



# 基于改进的 Mogrifier LSTM 算法水质浊度预测模型

杨博韬<sup>1,2</sup>, 刘黎志<sup>1,2</sup>✉

1. 武汉工程大学智能机器人湖北省重点实验室, 武汉 430205; 2. 武汉工程大学计算机科学与工程学院, 武汉 430205

**摘要** 水环境资源保护作为当下最重要的工作之一, 为了提高水质模型预测精度并制定更加完善的水资源管理策略, 现提出一种基于改进的 Mogrifier LSTM 算法水质浊度数据预测模型, 实现对水质数据的精准预测。该模型首先利用 CNN 卷积神经网络对复杂的水质数据进行特征提取, 有效解决了水质数据非线性不稳定的特点, 并对传统 Mogrifier 机制进行优化, 引入扩展系数, 利用 PSO 粒子群优化算法对超参数进行寻优操作, 通过改造后的 Mogrifier 机制对 LSTM 模型中不同时刻的上下文信息进行融合, 增强了水质数据的信息的交互。将预测结果与众多传统模型进行对比, 结果表明 CNN-改进 Mogrifier LSTM 模型具有更好的预测效果。

**关键词** 水质预测; 浊度; 卷积神经网络; 形变长短时记忆网络

近年来, 随着我国现代工业的发展, 全国各流域水质污染防治工作越来越受到重视<sup>[1]</sup>。在水环境保护工作中, 水质预测为流域规划及污染防治措施提供了有效预警<sup>[2]</sup>。提高水质预测模型的精确度和泛化性成为改善水环境质量至关重要的工作之一。

传统的水质预测工作大多使用基于线性分布的模型, 如自回归综合移动平均 (ARIMA) 方法和多元线性回归 (MLR) 模型<sup>[3-4]</sup>。由于水质系统具有非线性及不稳定的特性, 传统数学模型无法准确抓取水质数据的走势。为了解决传统模型的潜在问题, WANG 等<sup>[5]</sup>利用循环神经网络 (RNN) 模型预测位于太湖的三个点位的水质数据, 验证了 RNN 模型在水质预测工作方面的可行性。但 RNN 在训练长序列大规模的数据时会遇到梯度消失和梯度爆炸的问题。为了解决 RNN 存在的问题, 有相关研究人员提出了长短时记忆网络 (LSTM), 该网络是使用门控系统对输入输出进行加权学习, 使其对比 RNN 模型更加有效的捕捉长时间序列之间的信息关联。LSTM 模型已经被广泛用于水质预测领域<sup>[6-8]</sup>。LIU 等<sup>[9]</sup>将 LSTM 模型运用在饮用水水质预测中, 取得了不错的效果。然而单一的 LSTM 模型存在自身计算复杂度高和对大量数据的依赖问题, 在面对更加复杂的时序模型时, 无法保证模型的预测精度。

为了提高 LSTM 模型的性能, 众多学者继续对 LSTM 的融合模型进行研究<sup>[10-12]</sup>。BAEK 等<sup>[13]</sup>利用卷积神经网络 (CNN) 提取特征的能力配合 LSTM 的时序依赖学习预测河流水位及水质, 证明了 CNN-LSTM 模型的性能优于单一 LSTM 模型。KHULLAR 等<sup>[14]</sup>使用两个单向的 LSTM 模型分别从前后同时输入数据信息来预测河流水质因子, 更加精准的推导出未来时刻的水质信息。李浩等<sup>[15]</sup>研究发现 Mogrifier 机制通过新增额外的门控运算改善了上下文信息的交互能力, 有效提高了模型预测精度。由于 Mogrifier 门控机制为了保持经过 sigmoid 运算过后数值的稳定性通常会乘以一个固定常数, 难以应对更加复杂的数据序列。针对上述问题, 对 Mogrifier 机制进行改进, 赋予模型更丰富的超参数, 并通过粒子群优化算法 (PSO) 寻找最优的超参数组合。

收稿日期: 2023-10-19; 录用日期: 2024-01-13

基金项目: 智能机器人湖北省重点实验室创新基金资助项目 (HBIRL202207); 湖北省教育厅科学研究计划指导性资助项目 (B2017051)

