Short-term forecast of Yangtze River water level based on Long Short-Term Memory neural network

Shinan Chen1, * and Yunkai Qiao 2

1Xi'an Jiaotong University, College of Mathematics and Statistics, 710048, Xi'an, Shaanxi, China
2Hohai University, College of Computer Internet of Things Engineering, 213022, Changzhou, Jiangsu, China

* Corresponding author: chensn@stu.xjtu.edu.cn

Abstract. The water level fluctuation forecast of Yangtze River plays an important role in the navigation planning and other areas. This study used LSTM neural network to make short term forecast of the water level of Nanjing navigable river, focusing on the 2 days, 3 days and 5 days-forecast from the past 14 days. The error of mean square reached 0.064, 0.121 and 0.195, showing a rather accurate prediction. The model was also suitable for longer time series prediction. This study optimized the model by adjusting the hyper-parameters using qualitative and quantitative analysis and further decided that the batch size equals to 90 and the epoch equals to 250 in the case of 5 days-forecast from the past 14 days. The prediction accuracy increased by 21% with a good prediction performance, and the error of mean square was reduced to 0.153. It provided a reference for the selection of hyper-parameters for the prediction of water level in Nanjing station by LSTM neural network.

1. Introduction

The Yangtze River is the only waterway in China that traverses the eastern, central and western economic zones, which has a great impact on China's economy, energy, transportation, ecology and many other aspects. As the shipping of the Yangtze River plays an increasingly important role in economic activities, the need to improve shipping capacity and ensure shipping safety becomes increasingly urgent. In this context, the analysis and prediction of the water level characteristics of the Yangtze River will provide reliable reference suggestions for future shipping planning \[1\], so as to take precautions and avoid risks. At the same time, it also provides important data support for the prediction and analysis of the ecological environment of the Yangtze River, the horizontal displacement of the deep soil \[2\], the local epidemic disease \[3\], and the water environment governance.

In terms of research methods, traditional methods for studying water level characteristics are mainly divided into three types: theoretical analysis\[4\], experimental analysis\[5\] and numerical simulation based on dynamic system\[6\]. Among them, the theoretical method is usually not suitable to deal with the problem of water level prediction in the actual complex water environment. The experimental method has high cost, long time, and is not universal. Due to the lack of data support for some parameters of numerical simulation method, it is difficult to give reliable estimation, which increases the difficulty of research and model error. In recent years, deep neural network has shown a good application prospect in time series prediction. At present, there are a few researches on water level prediction using deep neural network at home and abroad. For example, Assem,H et al. \[7\] (2017) used deep convolutional
neural network (CNN) to predict the water level of three areas of the Shannon River, and the prediction accuracy was higher than that of ANN and SVM. Guo Yan et al. [8] (2021) established the daily scale water level prediction model of Dongting Lake by using LSTM and GRU neural networks respectively.

In terms of research objects, most of the research objects of water level prediction in China are groundwater or local lakes and rivers, but there are few researches on water level prediction at each station of the Yangtze River trunk road, and no unified and recognized research method has been formed.

Considering the availability of a large number of data and the high performance of deep neural network in dealing with time series prediction problems, this paper collected water level data of several stations along the Yangtze River trunk road by referring to the relevant studies above. Long Short-Term Memory (LSTM) was constructed to predict the water level of a single site.

2. Preparation for research

2.1. Selection of the study area
In this study, starting from the water level prediction of single station, the water level data of a region is input to the LSTM model for training, and the LSTM prediction model of the water level of the region is obtained. Firstly, the water level data of Nanjing is selected as the research object. Nanjing is located in the lower reaches of the Yangtze River, near the river and the sea, and is the political, economic and cultural center in the south of China. The Nanjing waterway plays a vital role in the main road of the Yangtze River both in terms of geography and economy, so the research on the water level of Nanjing station has far-reaching significance, and also provides a reference for the water level prediction of other important stations. Figure 1 shows the major cities passed by the main stream of the Yangtze River, with the geographical location of Nanjing marked out.

2.2. Model preparation
Long Short-term Memory (LSTM) is a special recurrent neural network, which can effectively solve the gradient vanishing problem of general Recurrent Neural Network (RNN). The theory based on it refers to the way that the human brain remembers things and makes inferences based on memory, and replaces the brain memory with the recurrent core to realize the information extraction of time series. On the basis of traditional RNN, LSTM network improves the way of memory information storage by recurrent core and adds the concepts of threshold control and long and short-term memory, which is more in line with the way of human brain memory information [9].

