

基于 ICS 优化 RBF 的水库水质三维预测方法

谢再秘¹ 贾宝柱¹ 王 骥² 莫春梅²

(1. 广东海洋大学船舶与海运学院, 湛江 524088; 2. 广东海洋大学电子与信息工程学院, 湛江 524088)

摘要: 针对已有水质预测模型在数据降噪、网络参数初始值设置和优化、精度提高等方面能力的不足, 构建了一种优化的水质三维预测模型。利用主成分分析算法筛选出水质关键参数, 并基于自适应噪声的完全集合经验模态分解算法结合小波阈值模型对三维水质参数和气象数据降噪处理, 使用 3 维卷积神经网络 (Three-dimensional convolutional neural networks, 3-D CNN) 提取出特征数据集, 自编码器 (Autoencoder, AE) 获得径向基函数 (Radial basis function, RBF) 网络参数初始值, 改进布谷鸟搜索算法 (Improved cuckoo search, ICS) 优化更新网络中超参数动态初始值。广东省湛江市徐闻县大水桥水库区域 22 个典型在线监测站点以及 6 个手持监测点的实测数据对比验证结果表明, 浊度和藻密度分别与总氮含量强正相关, 叶绿素含量与气温强正相关, 所提出的水质预测模型在 5 个典型精准性评价指标方面优于已有文献方法。研究成果可为管理部门和研究者对水质监测提供参考。

关键词: 水质三维预测; 改进布谷鸟搜索算法; 卷积神经网络; 自编码器; 径向基神经网络

中图分类号: TP391

文献标识码: A

文章编号: 1000-1298(2024)02-0306-09

OSID:



Reservoir Water Quality Three-dimensional Prediction Method Based on ICS Optimization RBF

XIE Zaimi¹ JIA Baozhu¹ WANG Ji² MO Chunmei²

(1. College of Shipping and Maritime Transport, Guangdong Ocean University, Zhanjiang 524088, China

2. College of Electronic and Information Engineering, Guangdong Ocean University, Zhanjiang 524088, China)

Abstract: Aiming at the shortcomings of existing water quality prediction models in data noise reduction, initial setting and optimization of network parameters, and accuracy improvement, an optimized three-dimensional water quality prediction model was constructed. The key parameters of water quality were screened by using the principal component analysis algorithm, the three-dimensional water quality parameters and meteorological data were de-noised based on the fully set empirical mode decomposition algorithm based on adaptive noise combined with the wavelet threshold model, the feature data set was extracted by using the three-dimensional convolutional neural network (3-D CNN) and the dynamic initial values of hyperparameters in radial basis function (RBF) neural networks were optimized by an improved cuckoo search algorithm (ICS) based on autoencoder (AE). The comparison and verification results of the measured data from twenty-two typical online monitoring stations and six handheld monitoring stations in the Dashuiqiao Reservoir area of Xuwen County, Zhanjiang City, Guangdong Province showed that turbidity and algae density were positively correlated with total nitrogen, and chlorophyll was positively correlated with temperature, the proposed water quality prediction model was superior to the existing literature methods in five typical accuracy evaluation indicators. The research results can provide reference for management departments and researchers to monitor water quality. Introducing inertial weight and adjusting position parameters to improve CS to speed up the convergence of RBF network. The autoencoder was used to initialize the initial values of network parameters to avoid the defects of artificial random setting. Adding WT algorithm can effectively reduce the white noise in the decomposition and reconstruction process of the fully set empirical mode decomposition algorithm based on adaptive noise.

Key words: water quality three-dimensional prediction; improved cuckoo search algorithm; convolutional neural network; autoencoder; radial basis neural network

收稿日期: 2023-10-30 修回日期: 2023-12-11

基金项目: 国家自然科学基金项目(52071090)和广东省科技计划项目(2019KZDZX1046)

作者简介: 谢再秘(1996—), 男, 博士生, 主要从事海洋工程与技术研究, E-mail: 1778037486@qq.com

通信作者: 贾宝柱(1974—), 男, 教授, 博士, 主要从事船舶与海洋工程等研究, E-mail: jiazbzh@gdou.edu.cn

0 引言

水库属于封闭式水源水域,其水质受水源质量和气象等多种因素耦合作用,环境敏感度高且难以实现准确的水质变化预测。相关研究表明,封闭水域的浊度、藻密度和叶绿素含量等水质参数是衡量水库水质富营养化现状和健康水平的主要因素^[1-2]。水质参数受气象、环境、人类活动等影响较大,且具有较强的不确定性。因此,构建水质参数与其影响因素之间的评估及预测模型对精准预测和分析水质变化趋势具有重要意义。

水质监测数据会直接影响模型预测精度^[3],监测数据中含有大量噪声及模态混叠会降低数据的可用性,PANDEY 等^[4]使用集合经验模态分解(Ensemble empirical mode decomposition, EEMD)算法应对信号分解过程中存在模态混叠问题,由于引入的白噪声在求和取平均后不能被完全抵消,ZHANG 等^[5]提出完全集合经验模态分解(Complementary ensemble empirical mode decomposition, CEEMD)算法处理分解过程存在残留的噪声,但同时引起模态效应减小以及收敛性能差等问题。SHA 等^[6]通过加入自适应白噪声减小模态效应,且通过采用相对较少的求和取平均次数使其具有更好的收敛性。

在水质预测模型精度方面,符东等^[7]提出了基于 BP 神经网络的水质预测模型,改善预测准确性。石庆兰等^[8]结合双向长短期记忆(Bidirectional long short-term memory, BiLSTM)神经网络与门控循环单元(Gated recurrent unit, GRU)神经网络,提高预测值与实际值吻合度。陈英义等^[9]利用卷积神经网络(Convolutional neural network, CNN)提高 LSTM 特征提取能力来提升模型预测精度。KIM 等^[10]将经验模态分解和自适应降噪方法结合起来改善传统模型的预测精度。PENG 等^[11]提出深度迁移学习模型提升模型的预测精度。许佳辉等^[12]提出图神经网络用于提升水质预测模型性能。以上研究没有考虑预测模型中网络参数对其性能的影响,当模型参数选择不当时会导致模型预测精度降低。

