

Multi-time Scale Prediction of Reservoir Water Level Based on ARIMA-LSTM Combined Model

Xuesong Jin¹ Tincheng Wu^{1*} Qingyu Lin² Lingjun Xiao³

1. Shaoxing Metro Products Real Estate Co., Ltd., Shaoxing, Zhejiang, 312099, China

2. Anhui Ningguo Pumped Storage Co., Ltd., Ningguo, Anhui, 242300, China

3. Shenzhen Yuetong Construction Engineering Co., Ltd., Shenzhen, Guangdong, 518019, China

Abstract

In order to predict the water level of reservoir quickly and accurately, and improve the early warning ability of reservoir dispatching. Therefore, this paper adopts a prediction method based on the combined model of differential autoregressive moving average (ARIMA) and long short-term memory network (LSTM) to predict the water level of the Three Gorges reservoir at multiple time scales based on time series. The mean absolute error (MAE), root mean square error (RMSE) and mean absolute percentage error (MAPE) were used as model evaluation indexes to evaluate the accuracy of the model. LSTM combination results showed that the ARIMA model can accurately track the projections for the three gorges reservoir water level, and the accuracy of the combined model is better than that of ARIMA and LSTM any single model, especially in the long term forecast, the advantage of the combined model more apparent, can quickly predict the reservoir water level, timely provide decision-making basis for reservoir dispatching.

Keywords

ARIMA; LSTM; combinatorial mode; water level prediction

基于 ARIMA-LSTM 组合模型的水库水位多时间尺度预测

金雪松¹ 吴铁城^{1*} 林清宇² 肖凌军³

1. 绍兴市地铁物产置业有限公司, 中国·浙江 绍兴 312099

2. 安徽宁国抽水蓄能有限公司, 中国·安徽 宁国 242300

3. 深圳市粤通建设工程有限公司, 中国·广东 深圳 518019

摘要

为了对水库水位进行快速精准预测, 进而提高水库调度的预警能力。因此论文采用一种基于差分自回归移动平均 (ARIMA) 和长短期记忆网络 (LSTM) 组合模型的预测方法, 采用该方法基于时间序列对三峡水库的水位进行多时间尺度预测, 以平均绝对误差 (MAE)、均方根误差 (RMSE) 以及平均绝对百分比误差 (MAPE) 作为模型评价指标, 采用误差评价模型的准确度。结果表明, 采用 ARIMA-LSTM 组合模型能够对三峡水库水位进行准确地追踪预测, 并且组合模型的准确度优于 ARIMA 和 LSTM 任意单一模型, 尤其是在长期预测中, 组合模型的优势更加明显, 可以快速预测水库水位, 及时为水库调度提供决策依据。

关键词

ARIMA; LSTM; 组合模型; 水位预测

1 引言

大型水库的水位预测是水库防洪调度的重要影响因素, 目前常用的水位预测方法分为传统的水动力学模型和人工智能为主的机器学习两个方向^[1]。大型水电站常基于干流和较大支流流量建立与水库水位之间的关系完成预测, 水动力

学方法在计算时的精度难以满足要求, 因此目前多数学者开始探索将机器学习等人工智能技术应用到水位预测中去, 诸如灰色模型^[2]、支持向量机^[3]、时间序列^[4]和神经网络等。神经网络模型是水位预测方面主流的机器学习方法, 马辉、纪国良等学者的研究都取得了不同的进展, 并对长短时记忆神经网络不断改进, 使预测结果精度进一步提升。不论是统计模型还是机器学习, 都会存在自身不足。因此结合两者之间的优势可以构建一种简单、快速、预测效果更好的组合模型 ARIMA-LSTM, 罗龙、李蓉蓉等运用该组合模型预测均取得了较好的效果, 因此论文选用 ARIMA-LSTM 组合模型预测水库水位值。

【作者简介】金雪松 (1997-), 男, 中国浙江绍兴人, 硕士, 从事工程管理研究。

【通讯作者】吴铁城 (1987-), 男, 中国浙江绍兴人, 本科, 工程师, 从事建设公司技术管理研究。

2 研究区域概况

三峡水库，三峡水电站蓄水而成，跨鄂渝 21 县市，水位年调于 145~175m 间，正常蓄水 175m，汛期降至 145m，总蓄水 393 亿 m³。水库运行改变库区水文，汛期 6—9 月水位 145m，非汛期 175m。汛期水位波动大，论文研究汛期水位预测，对防洪、发电、节水及调度具有重要意义。

3 研究方法

3.1 ARIMA 模型

ARIMA 模型全称为差分自回归移动平均模型，由自回归模型 (AR) 与滑动平均模型 (MA) 组合而成，是在 ARMA 模型基础上通过进一步优化实现的，常用于非平稳时间序列预测模型。ARIMA 模型一般表示为 ARIMA (p, d, q) (p, d, q 为模型的阶数)。ARIMA 模型通过 d 阶差分将非平稳时间序列 Y_t 转化为平稳时间序列 X_t ，然后用 ARMA (p, q) 模型对平稳时间序列 X_t 进行运算，ARMA (p, q) 模型为：

$$X_t = \sum_{i=1}^p \varphi_i X_{t-i} + \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \varepsilon_t \quad (1)$$