The working principle of LSTM neural network is shown in Figure 2 and Figure 3. Fig. 2 shows the working principle of a time-stepping recurrent neural network, highlighting the internal structure of the recurrent core. Among them, \(c_t, h_t, n_t\) are called cell state, hidden state and candidate state; \(i_t, f_t, o_t\) are input gate, forgetting gate and output gate respectively. They are all functions of the input characteristics \(x_t\) of the current time step and the memory information of the previous time step \(h_{t-1}\). Input gate represents the process of storing newly acquired information into long-term memory;
Forgetgate simulates the process of forgetting in the human brain over time. Output gate represents the process of extracting short-term memory from long-term memory, specifically as:

\[ i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i) \] (1)

\[ f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f) \] (2)

\[ o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_o) \] (3)

Where \( \sigma \) is Sigmoid activation function, \( W_{xi}, W_{hi}, W_{xf}, W_{hf}, W_{xo}, \) and \( W_{ho} \) are weight matrices, \( b_i, b_f, \) and \( b_o \) are bias items. Cellular \( c_t \) represents long-term memory, memory \( h_t \) represents short-term memory, and candidate state \( n_t \) represents newly acquired information. The working process of the three thresholds shown in Figure 2 is as follows.

\[ c_t = f_t c_{t-1} + i_t n_t \] (4)

\[ h_t = o_t \tanh(c_t) \] (5)

\[ n_t = \tanh(W_{xn}x_t + W_{hn}h_{t-1} + b_n) \] (6)

In the formula, \( \tanh \) is the activation function, \( W_{xn} \) and \( W_{hn} \) are weight matrices, and \( b_n \) is the bias term.

Figure 3 shows the time step expansion model construction of LSTM recurrent core, in which \( x_t \) is the feature of \( t \) time step input network. Since this paper only focuses on the prediction of water level, the input of each step is only the water level data of a certain day, so the input feature of each time step is 1, that is, \( x_t \) is one-dimensional. \( y_t \) is the prediction of the next step based on the storage information of the current recurrent core, namely the output of the LSTM network, which is a fully connected layer, as shown in Equation (7). Each \( h_t \) contains a set of memory.

\[ y_t = \text{softmax}(h_t W_{hy} + b_y) \] (7)

In the equation, \( \tanh \) and \( \text{softmax} \) are activation functions, which increase the nonlinearity of the neural network model, so that the neural network can approximate any function.

2.3. Data preparation
Firstly, the water level data of Nanjing station from 2010 to 2019 were collected, and a total of 3,545 valid data were obtained. The water level variation curve is shown in Fig. 4.

According to Figure 4, the water level in Nanjing presents obvious periodicity and seasonality, and the change period is about one year. These 3,545 numbers are divided into two parts. One part contains 3,161 data, which is used to generate the training set. The other part is 384 pieces of data, about one full
cycle, which is used to generate the test set. Assuming that the \( n_{\text{in}} \) day data is used to predict the day after \( n_{\text{out}} \) day data, the method of training set generation is shown in Figure 5,

![Fig.4. Changes of water level in Nanjing station for ten years](image1)

and the test set is generated in the same way.

Therefore, the number of training sets generated is:

\[
\hat{n} = 3162 - n_{\text{in}} - n_{\text{out}}
\]

Therefore, 3162 data were reconstructed into a three-dimensional array with the structure of \([3162 - n_{\text{in}} - n_{\text{out}}, n_{\text{in}}, 1]\), where the first dimension represents the number of samples, the second dimension represents the number of recurrent core time expansion steps, and the third dimension represents the number of features sent into the network by each time step. Similarly, the test set was reconstructed into a three-dimensional array with the structure of \([385 - n_{\text{in}} - n_{\text{out}}, n_{\text{in}}, 1]\).

3 Results analysis

3.1. Comparison of forecast results

Set the recurrent core expansion step size \( n_{\text{in}} \) as 14, which means that the 14-day data is used for prediction and the output value \( n_{\text{out}} \) is changed to compare the prediction effect of the 14-day data on the future water level data. The last day of the forecast, namely the \( n_{\text{out}} \)th output data, was compared with the real water level data to calculate the mean square error (MSE), and the predicted results were shown in Figure 6. In the training process, Epoch=50, batch size =60, number of memory =80, and number of memory in the last step =100.

![Fig.5. Method of training set generation](image2)

(a) \( n_{\text{out}}=2 \)  MSE=0.064

(b) \( n_{\text{out}}=3 \)  MSE= 0.121
According to Figure 6, the prediction on the second, third and fifth day after 14 days was made, and the mean square errors on the test set were 0.064, 0.121 and 0.195, respectively. It is found that the longer the interval between the predicted days and the input days is, the worse the prediction effect is when the LSTM has a fixed recurrent core unfurl step size. According to the image analysis, although the difference between the predicted value and the real value is more and more obvious with the increase of the number of days, the overall trend is roughly the same, which indicates that the model has a strong stability, and the prediction of a longer time span is also of certain reference significance for the sailing plan.