第一作者: 杨博韬 (2001—), 男, 硕士研究生, 1793009453@qq.com ✉通信作者: 刘黎志 (1973—), 男, 硕士, 副教授, llz73@163.com

基于上述研究方法，本研究提出基于 CNN-改进 Mogrifier LSTM 的水质浊度预测模型，可以利用 CNN 提取特征配合 LSTM 长时间记忆的优势，并加入改进后的 Mogrifier 门控机制，进而提高水环境质量预测模型的精度。

## 1 数据来源及预处理

本研究使用的数据来自于湖北省荆州市江陵县环境监测中心对柳口站的水质监测数据。为确保最终模型的实用性和可靠性，本研究选用了柳口站 2019 年 1 月 1 日至 2021 年 12 月 31 日的水质监测数据。数据每 4 个小时收集 1 次，评测标准按照《地表水环境质量标准》执行，总共监测 9 项评测指标：水温 (Temp)、pH、溶解氧 (DO)、电导率 (EC)、浊度 (THM)、高锰酸盐指数 (COD<sub>Mn</sub>)、氨氮 (NH<sub>3</sub>-N)、总磷 (TP) 及总氮 (TN)。本研究采用浊度作为输出变量来展示模型的预测效果。表 1 显示了全部数据的描述性统计。

表 1 水质监测数据的描述性统计  
Table 1 Descriptive statistics of water quality monitoring data

统计数值	水温/℃	pH	溶解氧/ (mg·L <sup>-1</sup> )	电导率/ (μS·cm <sup>-1</sup> )	浊度/NTU	高锰酸盐指数/ (mg·L <sup>-1</sup> )	氨氮/ (mg·L <sup>-1</sup> )	总磷/ (mg·L <sup>-1</sup> )	总氮/ (mg·L <sup>-1</sup> )
平均值	18.73	7.86	9.19	380.76	23.36	1.51	0.03	0.08	1.69
最大值	35.00	9.06	25.70	543.00	415.30	8.03	0.55	0.38	4.64
最小值	5.00	6.00	3.61	64.60	0.60	0.20	0.00	0.00	0.11

通过观察表 1 可发现，由于检测工具故障或数据传输错误等因素的影响。高锰酸盐指数及总磷指标的最低值出现了 0 或者小于 0 的情况，将此类数据进行删除处理。同时由于数据在收集传输过程中会存在数据丢失的问题，缺失值会对数据分析及水质预测模型的搭建造成影响。通过观察原始数据，发现浊度缺失值占比仅为 0.57%，为了提高模型的精准度和数据的真实性，本研究选择直接删除缺失值，避免人为填充数据与真实值之间误差较大。同时使用标准差法对数据中出现的异常极值进行筛选。在处理完缺失值和异常极值后，为消除数据之间量纲与单位的影响，对数据进行归一化处理。通过对特征进行归一化处理，使得各个特征拥有相似的尺度。本研究选取最大最小归一化，利用每列特征中的最大值和最小值对数据进行标准化处理。

## 2 理论与研究方法

### 2.1 改进 Mogrifier LSTM

在传统 LSTM 模型中，上一时刻的单元状态和当前输入值在进入下一时刻时是相互独立的，并没有互相交换信息。而这种上下文独立的输入会导致部分信息的流失，从而影响模型精度的提升。为了解决上述问题，在原始的 LSTM 模型上建立了 Mogrifier 机制，增加了信息进入细胞状态前的信息交互能力。

在 Mogrifier 结构中，每一次迭代都会将一项信息与通过 sigmoid 激活函数的另一项相乘形成新的信息，并且根据当前的迭代次数来决定上一时刻单元状态和当前输入值的轮替交互信息。由于经过 sigmoid 运算后，结果会分布于 0 到 1 之间，经过几次迭代后结果会逐渐变小，从而导致梯度消失的问题。所以为了解决上述问题，会在运算后将结果乘 2，保证数值的稳定性。具体公式如下式 (1)、(2) 所示。

$$x^i = 2\sigma(Q^i h^{i-1}) \odot x^{i-2}, \text{ for odd } i \in [1, \dots, r] \quad (1)$$

$$h^i = 2\sigma(R^i x^{i-1}) \odot h^{i-2}, \text{ for even } i \in [1, \dots, r] \quad (2)$$

