

文章编号: 1006-3080(2026)03-0428-11

DOI: 10.14135/j.cnki.1006-3080.20250703001

基于 BPNN 的焦化废水生化处理的智慧预测和调控

常定明^{1,2}, 雷 真³, 李 杰¹, 胡 鹏¹, 杨正武¹, 张芯婉¹, 胡华为¹, 张乐华¹(1. 华东理工大学资源与环境工程学院, 上海 200237; 2. 江苏建霖环保科技有限公司, 江苏 扬州 225100;
3. 华东理工大学国际卓越工程师学院, 上海 200237)

摘要:焦化厂污水生化处理过程复杂, 难以精准调控药剂投加量达到稳定出水的效果。本研究采用基于反向传播神经网络(BPNN)的智慧调控方法, 实现对焦化废水生化出水水质和药剂投加量的精准预测。结果表明, 水质预测模型的决定系数(R^2)达到 0.891, 药剂投加量预测模型的 R^2 分别为 0.605、0.798 和 0.809。数据清洗和特征选择显著提升了模型的预测性能, 双输出模型的预测效果优于单输出模型。本研究可为焦化行业生化处理过程及其加药装置的智慧调控提供理论参考。

关键词:神经网络; 水质预测; 智慧调控; 数据预处理; 焦化废水

中图分类号:X703.1

文献标志码:A

焦化废水水质水量波动大, 并含有大量硫氰化物、氨氮、硫化物、酚等有毒有害物质, 是一种典型的难降解废水。焦化废水常采用预处理-缺氧/好氧-深度处理工艺, 出水可达到《炼焦化学工业污染物排放标准》(GB16171—2012)直接排放标准或者进入中水单元进行回用处理。由于大量氨氮存在, 好氧段发生硝化反应, 需要投加碱来保持好氧池的 pH 值。另外, 在深度处理过程中, 通常投加氧化剂与吸附剂用于保证出水化学需氧量(COD)达到排放标准。然而, 蒸氨废水水量和水质波动较大, 为保证出水水质达标, 需要投加过量的碱和水处理药剂, 造成运行成本较高^[1-3]。

机器学习主要包括监督学习、无监督学习和强化学习, 核心优势在于能够从大量数据中提取关键特征, 而无需依赖传统的机理模型^[4-5], 广泛应用于水质预测和药剂投加量调控^[6-8]。例如, 水质预测方面, 支持向量机(SVM)、随机森林(RF)和人工神经网络(ANN)等算法已被用于预测污水处理厂的出水水质^[9-11]。深度学习技术如卷积神经网络(CNN)和长短期记忆网络(LSTM)在处理复杂非线性问题时表

现出色^[12-14], 复合模型和优化算法的引入则进一步提升了预测精度和模型的泛化能力^[15]。在药剂投加量调控方面, 阳艳芳^[16]研制出基于单片机的水处理自动加药系统, 该系统通过引入反馈控制逻辑, 能够适应不同水质条件和工艺流程, 有效降低了约 30% 的药剂消耗。此后, 国内相继发展出两类水处理自动加药系统: 一类是基于在线水质参数传感的定值调节系统, 另一类则是基于 DSC 或 PLC/based-PC 的过程控制系统, 相关技术持续迭代升级, 推动了自动加药技术的应用与推广^[17-18]。尽管机器学习在水处理中的应用取得了显著进展^[19-22], 但仍面临一些挑战。首先, 数据质量对模型性能的影响较大, 数据清洗和特征选择是提升预测精度的关键步骤。其次, 单一模型往往存在局限性, 复合模型和集成学习方法的研究仍需进一步深入。

本研究旨在通过建立反向传播神经网络(BPNN)模型, 实现焦化废水生化出水水质和药剂投加量的精准预测; 探究数据清洗和特征选择对模型性能的影响, 对比双指标输出与单指标输出预测的性能差异。通过上述研究内容为水处理加药装置的智慧调

收稿日期: 2025-07-03

基金项目: 2023 年鸡西市石墨产业揭榜挂帅项目(JKJB2023H03); 新疆天山英才项目

作者简介: 常定明(1989—), 男, 江苏盐城人, 博士, 高级工程师, 研究方向为工业水污染智慧控制与资源化; E-mail: changdingming123@163.com

通信联系人: 张乐华, E-mail: lezhanghua@ecust.edu.cn

引用本文: 常定明, 雷 真, 李 杰, 等. 基于 BPNN 的焦化废水生化处理的智慧预测和调控[J]. 华东理工大学学报(自然科学版), 2026, 52(3): 428-438.

Citation: CHANG Dingming, LEI Zhen, LI Jie, *et al.* Intelligent Prediction and Control Technology of Coking Wastewater Bio-Treatment Based on BPNN Model[J]. Journal of East China University of Science and Technology, 2026, 52(3): 428-438.

控提供理论支持和技术参考。

1 材料与方法

1.1 数据集来源

数据集来源为山西某焦化厂污水处理系统2024年1~10月的每日数据(除流量外,其他数据为日均值),共296条数据,每条包含12项特征参数。其中,进水水质和流量参数6项(进水COD、进水硫氰化物质量浓度、进水氨氮质量浓度、进水硫化物质量浓度、进水pH、进水流量),药剂投加量参数3项(氢氧化钠、净水剂02、净水剂904),出水水质和流量参数3项(出水氨氮质量浓度、出水COD、出水流量)。

1.2 水质水量检测方法

氨氮的检测方法:《水质 氨氮的测定 纳氏试剂分光光度法》(HJ 535—2009);COD的检测方法:《水质 化学需氧量的测定 重铬酸盐法》(HJ 828—2017);硫氰化物的检测方法:《焦化废水中硫氰酸盐含量的测定》(HG/T 6072—2022);pH的检测方法:《水质

pH值的测定 电极法》(HJ 1147—2020);硫化物的检测方法:《水质 硫化物的测定 亚甲基蓝分光光度法》(HJ 1226—2021);流量的检测方法:电磁流量计(型号LDG-6-10-00,上海光华仪表有限公司)。

1.3 数据清洗

中位数绝对偏差(MAD)计算:

$$MAD = b \times \text{Median}(|x - \mu|) \quad (1)$$

式中, μ 是样本参数的中位数值, b 为常数。

基于中位数、MAD值和阈值(threshold)设置达到离群点的判断边界(式2)。

$$\mu - (\text{threshold} \times MAD) < x < \mu + (\text{threshold} \times MAD) \quad (2)$$

对于识别出的离群点,使用3次样条插值法进行修正,得到完成清洗后的数据序列。

1.4 BPNN模型构建

为训练机器学习模型,一般将数据分为训练集、测试集和验证集3个部分^[23]。在本模型中,数据集以70:15:15的比例划分为训练集、验证集和测试集。原始数据集的统计信息如表1所示(文中氨氮用AN表示,余同)。