文献[13-14]采用遗传算法优化支持向量回归超参数,文献[15-16]提出通过智能算法对自适应神经模糊推理系统(Adaptive-network-based fuzzy inference system, ANFIS)的超参数进行优化以提高模型的预测性能,AZAD 等^[17]提出基于进化算法(Evolutionary algorithms, EA)的 ANFIS 预测模型参数。朱成云等^[18]构建基于混沌变异的分布估计(Chaotic mutation to improve the estimation of

distribution algorithm, CMEDA)算法优化混沌变异的分布估计(Least squares support vector regression, LSSVR)预测模型,提高其预测精度。ZHANG 等^[19]通过粒子群优化(Particle swarm optimization, PSO)门控循环单元神经网络预测模型来提高其拟合度。周新辉等^[20]提出自适应增强的粒子群优化算法对极限学习机超参数进行优化来提高模型预测精度。预测模型参数优化过程易陷入局部最优,SHAO 等^[21]针对这一问题,提出了布谷鸟搜索算法(Cuckoo search, CS)优化水质预测模型参数,该方法具有参数少、寻优能力强等特点。本文重点研究基于改进布谷鸟搜索算法(Improved cuckoo search, ICS)的水质预测模型参数优化方法来提高预测模型的整体性能。

已有研究多聚焦于水质一维预测研究,较少考虑水源水域浊度、叶绿素等随着深度变化呈现明显差异。文献[22-23]提出的 BRT 模型平均决定系数 R^2 为 0.7 左右,ZHANG 等^[24]构建 ANFIS 预测模型实现更加准确的水质三维动态变化预测。LIU 等^[25]提出三维循环模型能够预测水质时空变化,CHEN 等^[26]构建基于 K-means 算法的 RBF 神经网络方法预测水质三维变化,在典型评价指标上优于传统方法。RBF 网络具有结构简单、收敛速度快的优点^[27],但对深层水质空间特征提取存在不足。同时,多数研究中没有考虑气象等数据参数,网络预测模型输入不足导致其预测性能与实际情况有所偏差。

本文提出基于 AE 改进布谷鸟搜索算法(AE-ICS)对 CNN-RBF 的水质预测模型网络参数进行优化,有效避免 RBF 神经网络参数初始值随机设置的缺陷,通过结合小波变换(Wavelet transform, WT)算法对输入水质参数进行去噪,克服 CEEMDAN 算法重构后遗留的自适应白噪声对水质原始信号的干扰。将 CNN 提取获得的水质和气象数据特征输入 AE 初始化网络参数,利用 ICS 算法来优化网络超参数的动态初始值,并挖掘水质参数、气象参数之间关系,构建封闭水源水域的高精度水质预测模型,通过消融实验和对比实验验证模型的有效性。

1 水质预测模型构建

1.1 水质三维预测方法

基于数据预处理算法、3-D CNN、ICS、AE 和 RBF 神经网络,构建水质三维预测模型,包含 4 个模块,分别为网络输入模块、特征提取模块、优化模块和网络输出模块,流程图如图 1 所示。网络输入模块主要对原始水质数据降噪和降维,提高网络对水

质特征的学习和训练效率;特征提取模块主要运用 3 维卷积神经网络提取水质、气象等局部特征;优化模块利用易于与其他算法耦合的 ICS 算法,结合 AE

更新优化网络超参数动态初始值,提高网络的全局寻优能力;网络输出模块主要利用训练好的网络对水质进行三维预测。

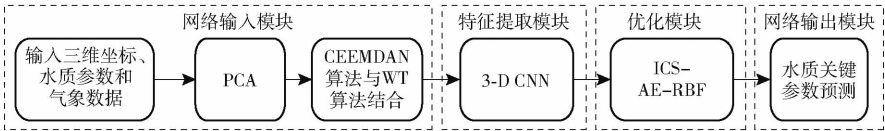


图 1 预测模型流程图
Fig. 1 Prediction model flowchart

1.1.1 网络输入模块

网络输入动态收集的水质数据具有波动性、季节周期性、趋势性、低相关性、非线性和维度不一致

等特点,因此,利用网络输入模块中的 PCA 算法和 CEEMDAN 算法分别对水质数据进行降维和降噪处理,提高水质稳定性,结构如图 2 所示。

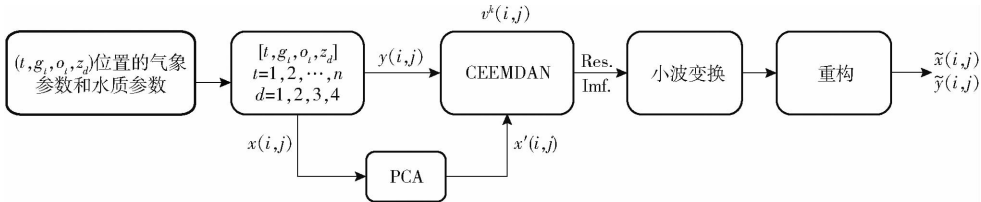


图 2 网络输入模块流程图
Fig. 2 Network input module flowchart

原始水质数据坐标位置为 4 维向量,即 (t, g_i, o_i, z_d) ,在时间 t 和深度 z_d 处获取水平位置 (g_i, o_i) 中包含水质参数与气象参数的 $x_n(i, j)$ 、水质预测参数 $y_n(i, j)$,利用 PCA 算法对水质参数进行降维,将降维处理后得到的 3 个主成分数据通过主成分分析获得降维后水质参数,进一步分析其与水质预测参数相关性来获得模型输入的水质参数,再利用能够解决 EEMD 算法加入白噪声后引入重构误差问题的 CEEMDAN 算法处理,水质数据被分解为 7 个本征模态函数(Inherent mode function, IMF),结合 WT 算法对每个特征信号进行降噪处理,再对降噪后的信号分量重构,得到降噪后的水质数据 $\hat{x}(i, j)$ 。 $\hat{y}(i, j)$ 与 $\hat{x}(i, j)$ 的处理过程一致,具体处理步骤如下:

对于输入包含原始水质参数与气象参数的 $x_n(i, j)$,为了解决一个 IMF 分量存在多个尺度成分或者一个尺度成分在多个 IMF 分量上存在,加入高斯白噪声信号 $v^k(i, j)$ 来改变极值点分布,得到新信号 x_{n+1} ,公式为

$$x_{n+1}(i, j) = x_n(i, j) + (-1)^q \varepsilon v^k(i, j) \quad (q = 1, 2) \tag{1}$$