式中： $x^i$  为当前输入值及后续在 Mogrifier 结构迭代的中间值； $h^i$  为当上一时刻单元状态及后续在 Mogrifier 结构迭代的中间值； $\sigma$  为 sigmoid 运算； $Q^i$  和  $R^i$  是为了帮助  $x^i$  和  $h^i$  交互而额外设立的矩阵。Mogrifier 门控结构如图 1 所示。

在传统 Mogrifier LSTM 模型中，存在 Mogrifier 机制迭代次数  $i$  这一超参数，同时原作者为了解决多次 sigmoid 运算结果逐渐变小的问题，简单将运算结果乘上 2。本研究为了让模型更加贴合水质数据非线性

不稳定的特性，提出改进 Mogrifier LSTM 网络模型，将固定系数修改为扩展系数  $n$ ，赋予传统 Mogrifier 机制更加丰富的超参数来面对复杂的水质数据，并通过 PSO (粒子优化算法) 来同步选取迭代次数  $i$  与扩展系数  $n$  的最优值，改进后的 Mogrifier LSTM 公式如式 (3)、(4) 所示。

$$x^i = n\sigma(Q^i h^{i-1}) \odot x^{i-2}, \text{ for odd } i \in [1, \dots, r] \quad (3)$$

$$h^i = n\sigma(R^i x^{i-1}) \odot h^{i-2}, \text{ for even } i \in [1, \dots, r] \quad (4)$$

模型中，如果迭代次数为 0 时，模型为最初始的 LSTM 模型。通过迭代次数  $i$  与未知数  $n$  来修改模型结构。改进后的 Mogrifier 门控结构如图 2 所示。

在改进 Mogrifier 机制中，将迭代次数  $i$  与扩展系数  $n$  加入粒子群算法 (PSO) 中，通过模拟鸟群的捕食行为来不断调整粒子的位置与速度，在全局寻找所寻参数的最优值。