表1 原始数据统计信息表

Table 1 Original data statistical information table

Statistical indicator	$\rho(\text{Inlet AN}/(\text{mg} \cdot \text{L}^{-1}))$	Inlet pH	Inlet COD/(\text{mg} \cdot \text{L}^{-1})	$\rho(\text{Inlet sulfide}/(\text{mg} \cdot \text{L}^{-1}))$	$\rho(\text{Inlet thiocyanide}/(\text{mg} \cdot \text{L}^{-1}))$	Inlet flow/t	Outlet flow/t	$m(\text{NaOH dosing})/\text{kg}$	$m(\text{Purifier 904})/\text{kg}$	$m(\text{Purifier 02})/\text{kg}$	$\rho(\text{Outlet AN})/(\text{mg} \cdot \text{L}^{-1})$	Outlet COD/(\text{mg} \cdot \text{L}^{-1})
Minimum	83.067	6.000	3187.000	3.700	312.600	521.000	636.000	650.000	0	1000.000	0	44.000
Maximum	293.958	9.625	5920.000	339.590	682.400	1261.000	1860.000	2250.000	7516.800	2100.000	127.175	115.650
Average	169.599	7.705	4470.789	109.976	486.458	1008.759	1362.700	1272.154	1262.487	1692.877	12.731	73.817
Standard deviation	32.478	0.834	450.294	98.030	66.394	98.756	209.979	436.407	1375.202	281.643	22.993	13.970

在前向传播阶段,输入数据通过输入层进入网络后逐层计算神经元的输出,直到产生输出层的输出。具体过程如下:

步骤1,对于隐藏层中的每个神经元,其输入为前一层所有神经元输出的加权和,即:

$$z_j = \sum_i w_{ij} a_i + b_j \quad (3)$$

其中, z_j 表示第 j 个神经元的输入, w_{ij} 表示第 i 个神经元到第 j 个神经元的权重, a_i 表示第 i 个神经元的输出(对于输入层, a_i 即为输入数据), b_j 表示第 j 个神经元的偏置。

步骤2,通过激活函数计算神经元的输出:

$$a_j = f(z_j) \quad (4)$$

其中, f 为激活函数,常用的激活函数有 sigmoid、

ReLU等。

在反向传播阶段,首先计算网络的实际输出与期望输出之间的误差,然后将误差反向传播到网络中,通过调整权重和偏置来减小误差。具体过程如下:

步骤1,计算输出层的误差:

$$E = \frac{1}{2} \sum_k (y_k - \hat{y}_k)^2 \quad (5)$$

其中, E 表示误差, y_k 表示第 k 个期望输出, $\hat{y}_k$ 表示第 k 个实际输出。

步骤2,对于输出层的每个神经元,计算其误差项:

$$\delta_k = \frac{\partial E}{\partial z_k} = (y_k - \hat{y}_k) \cdot f'(z_k) \quad (6)$$

其中, δ_k 表示第 k 个神经元的误差项, $f'(z_k)$ 表示激活函数 f 在 z_k 处的导数。

步骤 3, 对于隐藏层的每个神经元 δ_j , 计算其误差项:

$$\delta_j = \sum_k w_{jk} \delta_k \cdot f'(z_j) \quad (7)$$

其中, w_{jk} 为误差从输出层传回隐藏层时的权重。

步骤 4, 根据误差项更新权重和偏置:

$$w'_{ij} = w_{ij} - \eta \cdot \delta_j \cdot a_i \quad (8)$$

$$b_j = b_j - \eta \cdot \delta_j \quad (9)$$

其中, w'_{ij} 为根据隐藏层误差实际被修正的权重, a_i 表示第 i 个神经元的输出, η 表示学习率, 控制权重和偏置更新的步长。

步骤 5, 重复以上步骤, 直到满足停止条件(误差小于阈值或达到最大迭代次数)^[24]。

1.5 出水水质预测

为了探究数据清洗程度对模型性能的影响, 设定两个同源数据集: 其中一个为仅接受缺失值补充操作的数据集; 而另一个为完成了离群值替换工作的数据集。选择进水指标和药剂指标(进水氨氮质量浓度、进水 COD、进水硫氰化物质量浓度、进水 pH、进水硫化物质量浓度、进水流量、出水流量、氢氧化钠质量、净水剂 02 质量、净水剂 904 质量)作为输入参数, 将两个出水指标(出水氨氮质量浓度和出水 COD)定为输出参数。

为了找到性能最好的 BPNN 模型结构, 使用网格搜索法分别在隐含层数为 1、2、3 时对各层隐含层的神经元数进行遍历。对于每个隐含层, 神经元数目的遍历范围遵循的经验公式为:

$$N_h = \sqrt{N_{in} + N_{out}} + \alpha \quad (10)$$

式中, N_h 是神经元数目, N_{in} 是输入参数数量, N_{out} 是输出参数数量, $\alpha \in [1, 10]$ 。

1.6 药剂投加量预测

针对药剂(氢氧化钠、净水剂 02、净水剂 904)投加量的预测, 需要对数据集进行调整。使用氨氮变化量和 COD 变化量取代进、出水氨氮质量浓度和 COD 作为输入参数, 其余输入参数不变, 以便于实际应用。在模型训练过程中, 这两个指标分别通过进、出水氨氮质量浓度、进、出水 COD 的差值来获得。在实际中, 使用相关规定中的氨氮质量浓度和 COD 排放标准来代替出水氨氮质量浓度和 COD。

为了寻找最优的药剂投加量预测模型, 使用网格搜索法, 在隐含层数目不同的条件下(1、2、3 层)分别进行遍历, 找寻最优模型结构。

2 结果与讨论

2.1 数据分析与清洗

图 1 是进水水质与流量参数的相关性图, 图中每个色块的数字代表对应横纵坐标参数之间的相关性强弱。数值的正负性代表两参数之间的正负相关性, 其绝对值大小则代表相关性的强弱, 绝对值为 0 表示两者无关, 绝对值为 1.00 表示两者完全相关。从图 1 中可以看出所选择的 7 项参数中, 只有出水流量分别与进水硫化物浓度和进水流量的相关性指数达到 0.60, 其余参数彼此之间的相关性指数均小于

