DEEP LEARNING COMBINED WITH ATTENTION MECHANISM FOR MULTI-TIME-STEP RUNOFF PREDICTION MODELING

Jie LI[1], Linli JIANG[2,*], Xing ZHANG[3], Rongchuang YU[4], Liqiang ZHANG[5], Jiansheng WU[6]

Accurate runoff forecasting is crucial for flood prevention, drought management, and overall water resources management. To enhance the understanding of complex relationships within runoff data and improve forecasting accuracy, we introduce a multi-time-step attention mechanism for the hidden layer output of Long Short-Term Memory (LSTM) networks. This addresses the limitation of traditional LSTM models, which cannot allocate weights based on the importance of different time steps. Our proposed model, LSTM-TIMESTEP-ATT, is designed to predict the daily runoff process for the next year. Using daily runoff and water level data from the Guangxi Liujiang Bridge Hydrological Station between 2001 and 2010, we compared our model with standard LSTM, multiple linear regression, and support vector machine (SVR) models with a radial basis kernel. The results indicate that the LSTM-TIMESTEP-ATT model exhibits superior explanatory power and predictive accuracy. Specifically, it achieved the highest number of days with a relative error within 20%, outperforming the other three models. These findings provide a valuable reference for runoff predictions in the Liujiang River Basin, Guangxi, supporting effective water resource management and flood forecasting efforts.

Keywords: long short-term memory network (LSTM), attention mechanism, multi-step time series prediction, runoff forecasting

1 Introduction

Runoff forecasting is an important issue in hydrology, which is very important for flood prevention and control, drought and other water resources

[1] Associate professor, School of Mathematics and Computer Science, Guangxi Science & Technology Normal University, Laibin, China, e-mail: lijie980522@163.com
[2] Professor, School of Mathematics and Computer Science, Guangxi Science & Technology Normal University, Laibin, China, e-mail: jll200@163.com
[3] School of Mathematics and Computer Science, Guangxi Science & Technology Normal University, Laibin, China
[4]School of Mathematics and Computer Science, Guangxi Science & Technology Normal University, Laibin, China
[5] School of Mathematics and Computer Science, Guangxi Science & Technology Normal University, Laibin, China
[6] School of Mathematics and Computer Science, Guangxi Science & Technology Normal University, Laibin, China

management [1]. The runoff process is influenced by a variety of factors, such as precipitation patterns, basin characteristics, topography, and human activities, and is highly nonlinear and unstable. Traditional runoff forecast models often struggle to accurately capture the complex relationships and nonlinear features in time series data, leading to unstable forecast results or large errors [2,3].

Currently, runoff forecasting models are mainly categorized into process-driven models and data-driven models [4]. The former possess certain tangible significance, but the practical application is sometimes limited by the lack of understanding of the hydrological process mechanism and the high data requirements of the model [5]. The latter does not need to consider the actual physical significance of the runoff process, however, they rely solely on examining the correlation between input variables and output data to achieve runoff prediction, research shows that the machine learning model and the deep learning model have better results for runoff prediction [6-8]. Among them, Recurrent Neural Networks (RNNs) [9] exhibit specific advantages in dealing with temporal sequence prediction problems, but the relationship between the data before and after the day-by-day runoff process tends to have a strong correlation, and there is a certain level of association with the runoff of the present day and influencing factors of runoff before the multi-day runoff, and the traditional RNN was found to be unable to store the data for a long period of time during the actual training process, and it was also found that the traditional RNN could not store the data for a long period of time. for long-term preservation, and it is also found that the traditional RNNs are prone to issues of gradient vanishing and explosion phenomena, which limits use of long-term memory in practical applications [10]. In recent years, the proposed Long Short- Term Memory (LSTM) neural network changes neural structure of traditional recurrent neural network and solves the problems of long-term memory loss and gradient instability of traditional RNN, thus, LSTM is adaptable to deal with runoff forecasting and has received attention in hydrological forecasting [11-13].

