

张海霞,李瑞,王霞,等.基于VMD和GA-BiLSTM组合模型的河流水质预测[J].环境工程技术学报,2025,15(4):1181-1188.

ZHANG H X, LI R, WANG X, et al. River water quality prediction based on the combined model of VMD and GA-BiLSTM[J]. Journal of Environmental Engineering Technology, 2025, 15(4): 1181-1188.

基于VMD和GA-BiLSTM组合模型的河流水质预测

张海霞^{1,2}, 李瑞^{1,2}, 王霞^{1,2}, 赵泽霏^{1,2}, 康彦付³, 孙岩^{1,2*}

1.河北工程大学能源与环境工程学院

2.河北省大气污染成因与影响重点实验室

3.河北省邢台水文勘测研究中心

摘要 溶解氧(DO)是河流水质监测的关键指标之一,为了精准预测河流水体中DO浓度,融合变分模态分解(VMD)、遗传算法(GA)和双向长短期记忆神经网络(BiLSTM),构建了VMD-GA-BiLSTM深度学习组合模型,对邢台市艾辛庄断面2020—2023年的DO浓度数据进行了训练与测试,并与多个经典的深度学习模型(BiLSTM、GA-BiLSTM、EMD-GA-BiLSTM)预测结果进行对比。结果表明:VMD-GA-BiLSTM模型在测试集上的均方根误差(RMSE)、平均绝对误差(MAE)、决定系数(R^2)分别为0.149、0.135和0.974,相较于BiLSTM、GA-BiLSTM、EMD-GA-BiLSTM模型, RMSE分别降低0.464、0.307、0.290, MAE分别降低0.413、0.173、0.239, R^2 分别提升了0.216、0.133、0.088,表明构建的模型预测精度最高。将构建模型应用于邢台市后西吴桥断面对pH、DO和氨氮3项水质指标进行验证,与其他经典模型相比, VMD-GA-BiLSTM模型的RMSE、MAE最小且 R^2 最大,可见其在水质时间序列数据预测方面具高度的通用性和稳定性。VMD-GA-BiLSTM模型能够准确预测DO浓度以及其他水质指标浓度,为水资源的可持续利用和水环境保护提供科学依据。

关键词 河流;水质预测;溶解氧;变分模态分解;遗传算法;双向长短期记忆神经网络(BiLSTM)

中图分类号: X522 文章编号: 1674-991X(2025)04-1181-08 doi: 10.12153/j.issn.1674-991X.20240630

River water quality prediction based on the combined model of VMD and GA-BiLSTM

ZHANG Haixia^{1,2}, LI Rui^{1,2}, WANG Xia^{1,2}, ZHAO Zefei^{1,2}, KANG Yanfu³, SUN Yan^{1,2*}

1.School of Energy and Environmental Engineering, Hebei University of Engineering

2.Hebei Key Laboratory of Air Pollution Cause and Impact

3.Xingtai Hydrology Survey and Research Center of Hebei Province

Abstract Dissolved oxygen (DO) is a pivotal indicator in river water quality monitoring. In order to accurately predict the DO concentration in river water bodies, we developed a VMD-GA-BiLSTM deep learning integration model, combining Variational Mode Decomposition (VMD), Genetic Algorithm (GA), and Bidirectional Long Short-term Memory (BiLSTM). We also conducted training and testing on DO concentration data from Aixinzhuang section in Xingtai City for the years 2020-2023, with comparisons made against the prediction results of multiple classic deep learning models (BiLSTM, GA-BiLSTM and EMD-GA-BiLSTM). The results revealed that the VMD-GA-BiLSTM model achieved remarkable performance on the test set, with RMSE of 0.149, MAE of 0.135, and R^2 of 0.974. When compared to BiLSTM, GA-BiLSTM and EMD-GA-BiLSTM models, the model demonstrated significant improvements: RMSE was reduced by 0.464, 0.307, and 0.290, respectively; MAE was decreased by 0.413, 0.173, and 0.239, respectively; and R^2 was increased by 0.216, 0.133, and 0.088, respectively. These findings demonstrated the superior prediction accuracy of the constructed model. To further validate the versatility and stability of the model, we applied it to predict three water quality indicators: pH, DO, and ammonia nitrogen, in Houxiwuqiao section of Xingtai City. The results indicated that the VMD-GA-BiLSTM model outperformed other classical models, achieving the lowest RMSE and MAE values and the highest R^2 score. This demonstrated the

收稿日期:2024-09-24

基金项目:河北省自然科学基金青年基金项目(B2024402004);河北省高等学校科学研究项目(QN2023139)

作者简介:张海霞(1979—),女,教授,博士,主要从事水环境机器学习模型研究, zhanghaixia@hebeu.edu.cn

*通信作者:孙岩(1994—),女,讲师,博士,主要从事机器学习模型分析污染物研究, sunyan@hebeu.edu.cn

model's high adaptability and robustness in predicting water quality time series data. The research results showed that the VMD-GA-BiLSTM model can accurately predict concentrations of DO and other water quality indicators. This model can serve as a scientific basis for the sustainable utilization of water resources.