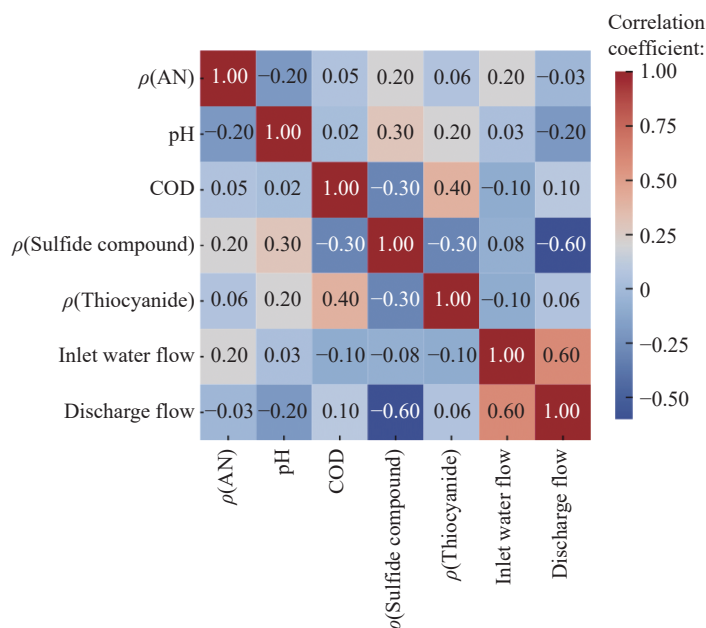


图 1 进水水质与流量参数相关性图

Fig. 1 Correlation diagram of influent water quality and flow parameters

0.50,说明进水水质与流量参数有较高的独立性,作为输入参数是可靠的。数据清洗前后的对比图和数据归一化后的各参数分布情况见图2和图3。

2.2 出水水质预测

2.2.1 数据清洗程度对模型性能的影响 表2示出了输出预测模型的最优结构和性能。从表2中可以

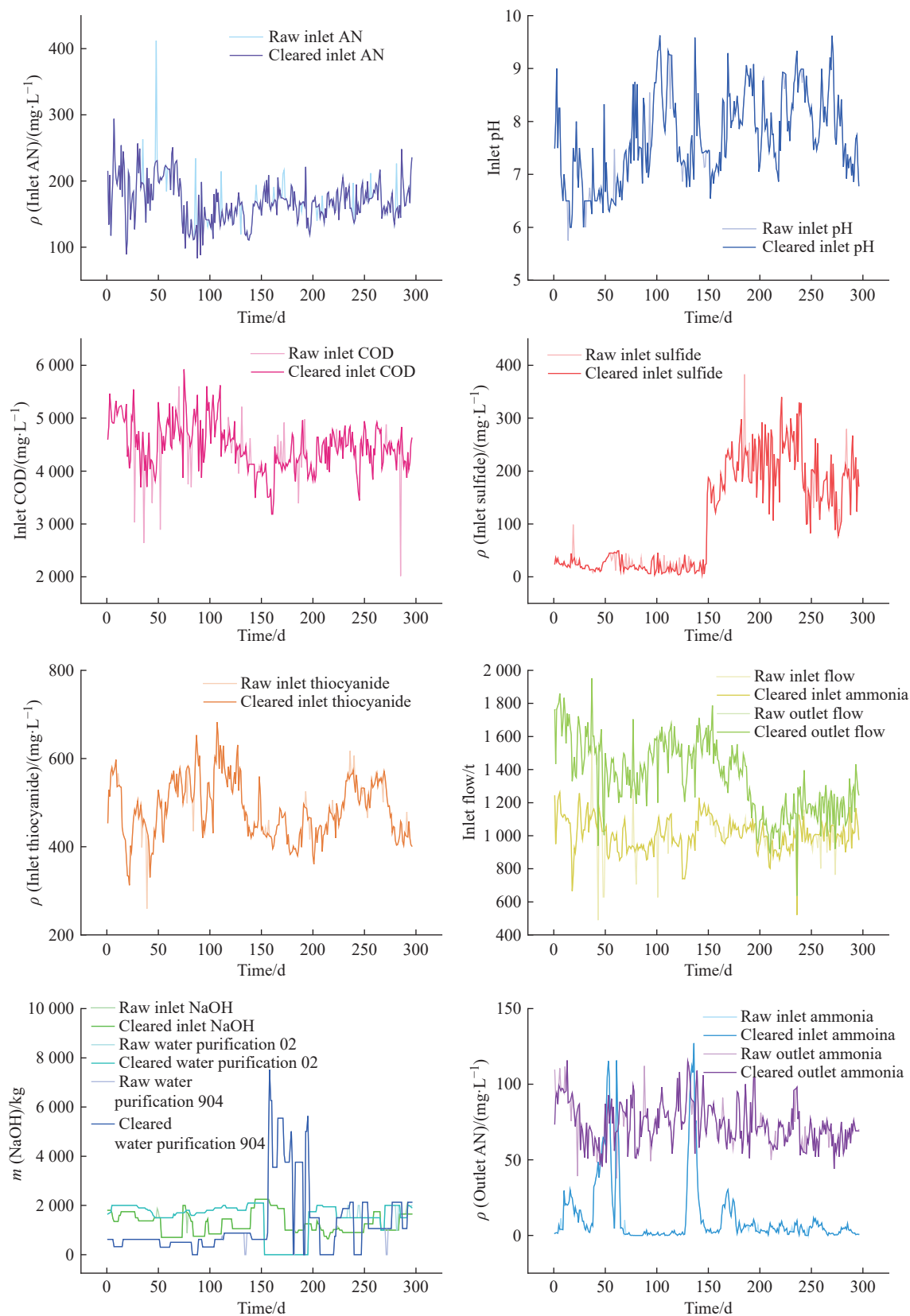


图2 数据清洗前后的数据对比图

Fig. 2 Data comparison diagram before and after data cleaning

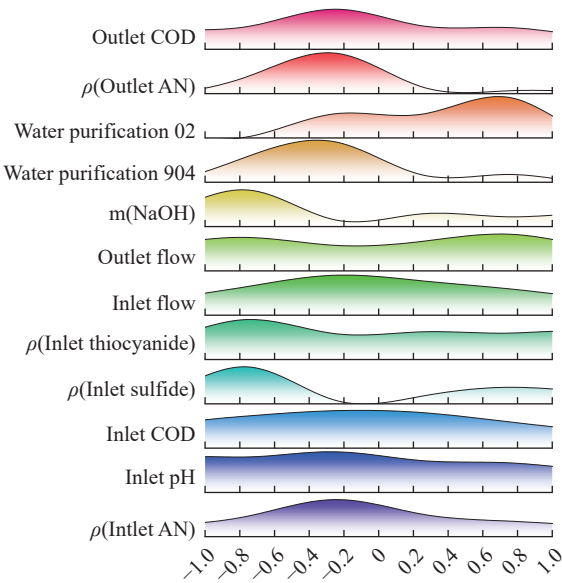


图 3 数据归一化后的各参数分布情况

Fig. 3 Distribution of the parameters after data normalization

看出,在所有的模型中,完成替换离群值且隐含层层数为 2 的结构拥有最低的均方根误差(RMSE, 8.229)和最高的测试集决定系数(R^2 , 0.891)。说明此模型的拟合优度最高,对出水水质的预测能力最可靠。确定该模型为最优模型,各层参数或神经元数目依次为“10-13-12-2”,图 4 为该最优模型的结构图(w 为权重矩阵, b 为偏置向量)。

