

基于双向 LSTM 与注意力机制卷积网络的供水量预测模型

邓云虎, 王婷, 陈国炜, 刘 丽

(合肥工业大学 土木与水利工程学院, 安徽 合肥 230009)

摘要: 【目的】供水量预测是智慧水务系统中保障调度合理性与运行效率的有效环节。针对传统长短期记忆网络 (Long Short Term Memory, LSTM) 在局部特征提取不足和时序建模单向性方面的局限, 本文提出了一种基于 CNN-BiLSTM-Attention 的叠加预测模型。【方法】该模型首先利用双向门控循环单元 (BiLSTM) 构建时序解析网络, 深度挖掘供水量序列的前后依赖关系; 其次, 在输入层引入卷积神经网络 (Convolutional Neural Networks, CNN) 模块, 增强模型对供水量数据周期性等局部特征的提取能力; 最后, 在输出层集成注意力机制 (Attention), 通过动态权重分配强化模型对节假日供水量突变等特殊事件的表征能力。【结果】以安徽省农村与城市水厂供水量数据为例, 实验结果表明: 与传统 LSTM 相比, 本模型在农村水厂预测精度提升 55% (MAPE), 在城市水厂提升 41% (MAPE)。参数敏感性分析进一步发现, 不同水厂对模型参数的响应存在差异: 农村水厂对学习率更敏感, 城市水厂则主要受 CNN 卷积核层数的影响。【结论】本研究提出的 CNN-BiLSTM-Attention 模型显著提升了供水量预测精度, 并揭示了不同类型水厂在模型参数敏感性上的差异, 为智慧水务系统的时序预测提供了有效工具与架构参考。

关键词: 用水量预测; 长短期记忆网络; 卷积神经网络; 注意力机制

中图分类号: TU991.31

文献标识码: A

供水量预测不仅有助于实现水资源的合理调配, 保障供需平衡, 还能有效降低冗余供水带来的经济成本。然而, 供水量数据往往受到地区差异、季节变化以及用户用水习惯等多重因素的影响, 呈现出复杂的动态特征。传统依赖经验的方法难以精准刻画这一复杂规律, 因此, 如何实现高精度的供水量预测成为当前供水领域亟待解决的关键问题^[1-2]。

现有的供水量预测方法主要包括回归分析模型与时间序列模型等。这些方法能够在一定程度上实现对需水量的预测。例如翁建华等^[3]采用时间序列中的自回归模型和自回归移动平均模型对城市某区域进行了短期预测, 结果显示预测值相对误差控制在 $\pm 10\%$ 以内, 精度较高。但回归模型在刻画非线性特征时存在局限。ALNSOUR M A 等^[4]基于时间序列方法, 并综合考虑人口、气候等因素开展预测, 验证了该方法的有效性。然而, 该类方法对数据质量要求较高, 易受异常值与缺失值干扰。

近年来, 随着人工智能与机器学习的发展, 深度学习模型在供水量预测中展现更优性能, 尤其在复杂数据特征提取方面表现突出^[5]。例如, NAMDARI H 等^[6]提出将门控循环单元 (Gated Recurrent Unit, GRU) 与卷积神经网络 (Convolutional Neural Network, CNN) 相结合, 并与 CNN 及 CNN-LSTM 模型对比, 结果表明其所构建的 CNN-GRU 模型预测精度最佳。在深度学习框架中, LSTM 凭借其独特的门控机制, 在供水量预测任务中得到了广泛应用。MU L 等^[7]利用 LSTM 模型进行短期预测, 结果表明其预测精度优于 ARIMA、SVR 和随机森林等传统方法。为解决短期预测不稳定性问题, SHAN S 等^[8]提出一种结合注意力机制与 BiLSTM 模型, 并通过 XGBoost 算法优化超参数, 实验结果显示该混合模型优于 LSTM、BiLSTM 及基于注意力的 LSTM 模型。为进一步提升模型的特征提取能力, ABBASIMEHR H 等^[9]进一步提出多层 LSTM 模型, 以增强对非线性特征的识别能力。ZANFEI A