Key words river; water quality prediction; dissolved oxygen (DO); variational mode decomposition; Genetic Algorithm (GA); Bidirectional Long Short-term Memory (BiLSTM)

水污染影响人类健康和生态系统,水质参数的准确预测对有效保护水环境至关重要^[1]。溶解氧(DO)作为河流水质的关键指标之一,是衡量水体自净能力和生态系统健康状态的重要指标^[2],准确预测河流中 DO 的浓度对于制定水资源管理策略、保护水生态环境具有重要意义^[3]。

随着人工智能与大数据技术的飞速发展,机器学习模型在水质预测领域展现出了巨大的潜力^[4-7]。其中长短期记忆网络(Long Short-term Memory, LSTM)模型能够有效处理时间序列数据中的长期依赖问题,对于这种连续、多变量且包含大量噪声的水质数据集尤为适用^[8]。但 LSTM 模型只能对水质序列进行正向处理,随着时间步长的增加,模型捕捉数据长期依赖性的能力下降,容易造成一些重要特征被遗忘,因此对长期时序规则的挖掘有待优化^[9]。双向长短期记忆神经网络(Bidirectional Long Short-term Memory, BiLSTM)作为 LSTM 模型的变体,通过结合前向和后向 2 个方向的 LSTM 网络,能够更全面地利用序列中的前后数据信息,进一步提高水质预测的准确性^[10-12]。这一特性在水质时间序列分析中非常重要,因为水质的变化往往受到前后数据的共同影响^[13]。姚晔等^[14]基于 BiLSTM 模型设计站点周边水位预测系统,双向 LSTM 解码器在信息流动中形成闭环连接,与 LSTM 等模型相比预测性能和稳定性都得到了提高。

神经网络模型研究一般利用试错法等手动方式调整超参数^[15-16],此过程充满随机性,不利于模型构建的标准化,且限制了模型性能。将超参数调整转化为最优化问题,并应用适当的优化算法实现自动化寻优,能有效规范神经网络在水环境问题中的应用流程,确保模型预测能力的稳定提升。尚旭东等^[17]研究提出了一种基于 BiLSTM 的组合水质预测方法,通过麻雀搜索算法优化 BiLSTM 的最大训练次数、最大隐藏层数和学习率参数,与其他模型相比提升了 DO 的预测精度。遗传算法(Genetic Algorithm, GA)具有全局搜索能力强、适应性好、鲁棒性高以及可并行实现的优点^[5],特别适用于解决复杂的非线性、非凸和多峰优化问题,能够根据水质数据特征调整模型参数。河流水质数据往往具有非线性、不确定

性等特点^[18-19],已有研究^[20-22]提出了一种基于融合变分模态分解(Variational Mode Decomposition, VMD)的混合预测模型,水质数据经过 VMD 分解,输入神经网络后更容易提取水质特征,此种模型相比于单一模型以及其他组合模型具有更高的精确度以及更快的收敛速度。其中还对比了 VMD 与小波变换在信号分析领域的性能,结果表明 VMD 模型在处理非平稳、非线性信号时具有更高的分辨率和鲁棒性。

为了提升河流水体中 DO 的预测精度,先通过 VMD 分解水质数据,降低数据复杂性和非平稳性,再引用 GA 对 BiLSTM 进行优化,对各分解信号进行预测,结合 BiLSTM 的时序特征捕捉能力,提升模型水质预测性能。旨在规范神经网络模型应用于水环境问题的流程构建,确保模型的预测效果,以此促进神经网络在水环境领域的广泛应用。

1 数据来源与研究方法

1.1 数据来源

采用邢台市滏阳河段艾辛庄断面(国考监测断面)的水质数据进行分析,艾辛庄断面受上游和支流来水水质影响较大,另外汛期农田退水产生的污染物也随径流进入河道。艾辛庄断面主要监测的是河流地表水水质,包括水温、pH、浊度、电导率、DO、氨氮、高锰酸盐指数、总磷、总氮共 9 项水质指标,监测间隔为 4 h。数据来源于中国环境监测总站发布的国家地表水水质自动监测站数据(<https://szddjc.cnemc.cn8070/GJZ/Business/Publish/Main.htm>)。DO 作为预测指标,数据集为 2020-01-01—2023-12-31 的监测数据,共实测 8 215 条数据。

1.2 数据预处理

本研究使用的监测数据集某些时间点的水质指标存在缺失值。缺失值的存在会削弱序列信息的完整性,同时会扰乱数据挖掘流程。预先处理缺失值是确保数据质量和分析有效性的关键步骤。KNN(K-nearest neighbors)插值法^[23]是一种广泛使用的缺失值插补方法,使用 KNN 插值法来填补 DO 浓度数据缺失值,共补插值 540 个,填补插值后共有 8 755 条数据。对于异常值,采用前后两数的平均值进行替换。

数据归一化处理有助于简化计算、提高数据可比性和模型性能。本研究采用最大最小标准化对数据集进行归一化处理,使结果映射到 [0,1],具体计算公式^[24]如下:

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

式中: x 为水质指标的实测值; $\max(x)$ 、 $\min(x)$ 分别为水质指标实测值的最大值、最小值; x' 为水质指标归一化后的结果。

1.3 变分模态分解

VMD 是一个解决变分问题的过程^[25],基本思想是将信号分解为一系列具有不同频率和幅度的模态分量,从而降低水质要素预测的复杂度。VMD 是在经验模态分解(Empirical Mode Decomposition, EMD)的基础上提出的。EMD 依赖信号本身特征自适应分离,在分离过程中需要进行多次迭代,并且缺乏统一的停止迭代标准,每次得到的本征模函数(intrinsic mode function, IMF)各不相同^[26]。而 VMD 需要先迭代得到变分函数的最优解,然后根据解来确定每个分量的中心频率和带宽,再自适应分离信号,有效地抑制了 EMD 的模态混叠现象。本研究假设初始信号为 f ,构造得到约束变分模型,如下式所示:

$$\begin{cases} \min_{(u_k, \omega_k)} \left\{ \sum_k \left\| \partial_t \left\{ \left[\delta(t) + \frac{j}{\pi t} \right] u_k(t) \right\} e^{-j\omega_k t} \right\|_2^2 \right\} \\ \text{s.t.} \quad \sum_k u_k = f \end{cases} \quad (2)$$

式中: ∂_t 为偏导运算符; $\delta(t)$ 为狄拉克分布函数,表示时间 t 上的脉冲信号; j 为虚数单位; u_k 和 ω_k 分别为分解后子序列与中心频率的集合^[27]。