图 5 所示本文模型的训练集、验证集、测试集的回归相关系数分别为 0.959、0.930、0.894 和 0.944。

这些结果不仅说明模型训练过程稳定,未出现过拟合现象,更重要的是,测试集上的高性能表明模型对未知数据具有良好的泛化能力,预测输出与真实值相关性很强(图 6),充分证明了其在实际应用中具备较高的可靠性。若模型应用场景下训练效果未达到预期,可通过调整隐含层数和各层神经元数量等结构超参数进行优化迭代。

2.2.2 单输出和双输出预测模型的对比 选择双输出模型中预测效果最优的双隐含层结构进行单输出预测,模型包含一层输入层、两层隐含层和一层输出层,通过网格搜索法优化模型参数,确定隐含层神经元数分别为 13 和 6 个。在单输出预测输入层中输入与双输出预测同样的参数,将输出层调整为单输出,对出水氨氮质量浓度和出水 COD 分别进行预测。表 3 所示为在不同数据清洗程度条件下,单输出预测模型训练得到的最优模型结构和模型性能评估值。可以得到出水氨氮质量和 COD 的预测值在测试集决定系数 R^2 分别只有 0.334 和 0.278,说明此条件下模型的预测能力非常差,无法进行预测工作。

图 7 是单输出双隐含层模型的结构图,经过网格搜索法遍历后,隐含层的神经元数最终确定在 13 和 6 个。图 8 和图 9 分别是出水氨氮质量浓度和 COD 预测值与实际值的回归图,其中氨氮质量浓度的预测模型在训练集上的相关系数达到了 0.920,但在验证集和测试集上却分别只有 0.725 和 0.576,因此推测单输出模型在训练时发生了严重的过拟合,

表 2 输出预测模型的最优结构和性能

Table 2 Optimal model structure and model performance of the output prediction model

Degree of data cleaning	Number of hidden layer	Number of neuron in each layer	Performance evaluation		
			MSE	RMSE	Test set R^2
Fill missing value (Replacement of outlier)	1	10-10-2	95.684	9.782	0.860
Fill missing value (Unreplaced outlier)	1	10-11-2	100.440	10.022	0.768
Fill missing value (Replacement of outlier)	2	10-13-12-2	67.716	8.229	0.891
Fill missing value (Unreplaced outlier)	2	10-12-7-2	76.886	8.768	0.814
Fill missing value (Replacement of outlier)	3	10-13-12-11-2	88.697	9.418	0.795
Fill missing value (Unreplaced outlier)	3	10-12-11-4-2	89.980	9.486	0.782

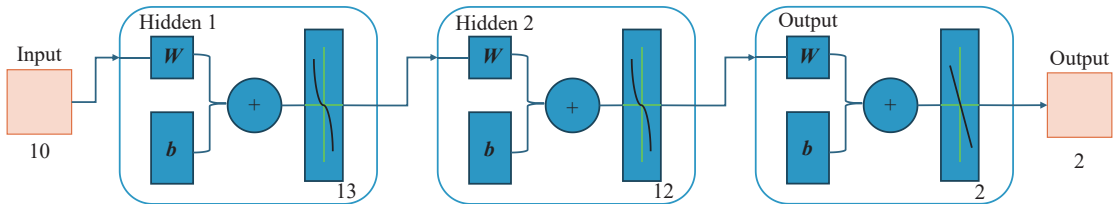


图 4 最优双输出双隐含层的最优模型结构图

Fig. 4 Optimal model structure diagram of the optimal double output double hidden layer model