### 2.2 CNN-改进 Mogrifier LSTM 模型

基于 CNN、改进 Mogrifier 机制和 LSTM 等模型的理论基础，建立 CNN-改进 Mogrifier LSTM 模型，模型的结构图如图 3 所示。

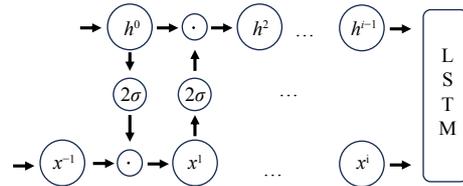


图 1 Mogrifier 基本结构图

Fig. 1 Mogrifier basic structure diagram

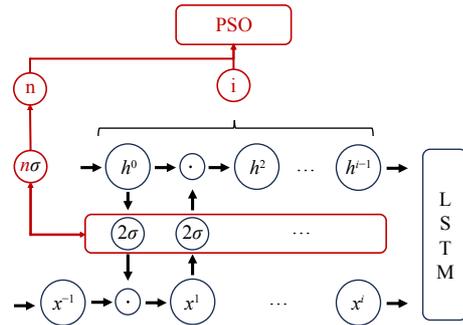


图 2 改进型 Mogrifier 基本结构图

Fig. 2 Basic structure diagram of improved Mogrifier

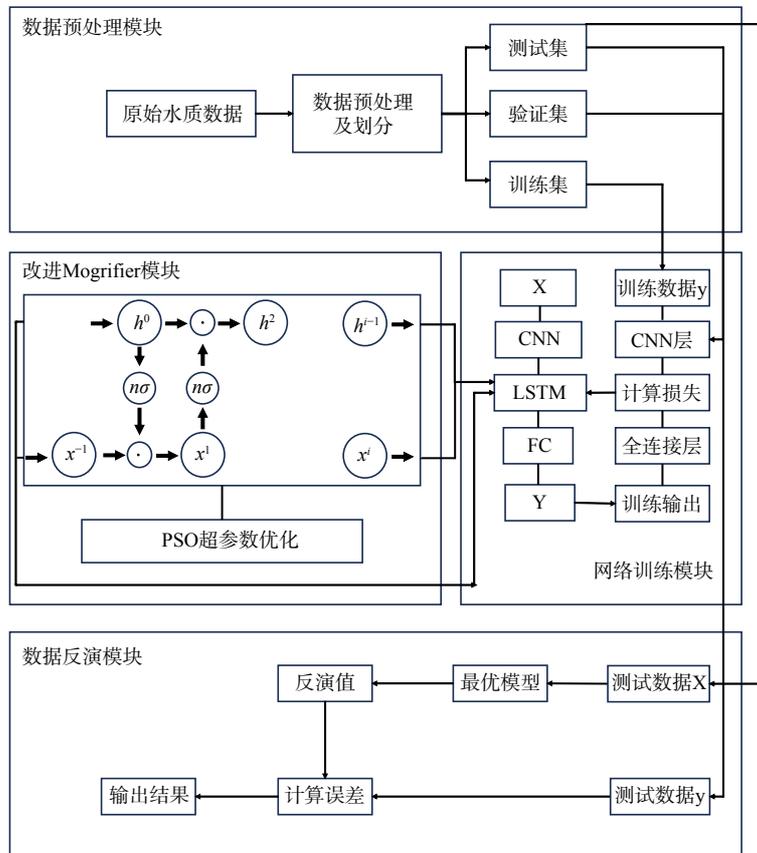


图 3 CNN 改进的 Mogrifier LSTM 神经网络模型结构

Fig. 3 CNN Improved Mogrifier LSTM Neural Network Model Structure

模型的具体实施步骤如下。

1) 数据预处理：对水质数据的缺失值进行剔除，将剔除后的数据标准化后按照 6:2:2 的比例划分数据集为训练集、验证集和测试集。

2) CNN 层：将水质数据作为 CNN 层的输入数据，提取水质数据的特征，构造对应的特征向量，并将特征向量输入 LSTM 层。

3) LSTM 层：接收 CNN 层的输出数据，对数据进行时序预测。LSTM 中的细胞状态容纳了历史网络中所有信息，并在进入新阶段时根据设立的算法对数据信息进行筛选。LSTM 层通过 3 种门控状态来对数据信号进行遗忘或记忆功能。

4) 改进 Mogrifier 层：在 LSTM 层中，每一时刻的单元状态和当前输入值会先进入 Mogrifier 层进行交互信息的过程，并通过 PSO 优化算法获取超参数的最优值。

5) 评价标准：将模型的最终输出与真实测量水质数据进行评价，计算出 RMSE、MAE、Mape 等评价指标，并与其他模型进行对比。

### 3 结果

#### 3.1 改进 Mogrifier 机制模型与其他模型的对比

在将预处理后的水质数据按照 6:2:2 的比例划分为训练集、验证集和测试集后。首先将训练集的水质数据输入到模型的 CNN 层，对水质特征进行提取。处理后的特征向量会经过改进后的 Mogrifier LSTM 模型的时序数据预测。同时在模型的改进 Mogrifier 层中，经过多轮筛选，选择出了 Mogrifier 门控机制的最优交互轮次  $i$  与最优扩展  $n$ ，结果分别为 5 和 2.035。将最优参数与模型用来训练测试集数据得到水质浊度预测曲线如图 4 所示，可以发现预测曲线 pred 能够准确把握真实值曲线 true 的走势。

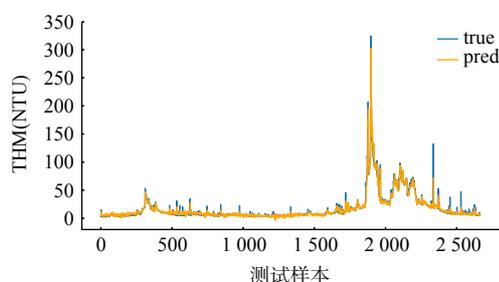


图 4 CNN-改进 Mogrifier LSTM 模型预测结果  
Fig. 4 CNN Improved Mogrifier LSTM Model Prediction Results