为了解决该模型的约束最优化问题,利用二次惩罚项和拉格朗日乘子法,将约束变分问题转变为非约束变分问题,最后使用交替方向乘子法(ADMM)来求解。

1.4 双向长短期记忆网络

1997 年 Hochreiter 等^[28]提出 LSTM 模型,创造性地使用了门控单元结构,为时间序列建模的发展奠定了重要基础。1999 年, Gers 等^[29]将一种改进后的遗忘门引入 LSTM 中,使其可以自动重置神经元。LSTM 作为循环神经网络中一种特殊网络结构,其网络结构如图 1 所示。

图 1 中,从左到右依次是遗忘门、输入门和输出门。 x_t 为 t 时刻的输入状态; h_t 为 t 时刻隐含层输出状态; C_t 为 t 时刻记忆单元状态; f_t 、 i_t 和 o_t 分别为遗忘门、输入门以及输出门状态计算结果; σ 和 $\tan h$ 为激活函数。这些门控机制使用 Sigmoid 函数来控制

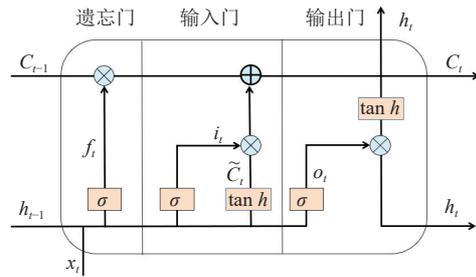


图 1 LSTM 网络结构

Fig.1 LSTM network architecture

制信息流动的程度。

BiLSTM 通过将 LSTM 层沿着时间轴前向和后向运行来计算双向隐藏状态^[30],其神经网络的结构如图 2 所示。前向 LSTM 从序列的第一个元素到最后一个元素顺序计算,而后向 LSTM 则相反,结合前后 2 个方向上的数据特征对当前时刻的数据进行预测^[31], \vec{h}_t 和 \overleftarrow{h}_t 分别为向前和向后隐藏状态,计算过程见式(3)~式(5)。

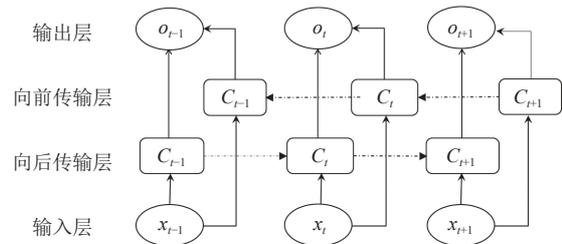


图 2 BiLSTM 网络结构

Fig.2 BiLSTM network architecture

向前传输层:

$$\vec{h}_t = \overrightarrow{\text{LSTM}}(h_{t-1}, x_t, C_{t-1}) \quad (3)$$

向后传输层:

$$\overleftarrow{h}_t = \overleftarrow{\text{LSTM}}(h_{t+1}, x_t, C_{t+1}) \quad (4)$$

t 时刻的隐藏状态:

$$H_t = \left[\vec{h}_t + \overleftarrow{h}_t \right] \quad (5)$$

1.5 构建河流水质预测模型 VMD-GA-BiLSTM

基于梯度在局部范围内搜索优化求解时,随应用范围扩大,BiLSTM 虽具有一定自适应学习、非线性复杂映射、概括推广等能力,但目标函数对象复杂的训练过程中易出现搜索能力较差、陷入局部极值点、收敛学习速度慢、训练效果不佳等问题。GA 通过模拟自然选择和遗传机制在解空间中进行全局搜索,能够自动调整 BiLSTM 神经网络的隐含层神经元个数、学习率和训练次数,以找到最适合当前水质预测任务的模型配置^[32-33]。这种优化方法不仅提高了模型的预测精度和泛化能力,还减少了人工调参的复杂性和不确定性。将 GA 中匹配能力和自适应能力最强的参数引入 BiLSTM 模型进行迭代强化训

练,当迭代次数无法更新时,强化训练终止。GA 改进 BiLSTM 模型的基本流程如图 3 所示。

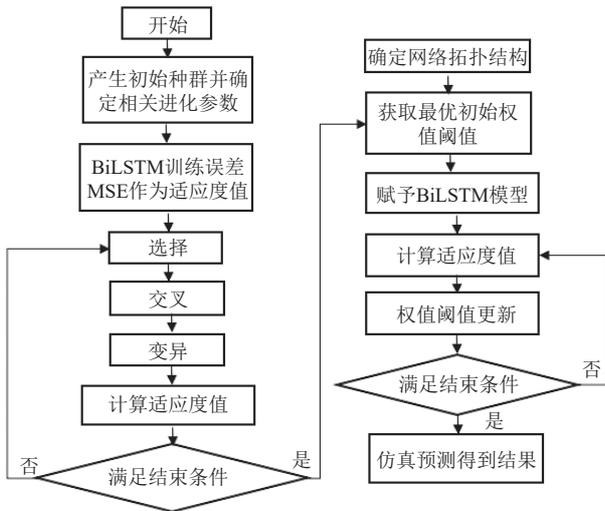


图 3 GA-BiLSTM 预测模型

Fig.3 GA-BiLSTM forecasting model

由 BiLSTM 模型随机初始化一组参数,利用 GA 算法生成初始种群,每个种群含 BiLSTM 完整参数,并赋给 BiLSTM 模型进行训练。将 BiLSTM 模型训练所得均方根误差(MSE)作为个体适应度值,并选择高适应度个体进行交叉变异,生成新种群。当满足结束条件时,选择适应度值最好的个体的参数赋予 BiLSTM 模型,然后对 BiLSTM模型进行训练和测试,输出预测结果。