Due to the shortcomings of the LSTM model in dealing with long sequential data forgetfulness and the inability to assign weights according to the importance during the training process, an optimized LSTM incorporating attention mechanism was constructed [14]. Attention mechanisms can selectively weigh the inputs according to their different degrees of attention to effectively identify significant features within inputs. Therefore, the incorporation of the attention mechanism in runoff forecasting modeling can improve the model's attention to important time steps and further enhance characterization ability and prediction performance of this model.

Utilizing the daily water level and flow data measured at the Liujiang River Bridge hydrological station, we have constructed a predictive model using a multi-time step attention mechanism and LSTM deep learning neural network method, which is applied to the one-year daily runoff prediction of the Liujiang River in Guangxi. This model, by incorporating the multi-time step attention mechanism to highlight significant time steps, can better capture the long-term dependencies and important features within time series data, thereby enhancing the accuracy and reliability of runoff forecasting. It provides reference significance for the runoff prediction of the Liujiang River in Guangxi.

2 Deep Learning Model Principles and Methods

2.1 Basic principles of the model

Recurrent Neural Networks (RNNs) is a mathematical model born stemming from way of connection between neurons in human brain, which is very good at processing sequence data. Although theoretically RNN is able to deal with infinite long sequence information, in practice, due to problems such as gradient vanishing and gradient explosion, it usually can only capture the information of the last few time steps in the actual operation process, and this limitation makes the RNN's effectiveness in dealing with long sequences limited.

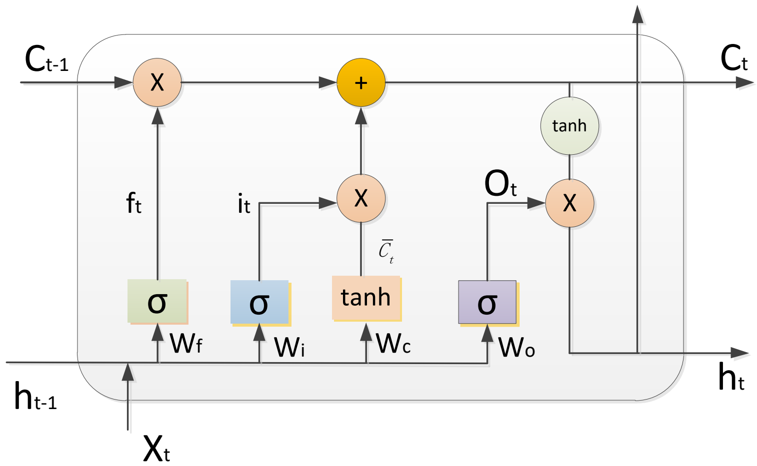

Fig. 1. Internal structure of LSTM cell

LSTM is a neural network modeling technique rooted in further development of RNNs. LSTM controls the information by changing the hidden layer structure, adding a conveyor belt"-like cellular unit state design, and allowing the information to selectively pass through a "gate" design. LSTM adeptly addresses prevalent issues of gradient vanishing and explosion encountered in RNNs, ensuring model stability and accuracy when handling extensive temporal sequences. Its proficiency in capturing long-term dependencies within sequential data renders it an optimal modeling instrument for time series analysis, particularly in applications like hydrological series prediction. Fig. 1 shows the internal structure of the LSTM cell unit.

The LSTM structure consists of three core gates, namely the forgetting gate, input gate and output gate, in which forgetting gate reads the preceding output and current input, Sigmoid activation function is applied to output a forgetting factor ranging from 0 to 1. In this context, 0 represents 'fully discarded' while 1 represents 'fully retained'. This forgetting factor is utilized to regulate the extent of forgetting the unit's state from the preceding moment [15, 16].

Forgetting gate formula is present in equation (1):

$$f_t = \sigma\left(W_i \bullet \left[h_{t-1}, X_t\right] + b_f\right) \tag{1}$$

Where W_i is weight matrix input, b_f is offset, and σ is sigmoid activation function.

Equation (2) represents input gate calculation:

$$i_t = \sigma\left(W_i \bullet \left[h_{t-1}, X_t\right] + b_i\right) \tag{2}$$

Where b_i controls current input's effect on cell state.