为了验证 CNN-改进 Mogrifier LSTM 预测模型的精度，本研究采用 LSTM 模型、CNN-LSTM 模型、Bi-LSTM 模型、CNN-Mogrifier LSTM 模型、Mogrifier LSTM 模型来预测水质数据中浊度指标，图 5 为各类模型的预测结果与真实值之间的误差对比图。

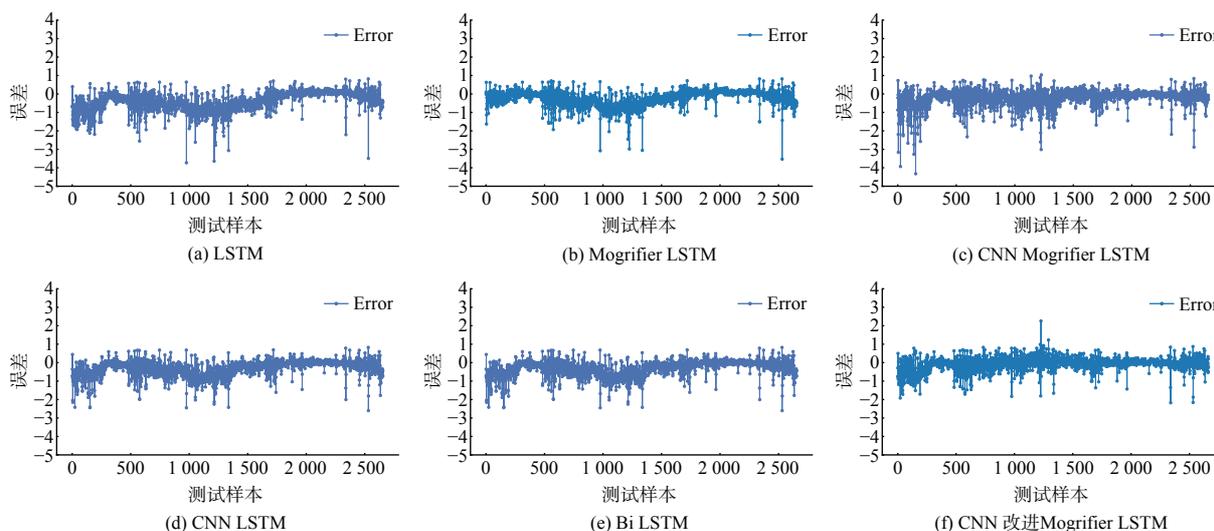


图 5 各模型的预测结果误差对比

Fig. 5 Comparison of prediction results errors among different models

通过观察 CNN-改进 Mogrifier LSTM 模型与上述其他模型的对比结果, CNN-改进 Mogrifier LSTM 模型的拟合效果最好, 测试数据的误差均较小且稳定。说明无论是在面对日常值的预测还是突发污染物的预警都表现出不错的水平。其他模型在面对日常值的预测中往往会有部分偏差。为了更加直观展示各类模型的拟合效果, 利用均方根误差 (RMSE)、平均绝对误差 (MAE)、平均绝对百分比误差 (MAPE)、决定系数 ( $R^2$ ) 各项评价指标对模型好坏进行具体评估, 统计结果如表 2 所示。CNN-改进 Mogrifier LSTM 模型的决定系数  $R^2$  最接近于 1, 并且 MAE、RMSE 和 MAPE 都有明显提升, 其中平均绝对百分比误差 MAPE 提升最为显著。可见 CNN-改进 Mogrifier LSTM 模型总体的拟合效果最好, 性能最佳。对比效果最好的 CNN- Mogrifier LSTM 模型, 均方根误差 RMSE、平均绝对误差 MAE 和平均绝对百分比误差 MAPE 分别减低了 2.7%、16.5%、37.0%, 决定系数  $R^2$  提升了 0.4%。这说明了改进后的 Mogrifier 机制对模型预测精度和泛化性的增强均有实际价值。