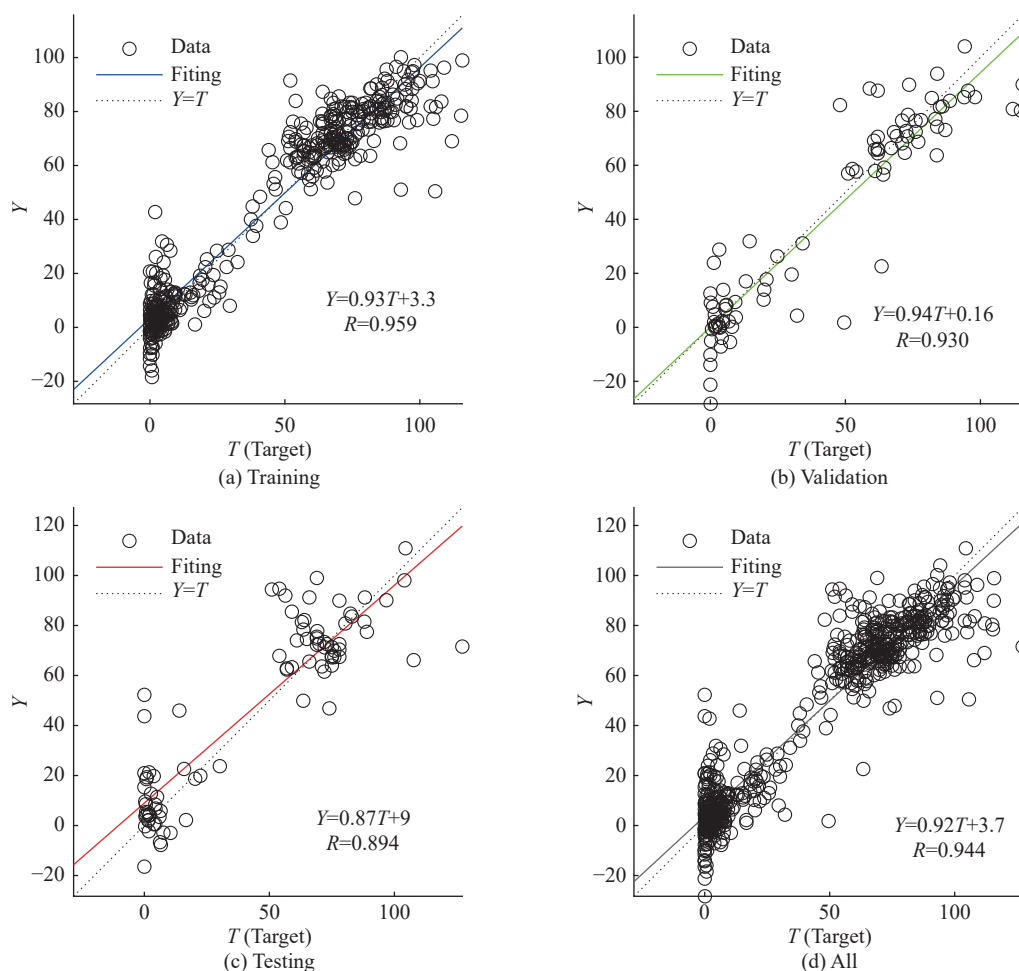


图5 本文模型预测值对实际值的回归图

Fig. 5 Regression plot of predicted vs. actual of the model in this paper

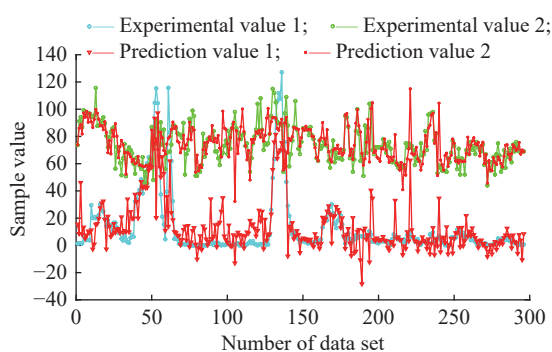


图6 模型预测值与实际值对比图

Fig. 6 Comparison diagram between the predicted value and the actual value of the model

导致其在验证和测试时表现极差。

2.3 药剂投加量预测

表4汇总了加药量单输出预测模型在不同隐含层数条件下的最优模型结构和预测性能。图10绘制了各模型测试集决定系数 R^2 和均方根误差RMSE的二维评价图。结合图表可以得知,针对NaOH投加量的预测模型在隐含层层数为2时取得最高决定

系数(0.605)和最低均方根误差(191.379),模型各层组成为7-12-12-1结构。针对净水剂904投加量的预测模型在隐含层层数为3时取得最高决定系数(0.798)和最低均方根误差(502.114),模型各层组成为7-12-12-12-1结构。针对净水剂02投加量的预测模型也在隐含层层数为3时取得最高决定系数(0.809),同时均方根误差为502.114,模型各层组成为7-12-12-12-1结构。结果表明,使用单输出预测模型能够对药剂投加量进行预测,但对NaOH投加量的预测效果较差。

2.4 出水水质与药剂投加量的关联分析

图11展示了出水水质随药剂投加量变化的情况。由图11(a)可得,随着NaOH投加量的增加,氨氮质量浓度先上升后下降,在1100 kg时达到峰值,且氨氮质量浓度随净水剂904投加量增加而升高。由图11(b)可得,当净水剂904投加量小于1000 kg时,COD随NaOH投加量先下降后上升,在1500 kg时达到最低;超过1000 kg时,COD先上升后下降。由图11(c)可得,随着净水剂02的增加,氨氮质量浓

表 3 单输出预测模型训练得到的最优模型结构和模型性能表

Table 3 Optimal model structure and model performance table obtained by training the single output prediction model

Parameter	Degree of data cleaning	Number of hidden layer	Number of neuron in each layer	Performance evaluation		
				MSE	RMSE	Test set R^2
AN	Fill missing value (Replacement of outlier)	2	10-13-6-1	101.550	10.077	0.334
COD	Fill missing value (Replacement of outlier)	2	10-13-6-1	54.650	7.393	0.278