式中 i, j, n ——序号 ε ——噪声控制系数
 q ——噪声序列正负号指数
对 x_{n+1} 进行 EMD 分解,得到一阶本征模态分量 C_1 ,经过 EMD 分解后得到本征模态分量为
 $E_1(x_n(i, j) + (-1)^q \varepsilon v^k(i, j)) = C_1^k(i, j) + r^k$
($k = 1, 2, \dots, N$) \tag{2}

式中 $E_i(\cdot)$ ——第 i 个本征模态分量

r ——信号残差
 N ——本征模态分量个数

对产生的 N 个模态分量进行总体平均得到 CEEMDAN 分解的第 1 个本征模态分量为

$$\overline{C_1(i, j)} = \frac{1}{N} \sum_{i=1}^N C_1^i(i, j) \tag{3}$$

计算去除第 1 个模态分量后的残差为

$$r_1(i, j) = x_n(i, j) - \overline{C_1(i, j)} \tag{4}$$

在 $r_1(i, j)$ 加入高斯白噪声得到新信号,以新信号为载体进行 EMD 分解,得到第 1 阶模态分量 D_1 ,由此可以得到 CEEMDAN 算法分解的第 2 个本征模态分量为

$$\overline{C_2(i, j)} = \frac{1}{N} \sum_{i=1}^N D_1^i(i, j) \tag{5}$$

计算去除第 2 个模态分量后的残差为

$$r_2(i, j) = r_1(i, j) - \overline{C_2(i, j)} \tag{6}$$

重复上述步骤,直到残差信号为单调函数,不能继续分解,算法结束。此时得到的本征模态分量数量为 L ,使用小波阈值对 L 个本征模态分量进行去噪,将所有的 IMF 重构得到降噪后的数据,即为 $\hat{x}(i, j)$,公式为

$$\hat{x}(i, j) = \sum_{s=1}^L \overline{C_s(i, j)} + r_s(i, j) \tag{7}$$

其中

$$\overline{C_s(i, j)} = \begin{cases} \text{sgn}(\overline{C_s(i, j)}) (|\overline{C_s(i, j)}| - V) & (|\overline{C_s(i, j)}| \geq V) \\ 0 & (|\overline{C_s(i, j)}| < V) \end{cases} \tag{8}$$

$$V = E_s \sqrt{2 \ln p} \quad (9)$$

$$E_s = (\text{median}(|C_s(i, j)|) / 0.6745)^2 \quad (10)$$

(s = 1, 2, \dots, L)

式中 $\overline{C_s(i, j)}$ ——第 s 个 IMF

V——IMF 对应的阈值

E_s ——第 s 个噪声方差

p——信号长度

$r_s(i, j)$ ——第 s 个信号残差

1.1.2 特征提取模块

特征提取模块主要提取水质局部特征,使卷积神经网络中卷积层能够学习水质局部特征信息。将处理后的原始数据 $\hat{x}(i, j)$ 及其对应的坐标位置数据 (g_i, o_i, z_i) 作为 3-D CNN^[28] 的输入,卷积网络层数设置为 3 层,第 1 层和第 2 层网络结构包括卷积层和池化层,第 3 层去除池化层和全连接层,保留卷积层,目的是提高网络的计算效率,其中网络输入包含 4 个通道,作为不同深度 z 切片,每个通道的长宽分别为 84 和 16,卷积层采用卷积核大小为 3 × 3,步长为 1,池化层采用平均池化,卷积核大小为 2 × 2,步长为 2。若 3 维卷积层的输入为 $\hat{x}(i, j)$,则通过信号特征 $\hat{x}(i, j)$ 与 a × b 的卷积核 w(i, j) 卷积得到 z(i, j),网络最终输出大小为 4 × 84 × 16,公式为

$$z(i, j) = \hat{x}(i, j) w(i, j) = \sum_{g=0}^{a-1} \sum_{h=0}^{b-1} x(g, h) w(i-g, j-h) \quad (11)$$

1.2 优化模块

1.2.1 ICS 算法

传统 CS 算法^[29]采用莱维飞行算法能够增强算法全局搜索能力,但会产生较大的随机行走步长,影响算法的局部搜索和收敛,同时对参数设置比较敏感,通过最大迭代次数调整步长缩放因子 α 和参数 β^[30-31]。为了减小最大迭代次数对 CS 优化性能的影响,提高 CS 的收敛速度,引入带惯性权重 w'_i 以及新的参数 α 和 β,公式为

$$l_i^{(t+1)} = w'_i l_i^{(t)} + \alpha_i^{(t)} \odot \text{Levy}(\beta_i^{(t)}) \odot (l_i^{(t)} - x_{\text{best}}^{(t)}) \quad (12)$$

$$\text{其中 } x_{\text{best}}^{(t)} = \text{nest}_i^{(t)}(x_1, x_2, \dots, x_i) \quad (13)$$

$$w'_i = 1 - e^{-\frac{1}{t}} \quad (14)$$

$$\alpha_i^{(t)} = 0.8 + 2.4 \left(\frac{1}{\sqrt{t}} \right) \left| \frac{F_{\text{best}}^{(t)} - F_i^{(t)}}{F_{\text{best}}^{(t)} - F_{\text{worst}}^{(t)} + \gamma} \right| \quad (15)$$

$$\beta_i^{(t)} = 1.0 + 0.2 \left| \frac{F_{\text{best}}^{(t)} - F_i^{(t)}}{F_{\text{best}}^{(t)} - F_{\text{worst}}^{(t)} + \gamma} \right|^{0.2t} \quad (16)$$

