**CNN-BiLSTM water level prediction method with attention mechanism**

Qingqing Nie, Dingsheng Wan* and Rui Wang

College of Computer and Information, Hohai University, Nanjing Jiangsu 211100, China

*Corresponding author’s e-mail: dshwan@hhu.edu.cn

**Abstract.** Hydrological time series data is stochastic and complex, and the importance of its historical features is different. A single model is difficult to overcome its own limitations when dealing with hydrological time series prediction problems, and the prediction accuracy of a single model can be further improved. According to the characteristics of hydrological time series data, a CNN-BiLSTM water level prediction method with attention mechanism is proposed. In this paper, CNN extracts the spatial characteristics of water level data and BiLSTM learns the time period characteristics by combining the past and future sequence information, attention mechanism is introduced to focus the salient features in the sequence. Taking the hourly water level data of Pinghe basin in China as experimental basis, experimental result shows that this method is more accuracy than Support Vector Machine (SVM), Temporal Convolutional Neural network (TCN), and Bidirectional Long Short-Term Memory network (BiLSTM) model.

1. Introduction

Hydrological data is a discrete record of time-varying hydrological phenomena [1]. Hydrological information such as water level, streamflow and rainfall is typical time series data, which contains a lot of useful information. Analyzing and mining hydrological data information and using it for hydrological forecasting [2] has great research significance and application value. Complex time series prediction is an important research direction of time series data mining. In recent years, various neural network models and support vector machine [3-8] are applied to high-dimensional and complex nonlinear time series prediction problems, among which the methods based on ConvLSTM model [7] and TCN model [8] are often applied to traffic flow prediction and weather prediction, etc. However, the single model is difficult to deal with the characteristics of time series data adequately and overcome its limitations effectively. Therefore, the accuracy of prediction is not enough, and the result is not reliable.

The bidirectional LSTM developed from LSTM model [9] can process time series data from both forward and reverse aspects [10] and provide past and future sequence information for each moment in the input. It has shown strong ability in many application fields [11-13], so it is introduced into the research of water level prediction. Reasonable combination of different models can integrate the advantages of each method [14,15] and improve the predicted accuracy, which has also attracted the attention of many researchers. In this paper, the method that combines CNN, BiLSTM and attention mechanism is proposed. Historical data of Pinghe basin in Zhangzhou region is tested, the model structure is designed and constantly adjusted, and finally an effective water level prediction model for
small and medium-sized watershed is established. The predicted performance in the experiment is better than that of Support Vector Machine and other single neural network models.

2. Related work

2.1. Convolutional Neural Network

Convolutional Neural Network (CNN) is not limited to the application in the fields of vision, but also can process time series data [16] and obtain the predictive features of data. The classical CNN includes convolution layer, activation layer and pooling layer [15]. Appropriate convolution operation is used to abstract the original input and express data features in a higher-level form.

The formula of convolution operation is:

$$ O_y = f(\sum_{r=0}^{R} \sum_{c=0}^{C} W_{r,c} P_{i,j} + W_b) $$

(1)

$P_{i,j}$ represents the input elements in row $i$ and column $j$, $W_{r,c}$ represents the weight in row $r$ and column $c$ (a total of $R$ rows and $C$ columns), $W_b$ represents the bias of the filter, $f(.)$ represents the activation function. The convolution operation shown in formula (1) is used to capture the data features of the input.