面对原始时间序列数据的随机波动性,采用 VMD 算法捕捉时序数据不同时间尺度下的非线性特征,将原始水质数据分解为一系列更为简单平稳的子序列,并结合深度学习算法构建分量预测模型,整体适应性更强,模型结构如图 4 所示。根据算法流程图,将模型的具体实现分为 4 个部分,分别是: 1)通过皮尔逊相关系数法,选取 9 个水质指标中与 DO 相关性较强的指标作为模型的输入变量。2)采用 VMD 将选取的水质指标数据按时序分解为一系列较为平稳的模态分量,从而降低预测难度。利用对中心频率的计算确定其分解层数(K)。3)使用 GA 对超参数进行调优,建立 GA-BiLSTM 模型。堆叠两层 BiLSTM 充分挖掘时序关联性,有效解决长期依赖问题,并设置随机失活层,随机减少网络中相互连接的神经元数量,以防止过拟合现象。4)将 VMD-GA-BiLSTM 模型预测的各分量结果进行叠加,得到最终的 DO 预测值,并与实际数据进行对比,分析预测误差状况。

1.6 评价指标

水质预测属于回归预测,本研究主要通过均方

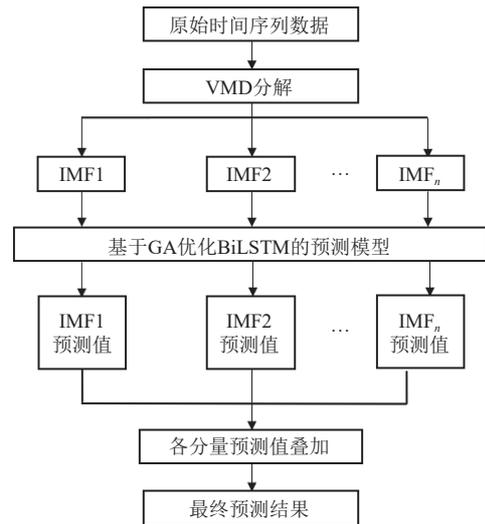


图 4 组合预测模型

Fig.4 Combined forecasting model

根误差(RMSE)、平均绝对误差(MAE)和决定系数(R²)等评价指标进行预测模型评价,计算公式^[34]如下:

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (f_i - x_i)^2} \quad (6)$$

$$MAE = \frac{1}{m} \sum_{i=1}^m |f_i - x_i| \quad (7)$$

$$R^2 = 1 - \frac{\sum_{i=1}^m (f_i - x_i)^2}{\sum_{i=1}^m (\bar{x}_i - x_i)^2} \quad (8)$$

式中: m 为测试集样本数量(1 280); f_i 为第 i 个样本的预测值; x_i 为第 i 个样本的实测值; \bar{x}_i 为所有实测值的平均值。 R^2 介于 0~1, 越接近 1, 回归拟合效果越好, 一般认为 R^2 超过 0.8, 模型具有较高的拟合优度。

2 结果与讨论

2.1 模型超参数设置

艾辛庄监测断面的 DO 实测浓度数据如图 5 所示。从图 5 可以看出 DO 浓度展现出显著的周期性与季节性特征, 波动性较强。将 2020-01-01—2023-05-31 的 DO 浓度数据作为训练集, 2023-06-01—2023-12-31 的数据作为测试集, 训练集和测试集的数据量分别为 7 475 和 1 280 条。数据经过 VMD 分解后输入预测模型进行训练。在预测模型训练阶段, 设定 GA 模型的种群规模为 20, 交叉概率为 0.5, 变异概率为 0.1, 最大进化次数为 10。经反复训练, 设置 BiLSTM 模型中隐藏层为 128 个神经元, 200 次迭代训练, 学习率为 0.004, 优化器为 Adam, 迭代步数 batch-size 为 32。最后, 为了防止模型在训

练过程中出现过拟合, 引入了 Dropout 机制, 并将其比例设置为 0.3。这一比例在减少神经元间依赖性的同时, 保持了模型足够的表达能力, 有助于提升模型的泛化能力。考虑到时间序列输入长度对神经网络性能的影响, 需要在两方面取得平衡: 如果长度太短, 神经网络可能无法充分利用历史信息; 反之, 可能由于历史信息过多难以从中提取关键信息。因此, 本研究将输入时间序列长度设为 10。

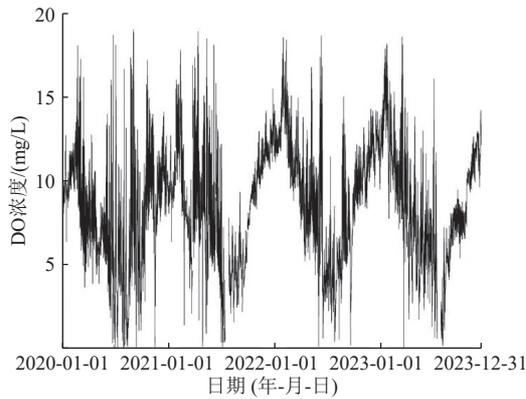


图 5 艾辛庄监测断面 DO 浓度数据

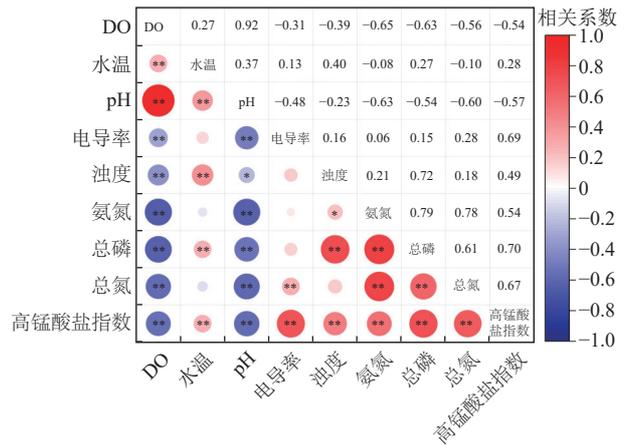
Fig.5 DO concentration data at Aixinzhuang monitoring section