式中 $l_i^{(t)}$ ——第 i 个粒子在第 t 次迭代时决策变量

⊙——点乘运算符号

$\alpha_i^{(t)}$ ——第 i 个粒子在第 t 次迭代时步长缩放因子

$\beta_i^{(t)}$ ——第 i 个粒子在第 t 次迭代时参数

$\text{Levy}(\beta_i^{(t)})$ —— $\beta_i^{(t)}$ 服从 Levy 的随机搜索路径

$\text{nest}(\cdot)$ ——获取决策变量中所有粒子值的函数

$x_{\text{best}}^{(t)}$ ——在第 t 次迭代适应度最好的决策变量 c_j, δ_j, m_j 和 w_j 相关值

$F_i^{(t)}$ ——第 i 个适应度函数值

$F_{\text{best}}^{(t)}$ ——第 t 次迭代中最好的适应度函数值

$F_{\text{worst}}^{(t)}$ ——第 t 次迭代中最差的适应度函数值

γ——用于避免除零误差

1.2.2 AE 初始化网络参数

利用 AE^[32] 优化 RBF 网络超参数动态初始值,使用无监督特征数据 z(i, j) 训练的权重参数作为网络的初始化值,结构图如图 3 所示。利用 AE 对特征提取模块中提取的特征 z(i, j) 进行学习,经过反向传播过程,获得 RBF 神经网络超参数的初始化值 l_i ,其中 $l_i = \{c_i, \delta_i, m_i, w_i\}$,使用 l_i 替代随机初始化 RBF 神经网络超参数值。一方面为了防止 RBF 网络训练不稳定,易处于局部最优,另外一方面使 ICS 算法加快梯度下降收敛,公式分别为

$$\text{Loss} = \text{dist}(x_i, x_i^{(1)}) \quad (17)$$

$$\text{其中 } x_i^{(1)} = \exp\left(-\frac{\|h_i - c_i\|}{2\delta_i^2}\right) w'_i + b_2 \quad (18)$$

$$h_i = \psi_i w_i + b_1 = \exp\left(-\frac{\|x_i - c_i\|}{2\delta_i^2}\right) w_i + b_1 \quad (19)$$

$$c_i = \{c_{i1}, c_{i2}, \dots, c_{il}\} \quad (i = 1, 2, \dots, m; l = 1, 2, 3)$$

$$\delta_i = \{\delta_{i1}, \delta_{i2}, \dots, \delta_{il}\}$$

式中 Loss—— x_i 与 $x_i^{(1)}$ 的损失值

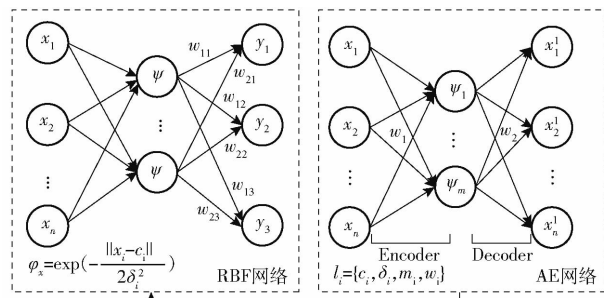
w'_i ——第 i 个解码层权重

b_2 ——AE 网络偏置

w_i ——第 i 个编码层权重,即 RBF 参数权重

b_1 ——RBF 偏置

h_i ——第 i 个 RBF 隐藏层结果



提供网络参数初始化值

图3 网络参数初始化结构图

Fig. 3 Network parameter initialization structure

ψ_i ——第 i 个 RBF 的计算结果
 c_i ——第 i 个高斯径向基核函数中心
 δ_i ——第 i 个高斯径向基核函数尺度宽度
 $\text{dist}(\cdot)$ ——二者的距离度量函数,通常用 MSE

1.2.3 ICS 优化网络参数

利用 ICS 算法进一步更新调整 RBF 中超参数

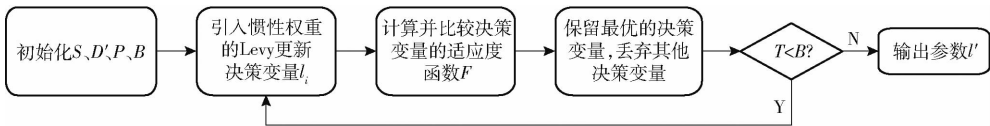


图 4 ICS 优化网络参数流程图
Fig. 4 Flowchart of ICS optimizing network parameters

网络参数优化步骤如下:

- (1) 数据特征提取。通过 1.1.2 节提取的特征作为 RBF 神经网络的学习和训练。
- (2) 初始决策变量。在每个决策变量取值范围内,通过 AE 对特征提取和学习来获得 S 个粒子(巢)的初始值,每个粒子代表一个独立决策变量的值,中心 c_i 、尺度宽度 δ_i 、隐藏单元数量 m_i 和权值 w_i 的 4 个决策变量在 S 个粒子中获得 S 个值。
- (3) 建立适应度函数。寻找 c_i 、 δ_i 、 m_i 和 w_i 中的最优解,构建 RBF 神经网络输出预测结果。根据设计的适应度函数计算模型中输出层的预测值与实际值之间的误差,然后比较 4 个决策变量不同组合值的模型适应度函数误差,确定决策变量的最优初始值。
- (4) 采用莱维飞行算法更新每个决策变量解。获得决策变量的最优解将保留给下一代,然后由莱维飞行操作员更新决策变量解,并计算更新后决策变量的适应度。比较每个决策变量更新后的模型适应度和当前的模型适应度,如果当前决策变量的适应度比更新后的决策变量更好,则保留当前决策变量值,如果不是,则保留更新后决策变量值。
- (5) 当每个决策变量经过第 t 次迭代后,比较每个决策变量的适应度。
- (6) 当满足终止条件时停止迭代,找到最优解,否则返回步骤(2)继续优化,此时最佳嵌套对应优化后的 RBF 神经网络权值等超参数,得到最优决策变量值 l' 。

1.3 网络输出模块

网络输出模块主要利用优化模块获得超参数 c 、 δ 、 m 、 w 的最优值,来构建 RBF 神经网络预测模型,对浊度等水质参数进行预测。网络的输出层为 3 个单元节点,均方误差(Mean square error, MSE)作为网络损失函数。RBF 神经网络通过 MSE 度量指

初始化值 l_i 来获得最优解,流程如图 4 所示。将高斯核函数作为 RBF 神经网络隐藏层的激活函数,计算得到 ψ_i ,公式为

$$\psi_i = \psi(\|z_i - c_i\|_2, \delta_i) = \exp\left(-\frac{1}{2\delta_i^2} \|z_i - c_i\|^2\right)$$

(20)

式中 $\|\cdot\|$ —— l_p 范数

标计算每一时刻水质参数的预测值与真实值之间误差平方的期望值,利用 ICS 算法对 RBF 神经网络超参数进行寻优,设置 ICS 算法适应度函数 F 。当网络训练能力达到损失降到最低时,获得 RBF 神经网络模型最优超参数 c 、 δ 、 m 、 w 。将 1.2.3 节所获得 RBF 结果 ψ_i 结合隐藏层与输出层之间的权值 w_{ij} ,最终计算得到预测值 Y_j 。计算公式分别为

$$MSE = \frac{1}{n} \sum_{j=1}^n (Y_j - \tilde{y}_j)^2$$

(21)

$$F = \frac{1}{MSE}$$

(22)

$$Y_j = \sum_{i=1}^m w_{ij} \psi_i(z) \quad (j=1, 2, 3)$$

(23)