2.2. Bidirectional LSTM recurrent neural network

LSTM unit includes three gates [9]: forgetting gate, input gate and output gate. Forgetting gate is a key structure for LSTM to learn long-term dependency, input gate determines how much the current input remains in the current memory unit state and output gate determines the extent to which the state of memory unit is transferred to the output of hidden state. Each LSTM neuron contains three inputs: $X_t$ is the input of the unit, at the current moment $t$, $h_{t-1}$ is the output of hidden state at the previous moment $t-1$, $C_{t-1}$ is the state of memory unit at $t-1$. There are two outputs of LSTM neurons: $h_t$ is the output of hidden state at $t$, $C_t$ is the state of memory unit at $t$. Formulas are:

$$ f_t = \sigma(w_{xf}x_t + w_{fh}h_{t-1} + b_f) $$

(2)

$$ i_t = \sigma(w_{xi}x_t + w_{hi}h_{t-1} + b_i) $$

(3)

$$ z_t = \tanh(w_{xz}x_t + w_{hz}h_{t-1} + b_z) $$

(4)

$$ C_t = f_t \odot C_{t-1} + i_t \odot Z_t $$

(5)

$$ O_t = \sigma(w_{ox}x_t + w_{ho}h_{t-1} + b_o) $$

(6)

$$ h_t = O_t \odot \tanh(C_t) $$

(7)

$w$ represents the corresponding weight parameters, and $b$ represents the corresponding bias. $\sigma$ is the sigmoid function. “$\odot$” means multiplication by elements.

Bidirectional LSTM prediction is based on the whole time series, and takes into account the positive and negative direction of historical data in the time dimension. Bidirectional LSTM structure consists of two layers of one-way LSTM, in which hidden layer in the positive time direction includes the past sequence information and calculates the current sequence information. Hidden layer in the reverse time direction reads the future sequence information in the input and adds the reverse sequence information in the calculation. Then, the values determined by the two LSTM networks are fed forward to output layer.
3. Water level prediction based on CNN-BiLSTM

3.1. Attention mechanism

There are some differences in the importance of hydrological time series features. In water level prediction, the past and future near the prediction time often contain more effective information, which has a greater impact on the current water level trend prediction. Therefore, on the basis of CNN and BiLSTM, this paper introduces attention mechanism [17] to assign different weights to the elements in the feature sequence. Attention mechanism highlights more significant features, reduces the influence of weak correlated features, and helps the model make more accurate evaluation. Formula is:

\[
\text{Att} = \sum_{i=1}^{n} \text{soft max}(f(x_i)) \cdot x_i
\]

where \(x_i\) is the feature sequence inputs to attention mechanism layer, and the attention mechanism is the weighted sum of weight.

3.2. CNN-BiLSTM model with attention mechanism

3.2.1. Model framework. Based on CNN, BiLSTM and attention mechanism, this paper constructs a hydrological time series prediction model, hereinafter referred to as "CNN-BiLSTM-Att" model. The model framework is shown in Figure 1.

CNN-BiLSTM-Att model is mainly composed of data preprocessing, convolution unit, BiLSTM unit, attention mechanism unit and output unit, etc.

(1) Preprocessing. During data preprocessing, the missing data of water level is filled. The abnormal water level is replaced by the mean value of the approaching time, and the missing part of rainfall is supplemented by the data of adjacent rainfall stations. Due to the large difference in the value range of features such as water level and rainfall, the dataset must be normalized before inputting into the prediction model. The one-dimensional water level and rainfall data are processed into a high-dimensional input matrix by using the dimensional reconstruction method. Then the dataset is divided into training set and test set according to the proportion.

(2) Convolution unit. The input of this part is the measured and preprocessed data. The influence of each rainfall station in Pinghe basin on the predicted station water level is different in space and time. CNN considers the input features from a global perspective, captures the spatial correlation between...
stations through autonomous learning, and extracts important spatial features for the subsequent model prediction.

(3) BiLSTM unit. The middle part of the model is bidirectional LSTM unit. One-way LSTM can not capture the intrinsic relationship between the characteristics of the future moment and the present moment, and only takes the past information as the basis of the possible future situation in the prediction process, which is easy to cause the lag problem. Therefore, BiLSTM is used to read the time series in reverse, and the reverse order information is considered in the calculation. In this part, the output of CNN unit are used as input to capture the time period characteristics of hydrological data.