收稿日期/Received: 2025-06-06

修回日期/Revised: 2025-10-22

基金项目: 安徽省自然科学基金项目 (2208085US13); 国家自然科学基金项目 (51479046)。

第一作者: 邓云虎 (2001—), 男, 瑶族, 硕士研究生, 从事农村管网安全理论与技术方面的研究。

E-mail: 1198795352@qq.com。

通信作者: 刘丽 (1972—), 女, 教授, 硕导, 博士, 从事城市管网安全理论与技术方面的研究。

E-mail: lliu@hfut.edu.cn。

等^[10]则将气象数据引入 LSTM 框架, 构建多变量预测模型, 并证明其在精度上明显优于多层感知器和季节性 ARIMA 模型。

尽管上述方法取得了显著进展, 但 LSTM 模型仍存在不足, 其局部特征捕捉能力有限、且单向数据流的建模方式难以全面反映时间序列的双向依赖关系, 从而影响预测精度。为此, 本文提出一种基于卷积神经网络、双向 LSTM 与注意力机制的复合预测模型 (CNN-Bidirectional LSTM-Attention), 简称 CNN-BiLSTM-ATT。该模型通过 BiLSTM 结构捕捉时间序列的前后向依赖关系, 在输入端融入 CNN 模块以提取周期性等局部特征, 并在输出端引入注意力机制对结果进行加权, 强化模型对节假日突增等特殊事件的响应能力。本文将该模型应用于两个农村水厂和两个城市水厂的供水量预测任务, 并与 LSTM、CNN-LSTM 和 CNN-LSTM-ATT 模型进行对比实验, 以验证其有效性和优势。

1 材料与方法

1.1 长短期记忆神经网络

长短期记忆网络是一种特殊的循环神经网络, 它的发明灵感来源于人类的记忆机制, 旨在解决传统 RNN 在处理长序列数据时出现的梯度消失和梯度爆炸问题。LSTM 的核心思想是通过引入“细胞状态”和“门”结构来控制信息的流动, 从而有效捕捉长期依赖关系。门结构包括遗忘门, 输入门和输出门。具体计算公式如下:

$$f_t = \text{sigmoid}(w_f \cdot [h_{t-1}, X_t] + b_f) \#(1)$$

$$i_t = \text{sigmoid}(w_i \cdot [h_{t-1}, X_t] + b_i) \#(2)$$

$$\tilde{C}_t = \tanh(w_c \cdot [h_{t-1}, X_t] + b_c) \#(3)$$

$$O_t = \text{sigmoid}(w_o \cdot [h_{t-1}, X_t] + b_o) \#(4)$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \#(5)$$

$$h_t = O_t \cdot \tanh(C_t) \#(6)$$

式中: f_t 表示遗忘门计算得出的向量值; h_{t-1} 表示上一个单元的输出; X_t 表示当前单元的输入; w_f 表示权重参数; b_f 表示偏差参数; i_t 表示输入门计算得出的向量值; $\tilde{C}_t$ 表示输入门计算得出的当前细胞状态; O_t 表示输出门计算得出的向量值; C_t 表示传入下一单元的细胞信息; C_{t-1} 表示上一单元传入的细胞信息; h_t 表示模型单元的最终输出。

1.2 融合注意力机制卷积神经网络的双向长短期记忆网络

融合注意力机制卷积神经网络的双向长短期记忆网络通过对 BiLSTM 模型耦合 CNN 模块以及注意力机制对模型进行优化。其中 BiLSTM, 全称为双向长短期记忆网络, 是一种基于 LSTM 的改进型循环神经网络。BiLSTM 包含两个 LSTM 层, 一个负责从前向后处理输入序列, 另一个则从后向前处理输入序列, 从而同时捕捉到序列中每个时间步之前和之后的上下文信息。

卷积神经网络主要是通过卷积操作来提取输入数据中的局部特征, 然后通过池化层进行数据降维, 提高计算效率。通过 CNN 可以学习用水量数据周期性等局部特征, 使得模型能够更好地理解数据中的局部结构。