式中 $\tilde{y}_j$ ——实际值 n ——样本数量

1.4 模型参数设置与评价指标

学习率(初始值为 0.01)是通过模型在训练过程中自动调整来使梯度下降,每迭代 10 次降低 100 倍;隐藏层节点数通过经验法确定,当隐藏层单元数设置为 25 时,PCA-CEEMDAN-CNN-AE-ICS-RBF 模型的 MSE 等 5 个度量指标性能表现最优(表 1);模型在所有数据集的训练次数为 50 时, MSE 达到最低;CEEMDAN 算法在数据降噪过程中设置原始数据 95% 的信息量被保留;ICS 算法中粒子规模 S 为 20,决策变量维度 D' 为 4,发现概率 P 为 0.25,迭代次数 B 为 2 500。

表 1 不同隐藏单元数量模型的性能比较

Tab. 1 Performance comparison of different hidden units models					
隐藏单元数量	MSE/ ($\text{mg}^2 \cdot \text{L}^{-2}$)	MAPE	RMSE/ ($\text{mg} \cdot \text{L}^{-1}$)	R	$D/\%$
18	0.134	0.224	0.307	0.900	99.5
25	0.059	0.118	0.201	0.920	99.9
40	0.102	0.194	0.239	0.916	99.7
38	0.172	0.228	0.319	0.889	98.9
75	0.204	0.330	0.420	0.875	98.1

采用 MSE、均方根误差 (Root mean square error, RMSE)、平均绝对百分比误差 (Mean absolute percentage error, MAPE)、相关系数 R (Coefficient of correlation) 和威尔莫特一致性指数 D (Willmott index of agreement) 5 个度量指标来评价组合模型的预测性能。MSE、MAPE 和 RMSE 越小, R 和 D 越大, 组合模型的预测性能越好。

2 结果与分析

2.1 实验方法

数据集收集于广东省湛江市徐闻县大水桥水库区域, 区域面积大约 900 m^2 , 长度大约为 30 m , 宽度大约为 30 m , 形状为正方形, 拥有良好的空间环境, 实验基地受周边污染程度低。将传感器节点分别布置于水库区域的周围和中间, 放置于 4 个不同深度, 分别为 0 、 0.6 、 1.2 、 1.8 m , 总共布置 22 个监测点, 分别为 $A7 \sim A10$ 、 $B1 \sim B6$ 、 $C1 \sim C6$ 、 $D1 \sim D6$, 以及手持监测仪放置于 0 m 水平面处, 共有 6 个手持监测点 ($A1 \sim A6$), 区域旁边放置气象监测站, 收集气象数据, 如图 5 所示。数据样本收集于 2023 年 5 月 16 日—6 月 15 日, 采集周期为每 1 h 采集 1 次, 将样本集按比例 7: 3 分为训练集和测试集, 其中训练集中 20% 数据作为验证集, 用来验证模型性能以及网络参数初始化和优化。

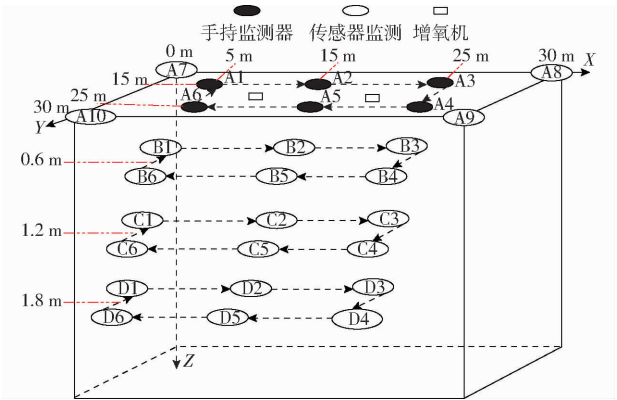


图5 水库采集点分布图

Fig. 5 Distribution of reservoir collection sites

2.2 数据统计

收集的水质参数包括水温、pH 值、电导率、浊度、溶解氧含量 (质量浓度)、氨氮含量、高锰酸盐指数、总磷含量、总氮含量以及叶绿素含量、藻密度等 11 项水质参数指标以及气温等气象参数, 水质参数和气象参数的均值等参数数据描述见表 2。若将收集的水质数据直接作为模型输入, 将会增加模型复杂度, 影响模型预测精度。因此, 利用 PCA 算法对获取的水质参数进行降维, 水质参数的主成分系数矩阵见表 3。从表 3 可知, 前 3 个主成分累积贡献

表 2 水质、气象数据描述

Tab.2 Description of water quality and meteorological data

指标		平均值	标准差
水质参数	水温 (WT)/ $^{\circ}\text{C}$	28.451	2.333
	pH 值 (pH)	8.086	0.206
	电导率 (EC)/ $(\mu\text{S}\cdot\text{cm}^{-1})$	155.014	2.953
	溶解氧含量 (DO)/ $(\text{mg}\cdot\text{L}^{-1})$	7.930	0.717
	氨氮含量 (AN)/ $(\text{mg}\cdot\text{L}^{-1})$	0.084	0.017
	高锰酸钾盐指数 (CODMn)/ $(\text{mg}\cdot\text{L}^{-1})$	1.856	0.030
	总磷含量 (TP)/ $(\text{mg}\cdot\text{L}^{-1})$	0.089	0.036
	总氮含量 (TN)/ $(\text{mg}\cdot\text{L}^{-1})$	0.905	0.027
	浊度 (T)/NTU	0.814	0.027
	叶绿素含量 (C)/ $(\mu\text{g}\cdot\text{L}^{-1})$	0.006	3.030×10^{-4}
	藻密度 (AD)/ $(\text{个}\cdot\text{L}^{-1})$	9.846×10^5	3.117×10^3
气象参数	气温 (AT)/ $^{\circ}\text{C}$	29.730	2.451
	降水量 (R)/mm	1.383	0.571
	相对湿度 (AH)/%	85.582	9.252