The candidate memory cell calculation formula is displayed in equation (3):

$$\overline{C_t} = \tanh\left(W_c \bullet \left[h_{t-1}, X_t\right] + b_c\right) \tag{3}$$

Activation function takes charge in formulating a list of new, tentative memory constructs tailored for this precise moment.

The formula for updating cell state at the current time is detailed in equation (4):

$$C_t = f_t \bullet C_{t-1} + i_t \bullet \overline{C_t} \tag{4}$$

New cell state at this juncture, multiply immediately beforehand cell state point by point by forgetting factor, if the point-by-point value is a value close to 0, it means that the latest information is excluded from cell state's content. Current new cell state value is obtained by obtaining the input gate output for point-by-point summation.

Output gate supervises output of this cell state value and is calculated as presented in equation (5):

$$\begin{aligned} O_t &= \sigma\left(W_o \bullet \left[h_{t-1}, X_t\right] + b_o\right) \\ h_t &= O_t \bullet \tanh\left(C_t\right) \end{aligned} \tag{5}$$

Current yielded data is calculated by an activation function (Sigmoid) and then transformed by a function (tanh), hidden layer's ultimate output is forwarded to next LSTM layer as its initial input.

As seen in Fig. 1 and the LSTM formulation, LSTM is designed to be more adept at dealing with the problem of exploding or vanishing gradients. Although LSTM has some advantages in areas such as time series prediction, LSTM may still face the challenge of vanishing or exploding gradients when dealing with very long sequence data, which may sometimes exist insufficient to capture short-term dependencies.

2.2 Time-step attention mechanisms

Attention mechanism serves as a useful tool in the domains of machine learning and natural language processing, which simulates the process of human attention and permits selective attention allocation to input segments by model [17]. In machine learning, if input series extends over a considerable length, it is difficult for model to directly capture the important information at each position. By introducing attention mechanism, this model is able to dynamically adjust the attention according to different parts of input sequence with the intention of better represent important content [18].

The time step attention mechanism is a method to enhance the attention to different parts of the time series in a sequence model, and its core principle is to acquire the importance or weight of every time step so that essential time steps for prediction accuracy receive greater emphasis from the model. The introduction of time-step attention mechanism allows for a more flexible and targeted approach to capturing dynamic changes and trends in the time series. There are three key steps in the model[19]: first, the model needs to calculate the relationship between each time step and other time steps in the sequence, and use the score to indicate the importance of the time step for the current prediction result; next, attention weights for all time steps are normalized through the application of Softmax function, ensuring their sum equals 1, indicating that different attention is assigned to different time steps; finally, the model allocates different attention to different time steps based on the calculated attention weights are weighted and summed to input sequence to get ultimate output, so as to find out more attention to those time steps that affect the prediction results to a greater extent. Fig. 2 illustrates the fundamental principle of the time-step attention mechanism.

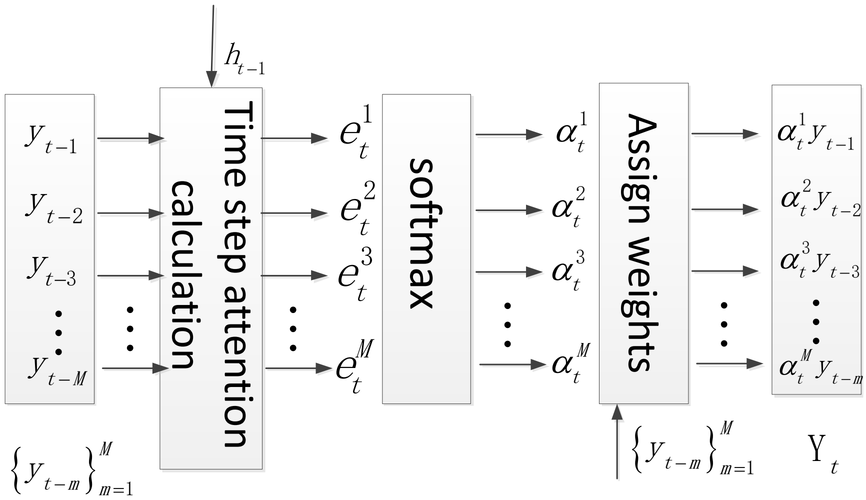

Fig. 2. Time step attention mechanism

The formula for calculating the time step attention weights is shown in equation (6).