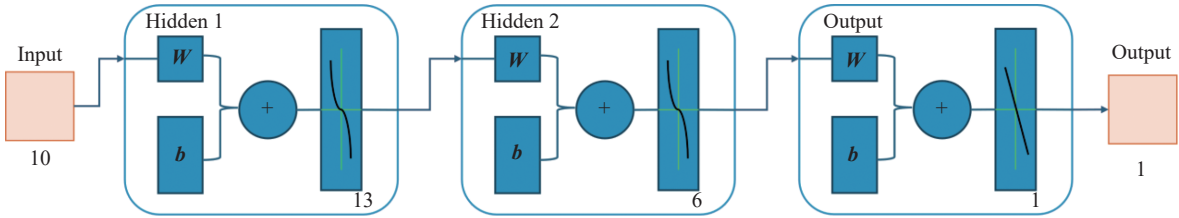


图 7 单输出双隐含层模型结构图

Fig. 7 Model structure diagram of single output double hidden layer

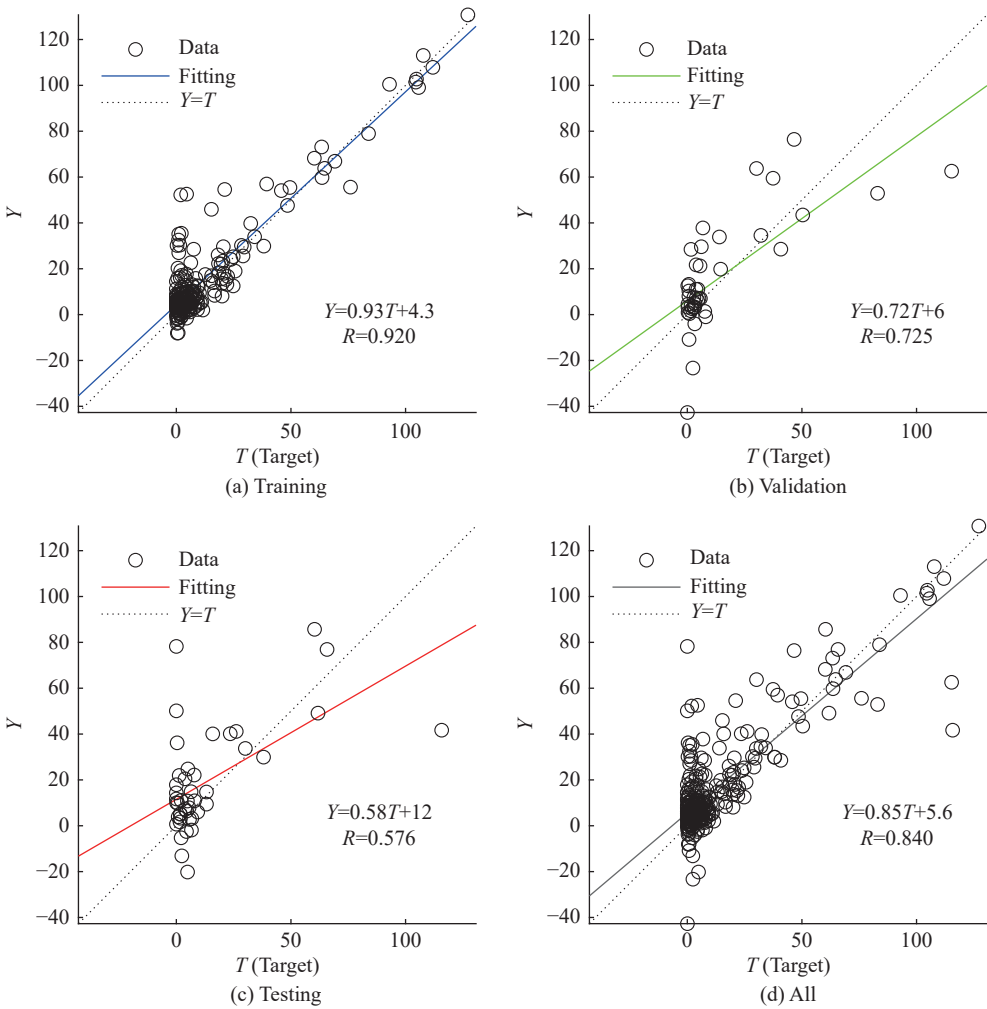


图 8 出水氨氮质量浓度预测值对实际值的回归图

Fig. 8 Regression plot of predicted vs. actual effluent AN mass concentration

度整体下降,但在 NaOH 投加量为 1 100 kg 时,所需净水剂 O2 最多。由图 11(d)可得,当净水剂 O2 投加量在 800~1 600 kg、NaOH 投加量在 1 400~2 000 kg 时,COD 最低。由图 11(e)可得,氨氮质量浓度随净水剂 904

增加而上升,随净水剂 O2 增加而下降。由图 11(f)可得,COD 随净水剂 904 增加而上升,当净水剂 904 投加量小于 1 500 kg 时,COD 随净水剂 O2 增加先下降后上升,在 1 100 kg 时达到最低。

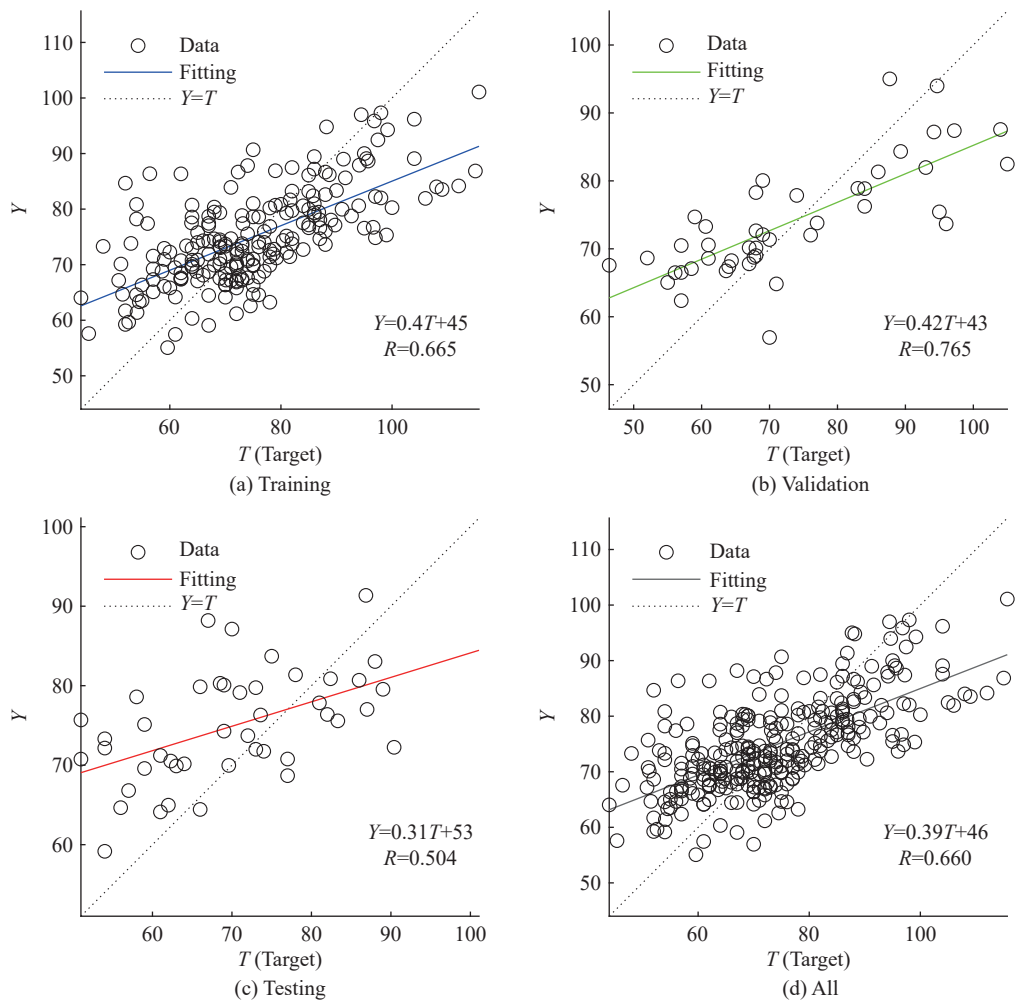


图 9 出水 COD 预测值对实际值的回归图

Fig. 9 Regression plot of predicted value vs. actual effluent COD

表 4 加药量单输出预测模型在不同隐含层数时的最优模型结构和性能

Table 4 Optimal model structure and performance of dosage single output prediction model at different hidden layers

Chemical agent	Number of hidden layer	Number of neuron in each layer	Performance evaluation		
			MSE	RMSE	Test set R^2
NaOH	1	7-12-1	41 332.230	203.303	0.522
Purifier 904	1	7-12-1	346 441.042	588.592	0.596
Purifier 02	1	7-12-1	63 468.463	251.930	0.709
NaOH	2	7-12-12-1	36 626.075	191.379	0.605
Purifier 904	2	7-12-12-1	255 470.358	505.441	0.614
Purifier 02	2	7-12-12-1	35 229.223	187.694	0.592
NaOH	3	7-12-12-12-1	36 867.625	192.009	0.593
Purifier 904	3	7-12-12-12-1	252 118.297	502.114	0.798
Purifier 02	3	7-12-12-12-1	38 945.682	197.347	0.809