(4) Attention mechanism unit. The inconsistency of the importance degree of time series features in time dimension is difficult to be reflected by CNN model or BiLSTM model. In the training process, it is difficult for the model to pay enough attention to these important features, or even ignore them, resulting in the decline of prediction accuracy. In order to highlight the influence of significant features in hydrological time series data and improve the prediction accuracy, attention mechanism unit is added to the latter part of the composite model. It assigns varying degrees of attention to input features, so that the composite model can focus on the most effective information under limited input conditions.

(5) Output unit is the last hidden layer in the network structure, which outputs the final prediction results of the model through the full connection layer.

3.2.2. Algorithm description. Data in this paper are recorded in hours. \( X_{m}^{i} \) is the input of hydrological time series. The superscript \( i \) represents the dimension of the input variable, for example, \( i=0 \) represents the actual historical water level. the subscript \( m \) represents the recording order with an interval of 1 hour. The input sequence is expressed as \( \{x_{1}^{i}, x_{2}^{i}, x_{3}^{i}, \ldots, x_{m}^{i}\} \) and the input form is the matrix of \( (m \times I) \). The convolution operation maps the input to the corresponding output \( \{c_{1}^{i}, c_{2}^{i}, c_{3}^{i}, \ldots, c_{m}^{i}\} \) and the output serves as the input of BiLSTM. \( W_{bi} \) is the calculation result of BiLSTM unit, and \( W_{aw} \) is the weight of sequence factors processed by the attention mechanism unit. Multiplying the two and the result \( O(i) \) will be transferred to subsequent output unit.

\[
0(i) = W_{bi}(i) \circ W_{aw}(i)
\]  

\( \circ \) means multiplication by elements. \( i \) is the corresponding position of time series in the feature layer. According to above process, the predicted sequence \( P_{m} \) is obtained. Then, the parameters of the model are updated by back propagation combined with the real sequence \( Y_{m} \) and the training process is cycled until the algorithm converges.

4. Experiment and result analysis

4.1. Data description and processing

The experimental basis of this paper is the hydrological time series data of Pinghe basin in Zhangzhou, Fujian Province. There are four stations in Pinghe basin, namely Pinghe, Gaokeng, Nansheng and Xiazhai. Pinghe station is a hydrological station, which records the water and rainfall information, while Gaokeng station, Nansheng station and Xiazhai station are rainfall stations, which record the rainfall information. Pinghe hydrological station is located at the confluence of Pinghe basin and the real-time water level will be affected by the rainfall of the other three stations in the basin. Therefore, the rainfall data of these stations are also input into the model.

In this paper, the water level of Pinghe hydrological station and the rainfall of Pinghe, Gaokeng, Nansheng and Xiazhai during 2010 to 2020 are selected as the experimental data. The data records the water level or rainfall value of the station every hour, with a total of 60,133. The preprocessed data
were divided into training data and test data in a ratio of 7:3. Pinghe basin is a small area, and the experiment shows that the real-time water level and rainfall in the first 2 hours are best used to predict the water level in the next 1-6 hours of Pinghe station. Water level and rainfall in the first 2 hours are taken as the input features, and the water level in the next 1-6 hours are taken as the predicted outputs, formula is:

$$Y_{t+m} = f_m(Y_{t-1}, X_{t-1}^1, X_{t-1}^2, X_{t-1}^3, X_{t-1}^4, X_{t}^1, X_{t}^2, X_{t}^3, X_{t}^4)$$

Y represents the water level, $m = (1, 2, ..., 6)$, X represents the rainfall, the superscript represents the different stations and the subscript represents the different moments recorded.

4.2. Experimental results and analysis

In this paper, the water level of Pinghe station in the next 1-6 hours is predicted. The specific experimental steps are as follows:(1) defining model (2) adjusting CNN layer and BiLSTM layer during the experiment (3) considering whether to add the Dropout layer according to the loss (4) the number of output neurons is defined as 1 (5) selecting appropriate activation function and learning rate to optimize the model (6) running the model and saving the result.