表 3 成分矩阵

Tab.3 Component matrix

指标	成分 1	成分 2	成分 3
WT	-0.605	0.043	-0.370
pH	-0.507	0.802	-0.113
EC	-0.855	-0.279	0.062
DO	-0.420	0.633	0.617
AN	0.870	0.323	0.107
CODMn	0.519	-0.259	0.858
TP	0.672	-0.026	-0.490
TN	0.446	0.805	-0.178
特征值	2.775	1.942	1.192
累积贡献率/%	39.646	67.395	84.427

率为 84.427%, 表明其能够反映原始数据的绝大部分信息, 将原始水质参数映射到 3 个主成分空间获得降维后的水质参数, 根据主成分分析来选取原始水质参数与 3 个主成分相关系数大于 0.8 的水质参数作为模型待输入的水质参数, 包括 pH 值、电导率、溶解氧含量、氨氮含量、高锰酸钾指数、总磷含量和总氮含量, 进一步分析模型待输入水质参数、气象因素分别与浊度、叶绿素含量、藻密度相关性, 目的是筛选出与模型预测水质参数强相关性的水质参数、气象参数, 见表 4。从表 4 可知, 气温、降水量、相对湿度、pH 值和总氮含量对浊度具有正向强相关性, 气温、pH 值、氨氮含量和总氮含量对叶绿素含量具有正向强相关性, 气温、降水量、氨氮含量和总磷含量对藻密度具有正向强相关性, 综合以上得出, 将气温、降水量、相对湿度、氨氮含量、总磷含量和 pH 值作为模型输入的气象参数和水质参数。

2.3 消融实验

为了验证提出的 PWCAIR (PCEMDAN - WT -

表 4 气象因素、水质参数与浊度等参数的相关系数
Tab.4 Correlation coefficients of meteorological factors, water quality parameters and turbidity parameters

指标	T	C	AD
AT	0.824 *	0.923 *	0.822 *
R	0.862 *	-0.288	0.814 *
AH	0.853 *	0.116	0.242
pH	0.812 *	0.882 *	-0.126
AN	-0.103	0.821 *	0.842 *
CODMn	-0.063	0.128	-0.049
TN	0.892 *	0.901 *	0.886 *

注：* 表示在 0.05 级别(双尾)相关性显著。

CNN-AE-ICS-RBF) 水质三维预测模型的有效性,以广东省湛江市徐闻县大水桥水库区域采集的样本集作为测试集,采用 4 种消融实验,即去除 AE 后的 PWCIR 方法、去除 CNN 后的 PWAIR、去除 ICS 优化算法后的 PWCAR 和使用未经过 PCEMDAN 算法和 WT 算法处理的原始数据。5 个典型度量指标的实验结果见表 5。

表 5 不同算法性能比较
Tab.5 Comparison of different algorithms

模型	MSE/ ($\text{mg}^2 \cdot \text{L}^{-2}$)	MAE/ ($\text{mg} \cdot \text{L}^{-1}$)	RMSE/ ($\text{mg} \cdot \text{L}^{-1}$)	<i>R</i>	<i>D</i> /%
PWCAIR	0.103	0.185	0.261	0.945	99.9
PWCIR	0.209	0.299	0.358	0.901	96.8
PWAIR	0.184	0.285	0.328	0.915	97.0
PWCAR	0.270	0.375	0.496	0.872	92.8
RBF	0.330	0.438	0.550	0.838	91.5

将所提出的预测模型与 PWCIR 预测模型对比,MSE、MAE、RMSE、*R*、*D* 分别提升 50.7%、38.1%、27.1%、4.9%、3.2%;与 PWAIR 预测模型对比,MSE、MAE、RMSE、*R*、*D* 分别提升 44.0%、35.1%、20.4%、3.3%、3.0%;与 PWCAR 预测模型对比,MSE、MAE、RMSE、*R*、*D* 分别提升 61.9%、50.7%、47.4%、8.4%、7.7%。与 RBF 预测模型对比,MSE、MAE、RMSE、*R*、*D* 分别提升 68.8%、57.8%、52.5%、12.8%、9.2%。通过消融实验结果得出,ICS 算法对所提出的模型预测性能影响最大,其次分别为 AE 算法、CNN 和数据预处理算法,证明提出的三维预测模型在数据预处理、算法优化实现以及

特征提取方面有优势。

2.4 对比实验

通过 2.3 节消融实验验证了提出的水质三维预测模型性能优越性,进一步利用对比实验比较 CS 和 ICS 等优化算法对模型预测性能的影响以及验证提出的水质三维预测模型有效性,将本文提出的预测模型与文献[25-28]所提出预测模型和 PWCACR(PCEMDAN-WT-CNN-AE-CS-RBF) 分别进行对比,结果见表 6。ICS 算法较 CS 算法对模型预测性能有一定提升,MSE、MAE、RMSE、*R*、*D* 分别提升 37.6%、26.3%、9.7%、2.5%、2.4%。对比文献[25-28]分别提出模型的预测性能,文献[27]提出模型的预测性能在 4 个预测模型中最优,本文提出的水质预测模型与文献[27]提出模型进行对比,MSE、MAE、RMSE、*R*、*D* 分别提升 23.1%、20.3%、10.9%、1.9%、1.5%。通过以上结果分析得出,ICS 优化算法能够较好提升模型的预测性能,同时也证明了提出的水质预测模型有效性,具有网络中超参数优化和特征提取上的优势。

表 6 不同模型性能对比
Tab.6 Performance of different models

模型	MSE/ ($\text{mg}^2 \cdot \text{L}^{-2}$)	MAE/ ($\text{mg} \cdot \text{L}^{-1}$)	RMSE/ ($\text{mg} \cdot \text{L}^{-1}$)	<i>R</i>	<i>D</i> /%
PWCAIR	0.103	0.185	0.261	0.945	99.9
PWCACR	0.165	0.251	0.289	0.921	97.5
BRT ^[22-23]	0.221	0.345	0.419	0.878	94.7
ANFIS ^[24]	0.134	0.232	0.293	0.927	98.4
三维循环模型 ^[25]	0.218	0.319	0.388	0.890	95.2

2.5 工程仿真实验

利用水下 1.8 m 的 1 d 内的数据集对浊度等水质参数进行测试,结果如图 6 所示。通过分析得出,预测模型获得浊度等水质参数的预测值与实际值误差较小,拟合度高,预测精度及可靠性均符合水质预测的要求。