2.2 输入因子集选择

本研究获得的数据包含 9 个水质指标, 若全部作为输入变量, 会增加模型复杂度, 从而降低模型预测能力。首先对原始数据集进行皮尔逊系数分析, 图 6 表示该数据集不同特征之间皮尔逊系数相关性 (图中圆形面积大小和颜色深浅代表相关性强弱; 红色代表正相关, 蓝色代表负相关)。可知各水质指标与 DO 浓度的关联性从高到低依次为 pH(0.92)、氨氮(0.65)、总磷(0.63)、总氮(0.56)、高锰酸盐指数(0.54)、浊度(0.39)、电导率(0.31)、水温(0.27)。根据皮尔逊相关系数法选取与 DO 浓度相关系数大于 0.5 的水质指标, 即选取 pH、氨氮、总磷、总氮、高锰酸盐指数 5 个指标作为多元变量输入至模型中对 DO 浓度进行预测。

2.3 VMD 分解

该研究通过深入分析相邻分解模态末尾序列的中心频率变化特征, 结合参数常规设定与实验验证, 最终确定 $K=10$ 为最佳分解值。原始数据经过 VMD 分解后提取到固有模态 IMF1~IMF10 (图 7), 这表示原始数据经过 VMD 分解后提取到了 10 个固有模态函数。每个 IMF 都代表了信号中的一个特定模态, 它们共同构成了原始信号的完整表示。在这 10 个序列中, 第 1、2 个序列变化缓慢, 波动性弱, 能够体现 DO 浓度的波动趋势; 第 3~10 个序列周期性明



注: *表示在 0.05 水平上显著相关; **表示在 0.01 水平上显著相关。

图 6 皮尔逊相关系数矩阵热力图

Fig.6 Heatmap of Pearson correlation coefficient matrix

显, 反映了 DO 的周期性。相比于直接采用水质数据进行预测, VMD 能够减少水质数据的非平稳性对预测精度的影响, 从而提升 BiLSTM 的预测准确度。

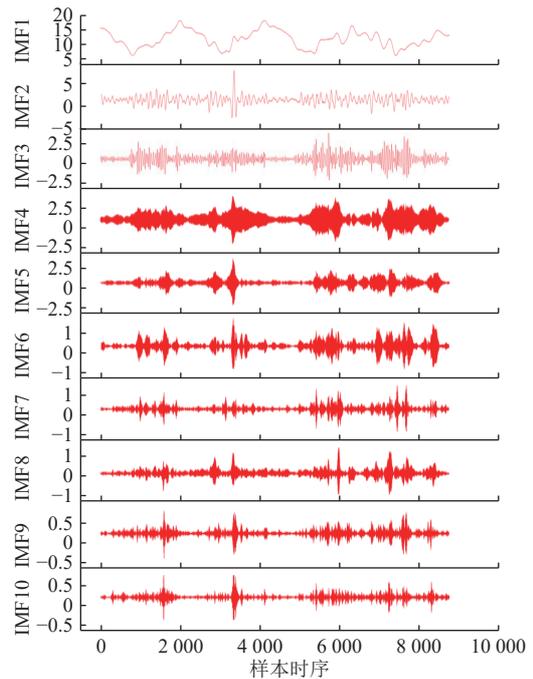


图 7 VMD 提取 DO 浓度模态分量

Fig.7 Modal components of DO concentration extracted by VMD

2.4 VMD-GA-BiLSTM 模型预测分析

采用 VMD-GA-BiLSTM 组合模型以邢台市艾辛庄断面 2020-01-01—2023-05-31 的 DO 浓度数据作为训练集进行训练, 将 2023-06-01—2023-12-31 的 DO 浓度数据作为测试集输入到模型中进行预测, 得到 DO 浓度的预测结果如图 8 所示。可以看出该模型能够较好地预测水质情况的走势, 预测值

和实测值波动性一致。这是因为 BiLSTM 独特的结构设计,能够较为有效地处理时间序列数据,考虑水质数据的前后信息。但同时不可避免地存在一定的滞后性问题,直观表现为图像上的平移错位。在测试集上各评价指标的结果, RMSE、MAE、 R^2 分别为 0.149、0.135 和 0.954, 预测值和实测值拟合度较好。

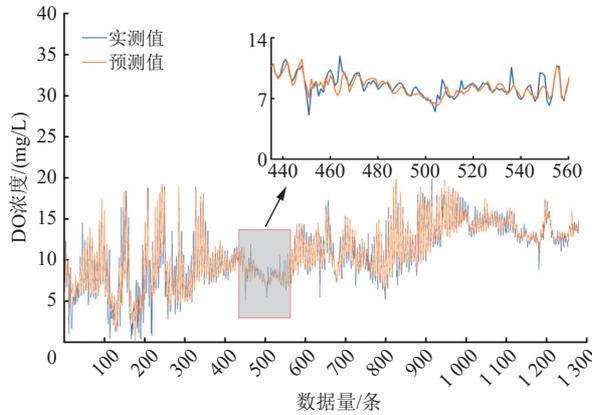


图 8 艾辛庄断面 DO 浓度实测值与 VMD-GA-BiLSTM 模型预测值对比

Fig.8 Comparison between the measured values of DO concentration at Aixinzhuang section and the predicted values of VMD-GA-BiLSTM model

由于 VMD 和 EMD 都是信号分解的方法,均可以与 BiLSTM 神经网络和 GA 结合使用。为验证 VMD-GA-BiLSTM 组合模型在预测水质指标方面的科学性和合理性,选取与构建模型相同的艾辛庄监测断面 2020-01-01—2023-12-31 的 DO 浓度监测数据作为对比模型的数据集,同样将 2020-01-01—2023-05-31 的 DO 浓度数据作为训练集,2023-06-01—2023-12-31 的 DO 浓度数据作为测试集,分别输入 SVM、LSTM、BiLSTM 这 3 种经典模型和 GA-BiLSTM、EMD-GA-BiLSTM 这 2 种组合模型中,预测结果如图 9 所示。从图 9 可以看出,相较其他模型,VMD-GA-BiLSTM 能够更快、更准确地捕捉水质参数的变化趋势,预测结果能更好地响应实际值波动。