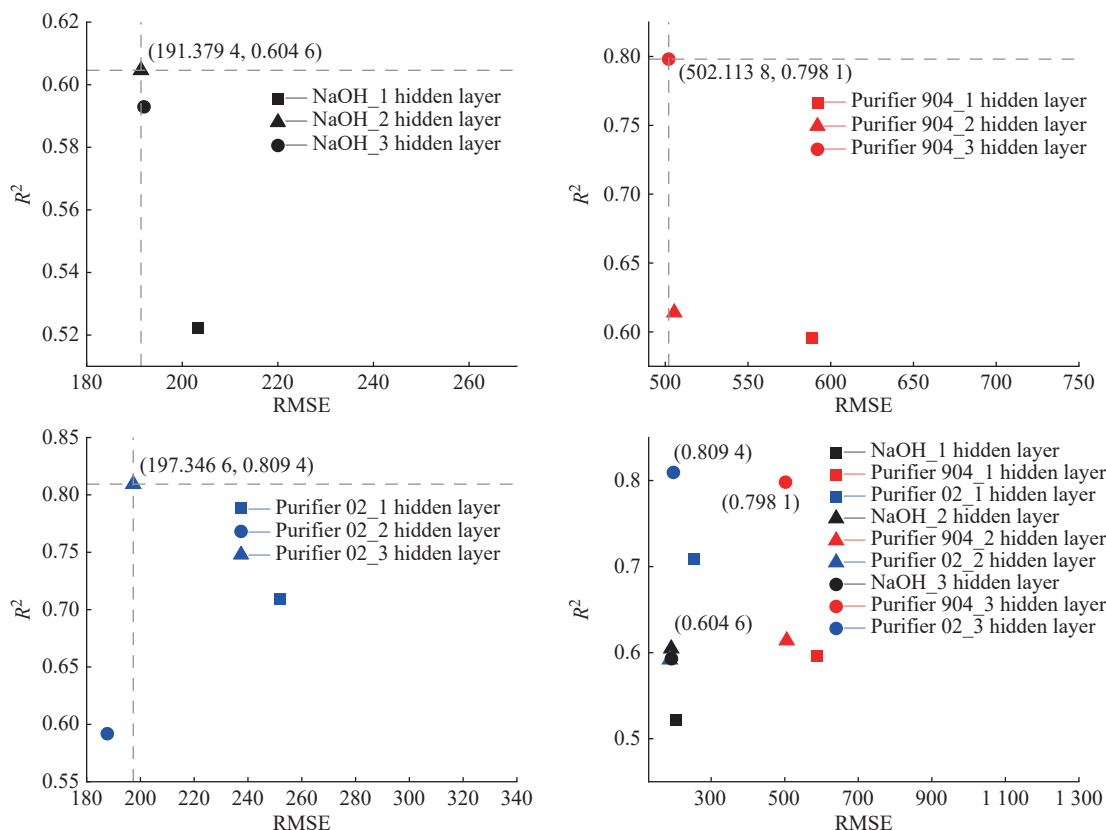
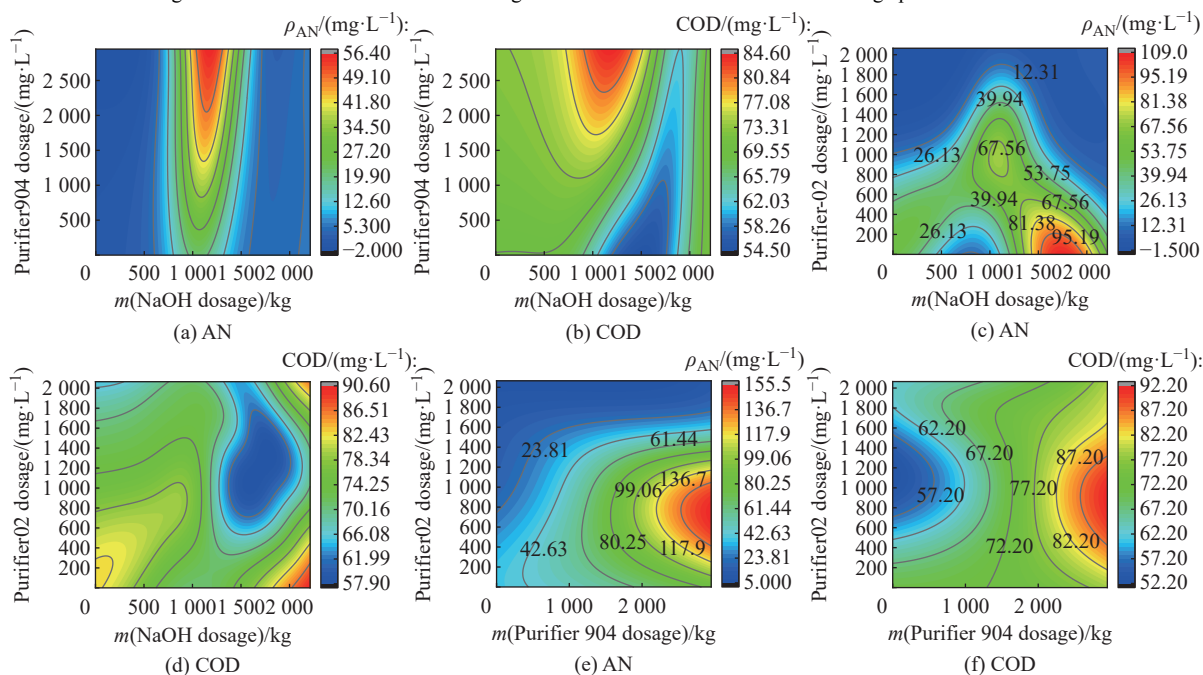
图 10 各模型药剂投加量测试集 R^2 和 RMSE 二维评价图Fig. 10 Two-dimensional evaluation diagram of R^2 and RMSE of each model dosage prediction test set

图 11 出水水质随药剂投加量变化的等高线图

Fig. 11 Contour diagram of effluent water quality with the charge of reagent dosage

3 结束语

本研究针对焦化行业污水处理厂进出水水质和水处理药剂投加量等 12 条特征参数建立了基于网格

搜索法优化的 BPNN 模型, 较好地实现对污水处理厂出水水质和药剂投加量的预测。水质预测模型的最佳模型结构为 10-13-12-2, 并能够用于探究出水水质与药剂投加量的变化关系。数据清洗和特征选择能够显著提升 BPNN 模型的预测性能。在出水水质

预测模型中,双输出预测模型的性能优于两个指标的单输出预测模型。本研究为焦化行业污水处理的智能化调控提供了理论基础和技术支撑,有助于推动污水处理系统的精细化管理与节能降耗。