注意力机制来自于对人脑注意力特性的模拟。其通过为输入数据的不同部分赋予不同的权重, 使模型能够动态地关注于与当前任务最相关的信息。它允许模型捕捉到长距离依赖关系, 并有效地处理信息过载的问题。注意力机制的核心思想可以概括为三个步骤: 计算注意力得分, 归一化注意力权重, 加权求和^[11,12]。具体计算公式如下:

$$S_i = O_t \cdot \tanh(W_H S_i + b_s) \#(7)$$

$$a_i = \text{softmax}(S_i) \#(8)$$

$$C_i = \sum_{j=1}^T a_j h_j \#(9)$$

式中: S_i 是注意力得分; a_i 是注意力权重; C_i 是加权求和后的结果; W_H 和 b_s 是权重矩阵和偏置项。

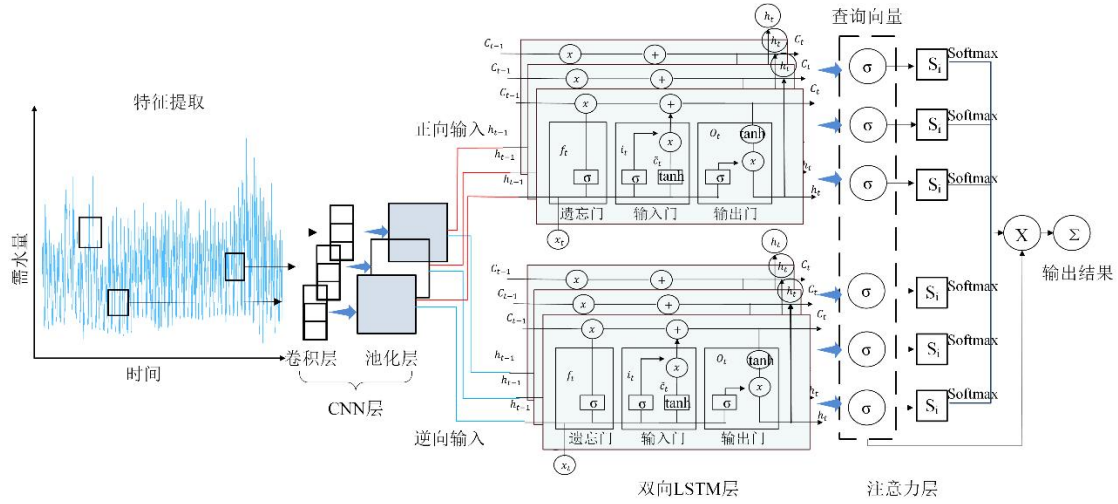


图 1 CNN-BiLSTM-ATT 模型结构图

Fig.1 CNN-BiLSTM-ATT Model Architecture

1.3 数据预处理

本文选取四个供水量数据集，分别来自两个农村水厂 A、B 和两个城市水厂 C、D。各数据集分别有 2880 条数据，时间跨度从 2024 年 2 月 22 日至 2024 年 6 月 20 日，采集时间步长为 1 小时。本实验采用 80%训练集、20%测试集的划分方式：训练集用于模型拟合，测试集用于训练模型后测试模型性能。

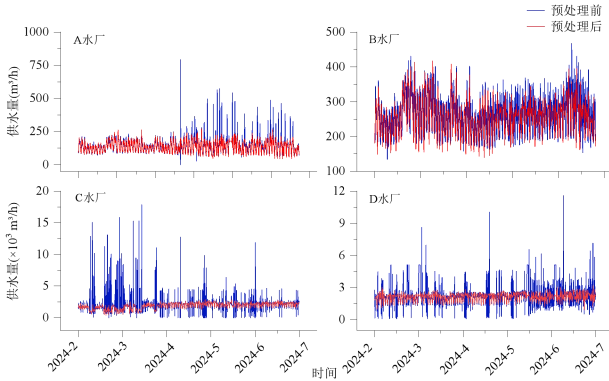


图 2 水厂供水量分布图(左上: A 水厂, 右上: B 水厂, 左下: C 水厂, 右下: D 水厂)

Fig.2 Water Supply Distribution of the Water Plant
(Topleft: Plant A, Top right: Plant B, Bottom left: Plant C, Bottom right: Plant D)

数据集利用箱线图法与小波降噪对进行异常值修正和降噪处理，结果如图 2 所示。本文采用滑窗法进行数据的输入，滑窗法是时间序列预测中一种常用方法，通过将数据分割为固定长度的

窗口，然后以窗口为单位进行数据的输入和训练。从而更好提取每个窗口中数据特征，捕捉数据的变化特征及趋势。

1.4 评价指标

本文采用了均方误差 (Mean Squared Error, MSE) 作为训练集的损失函数。MSE 是回归问题中常用的损失函数，它计算了模型预测值与真实值之间差值的平方的平均值，能够反映模型预测精度的高低。其计算公式如下：