由表 1 可以看出,VMD-GA-BiLSTM 模型相较于 SVM、LSTM、BiLSTM、GA-BiLSTM、EMD-GA-BiLSTM 模型, RMSE 分别降低 0.541、0.510、0.464、0.307、0.29, MAE 分别降低 0.474、0.444、0.413、0.173、0.239, R^2 分别提升 0.267、0.245、0.216、0.133、0.088。由各评价指标可看出,使用 BiLSTM 相较于 LSTM 预测效果更好;使用 GA 优化后的 BiLSTM 相较于优化的 BiLSTM 预测效果更好, R^2 提升了 0.083,并且在测试集上的性能及各评价指标结果均有所增强,表明优化后的超参数组合更能

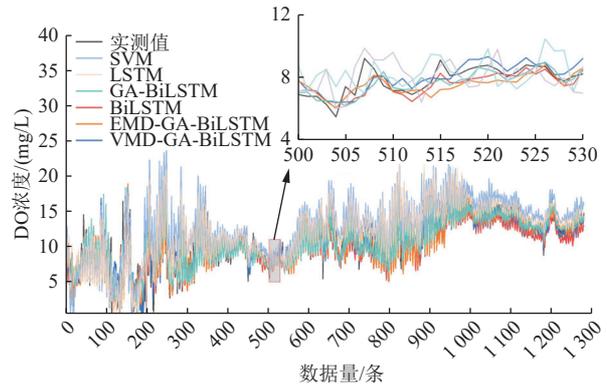


图 9 艾辛庄断面 DO 浓度实测值与不同模型预测值对比

Fig.9 Comparison between the measured values of DO concentration at Aixinzhuang section and the predicted values of different models

适应数据特征,从而提高了模型的泛化能力;使用 EMD 和 VMD 分解后的 GA-BiLSTM 相较于 GA-BiLSTM 预测效果更好, R^2 分别提升了 0.045 和 0.133,经过分解后相较于未经分解的模型预测精度有所提高。本研究构建的 VMD-GA-BiLSTM 模型对 DO 浓度的预测结果拟合度最高(R^2 为 0.954),表明该模型预测值更接近实际值。总体来看,构建的预测模型能够有效提高河流 DO 浓度的预测精度,预测效果优于经典单一模型和其他 2 种组合模型。

表 1 艾辛庄断面 DO 浓度预测模型评价指标对比

Table 1 Comparison of evaluation metrics for DO concentration prediction models at Aixinzhuang section

模型	RMSE	MAE	R^2
SVM	0.690	0.609	0.687
LSTM	0.659	0.576	0.709
BiLSTM	0.613	0.548	0.738
GA-BiLSTM	0.456	0.408	0.821
EMD-GA-BiLSTM	0.439	0.374	0.866
VMD-GA-BiLSTM	0.149	0.135	0.954

2.5 模型泛化能力试验

为了验证该研究模型对其他监测断面及其他水质指标的预测效果,选取邢台市后西吴桥监测断面进行模型泛化能力试验。邢台市牛尾河是海河流域子牙河系滏阳河的一条支流,后西吴桥断面是牛尾河上的国家地表水质考核断面。选取该断面的 pH、DO、氨氮浓度作为预测指标进行试验,将 3 个指标 2020-01-01—2023-12-31 的数据作为数据集,同样使用 2020-01-01—2023-05-31 的数据作为训练集,2023-06-01—2023-12-31 的数据作为测试集。根据皮尔逊相关系数法分析,预测 pH 时选取 DO 作为输入特征,预测 DO 时选取 pH、水温作为输入特征,预

测氨氮浓度时选取总磷浓度作为输入特征。选用 BiLSTM、GA-BiLSTM 和 VMD-GA-BiLSTM 模型进行预测及对比分析。由表 2 可知, 针对 pH、DO 和氨氮浓度 3 项水质指标, VMD-GA-BiLSTM 模型的 RMSE、MAE 均最小, R^2 最大, 充分说明该模型在预测不同站点的水质数据时具有泛化性, 证明了 VMD-GA-BiLSTM 模型预测水质时间序列数据具有通用性和稳定性。

表 2 后西吴桥断面不同指标的模型预测结果对比
Table 2 Comparison of model prediction results of different indexes at Houxiwuqiao section

水质指标	预测模型	RMSE	MAE	R^2
pH	BiLSTM	0.735	0.781	0.757
	GA-BiLSTM	0.523	0.489	0.831
	VMD-GA-BiLSTM	0.154	0.168	0.922
DO 浓度	BiLSTM	0.642	0.672	0.779
	GA-BiLSTM	0.445	0.421	0.840
	VMD-GA-BiLSTM	0.150	0.147	0.906
氨氮浓度	BiLSTM	0.744	0.548	0.689
	GA-BiLSTM	0.564	0.357	0.824
	VMD-GA-BiLSTM	0.169	0.137	0.931

基于试验数据分析, 构建的 VMD-GA-BiLSTM 预测模型针对不同水质指标均显示出卓越优异的预测效能, 其预测值与水质实际观测值之间呈现出良好的拟合性, 表明该模型在水质预测领域的高效性和稳定性, 为后续水质状态的精准评估提供了有效的模型。

3 结论与展望

(1) VMD-GA-BiLSTM 组合预测模型在邢台市艾辛庄断面的 DO 浓度预测任务中展现出了显著的优势, 实现了对非平稳时间序列数据的高效处理和精准预测。从模型效果来看, VMD-GA-BiLSTM 相较于 SVM、传统 LSTM、BiLSTM、GA-BiLSTM 以及 EMD-GA-BiLSTM 等模型, 在 RMSE、MAE 和 R^2 多个评价指标上均取得了更优的表现。RMSE 和 MAE 的显著降低以及 R^2 的显著提升, 表明 VMD-GA-BiLSTM 模型具有更高的预测精度和更强的拟合能力。