仿真实验测试数据采集于 2023 年 6 月 7—15 日,浊度等参数在水质三维坐标上的分布结果如图 7 所示。根据实验结果和实地分析,图 7a 中,水库区域 A 方向的浊度高,水库区域 B 方向的浊度低,A 方向区域主要受周围土壤泥沙影响较大;图 7b 中,

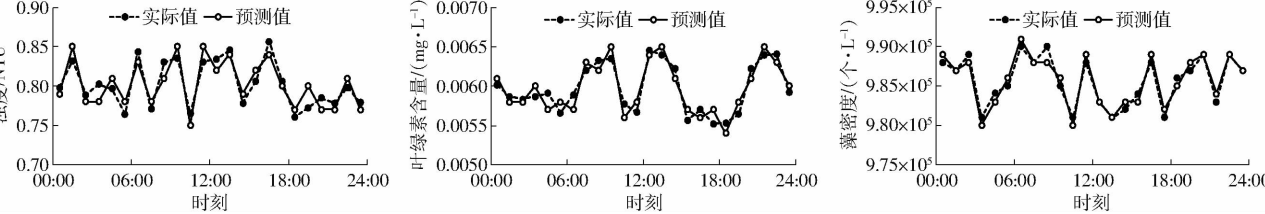


图 6 浊度、叶绿素含量和藻密度实际值与预测值
Fig.6 Actual and predicted values of turbidity, chlorophyll content and algal density

水库区域 C 方向的叶绿素含量低,水库区域 D 方向的叶绿素含量高,D 方向区域主要受周围农田化肥、农药等影响较大;图 7c 中,水库区域 E 方向的藻密度低,水库区域 F 方向的藻密度高,F 方向区域靠近

学校和生活区,受生活废水影响较大。综合分析浊度等水质参数在水库区域的分布规律可知,该水库区域水质参数控制在一定范围,水质状况良好,可为水库监管相关部门了解水质分布情况提供帮助。

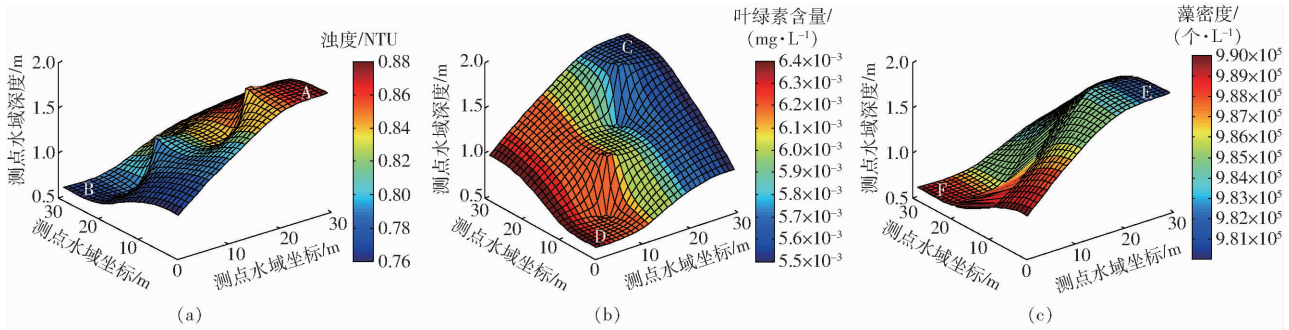


图7 水质参数三维分布图

Fig.7 Three-dimensional distributions of water quality parameters

3 结束语

针对水质预测模型存在精度差、收敛速度慢的问题,提出了基于改进布谷鸟算法结合自动编码器优化RBF-CNN的水质三维预测方法。该方法基于传统CS原理,引进惯性权重和调整位置参数改

进CS来加快RBF网络收敛速度,利用自动编码器初始化网络参数初始值来避免人为随机设置的缺陷,提升了水质参数预测精度,优于BRT、ANFIS和三维循环模型的预测性能,揭示了浊度、叶绿素含量和藻密度的变化规律,可为环境监测等部门提供借鉴。

参 考 文 献

[1] 胡圣,邱光胜,印士勇. 三峡水库香溪河库湾水体富营养化演化监测分析[J]. 人民长江,2012,43(12):10-13.
HU Sheng, QIU Guangsheng, YIN Shiyong. Analysis and monitoring on eutrophication evolution of waters in Xiangxi bay of Three Gorges Reservoir[J]. People's Yangtze River,2012,43(12):10-13. (in Chinese)

[2] 姚玲爱,赵学敏,马千里,等. 台风强降雨输入水源水库的泥沙沉降模拟分析[J]. 生态环境学报,2018,27(10):1900-1907.
YAO Ling'ai, ZHAO Xuemin, MA Qianli, et al. Simulation of the impact of typhoon-induced suspended solids precipitation on water quality in a reservoir for drinking water[J]. Journal of Ecological Environment, 2018, 27(10):1900-1907. (in Chinese)

[3] 王骥,谢再秘,莫春梅. 神经网络在养殖水质精准预测方面的研究进展[J]. 水产学报,2023,47(8):19-34.
WANG Ji, XIE Zaimi, MO Chunmei. Research progress in accurate prediction of aquaculture water quality by neural network[J]. Journal of Aquatic Sciences, 2023,47(8):19-34. (in Chinese)

[4] PANDEY P, BOKDE N D, DONGRE S, et al. Hybrid models for water demand forecasting[J]. Journal of Water Resources Planning and Management, 2021, 147(2):04020106.

[5] ZHANG J R, TANG H M, WEN T, et al. A hybrid landslide displacement prediction method based on CEEMD and DTW-ACO-SVR cases studied in the Three Gorges Reservoir area[J]. Sensors, 2020, 20(15):4287.

[6] SHA J, LI X, ZHANG M, et al. Comparison of forecasting models for real-time monitoring of water quality parameters based on hybrid deep learning neural networks[J]. Water, 2021, 13(11):1547.