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_{test} - y_{pred})^2 \quad (10)$$

式中: y_{test} 表示真实值; y_{pred} 表示模型的预测值; n 表示样本数量。

为评估预测结果的优劣，本文分别采用平均绝对误差 (Mean Absolute Error, MAE)、拟合优度 (R Squared, R^2)、平均绝对百分比误差 (Mean Absolute Percentage Error, MAPE)、均方根误差 (Root Mean Squared Error, RMSE) 四项指标^[13-15]，具体计算公式如下：

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_{pred} - y_{test}| \quad (11)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_{test} - y_{pred})^2}{\sum_{i=1}^n (y_{test} - \bar{y}_{test})^2} \quad (12)$$

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{y_{pred} - y_{test}}{y_{test}} \right| \quad (13)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_{test} - y_{pred})^2} \quad (14)$$

式中 $\bar{y}_{test}$ 代表监测值的平均值。 R^2 是表示一个随机变量与多个随机变量关系的数字特征，用以反映因变量变化可靠程度。这一指标越接近 1 表示模型拟合效果越好，MAE，RMSE 与 MAPE 均能描述数据的变化程度，其值越小表示模型的预测精度越高。

2 结果与讨论

2.1 基于不同预测模型的训练集效果对比

本文将构建的 CNN-BiLSTM-ATT 模型与三种对比模型（LSTM、CNN-LSTM 和 CNN-LSTM-ATT）分别应用于四个水厂的供水量数据，并在相同条件下进行 100 次迭代训练。结果显示，四种模型的迭代误差均逐渐收敛，表明它们均能够实现对供水量的有效预测。图 3 展示了经过迭代训练后训练集的迭代误差。以 A 水厂为例，CNN-BiLSTM-ATT 在训练集上的效果最优，其均方误差(MSE)由 0.12 降至 0.002，优于 LSTM（初始误差 0.13，最终误差 0.003），并在最终误差上分别比其他三种模型降低了 57% 和 50%。在 B、C 和 D 水厂的数据实验中，CNN-BiLSTM-ATT 同样展现出优势，最终 MSE 分别为 0.002、0.003 和 0.002，较 LSTM 分别降低了 50%、25% 和 50%。

总体来看，CNN-BiLSTM-ATT 在训练集效果上表现最为优异。无论在初始误差、最终误差还是误差收敛的平稳性方面，该模型均优于 LSTM。其优势可能源于双向时序建模与卷积特征提取的协同作用，使模型更充分地捕捉短期和长期数据关联性，从而在训练过程中获得更全面的信息并实现更充分的拟合。

图 3 模型训练集误差收敛图(左上: A 水厂, 右上: B 水厂, 左下: C 水厂, 右下: D 水厂)

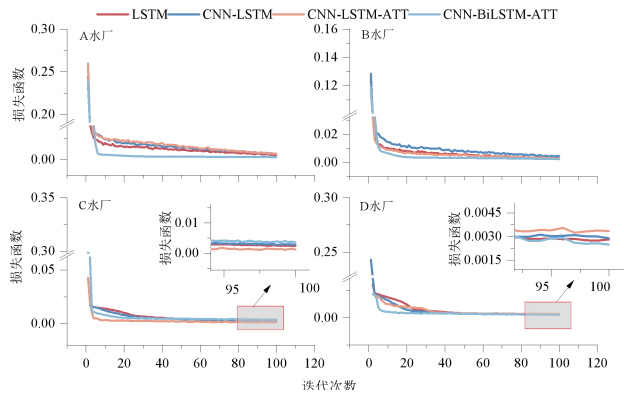
Fig.3 Model Training Error Convergence Curve (Top left: Plant A, Top right: Plant B, Bottom left: Plant C, Bottom right: Plant D)

2.2 基于不同模型的测试集效果对比

为直观比较四种模型在测试集上的表现，图 4 给出了预测效果对比。可以看出，CNN-BiLSTM-ATT 能够较好地捕捉供水量变化趋势，尤其在峰值预测上表现更为精准，其预测曲线与实际监测值高度吻合；相比之下，CNN-LSTM 和 CNN-LSTM-ATT 虽整体拟合度较好，但在局部峰值时段出现偏差；而 LSTM 模型的预测值整体偏离实际监测值，误差较大。