(2) 在模型泛化能力试验中, 选取邢台市牛尾河后西吴桥监测断面的 pH、DO、氨氮浓度 3 项指标, 将 VMD-GA-BiLSTM 组合模型和其他模型的预测结果进行对比, 结果该模型的预测值和实测值拟合度最高, 证明了 VMD-GA-BiLSTM 模型在预测河流

水质时间序列数据上具有通用性和稳定性。

VMD 分解与遗传算法(GA)优化的 BiLSTM 模型有效解决了邢台市艾辛庄断面数据不平稳以及神经网络模型的调参问题, 但其应用于河流水质指标预测时仍面临一些挑战。VMD 分解效果受参数选择的影响较大, 不当的参数设置可能导致信息丢失或过度分解, 影响模型性能。遗传算法虽然具有强大的全局搜索能力, 但其收敛速度和局部搜索能力仍有待提升, 容易陷入局部最优解。此外, 河流水质数据的复杂性和多变性也对模型的泛化能力提出了更高要求。后续对河流水质预测的研究包括探索更精准、更高效的信号分解技术, 以提升数据处理能力, 确保数据分解能够准确提取水质时间序列的特征信息; 通过纳入更多水质参数及环境因素增强模型的泛化性, 构建更加全面、综合的水质指标预测模型, 以应对复杂多变的数据, 进一步提升河流水质指标预测的精度和可靠性。

参考文献

- [1] SHAH M I, JAVED M F, ALQAHTANI A, et al. Environmental assessment based surface water quality prediction using hyperparameter optimized machine learning models based on consistent big data[J]. *Process Safety and Environmental Protection*, 2021, 151: 324-340.
- [2] 黄炜惠, 李文攀, 霍守亮, 等. 美国溶解氧基准标准及其对我国的启示 [J]. *环境科学研究*, 2021, 34(6): 1338-1346.
HUANG W H, LI W P, HUO S L, et al. US dissolved oxygen criterion, standard and its revelation for China[J]. *Research of Environmental Sciences*, 2021, 34(6): 1338-1346.
- [3] 周添一, 徐庆, 刘振鸿, 等. 优化 NARX 神经网络对时间序列溶解氧的预测 [J]. *东华大学学报(自然科学版)*, 2022, 48(2): 105-110.
ZHOU T Y, XU Q, LIU Z H, et al. Time-series dissolved oxygen prediction based on optimized NARX neural network[J]. *Journal of Donghua University (Natural Science)*, 2022, 48(2): 105-110.
- [4] GRAVES A, FERNÁNDEZ S, SCHMIDHUBER J. Bidirectional LSTM networks for improved phoneme classification and recognition[C]// *Artificial Neural Networks: Formal Models and Their Applications: ICANN 2005*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005: 799-804.
- [5] 左朝晖, 李绍康, 杨津津, 等. 基于 GA-BP 神经网络的页岩气开发区域水资源承载力研究 [J]. *环境工程技术学报*, 2021, 11(1): 194-201.
ZUO Z H, LI S K, YANG J J, et al. Research on water resources carrying capacity of shale gas development area based on GA-BP neural network[J]. *Journal of Environmental Engineering Technology*, 2021, 11(1): 194-201.
- [6] 李未, 朱梦圆, 王裕成, 等. 基于 NARX 神经网络的千岛湖藻类短期预测模型构建 [J]. *环境科学研究*, 2022, 35(4): 918-925.
LI W, ZHU M Y, WANG Y C, et al. Short-term forecasting model for algae based on NARX neural network in Qiandaohu Reservoir[J]. *Research of Environmental Sciences*, 2022, 35(4):