[7] 符东,吴雪菲,易珍言,等. 沱江水质模糊综合评价及主要污染物的预测研究[J]. 农业环境科学学报, 2020,39(12):2844-2852.
FU Dong, WU Xuefei, YI Zhenyan, et al. Fuzzy comprehensive assessment of water quality and prediction of main pollutants in the Tuo River[J]. Journal of Agricultural and Environmental Sciences, 2019,39(12):2844-2852. (in Chinese)

[8] 石庆兰,束金阳,李道亮,等. 基于BiLSTM-GRU融合网络的稻虾养殖溶解氧含量预测[J]. 农业机械学报,2023,54(10):364-370.
SHI Qinglan, SHU Jinyang, LI Daoliang, et al. Dissolved oxygen prediction in rice and shrimp culture based on BiLSTM-GRU fusion neural networks[J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(10):364-370. (in Chinese)

[9] 陈英义,方晓敏,梅思远,等. 基于WT-CNN-LSTM的溶解氧含量预测模型[J]. 农业机械学报,2020,51(10):284-291.
CHEN Yingyi, FANG Xiaomin, MEI Siyuan, et al. Dissolved oxygen prediction model based on WT-CNN-LSTM[J]. Transactions of the Chinese Society for Agricultural Machinery,2020,51(10):284-291. (in Chinese)

- [10] KIM J, YU J, KANG C, et al. A novel hybrid water quality forecast model based on real-time data decomposition and error correction[J]. *Process Safety and Environmental Protection*, 2022, 162: 553–565.
- [11] PENG L, WU H, GAO M, et al. TLT: recurrent fine-tuning transfer learning for water quality long-term prediction[J]. *Water Research*, 2022, 225:119171.
- [12] 许佳辉, 王敬昌, 陈岭, 等. 基于图神经网络的地表水水质预测模型[J]. *浙江大学学报(工学版)*, 2021, 55(4): 601–607.
XU Jiahui, WANG Jingchang, CHEN Ling, et al. Surface water quality prediction model based on graph neural network[J]. *Journal of Zhejiang University (Engineering and Technology Edition)*, 2021, 55(4): 601–607. (in Chinese)
- [13] SINGH J, SWAROOP S, SHARMA P, et al. Real-time assessment of the Ganga River during pandemic COVID–19 and predictive data modeling by machine learning[J/OL]. *International Journal of Environmental Science and Technology*, 2022; 1–24. DOI:10.1007/s13762–022–04423–1.
- [14] QUAN Q, ZOU H, HUANG X F, et al. Research on water temperature prediction based on improved support vector regression [J]. *Neural Computing and Applications*, 2022, 34(11): 8501–8510.
- [15] KISI O, AZAD A, KASHI H, et al. Modeling groundwater quality parameters using hybrid neuro-fuzzy methods[J]. *Water Resources Management*, 2019, 33(2): 847–861.
- [16] AZAD A, KARAMI H, FARZIN S, et al. Prediction of water quality parameters using ANFIS optimized by intelligence algorithms (case study: Gorganrood River)[J]. *KSCE Journal of Civil Engineering*, 2018, 22(7): 2206–2213.
- [17] AZAD A, KARAMI H, FARZIN S, et al. Modeling river water quality parameters using modified adaptive neuro fuzzy inference system[J]. *Water Science and Engineering*, 2019, 12(1): 45–54.
- [18] 朱成云, 刘星桥, 李慧, 等. 工厂化水产养殖溶解氧预测模型优化[J]. *农业机械学报*, 2016, 47(1): 273–278.
ZHU Chengyun, LIU Xingqiao, LI Hui, et al. Optimization of prediction model of dissolved oxygen in industrial aquaculture [J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2016, 47(1): 273–278. (in Chinese)
- [19] ZHANG X H, CHEN X H, ZHENG G C. Improved prediction of chlorophyll-a concentrations in reservoirs by GRU neural network based on particle swarm algorithm optimized variational modal decomposition[J]. *Environmental Research*, 2023, 221:115259.
- [20] 周新辉, 黄琳, 樊宇星, 等. 基于建模预测与关系规则的养殖水体溶解氧含量调控方法[J]. *农业机械学报*, 2022, 53(6): 318–326.
ZHOU Xinhui, HUANG Lin, FAN Yuxing, et al. Dissolved oxygen control method based on modeling prediction and relation rule database[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2022, 53(6): 318–326. (in Chinese)
- [21] SHAO D G, NONG X Z, TAN X Z, et al. Daily water quality forecast of the south-to-north water diversion project of China based on the cuckoo search–back propagation neural network[J]. *Water*, 2018, 10(10): 1471.
- [22] KNIERIM K J, KINGSBURY J A, HAUGH C J, et al. Using boosted regression tree models to predict salinity in Mississippi embayment aquifers, Central United States[J]. *Journal of the American Water Resources Association*, 2020, 56(6): 1010–1029.
- [23] ROSECRANS C Z, NOLAN B T, GRONBERG J M. Prediction and visualization of redox conditions in the groundwater of Central Valley, California[J]. *Journal of Hydrology*, 2017, 546: 341–356.
- [24] ZHANG Z X, LIMAYE A B, KHOSRONEJAD A. Three-dimensional realizations of flood flow in large-scale rivers using the neural fuzzy-based machine-learning algorithms[J]. *Computers and Fluids*, 2022, 246: 105611.
- [25] LIU W C, CHEN W B. Prediction of water temperature in a subtropical subalpine lake using an artificial neural network and three-dimensional circulation models[J]. *Computers and Geosciences*, 2012, 45: 13–25.
- [26] CHEN Y Y, YU H H, CHENG Y J, et al. A hybrid intelligent method for three-dimensional short-term prediction of dissolved oxygen content in aquaculture[J]. *Plos One*, 2018, 13(2): e0192456.
- [27] PARK J, SANDBEG I W. Universal approximation using radial-basis-function networks[J]. *Neural Computation*, 1991, 3(2): 246–257.
- [28] ZHAO J F, MAO X, CHEN L J. Learning deep features to recognise speech emotion using merged deep CNN[J]. *IET Electric Power Applications*, 2018, 12(6): 713–721.
- [29] YANG X S, DEB S. Cuckoo search via lévy flights[C]//Coimbatore:NaBIC, 2009.
- [30] LI Y Y, FQ S Y, LI Z J. Reactive power voltage control on power distribution network with photovoltaic power generation based on improved cuckoo search algorithm[J]. *Guangdong Electric Power*, 2016, 29(8): 79–84.
- [31] LI R Y, DAI R W. Adaptive step-size cuckoo search algorithm[J]. *Computer Science*, 2017, 44(5): 235–240.
- [32] ZHANG Y F, THORBURN P J. A deep surrogate model with spatio-temporal awareness for water quality sensor measurement [J]. *Expert Systems with Applications*, 2022, 200:116914.