综上，CNN-BiLSTM-ATT 在四种模型中表现最优，不仅能合理反映供水周期特征，还在高峰需求预测中展现出更强的响应能力。

图 5 展示了各模型在测试集上预测值与真实值的拟合情况，数据点越接近对角线，模型预测性能越优。以 A 水厂为例，CNN-BiLSTM-ATT 预与监测值高度一致，拟合度达 0.97，而 LSTM、CNN-LSTM 和 CNN-LSTM-ATT 的拟合度分别为 0.87、0.92 和 0.95，均低于 CNN-BiLSTM-ATT。



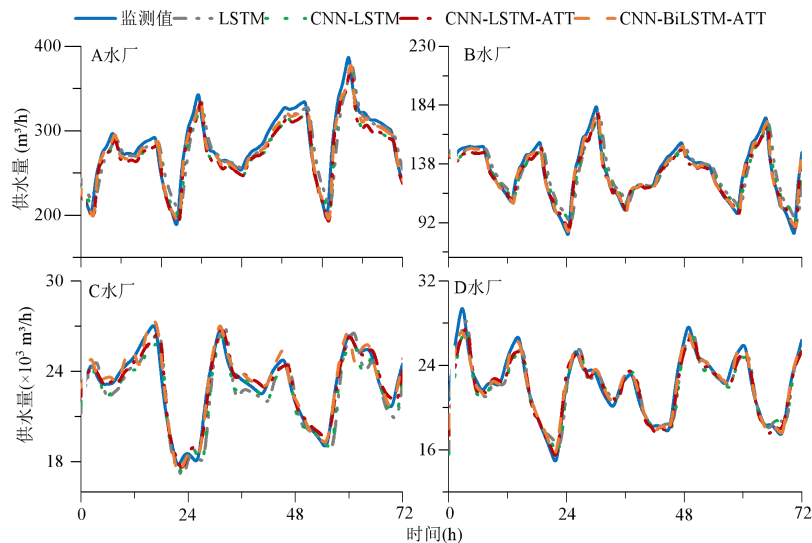


图 4 模型测试集预测结果对比图

Fig.4 Comparison of Model Predictions on the Test Set

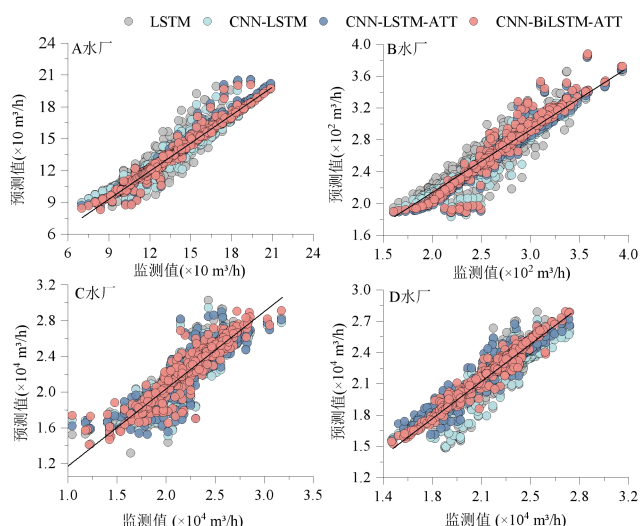


图 5 模型测试集分布对比图

Fig.5 Distribution Comparison on the Test Set

这表明 CNN 模块增强了模型对局部特征的提取能力；而双向 LSTM 克服了传统 LSTM 单向建模的局限，更好地捕获了前后时序依赖，注意力机制能够突出关键特征，从而进一步优化预测效果。在 D 水厂中尤为明显，LSTM、CNN-LSTM 和 CNN-LSTM-ATT 的表现接近，但在引入双向 LSTM 后，模型预测精度显著提升，印证了双向结构在复杂时序建模中的优势。