### 3.2 改进 Mogrifier 机制模型的多步预测比较

为了验证 CNN-改进 Mogrifier LSTM 模型在预测未来不同时间长度的性能和泛化能力, 本研究设定 4~24 h 的步长水质浊度预测实验。测试集在模型 CNN-改进 Mogrifier LSTM 和 CNN-Mogrifier LSTM 中的评价指标 RMSE、MAE、MAPE 如表 3 所示。

通过观察未来 4~24 h 水质数据的结果可以发现, 在不同步长中, 改进后的 Mogrifier 机制模型的 RMSE、MAE、MAPE 基本均优于传统 Mogrifier 机制模型。只有在步长为 20 h 时, 传统 Mogrifier 机制的预测效果较改进的 Mogrifier 机制预测效果略好一些。同时改进后的 Mogrifier 不同步长模型 RMSE、MAE、MAPE 较传统 Mogrifier 机制模型分别降低了 2.31%、12.32%、14.50%。但随着预测步骤的增加, 预测效果也都出现了下降趋势。

## 4 结论

1) 本研究提出 CNN 模型对复杂水质数据进行特征提取, 有效解决了水质数据非线性不稳定的特点, 并结合 LSTM 对长期水质数据进行预测, 再结合改进后的 Mogrifier 门控机制对 LSTM 模型中不同时刻的上下文信息进行融合, 增强了水质数据的信息的交互。

2) CNN-改进 Mogrifier LSTM 模型以明显优于传统模型和传统 Mogrifier 机制模型的精度预测了水环境资源中的浊度数据。相较于传统模型中效果最好的 CNN-Mogrifier LSTM 模型, 该模型 RMSE、MAE、MAPE 分别提升了 16.5%、2.7% 和 37.0%。

3) CNN-改进 Mogrifier LSTM 模型无论面对常规水质预测还是突发污染出现的水质指标极值的预测都拥有较好的效果。

### 参考文献

- [1] 徐祖信. 我国河流综合水质标识指数评价方法研究[J]. 同济大学学报(自然科学版), 2005(4): 482-488.

表 2 各模型评价指标的比较

Table 2 Comparison of evaluation indicators of various models

水质预测模型	MAE	RMSE	$R^2$	MAPE
LSTM	4.403 6	8.084 9	0.892 8	47.316 1
Bi-LSTM	3.670 1	7.008 8	0.919 4	39.335 3
CNN-LSTM	3.380 3	6.815 8	0.923 8	37.165 2
Mogrifier LSTM	3.321 9	6.826 5	0.923 6	32.646 8
CNN- Mogrifier LSTM	3.074 1	6.639 8	0.927 8	29.980 4
CNN-改进Mogrifier LSTM	2.566 8	6.463 7	0.931 4	18.880 4

表 3 各模型评价指标的比较

Table 3 Comparison of evaluation indicators of various models

时间/h	MAE		RMSE		MAPE	
	IMog	Mog	IMog	Mog	IMog	Mog
4	2.566 8	3.074 1	6.463 7	6.639 8	18.880 4	29.980 4
8	5.405 5	6.733 0	11.578 5	12.624 1	39.695 7	51.262 5
12	7.266 6	8.131 3	15.246 1	14.727 8	55.555 2	74.192 9
16	9.318 5	11.179 3	17.980 4	18.617 3	81.168 5	112.554 3
20	10.006 3	9.676 3	20.401 6	20.803 4	83.622 9	73.002 6
24	10.853 6	13.003 5	17.944 2	18.325 5	109.181 1	112.938 3