Making the comparison of experimental results more intuitive and obvious, 200 samples of hourly water level from 2018/9/25 6:00-2018/10/3 13:00 were presented as the experimental results. Figure 2 shows the predicted results of 1-6 hours in Pinghe basin. It can be seen from the figure that the model is basically correct in predicting the trend of water level change in the next six hours. When the water level changes gently, the degree of fitting between the predicted curve and the real curve is higher; when the water level changes rapidly, the gap between the predicted value and the real value is also more obvious.

![Figure 2](image)

**Figure 2.** Prediction results for 1-6 hours in Pinghe basin.

In order to verify the effectiveness of this model, the predicted results of CNN-BiLSTM-Att model are compared with those of SVM, TCN, BiLSTM and CNN-BiLSTM model. The predicted results are shown in figure 3.
Through comparison, it can be found that the predicted results of TCN and BiLSTM are relatively close, and improves greatly compared with the SVM. CNN-BiLSTM model integrates global and local features to some extent, so it also has good performance in prediction. However, CNN-BiLSTM, TCN and BiLSTM generally deviate from the observed data in terms of the degree of fitting with the increase of the predicted time. CNN-BiLSTM-Att model proposed in this paper has the highest degree of fitting with the observed data among all the comparison methods. The model is basically correct in predicting the trend of the wave series, and has improved the prediction accuracy compared with other models.

Root Mean Square Error (RMSE), $R^2$ score, and Nash–Sutcliffe efficiency coefficient (NSE) are selected as the evaluation indexes of the model. Comparison of prediction results between CNN-BiLSTM-Att and other models on the whole test set is shown in table 1.

| Model          | Evaluation Index | 1h   | 2h   | 3h   | 4h   | 5h   | 6h   |
|----------------|------------------|------|------|------|------|------|------|
| CNN-BiLSTM-Att | RMSE             | 0.029| 0.034| 0.030| 0.031| 0.032| 0.034|
|                | $R^2$            | 0.983| 0.976| 0.980| 0.980| 0.979| 0.976|
|                | NSE              | 0.983| 0.976| 0.980| 0.980| 0.979| 0.976|
|                | RMSE             | 0.055| 0.056| 0.041| 0.058| 0.058| 0.060|
| CNN-BiLSTM     | $R^2$            | 0.938| 0.937| 0.966| 0.935| 0.933| 0.932|
|                | NSE              | 0.938| 0.937| 0.966| 0.935| 0.933| 0.932|
|                | RMSE             | 0.054| 0.055| 0.054| 0.057| 0.058| 0.060|
| TCN            | $R^2$            | 0.941| 0.937| 0.943| 0.936| 0.932| 0.931|
|                | NSE              | 0.941| 0.937| 0.943| 0.936| 0.932| 0.931|
|                | RMSE             | 0.054| 0.055| 0.053| 0.057| 0.057| 0.059|
| BiLSTM         | $R^2$            | 0.940| 0.938| 0.942| 0.935| 0.933| 0.932|
|                | NSE              | 0.940| 0.938| 0.942| 0.935| 0.933| 0.932|
|                | RMSE             | 0.166| 0.166| 0.166| 0.166| 0.166| 0.166|
| SVM            | $R^2$            | 0.438| 0.438| 0.438| 0.438| 0.438| 0.438|
|                | NSE              | 0.438| 0.438| 0.438| 0.438| 0.438| 0.438|