为更全面评估预测性能，本文引入四个评价指标对模型预测效果进行评估，结果如图 6 所示。以 A 水厂为例，CNN-LSTM-ATT 的 MAE 为 3.56，

较 LSTM (7.57)、CNN-LSTM (6.07) 和 CNN-LSTM-ATT (3.94) 分别降低了 53%、42% 和 10%。在 B 水厂，该模型的 MAE 为 7.38，相较三种对比模型 (13.75、11.76 和 9.84)，分别优化 47%、38% 和 25%。在 C、D，CNN-BiLSTM-ATT 的 MAE 分别为 527.39 和 689.33，均显著优于对比模型（优化幅度在 5%-33%之间）。RMSE 与 MAPE 的结果也表现出一致趋势，进一步验证了 CNN-BiLSTM-ATT 在不同应用场景下的预测精度优势及稳定性。

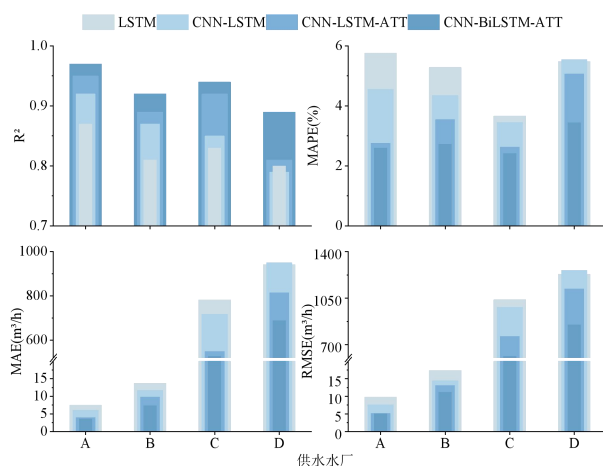


图 6 模型测试集指标对比图(左上: R^2 , 右上: MAPE, 左下: MAE, 右下: RMSE)

Fig.6 Comparison of Model Performance Metrics on the Test Set

2.3 参数敏感性分析

为探究关键参数对模型精度的影响，本文选取卷积核层数、LSTM 隐藏单元数、训练批次数和学习率四个核心参数进行参数敏感性分析，如表 1 所示。其中卷积层主要用于提取输入需水量数据的局部特征，捕捉数据中的空间相关性；而隐藏单元用于存储和处理用水量数据中的信息，是模型的核心模块；训练批次指每次迭代中用于计算梯度和更新参数的样本数量，其大小影响训练的稳定性与收敛速度；学习率控制了模型在每次参数更新时的步长，决定了模型在损失函数上的优化速度。数据集分别选取了农村 A 水厂、城市 C 水厂，并进行对比实验，实验结果如图 7 所示。

在农村水厂数据集时， R^2 在 0.88-0.97 之间波动，RMSE 范围为 3.56~8.34 m^3/h ，MAE 为 2.67-6.69 m^3/h ，MAPE 为 2.03-4.66%。其中，学习率对预测精度影响最大。 R^2 波动可达 10%。而其他参数的影响幅度约为 3%~5%。在城市水厂中， R^2 范围为 0.87-0.94，RMSE 为 612.01~745.68 m^3/h ，MAE 为 527.39~722.84 m^3/h ，MAPE 为 2.42~4.67%。与农村水厂下不同，城市场景中卷积核层数对精度影响最大， R^2 波动约 8%，其余参数的影响幅度较小（2%~4%）。分析其原因，是因为农村水厂供水量数据受季节性和随机波动影响较大，学习率过高可能导致梯度发散，过低则训练效率不足。而城市水厂用水规律性强，如日夜、工作日与周末差异明显，而 CNN 层在提取局部特征时发挥关键作用，其参数设置直接影响后续特征学习质量。

表 1 参数梯度设置

Tab.1 Parameter Gradient Settings

应用 场景	梯度	参数类别			
		卷积核 层数	LSTM 单元数	训练 批次	学习 率
农村	1	16	32	16	0.01
	2	32	64	32	10^{-3}
	3	64	128	64	10^{-4}
	4	128	256	128	10^{-5}
城市	1	16	64	16	0.01
	2	32	128	32	10^{-3}
	3	64	256	64	10^{-4}
	4	128	512	128	10^{-5}