$$e_t^m = \tanh\left(W_e h_{t-1} + U_e y_t^m + b_e\right) \tag{6}$$

The calculation formula of the attention weight coefficient is shown in equation (7):

$$\alpha_t^m = softmax\left(e_t^m\right) = \frac{\exp(e_t^m)}{\sum_{i=1}^{M}\exp\left(e_t^i\right)} \tag{7}$$

The output value calculation formula is shown in equation (8):

$$Y_t = \left(\alpha_t^1 y_{t-1}, \alpha_t^2 y_{t-2}, \cdots, \alpha_t^M y_{t-M}\right) \tag{8}$$

Where, e_t^m quantifies the extent to which current input is related to hidden layer's previous output, W_e, U_e, b_e is the weight and bias of the attention mechanism, α_t^m indicates to enhance or weaken the input time-step data, $\{y_{t-m}\}_{m=1}^{M}$ is the current moment of the multiple time-step inputs, Y_t is the value of the attention time-step data obtained by multiplying the weights with the corresponding time-step values, indicates the degree of influence of the adaptive optimization of the various input time-steps.

3 LSTM-TIMESTEP-ATT Forecasting Model

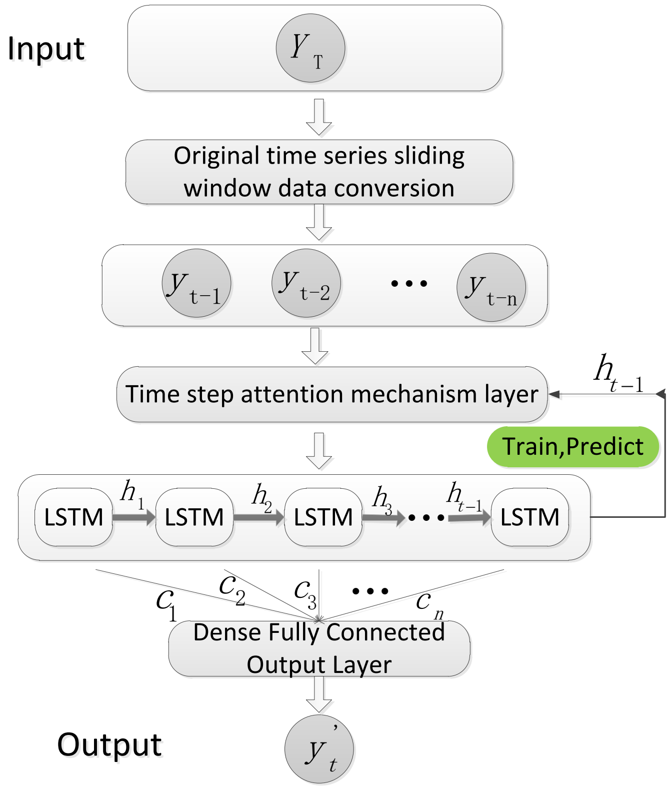

Fig. 3. LSTM-TIMESTEP-ATT network model structure

In this research, LSTM model combined with a multiple time step attention mechanism is employed to construct Guangxi Liujiang River runoff prediction model, referred to as the LSTM-TIMESTEP-ATT prediction model, which mainly contains an input layer, a time step attention layer, a hidden layer of LSTM network, and a fully connected output layer. The model first performs a sliding-window data transformation on the original time series, and takes the first four time-step sequence values as model inputs. Subsequently, the input multi-step time series values are fused with the LSTM network's preceding hidden state output, enabling the time-step attention layer to derive weight coefficients for each time step relevant to the current prediction, and then the optimized inputs combined with the improved multi-step attention are used to calculate LSTM hidden layer output. Conclusively, It is then passed through to fully connected layer for deriving ultimate prediction outcomes. LSTM-TIMESTEP-ATT can adaptively adjust the weights of different time steps in the input runoff sequence data through certain training, to ascertain relative importance of different temporal factors within input multi-time step dataset that affect runoff volume, the model is employed. The schematic representation of the LSTM-TIMESTEP-ATT network model is depicted in Fig. 3.