- 918-925.
- [7] 张梦迪, 徐庆, 刘振鸿, 等. 基于动态滑动窗口 BP 神经网络的水质时间序列预测 [J]. *环境工程技术学报*, 2022, 12(3): 809-815.
ZHANG M D, XU Q, LIU Z H, et al. Prediction of water quality time series based on the dynamic sliding window BP neural network model[J]. *Journal of Environmental Engineering Technology*, 2022, 12(3): 809-815.
- [8] RAMMURTHY S K, PATIL S B. An LSTM-based approach to predict stock price movement for IT sector companies[J]. *International Journal of Cognitive Informatics and Natural Intelligence*, 2021, 15(4): 1-12.
- [9] 谢小良, 吴琳琳. 基于 DA-CNN-BiLSTM 的河流溶解氧浓度预测 [J]. *人民黄河*, 2024, 46(7): 92-97.
XIE X L, WU L L. River dissolved oxygen prediction based on DA-CNN-BiLSTM[J]. *Yellow River*, 2024, 46(7): 92-97.
- [10] SCHUSTER M, PALIWAL K K. Bidirectional recurrent neural networks[J]. *IEEE Transactions on Signal Processing*, 1997, 45(11): 2673-2681.
- [11] 陈树龙, 黎志伟, 黄祖安, 等. 基于 BiLSTM 深度学习模型的污水处理厂水质预测 [J]. *当代化工研究*, 2024(5): 96-98.
CHEN S L, LI Z W, HUANG Z A, et al. Water quality prediction of sewage treatment plant based on BiLSTM deep learning model[J]. *Modern Chemical Research*, 2024(5): 96-98.
- [12] 商艳红, 张静. 基于局部化双向 LSTM 和状态转移约束的养殖水质分类预测 [J]. *渔业现代化*, 2019, 46(2): 28-34.
SHANG Y H, ZHANG J. Aquaculture water quality prediction based on local Bi-LSTM and state transformation constraint[J]. *Fishery Modernization*, 2019, 46(2): 28-34.
- [13] 邹青宏. 基于多时间尺度双向 LSTM 网络的水质预测方法研究 [D]. 重庆: 重庆大学, 2021.
- [14] 姚晔, 许锡伟, 管剑波, 等. 基于双向 LSTM 神经网络的站点周边水位预测系统设计 [J/OL]. *计算机测量与控制*, 2024: 1-12. (2024-02-27)[2025-02-05]. <http://kns.cnki.net/KCMS/detail/deta il.aspx filename=JZCK2024022600A&dbname=CJFD&dbcode=CJFQ>.
- [15] ZHANG J F, ZHU Y, ZHANG X P, et al. Developing a Long Short-Term Memory (LSTM) based model for predicting water table depth in agricultural areas[J]. *Journal of Hydrology*, 2018, 561: 918-929.
- [16] NAJAH A, EL-SHAFIE A, KARIM O A, et al. Application of artificial neural networks for water quality prediction[J]. *Neural Computing and Applications*, 2013, 22(1): 187-201.
- [17] 尚旭东, 段中兴, 陈炳生, 等. 基于双向长短期记忆网络组合模型的水质预测 [J]. *环境科学学报*, 2024, 44(7): 261-270.
SHANG X D, DUAN Z X, CHEN B S, et al. Water quality prediction based on a composite model of bidirectional long short-term memory networks[J]. *Acta Scientiae Circumstantiae*, 2024, 44(7): 261-270.
- [18] LI W S, WEI Y G, AN D, et al. LSTM-TCN: dissolved oxygen prediction in aquaculture, based on combined model of long short-term memory network and temporal convolutional network[J]. *Environmental Science and Pollution Research International*, 2022, 29(26): 39545-39556.
- [19] JIAO J G, ZHAO L Q, HUANG S J, et al. Water quality prediction based on IGRA-ISSA-LSTM model[J]. *Water, Air, & Soil Pollution*, 2023, 234(3): 172.
- [20] 白雯睿, 杨毅强, 郭辉, 等. 基于 VMD-CNN-LSTM 的珠江流域水质多步预测模型研究 [J]. *四川轻化工大学学报 (自然科学版)*, 2022, 35(4): 66-74.
BAI W R, YANG Y Q, GUO H, et al. Research of multi-step prediction model for water quality in Pearl River Basin based on VMD-CNN-LSTM[J]. *Journal of Sichuan University of Science & Engineering (Natural Science Edition)*, 2022, 35(4): 66-74.
- [21] 顾乾晖, 曾斌, 涂振宇. 基于 VMD-PSO-LSTM 水质预测模型的应用研究 [J]. *南昌工程学院学报*, 2021, 40(4): 23-29.
GU Q H, ZENG B, TU Z Y. Research on application of water quality prediction model based on VMD-PSO-LSTM[J]. *Journal of Nanchang Institute of Technology*, 2021, 40(4): 23-29.
- [22] CHEN S Y, HUANG J L, WANG P, et al. A coupled model to improve river water quality prediction towards addressing non-stationarity and data limitation[J]. *Water Research*, 2024, 248: 120895.
- [23] ZHANG W Z, CHENG A M K, SUBHLOK J. DwarfCode: a performance prediction tool for parallel applications[C]//IEEE Transactions on Computers. IEEE, 2016: 495-507.
- [24] 王晨, 李义轩, 吴其琦, 等. 基于 SVM-STL-LSTM 的区域短期电力负荷预测研究 [J]. *水电能源科学*, 2024, 42(4): 215-218.
WANG C, LIY X, WU Q Q, et al. Research on regional Short-term load forecasting based on SVM-STL-LSTM[J]. *Water Resources and Power*, 2024, 42(4): 215-218.
- [25] 滕陈源, 丁逸超, 张有兵, 等. 基于 VMD-Informer-BiLSTM 模型的超短期光伏功率预测 [J]. *高电压技术*, 2023, 49(7): 2961-2971.
TENG C Y, DING Y C, ZHANG Y B, et al. Ultra-short-term photovoltaic power prediction based on VMD-informer-BiLSTM model[J]. *High Voltage Engineering*, 2023, 49(7): 2961-2971.
- [26] 张鼎渊. 基于 EEMD: LSTM 水质预测方法的研究与应用 [D]. 合肥: 中国科学技术大学, 2021.
- [27] 杨语蒙, 李兴东. 基于 GA-LSTM 组合模型的股票价格预测 [J]. *现代计算机*, 2021, 27(33): 1-7.
YANG Y M, LI X D. Stock price prediction based on GA and LSTM combination model[J]. *Modern Computer*, 2021, 27(33): 1-7.
- [28] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. *Neural Computation*, 1997, 9(8): 1735-1780.
- [29] GERS F A, SCHMIDHUBER J, CUMMINS F. Learning to forget: continual prediction with LSTM[J]. *Neural Computation*, 2000, 12(10): 2451-2471.
- [30] PENG H C, XIE P, LIU L J, et al. Morphological diversity of single neurons in molecularly defined cell types[J]. *Nature*, 2021, 598(7879): 174-181.
- [31] ABBASIMEHR H, PAKI R. Improving time series forecasting using LSTM and attention models[J]. *Journal of Ambient Intelligence and Humanized Computing*, 2022, 13(1): 673-691.
- [32] 黄媛. 基于数据驱动的锂离子电池 SOH 与 RUL 预测研究 [D]. 重庆: 重庆理工大学, 2024.
- [33] 张甲甲, 万定生. 基于混合 GA 优化 LSTM 的中小流域流量预测研究 [J]. *计算机仿真*, 2022, 39(2): 283-287.
ZHANG J J, WAN D S. Improved hybrid genetic algorithm and its application in runoff prediction[J]. *Computer Simulation*, 2022, 39(2): 283-287.
- [34] 孔志浩, 卢鹤, 毛建华, 等. 基于多模态和图卷积的飞机部段形变预测方法 [J]. *电子测量技术*, 2023, 46(20): 177-183.
KONG Z H, LU H, MAO J H, et al. Prediction method of aircraft segment deformation based on multi-mode and graph convolution[J]. *Electronic Measurement Technology*, 2023, 46(20): 177-183. ◇