滑度和模型预测的准确性。

② 利用 GA 对 BP 神经网络进行优化,经优化的神经网络预测模型可获得最优权重与偏置。建立用于给水厂混凝剂投加量预测的 VMD-GA-BP 神经网络模型,该模型的 MAE 为 98.99 kg, RMSE 为 127.42 kg, MAPE 为 4.53%,  $R^2$  为 0.90。

③ 相比于单层 VMD-GA-BP 模型、模型输入指标未采用 VMD 算法的 GA-BP 模型与未使用 GA 优化的 VMD-BP 模型,搭建的多层 VMD-GA-BP 神经网络模型预测精度更高,可为给水厂混凝剂投加量预测提供有效参考。

### 参考文献:

- [1] 韩梅, 邹放, 苑蕊, 等. 基于机器学习的混凝智能投药模型应用效果及路径探讨[J]. 给水排水, 2022, 48(10): 180-184.  
HAN Mei, ZOU Fang, YUAN Rui, *et al.* Discussion on application effect and application path of machine learning-based coagulation intelligent dosing model[J]. Water & Wastewater Engineering, 2022, 48(10): 180-184(in Chinese).
- [2] 詹咏, 徐国勋, 吴文权, 等. 最佳混凝条件下混凝剂投量数学模型[J]. 中国给水排水, 2004, 20(1): 55-57.  
ZHAN Yong, XU Guoxun, WU Wenquan, *et al.* Mathematical model of coagulant dosage under the optimal coagulation conditions[J]. China Water & Wastewater, 2004, 20(1): 55-57(in Chinese).
- [3] 庾婧艺, 徐冰峰, 徐悦, 等. 优化 RBF 神经网络控制水厂混凝剂投加的研究[J]. 中国农村水利水电, 2021(8): 212-215, 220.  
TUO Jingyi, XU Bingfeng, XU Yue, *et al.* Optimized RBF neural networks predicts coagulation in waterworks[J]. China Rural Water and Hydropower, 2021(8): 212-215, 220(in Chinese).
- [4] 白桦, 李圭白. 基于神经网络的混凝投药系统预测模型[J]. 中国给水排水, 2002, 18(6): 46-47.  
BAI Hua, LI Guibai. Prediction model of coagulant dosing system established based on neural network[J]. China Water & Wastewater, 2002, 18(6): 46-47(in Chinese).
- [5] OLADIPUPO B, YSKANDAR H, KARIM D. Coagulation process control in water treatment plants using multiple model predictive control[J]. Alexandria Engineering Journal, 2014, 53(4): 939-948.
- [6] 刘旺, 魏郭子建, 施常洁, 等. 遗传算法优化的 BP 神经网络预测水厂絮凝剂投加量[J]. 净水技术, 2023, 42(1): 60-68.  
LIU Wang, WEI Guozijian, SHI Changjie, *et al.* Prediction of flocculant dosage in WTP with BP neural network optimized by genetic algorithm[J]. Water Purification Technology, 2023, 42(1): 60-68(in Chinese).
- [7] 刘洪波, 黄剑虹, 张国荣, 等. 给水厂混凝剂智能投加模型构建与应用[J]. 上海理工大学学报, 2022, 44(4): 351-356, 387.  
LIU Hongbo, HUANG Jianhong, ZHANG Guorong, *et al.* Modeling and application of intelligent coagulant dosing model in drinking water plant[J]. Journal of University of Shanghai for Science and Technology, 2022, 44(4): 351-356, 387(in Chinese).
- [8] 杨存满, 鞠佳伟, 袁芳, 等. 基于 PSO-BP 神经网络的水厂智能消毒预测模型[J]. 中国给水排水, 2022, 38(3): 57-61.  
YANG Cunman, JU Jiawei, YUAN Fang, *et al.* Research on intelligent disinfection prediction model of waterworks based on PSO-BP neural network[J]. China Water & Wastewater, 2022, 38(3): 57-61(in Chinese).
- [9] LI J S, WANG Y D, TIAN W L. Singular value detection of genetic algorithm optimizing RBF neural network[J]. Information Technology Journal, 2013, 12(11): 2201-2206.

作者简介:王坤(1980-),男,安徽亳州人,硕士,副教授,主要研究方向为水处理理论与技术。

E-mail:wangkun@ahjzu.edu.cn

收稿日期:2024-08-06

修回日期:2024-11-12

(编辑:任莹莹)