其中， φ_i 、 θ_j 、 ε_t 分别为自回归系数、滑动平均系数、白噪声序列。

3.2 LSTM 模型

LSTM 模型是 RNN 算法的改进，有效解除了 RNN 在处理时间序列问题时出现梯度消失和梯度爆炸的问题。LSTM 通过三种不同的门 (遗忘门、输入门和输出门) 控制历史信息对当前信息的影响，由许多记忆块构成，记忆块的功能可用以下公式表示：

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (2)$$

$$\begin{cases} i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \\ \tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \\ C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \\ o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \\ h_t = o_t \cdot \tanh(C_t) \end{cases} \quad (3)$$

其中， W_f 为遗忘门的权重； b_f 为遗忘门的偏置； W_i 为输入门 sigmoid 激活函数的权重； b_i 为输入门 sigmoid 激活函数的偏置； W_c 为输入门 tanh 函数的权重； b_c 为输入门 tanh 函数的偏置； C_t 为新细胞状态信息； W_o 为输出门的权重； b_o 为输出门的偏置。

3.3 ARIMA-LSTM 组合模型

将时间序列 Y_t 分解为线性部分 L_t 和非线性部分 N_t ，即 $Y_t = L_t + N_t$ 。ARIMA 模型对线性部分拟合，LSTM 模型对非线性部分拟合。

3.4 模型评价指标

为检验预测模型的拟合度，选取平均绝对误差 (MAE)、均方根误差 (RMSE) 以及平均绝对百分比误差 (MAPE)

作为模型评价指标，采用误差评价模型的准确性，具体公式如下：

$$\begin{cases} MAE = \frac{1}{n} \sum_{i=1}^n |x_i - y_i| \\ RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - y_i)^2} \\ MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|x_i - y_i|}{x_i} \times 100\% \end{cases} \quad (4)$$

其中， n 为样本数量； x_i 为实际水位值； y_i 为预测水位值；MAPE 用来评价模型的准确度，误差结果越小，模型越优，精度越高。

4 实证分析

4.1 数据选取

利用全国及三峡水库水雨情系统数据 (2015—2019 年的 6—9 月)，按 8 : 2 划分训练集与测试集。采用多时间尺度 (时、日、月) 预测水位，包括短期 (1 次 / 小时，预测 1 天)、中期 (4 次 / 天，预测 10 天)、长期 (15 次 / 月，预测 4 月)。以下以短期预测为例进行模型仿真说明。

4.2 ARIMA 模型仿真

由于水位序列具有平稳性和非平稳性的特性，因此需要检验水位时间序列的平稳性。如果时间序列是非平稳性，则需要平稳化处理，论文采用自相关系数 (AFC) 和偏相关系数 (PAFC) 验证时间序列的平稳性。若一个时间序列的自相关系数或偏相关系数表现为拖尾或截尾，那么该时间序列即为平稳；反之，时间序列为非平稳。

①平稳性检验。针对序列 X_t 前 20 个数据进行平稳性检验，根据其自相关系数和偏相关系数可知，序列 X_t 的自相关系数既不是拖尾又不是截尾，因此时间序列 X_t 是非平稳性。接下来对 X_t 进行一阶差分，得到其一阶差分结果也是非平稳，再进行二阶差分得到序列 Y_t 。针对序列 Y_t 前 20 个数据进行平稳性检验，观察其自相关系数和偏相关系数函数图，发现序列 Y_t 的相关系数函数图呈上下震荡特性，因此水位序列 Y_t 是平稳性。

②定阶：采用赤池信息准则 (AIC) 和贝叶斯信息准则 (BIC) 对序列 Y_t 进行定阶，最终确定序列 Y_t 为 5 阶。因此，得到序列 X_t 时序模型为 ARIMA (5, 2, 0)。采用 Yule-Walker 方程计算可得时间序列 X_t 模型的方程为：

$$(1 - 0.128B - 0.2388B^2 - 0.06451B^3 - 0.056B^4 - 0.0231B^5) kX(t) = a_t \quad (5)$$

其中， B 为后移差分算子； a_t 为模型残差。

③预测：将时间序列 X_t 方程展开后可得到预测方程为：

$$X_t = 0.128X_{t-1} + 0.2388X_{t-2} + 0.6451X_{t-3} + 0.056X_{t-4} + 0.0231X_{t-5} + a_t \quad (6)$$

选取三峡水库中心站的 120 个水位数据，每一个小时观测一次数据，进行短期预测，训练集和测试集之比设置为 8 : 2，即选取 4 天水位数据作为训练集，接下来 1 天水

位值作为测试集,预测未来1天的水位值。根据预测结果可知,采用ARIMA模型预测拟合得到的水位预测曲线并不是很平滑,且预测结果的误差比较大,因此有待进一步提高预测精度。

4.3 LSTM 模型仿真

LSTM模型的工作流程如下:

①输入数据预处理:原始数据中常会存在不正常的样本数据,奇异数据往往会引起训练时间增加,甚至造成训练网络无法收敛。为避免发生以上情况,方便处理拟合后的数据,加快模型训练速度,针对样本数据采取归一化处理。将样本数据均归一化处理到[0, 1]之间。具体见公式(7):

$$x'_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \quad (7)$$

其中, x'_i 为 i 时刻归一化后的数据; x_i 为 i 时刻样本数据; x_{\min} 、 x_{\max} 分别为样本数据的最小值和最大值。

②设置网络参数和训练:在实际的模型拟合中,LSTM模型设置一个输入层、一个输出层以及100个隐含层,选取三峡水库中心站的水位数据,进行短期预测,训练集和测试集之比设置为8:2,预测未来一天的水位值。

③输出结果:由于原始数据在预处理时采用归一化处理,因此采用LSTM模型预测的结果并不是所求最终结果,还需将LSTM模型的预测结果进行反归一化处理,最终输出模型的实际预测结果。

结合ARIMA模型预测结果,将LSTM模型与ARIMA模型的预测结果进行比较,LSTM模型预测结果更为接近于测试集,预测曲线变化的趋势基本保持一致,预测误差相对减小,LSTM模型优于ARIMA模型。

4.4 ARIMA-LSTM 组合模型仿真

ARIMA-LSTM组合模型的工作流程如下:

①ARIMA模型建模:使用ARIMA模型对原始数据进行时间序列 Y_t 预测,得到预测值 \hat{L}_t 和残差序列 e_t 。

②LSTM模型建模:将ARIMA模型输出的残差序列 e_t 作为非线性部分 N_t ,将非线性部分 N_t 作为LSTM模型的输入集进行训练。在实际模型拟合中,LSTM模型设置一个输入层、一个输出层以及80个隐含层,选取三峡水库中心站的水位数据,进行短期预测,训练集和测试集之比设置为8:2,预测未来一天的水位值。输入LSTM模型后得到输出结果预测残值 \hat{N}_t 。

③ARIMA-LSTM组合预测:将ARIMA模型得到的线性部分预测值 \hat{L}_t 和LSTM模型得到的非线性部分预测值 \hat{N}_t 求和,得到最终预测值 \hat{Z}_t ,即 $\hat{Z}_t = \hat{L}_t + \hat{N}_t$ 。

结合ARIMA模型和LSTM模型的预测结果,将ARIMA模型、LSTM模型和ARIMA-LSTM组合模型的预测结果进

行比较,ARIMA-LSTM组合模型预测结果更为接近测试集,预测曲变化趋势基本保持一致,预测误差相对减少。因此该组合模型是一种比较有效的时间序列分析模型。

5 模型结果对比分析

为了检验模型的适配度,论文分别选取短期、中期、长期的水位数据进行预测,通过多时间尺度的预测验证模型的准确度。分别利用ARIMA模型、LSTM模型以及ARIMA-LSTM组合模型对三峡水库中心站水位进行多时间尺度预测,预测结果评价验证指标如表1所示。

表1 ARIMA、LSTM、ARIMA-LSTM 三种模型预测结果比较

时间尺度	预测模型	MAE (m)	RMSE (m)	MAPE (%)
短期	ARIMA	0.087	0.958	1.324
	LSTM	0.073	0.082	0.856
	ARIMA-LSTM	0.062	0.068	0.758
中期	ARIMA	0.243	0.352	2.458
	LSTM	0.096	0.128	1.256
	ARIMA-LSTM	0.087	0.098	0.968
长期	ARIMA	0.687	0.886	4.357
	LSTM	0.368	0.468	2.669
	ARIMA-LSTM	0.312	0.389	1.325

根据上述结果,可以从中发现,组合模型ARIMA-LSTM对于短期、中期和长期的水位预测值与实际值适度比传统的时间序列模型和LSTM模型的适配度更高。随着预测时期的增加,所有模型的预测误差也随之增大,其中ARIMA-LSTM模型的预测效果最好,随着时间步长的增加预测结果略有上升,和其他模型比相对稳定。

6 结论

本研究通过对比时间序列模型和长短期记忆网络模型在三峡水库水位预测中的应用,发现单一模型的预测精度有限,进而提出并验证了一种融合ARIMA和LSTM的组合模型,该模型在短期、中期和长期预测中均展现出较单一模型更优的预测性能,对水库防洪调度具有实际应用价值。

参考文献

- [1] 刘志武,王菁,许继军.应用一维水动力学模型预测三峡水库蓄水位[J].长江科学院院报,2011,28(8):22-26.
- [2] 彭令,牛瑞卿,叶润青,等.基于进化支持向量机的滑坡地下水位动态预测[J].中南大学学报(自然科学版),2012,43(12):4788-4795.
- [3] 李蓉蓉,戴永.基于LSTM和时间序列分析法的短期风速预测[J].计算机仿真,2020,37(3):393-398.
- [4] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural Computation, 1997,9(8):1735-1780.