4 LSTM-TIMESTEP-ATT Model Experiment and Performance Analysis

4.1. Experimental data

The daily water level and flow data of Liujiang Bridge measured by hydrological stations were obtained from the observation files of Liuzhou Hydrological Management Information System. A total of ten years and 3650 data points from the period of January 1, 2001 to December 31, 2010 were accumulated. The data were divided into two parts; The cumulative number of data points is 3285, spanning from January, 2001 to December, 2009, were utilized for model training. The remaining 365 samples, covering period from January, 2010 to December, 2010, were employed for model testing to forecast the runoff for the entire year of 2010.

4.2. Prediction model performance metrics

In using the established LSTM-TIMESTEP-ATT Guangxi Liujiang River runoff forecasting model for forecasting test and forecasting accuracy computation and analysis process, the evaluation index calculation formulas used in this paper are presented below.

(1) Mean Square Error (MSE):

$$MSE = \frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2 \tag{9}$$

(2) Root Mean Square Error (RMSE).

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2} \tag{10}$$

(3) Mean Absolute Error (MAE):

$$MAE = \frac{1}{m}\sum_{i=1}^{m}|y_i - \hat{y}_i| \tag{11}$$

(4) Relative Error Percent (REP), Which quantifies the extent of divergence of each predicted value from the corresponding observation, is given by the formula:

$$REP = \frac{\hat{y}_i - y_i}{y_i} \times 100\% \tag{12}$$

(5) Mean Absolute Percent Error (MAPE), which assesses overall degree of divergence between forecasted values and actual values Eq(13):

$$MAPE = \frac{1}{m}\sum_{i=1}^{m}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100\% \tag{13}$$

Where m stands for total count of samples that are predicted, y_i signifies actual observations, and $\hat{y}_i$ indicates predicted value.

(6) coefficient of determination (R2), R2 is used to measure the ability of the model to explain the dependent variable, R2 is closer to 1, indicating that a superior model clarifies alterations in observation more effectively, yielding a better overall effect. The formula is:

$$R2 = 1 - \frac{\sum_{i=1}^{m}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{m}(y_i - \overline{y_i})^2} \tag{14}$$

Where, m denotes the count of predicted instances, y_i signifies factual observation from original dataset, while $\overline{y_i}$ represents mean value, $\hat{y}_i$ represents predicted value.

4.3 Forecasting model forecast test and error analysis

In this paper, LSTM is used for multiple time series forecasting. Considering that the runoff of the day may not be solely related to the runoff of the previous day, and that the single time-step prediction has certain defects, this paper conducts a multiple time-step prediction experiment. The model utilized runoff values from preceding four days as input to forecast the runoff for the subsequent day. Simultaneously, with the aim of analyzing the importance of different time-step runoff values to the forecast volume, further add the attention mechanism to calculate and analyze to be diverse crucial factors are assigned varying weighting coefficients for the forecasting experiment's computations.

The prediction model employs deep neural network architecture for each layer, utilizing MSE as the optimization objective. The experiments involve selecting a combination of hyperparameters, including a time step of 4, 4 hidden nodes, a batch size of 1, a learning rate set at 0.001, and 100 training epochs, to evaluate the model's performance.

This experiment examines the variation of MAE for different number of training rounds. When taking the number of training rounds Epoch tends to 100, the mean absolute error MAE basically tends to be stable, therefore, it is determined that the Epoch is 100.The error curves of the network model learning and training under above hyperparameter combinations and the change curves of forecasting model for predicted and actual runoff volume of the training samples from 2001 to 2009 are given in Figs. 4 and 5, respectively.

Fig. 5 displays the prediction model's performance on 2001-2009 samples, comparing predicted runoff volumes with actual changes. The trend is more similar, and the average absolute percentage error of its training runoff is 17.9%.