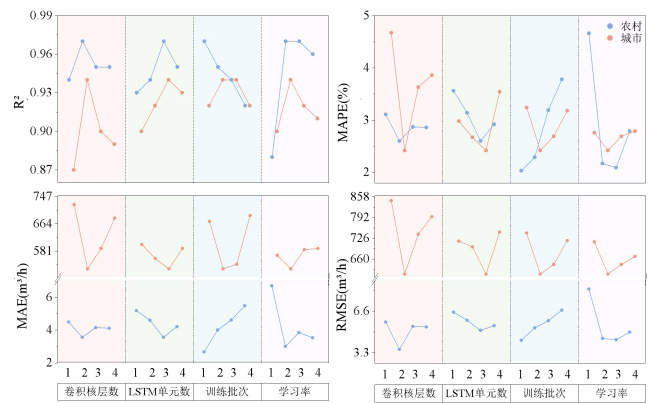


图 7 CNN-BiLSTM-ATT 模型参数敏感性分析(左上: R^2 , 右上: MAPE, 左下: MAE, 右下: RMSE)

Fig.7 Parameter Sensitivity Analysis of the CNN-BiLSTM-ATT Model(Top left: R^2 ,Top right:MAPE,Bottom left:MAE,Bottom right:RMSE)

3 结论

1) 本文提出的 CNN-BiLSTM-ATT 模型在四个水厂的供水量预测中均表现出最佳的预测效果。CNN-BiLSTM-ATT 模型在 A 水厂的供水量预测中， R^2 达到 0.97，比 LSTM 提升了 11%。同时，RMSE 由 9.77 m^3/h 降至 5.07 m^3/h ，降低 48%；MAE 由 7.57 m^3/h 降至 3.56 m^3/h ，下降 53%；MAPE 则由 5.76% 降至 2.60%，降低 55%。在其他水厂的数据中也得到了类似结果。总体而言，CNN-BiLSTM-ATT 在农村水厂的预测精度（MAPE）提升 55%，在城市水厂提升 41%。

2) 在多组不同水厂的数据验证中，CNN-BiLSTM-ATT 模型均展现出优异的泛化能力与稳定性，且预测精度整体维持在较高水平：拟合优度（ R^2 ）在 0.89~0.97 之间波动，平均绝对百分比误差（MAPE）则稳定在 2.6%~3.44%。这一结果表明，该模型在不同应用场景下均能保持良好的鲁棒性与可靠性。

3) 参数敏感性分析表明，CNN-BiLSTM-ATT 模型在不同类型水厂中的参数敏感性存在差异：农村水厂的预测精度对学习率更为敏感，而城市水厂的预测精度则主要受 CNN 卷积核层数影响。这提示在模型应用过程中，需根据不同场景的数据特征，重点调整对预测精度影响较大的关键参数，进而获得更优的预测性能。