3.2. Hyperparameter selection

In the process of model training, the selection of hyperparameters is very important. The hyperparameters involved in this study mainly include batch size of sample generation, times of iteration Epoch and number of memory. In this paper, hyperparameters were selected to predict the water level on the fifth day in the future based on the 14-day data. As the memory has little influence on the prediction effect within a certain range, and in general, the more the number of memory, the better the prediction effect on the test set. However, with the increase of the number of memory, the amount of computation becomes larger, the efficiency of computation becomes lower, and the problem of overfitting is prone to occur. Considering the above qualitative and quantitative analysis, the number of memory selected is 80, and the number of memory in the last step is 100.

Batch size represents the number of samples used in each calculation of gradient in the process of min-batch SGD. Epoch represents the number of iterations. After each iteration, it represents a forward propagation after gradient descent of all small batch samples. Since too small batch size leads to problems such as slow convergence speed and accuracy oscillation, etc., and too large batch size puts higher requirements on epoch, which leads to problems such as increased training time and poor generalization, etc., the value range of batch size is selected from 30 to 110. Change the value of the batch size and epoch to analyze the number of iterations required by different batch sizes to reach MSE=0.156[10]. The results are shown in Figure 7.

The results show that, with the increase of batch size, the requirement for Epoch becomes higher accordingly. This result provides the basis for determining the epoch range by different batch sizes. Search further for the optimal solution near MSE=0.156, and we will find the relationship between MSE and the number of iterations at a fixed batch size, as shown in Figure 8.
Fig. 7. The number of iterations required by different batch sizes to reach MSE=0.156

Fig. 8. MSE minimum value under different batch sizes

Figure 8 shows that when the batch size is 30, 50, 70 and 90, the minimum MSE values corresponding to them are 0.156, 0.154, 0.156 and 0.153 respectively, and the corresponding iteration times are 110, 210, 190 and 250 respectively. In addition, when the batch size is 110, the number of iterations is 270 and the MSE is 0.153. In order to reduce the operation time, the optimal solution was batch size=90 and epoch=250, that is, the optimal super-parameter selection was made in the fifth day of 14-day prediction, with MSE reaching 0.153. At this time, the prediction effect on the test set was shown in Figure 9. Compared with the prediction effect before adjusting parameters as shown in Fig. 6, the accuracy of the prediction effect was increased by 21.5%.
4. Conclusion

In this paper, the LSTM neural network algorithm is used to effectively predict the water level of Nanjing station of Yangtze River. According to the data of 14 days, the water level data of the second, third and fifth day are predicted. The model training effect is good, which has certain guiding significance for the short-term prediction of this station. Further, this paper carried out hyperparameter optimization for the data on the fifth day after the 14-day prediction, and obtained the batch size=90 and epochs=250 for the optimal hyperparameter combination. Under this hyperparameter combination, the mean square error on the test set was reduced to 0.153, which was 21.5% higher than the prediction accuracy before adjusting the parameters. It has a certain reference value to solve the problem of over-parameter selection in predicting water level of Nanjing station by LSTM.

References

[1] Liu Ting. Influence of Water Level Change on the Yangtze River Shipping Logistics Cost: Mechanism and Model Research [D]. Wuhan: Wuhan University of Technology, 2014.

[2] Hua Xinglin, Liu Jiangang. Study of Dynamic Relationship between Deep Horizontal Displacement and Water Level Near Underground Water Vertical Excavation of the Yangtze River Embankment Slope [J]. Exploration Science and Technology, 2016(03): 18-21.

[3] Liang Yousheng, Song Hongtao. Impact of Variation in Water Level of the Yangtze River After the Construction of the Three Gorge Dam on the Transmission of Schistosomiasis in Jiangsu, China [J]. Chinese Journal of Parasitic Diseases, 1999(04): 47-49.

[4] Aksoy H, et al. Stochastic modeling of Lake Van water level time series with jumps and multiple trends[J]. Hydrology and Earth System Sciences, 2013, 17(6): 2297-2303.

[5] Fischer P, U Öhl. Effects of water-level fluctuations on the littoral benthic fish community in lakes: a mesocosm experiment[J]. Behavioral Ecology, 2005, 16(4): 741-746.

[6] Wang Zhiyong et al. 1D-2D Coupled Hydrodynamic Simulation Model of River-Lake System[J]. Advances in Water Science, 2011, 22(04): 516-522.

[7] Assem H, Ghariba S, Makrai G, et al. Urban Water Flow and Water Level Prediction Based on Deep Learning[J]. Springer International Publishing: Cham, 2017, 317-329.

[8] Guo Yan, Lai Xijun. Research on Water Level Prediction of Dongting Lake Based on Recurrent Neural Network[J]. Resources and Environment in the Yangtze Basin, 2021, 30(03): 689-698.

[9] Hochreiter S, J Schmidhuber. Long Short-Term Memory[J]. Neural Computation, 1997, 9(8): 1735-1780.

[10] Bi Zhanjia. Research on Parallel Accelerate Training Algorih of Multilayer Neural Network Based on Multi-GPU [D]. Harbin: Harbin Institute of Technology, 2015.