焦化行业水质波动显著,药剂投加量大,因此对智慧加药模型的需求较为迫切。本研究构建的模型能够较好地预测焦化行业污水处理厂的出水水质和药剂投加量,在节能降耗、提升系统运行稳定性方面展现出较高的实际应用价值,具备良好的落地潜力。然而在目前实验与验证过程中,仍有以下几个方面需要进一步完善:(1)焦化行业水质波动剧烈,极端工况下的数据样本仍相对有限,需进一步扩充数据量以提升模型的泛化能力;(2)生化出水容易对电极或光学探头造成污染,传感器的长期稳定性和可靠性仍是实际部署中的关键挑战。未来我们将结合机理模型、数据驱动方法和容错控制策略,在实际项目中不断迭代与优化该系统,以推动其工程化应用。

参考文献:

- [1] 孙贤波,钱飞跃,庞丽娜,等.焦化废水生化出水总氰构成及去除特性[J].华东理工大学学报(自然科学版),2011,37(3): 335-339.
- [2] 张万辉,韦朝海.焦化废水的污染物特征及处理技术的分析[J].化工环保,2015,35(3): 272-278.
- [3] 郭忠良,李培,李森.焦化废水处理技术研究进展[J].中国资源综合利用,2023,41(12): 155-158.
- [4] SIGMUND G, GHARASOO M, HÜFFER T, *et al.* Deep learning neural network approach for predicting the sorption of ionizable and polar organic pollutants to a wide range of carbonaceous materials[J]. Environmental Science & Technology, 2020, 54(7): 4583-4591.
- [5] 陈春霞.浅析机器学习的发展与应用[J].信息系统工程,2017(8): 99-100.
- [6] LOWE M, QIN R W, MAO X W. A review on machine learning, artificial intelligence, and smart technology in water treatment and monitoring[J]. Water, 2022, 14(9): 1384.
- [7] 成睿,孟广源,殷瑶,等.神经网络 BPNN 模型机器学习方法应用于电化学去除氨氮过程的预测与优化[J].华东理工大学学报(自然科学版),2023,49(2): 202-210.
- [8] 李思敏,产青青,金鑫,等.机器学习在水务行业中的应用现状与发展前景[J].水电能源科学,2024(3): 43-48.
- [9] LIU M, LU J. Support vector machine: An alternative to artificial neuron network for water quality forecasting in an agricultural nonpoint source polluted river?[J]. Environmental Science and Pollution Research, 2014, 21(18): 11036-11053.
- [10] HAGHIABI A H, NASROLAHI A H, PARSIAIE A. Water quality prediction using machine learning methods[J]. Water Quality Research Journal, 2018, 53(1): 3-13.
- [11] CASTRASTRILLO M, GARCÍA Á L. Estimation of high frequency nutrient concentrations from water quality surrogates using machine learning methods[J]. Water Research, 2020, 172: 115490.
- [12] WU Y, ZHANG X, XIAO Y, *et al.* Attention neural network for water image classification under IoT environment[J]. Applied Sciences, 2020, 10(3): 909.
- [13] ŠILJIĆ TOMIĆ A, ANTANASIJEVIĆ D, RISTIĆ M, *et al.* A linear and non-linear polynomial neural network modeling of dissolved oxygen content in surface water: Inter- and extrapolation performance with inputs' significance analysis[J]. Science of the Total Environment, 2018, 610-611: 1038-1046.
- [14] MA J, DING Y X, CHENG J C P, *et al.* Soft detection of 5-day BOD with sparse matrix in city harbor water using deep learning techniques[J]. Water Research, 2020, 170: 115350.
- [15] CHEN K Y, CHEN H X, ZHOU C L, *et al.* Comparative analysis of surface water quality prediction performance and identification of key water parameters using different machine learning models based on big data[J]. Water Research, 2020, 171: 115454.
- [16] 阳艳芳.单片机在水处理自动加药系统中的应用[J].电子技术应用,1995(4): 6-7, 21.
- [17] 曹龙.给水处理厂加药系统的控制研究[D].沈阳:东北大学,2014.
- [18] 孙奎.污水处理自动加药系统设计与应用[J].科技创新与应用,2017(4): 27-28.
- [19] MAN Y, HU Y S, REN J Z. Forecasting COD load in municipal sewage based on ARMA and VAR algorithms[J]. Resources, Conservation and Recycling, 2019, 144: 56-64.
- [20] WANG Z F, MAN Y, HU Y S, *et al.* A deep learning based dynamic COD prediction model for urban sewage[J]. Environmental Science: Water Research & Technology, 2019, 5(12): 2210-2218.
- [21] BAGHERZADEH F, MEHRANI M J, BASIRIFARD M, *et al.* Comparative study on total nitrogen prediction in wastewater treatment plant and effect of various feature selection methods on machine learning algorithms performance[J]. Journal of Water Process Engineering, 2021, 41: 102033.
- [22] HEO S, NAM K, TARIQ S, *et al.* A hybrid machine learning-based multi-objective supervisory control strategy of a full-scale wastewater treatment for cost-effective and sustainable operation under varying influent conditions[J]. Journal of Cleaner Production, 2021, 291: 125853.
- [23] HUANG R X, MA C X, MA J, *et al.* Machine learning in natural and engineered water systems[J]. Water Research, 2021, 205: 117666.

[24] 安昱宁. 基于深度学习的污水处理厂出水水质预测建模

研究 [D]. 山东青岛: 青岛理工大学, 2024.

Intelligent Prediction and Control Technology of Coking Wastewater Bio-Treatment Based on BPNN Model

CHANG Dingming^{1,2}, LEI Zhen³, LI Jie¹, HU Peng¹, YANG Zhengwu¹, ZHANG Xinwan¹,
HU Huawei¹, ZHANG Lehua¹

(1. School of Resources and Environmental Engineering, East China University of Science and Technology, Shanghai 200237, China; 2. Jiangsu Jianlin Environmental Technology Co. Ltd, Yangzhou 225100, Jiangsu, China; 3. International Elite Engineering School, East China University of Science and Technology, Shanghai 200237, China)

Abstract: The biochemical treatment of coking wastewater is a complex process, so accurate regulation of chemical dosing to ensure stable effluent quality poses a great challenge. In this study, an intelligent control model based on a back propagation neural network (BPNN) was proposed to achieve precise prediction of effluent quality and chemical dosage in the biochemical treatment process of coking wastewater. The results demonstrated that the coefficient of determination (R^2) of the effluent quality prediction model reached 0.891. The R^2 values of the chemical dosage prediction models were 0.605, 0.798, and 0.809, respectively. Data preprocessing techniques, including data cleaning and feature selection, significantly enhanced the predictive performance of the models. Moreover, the dual-output model exhibited superior predictive accuracy than the single-output models. This study provides a theoretical reference for the intelligent regulation of biochemical treatment processes and chemical dosing systems in the coking industry

Key words: neural network; water quality prediction; intelligent control; data cleaning; coking wastewater

(责任编辑: 王晓丽)