Through the analysis of table 1, it can be seen that CNN-BiLSTM-Att model has good predicting effect in the next six hours, and the numerical performance of prediction in $R^2$, NSE and other evaluation indexes is the best. The results show that the predicted value of CNN-BiLSTM-Att model is closer to the real value, which reflects the rationality and effectiveness of the model. And CNN-BiLSTM-Att model can excavate the change rule of actual water level in Pinghe basin.
5. Conclusion
Aiming at the problem that a single model is not accurate enough when dealing with complex hydrological prediction problem, this paper constructs the CNN-BiLSTM water level prediction method with attention mechanism for small and medium-sized rivers and makes application analysis on the hourly water level data of Pinghe basin. The experimental results show that the CNN-BiLSTM water level prediction method with attention mechanism has a better prediction effect, and the error between the predicted value and the actual water level is small. By comparing with SVM, TCN, BiLSTM and CNN-BiLSTM model, it is proved that the CNN-BiLSTM method with attention mechanism is suitable for water level prediction, and the method has high predicted accuracy and a good application prospect. Future work will take into account more factors affecting hydrological data and expand the improved method to more catchment sites.

References
[1] Ai, P., Ni, W.X. (2003) Review and prospect of hydrological data mining technology research in China. Computer Engineering and Applications, 40(28):13–17.
[2] Zhao, Q., Zhu, Y., Shu, K. et al (2020) Joint spatial and temporal modeling for hydrological prediction. IEEE Access 8,78492–78503.
[3] Koprinska, I., Wu, D., Wang, Z. (2018) Convolutional neural networks for energy time series forecasting. In:2018 International Joint Conference on Neural Networks (IJCNN), pp.1–8.
[4] Lin, T., Feng, J.K., Hao, Z.X.et al (2020) Research on cloud computing resource load prediction based on combination prediction model. Computer Engineering and Science 42(07):1168–1173.
[5] Shi, J., Mao, J.L., Jin, C.Q. (2019) Spatiotemporal dependent travel time prediction of urban roads. Journal of Software,30(03):770–783.
[6] Meng, E., Huang, S., Huang, Q. et al (2019) A robust method for non-stationary streamflow prediction based on improved EMD-SVM model. Journal of hydrolog.568:462–478.
[7] Wang, Q.W., Chen, Y.R., Liu, Y.C. (2020) Short-term passenger flow prediction of urban rail transit based on convolution short-time memory neural network. Control and Decision Making, 1–10.
[8] Kong, Z., Zhang, H.L., Yue, S.K. et al (2020) A multi-scale bilinear weather prediction model based on temporal convolutional network. Journal of Graphics,41(05):764–770.
[9] Hochreiter, S., Schmidhuber, J. (1997) Long short-term memory. Neural Computation,9(8):1735–1780.
[10] Zeng, H.J., Guo, J.S. (2019) Failure prediction of aeroengine by bidirectional LSTM Neural Network. Journal of Air Force University of Engineering (Natural Science edition),20(04):26–32.
[11] Li, Y., Dong, H.B. (2018) Text Sentiment Analysis Based on CNN and BiLSTM Feature Fusion. Computer Applications,38(11):3075–3080.
[12] Kang, S.Q., Zhou, Y., Wang, Y.J. et al (2021) RUL prediction method for rolling bearings based on improved SAЕ and bidirectional LSTM.ACTA AUTOMATICA SINICA,1–11.
[13] Zang, R.Q., Zuo, M.Y., Guo, X.X. (2020) Study on the prediction of disease in elderly patients based on Doc2Vec and BiLSTM. Computer Engineering and Science,42(12):2273–2279.
[14] Wang, W., Sun, Y.X., Qi, Q.J. et al (2019) A text emotion classification model based on BiGRU- Attention neural network. Computer Application Research,36(12):3558–3564.
[15] Li, M., Ning, D.J., Guo, J.C. (2019) CNN-LSTM model based on attention mechanism and its application. Computer Engineering and Applications,55(13):20–27.
[16] Lee, W., Kim, K., Park, J. et al (2018) Forecasting solar power using long-short term memory and convolutional neural networks. IEEE Access 6,73068–73080.
[17] Ran, X., Shan, Z., Fang, Y. et al (2019) An LSTM-based method with attention mechanism for travel time prediction. Sensors,19(4):861.