参考文献

- [1] 王浏琳,陈家栋,张方敏.基于神经网络的南京市用水量预测[J].水电能源科学,2023,41(12):28-31.[WANG L L,CHEN J D,ZHANG F M.Water consumption prediction in Nanjing based on neural network[J].Water Power & Energy Science,2023,41(12):28-31.]
- [2] 王晓婷,吴以朋,刘书明,等.基于 FFB-LSTM 的供水计量区超短时水量预测方法研究[J].给水排水,2023,59(3):133-139.[WANG X T,WU Y P,LIU S M,et al.Ultra-short-term water consumption forecasting in water metering areas based on FFB-LSTM[J].Water & Wastewater Engineering,2023,59(3):133-139.]
- [3] 翁建华,廉东方,崔晓钰.几种 AR 和 ARMA 模型对城市用水量预测的比较[J].水科学与工程技
术,2014(5):55-57.[WENG J H,LIAN D F,CUI X Y.Comparison of several AR and ARMA models for urban water consumption forecasting[J].Water Science and Engineering Technology,2014(5):55-57.]
- [4] ALNSOUR M A,IJAM A Z.Specifying a cascade water demand forecasting model using time-series analysis:a case of Jordan[J].Sustainable Water Resources Management,2023,9(1):37.
- [5] SHAH S,BEN MILED Z,SCHAEFER R,et al.Differential learning for outliers:a case study of water demand prediction[J].Applied Sciences,2018,8(11):2018.
- [6] NAMDARI H,ASHRAFI S M,HAGHIGHI A.Deep learning-based short-term water demand forecasting in urban areas:a hybrid multichannel model[J].AQUA—Water Infrastructure,Ecosystems and Society,2024,73(3):380-395.
- [7] MU L,ZHENG F,TAO R,et al.Hourly and daily urban water demand predictions using a long short-term memory based model[J].Journal of Water Resources Planning and Management,2020,146(9):05020017.
- [8] SHAN S,NI H,CHEN G,et al.A machine learning framework for enhancing short-term water demand forecasting using attention-BiLSTM networks integrated with XGBoost residual correction[J].Water,2023,15(20):3605.
- [9] ABBASIMEHR H,SHABANI M,YOUSEFI M.An optimized model using LSTM network for demand forecasting[J].Computers & Industrial Engineering,2020,143:106435.
- [10] ZANFEI A,BRENTAN B M,MENAPACE A,et al.A short-term water demand forecasting model using multivariate long short-term memory with meteorological data[J].Journal of Hydroinformatics,2022,24(5):1053-1065.
- [11] NGUYEN T P,YEH C T,CHO M Y,et al.Convolutional neural network bidirectional long short-term memory to online classify the distribution insulator leakage currents[J].Electric Power Systems Research,2022,208:107923.
- [12] 张子瑞,郭伟.基于 CNN-LSTM-Attention 的公共建筑能耗预测[J].机电工程技术:1-8.[ZHANG Z R,GUO W.Energy consumption prediction of public buildings based on CNN-LSTM-Attention[J].Mechanical & Electrical Engineering Technology:1-8.]
- [13] KLEY-HOLSTEG J,ZIEL F.Probabilistic multi-step-ahead short-term water demand forecasting with lasso[J].Journal of Water Resources Planning and Management,2020,146(10):04020077.
- [14] WANG K,XIE X,LIU B,et al.Reliable multi-horizon water demand forecasting model:A temporal deep learning approach[J].Sustainable Cities and Society,2024,112:105595.
- [15] JIANG Q,GUO W,WANG Z,et al.Forecasting regional water demand using multi-fidelity data and harris hawks optimization of generalized regression neural network models – a case study of Heilongjiang Province,China[J].Journal of Hydrology,2024,634:13108

A Water Demand Forecasting Model Based on Bi-directional LSTM and

Attention-Enhanced Convolutional Neural Networks

DENG Yunhu, WANG Ting, CHEN Guowei, LIU Li

(School of Civil and Hydraulic Engineering, Hefei University of Technology, Hefei, Anhui 230009, China)

Abstract: **【Objective】** Water supply forecasting is an effective measure to ensure rational scheduling and operational efficiency in smart water systems. Aiming at the shortcomings of traditional Long Short-Term Memory (LSTM) networks—namely insufficient perception of local features and unidirectional temporal modeling—in water demand prediction, this study proposes an innovative hybrid prediction model based on CNN-BiLSTM-Attention. **【Methods】** The model first employs a Bidirectional Long Short-Term Memory (BiLSTM) network to construct a dual-channel temporal analysis framework, which captures forward and backward dependencies in the water supply time series. Second, a Convolutional Neural Network (CNN) module is introduced at the input layer to enhance the model's ability to extract local patterns such as periodicity in the data. Finally, an attention mechanism is integrated at the output layer to dynamically assign weights and strengthen the model's capability to represent abrupt changes in water demand during holidays and other special events. **【Results】** Experimental validation was conducted using water supply data from rural and urban water plants in Anhui Province, and the proposed model was compared with the traditional LSTM model. The empirical results demonstrate that: (1) Compared to the traditional LSTM, the CNN-BiLSTM-Attention model improves forecasting accuracy by 55% (MAPE) on rural water plant data and by 41% (MAPE) on urban water plant data. (2) In parameter sensitivity analysis, the model exhibited stable performance on urban datasets, with R^2 ranging from 0.88 to 0.97, RMSE between 3.56 and 8.34 m³/h, MAE between 2.67 and 6.69 m³/h, and MAPE consistently within 2.03% to 4.66%. Similar stability was observed for rural water plants. **【Conclusion】** The proposed composite prediction model not only improves forecasting accuracy but also validates the effectiveness of feature fusion strategies. It provides a scalable architecture reference for time series forecasting tasks in the domain of smart water management.

Keywords: water demand forecasting; Long Short-Term Memory; Convolutional Neural Network; attention mechanism