注: IMog为CNN-改进Mogrifier LSTM模型, Mog为CNN-Mogrifier LSTM模型。

- [2] HO Y J, AFAN A H, EL-SHAFA H A, et al. Towards a time and cost effective approach to water quality index class prediction[J]. *Journal of Hydrology*, 2019, 575: 148-165.
- [3] 石子泊, 邹志红. 基于小波变换的 ARIMA 模型在水质预测中的应用研究[J]. *环境工程学报*, 2014, 8(10): 4550-4554.
- [4] MARLON V, BORGES G S D, BRUNO V M. Multiple linear regression analysis (MLR) applied for modeling a new WQI equation for monitoring the water quality of Mirim Lagoon, in the state of Rio Grande do Sul—Brazil[J]. *SN Applied Sciences*, 2021, 3(1): 1-11.
- [5] WANG H, GAO Y. Elman's recurrent neural network applied to forecasting the quality of water diversion in the water source of lake Taihu[J]. *Energy Procedia*, 2011, 11: 2139-2147.
- [6] PANG J F, LOU W, YAO Z Y, et al. Water quality prediction in urban waterways based on wavelet packet denoising and LSTM[J]. *Water Resources Management*, 2024, 38(7): 2399-2420.
- [7] 郭利进, 许瑞伟. 基于改进果蝇算法的 LSTM 在水质预测中的应用[J]. *长江科学院院报*, 2023, 40(8): 57-63.
- [8] XU H, LV B, CHEN J, et al. Research on a prediction model of water quality parameters in a marine ranch based on LSTM-BP[J]. *Water*, 2023, 15(15): 2760-2760.
- [9] LIU P, WANG J, SANGAIAH K A, et al. Analysis and prediction of water quality using LSTM deep neural networks in IoT environment[J]. *Sustainability*, 2019, 11(7): 2058-2058.
- [10] 李余隆, 张兰, 李立. 基于 GCN-LSTM 的钱塘江南源水质预测研究[J]. *人民黄河*, 2023, 45(12): 83-87+95.
- [11] 苏辉锋, 丁乐声, 王绪旺, 等. 基于 CNN-GRU 混合模型的养殖工船水体溶解氧预测研究[J]. *南方水产科学*, 2023, 19(4): 174-180.
- [12] ALFWZAN W F, SELIM M, ALTHOBAITI S, et al. Application of Bi-LSTM method for groundwater quality assessment through water quality indices[J]. *Journal of Water Process Engineering*, 2023, 53: 103889-103889.
- [13] SANGSOO B, JONGCHEOL P, AHN J C. Prediction of water level and water quality using a CNN-LSTM combined deep learning approach[J]. *Water*, 2020, 12(12): 3399-3399.
- [14] SAKSHI K, NANHEY S. Water quality assessment of a river using deep learning Bi-LSTM methodology: forecasting and validation[J]. *Environmental Science and Pollution Research International*, 2021, 29(9): 1-15.
- [15] 李浩, 于志远, 尹业成, 等. 基于 CNN-Mogrifier LSTM 的人体运动模式识别算法[J]. *电子测量技术*, 2021, 44(21): 95-100.

(责任编辑: 金曙光)

## Water quality turbidity prediction model based on improved Mogrifier LSTM algorithm

YANG Botao<sup>1,2</sup>, LIU Lizhi<sup>1,2,\*</sup>

1. Hubei Key Laboratory of Intelligent Robot, Wuhan Institute of Technology, Wuhan 430205, China; 2. School of Computer Science and Engineering, Wuhan Institute of Technology, Wuhan 430205, China

\*Corresponding author, E-mail: [llz73@163.com](mailto:llz73@163.com)

**Abstract** Water environment and resource protection represent crucial tasks in the present era. To enhance water quality model prediction accuracy and develop more comprehensive water resource management strategies, a water quality turbidity data prediction model based on the improved Mogrifier LSTM algorithm was proposed to achieve accurate prediction of water quality data. Firstly, the model employed CNN to extract features from complex water quality data, effectively addressing the nonlinear and unstable characteristics of such data. Additionally, the traditional Mogrifier mechanism was optimized by introducing an expansion coefficient and hyperparameters was optimized using PSO algorithm. Through modified Mogrifier mechanism, context information from different moments in LSTM models was fused together to enhance interaction among water quality data. Comparison with many traditional models showed that CNN-improved Mogrifier LSTM model yielded better prediction results.

**Keywords** water quality prediction; turbidity; convolutional neural network; Mogrifier LSTM