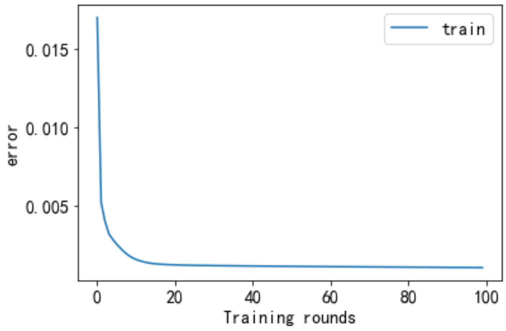

Fig. 4. Model training error curve

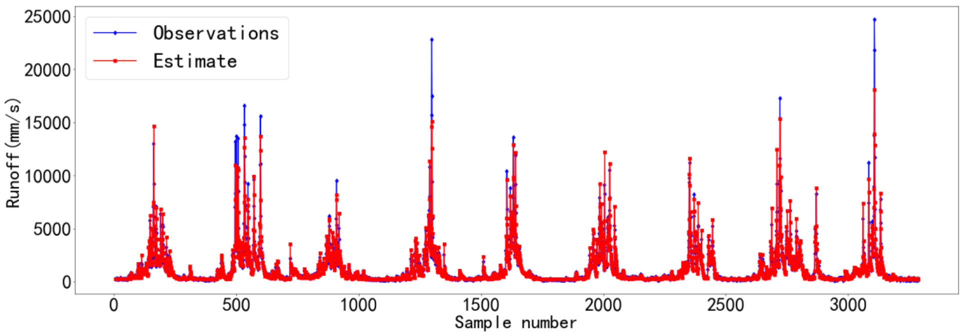

Fig. 5. Comparison of runoff prediction results with observations for a total of 10 years (training set) from 2001 to 2009

Fig. 6 gives the different average weight coefficients given by the model to the four time steps when incorporating the time step attention mechanism into LSTM.

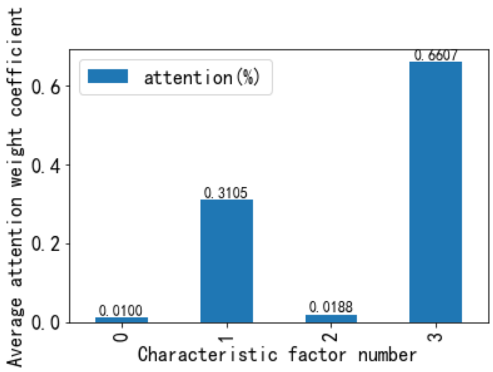

Fig. 6. Average Attention coefficient distribution histogram for the sequence of four time steps for forecasting 365 days in 2010

In Fig. 6, this model achieved a peak average attention weight coefficient at 0.6607 for the last time step with the highest average attention in the prediction of the 365 days of this experiment, followed by the second time step with an average attention weight coefficient of 0.3106, and the lowest average attention for the first time step with only 0.01. Attention coefficients vary daily, influenced by correlation between past outputs and current inputs across time steps.

4.4 Comparative analysis of LSTM-TIMESTEP-ATT forecast model with other models

For improved experimental result comparison, a comparative analysis of the forecasting performance of the established LSTM-TIMESTEP-ATT forecasting model is conducted, with respect to other nonlinear and linear forecasting models., such as with three models: multivariate linear regression forecasting model, LSTM forecasting model, and Support Vector Machines SVR (Radial Basis Kernel).

The experiment uses identical training and prediction data, while searching for optimal hyperparameters for the LSTM model under the same TensorFlow fixed seed values for training; Compare the predicted results of the LSTM-TIMESTEP-ATT forecasting model and the LSTM forecasting model without time step attention with the actual observed values.

Multiple linear regression forecasting model, for the four time steps of the prediction, establishes the equation of the quadratic linear regression forecasting:

$$Y = -0.034y_{step1} + 0.29y_{step2} - 0.62y_{step3} + 1.234y_{step4} + 153.91 \quad (15)$$

Similarly, support vector machine SVR (radial basis kernel) is used for modeling prediction and comparison with observed values.

A comparison was made between the results of the four models when the modeling sample was from 2001 to 2009 and the prediction sample was also from 2010. The comparison chart of multiple models is shown in Figure 7.

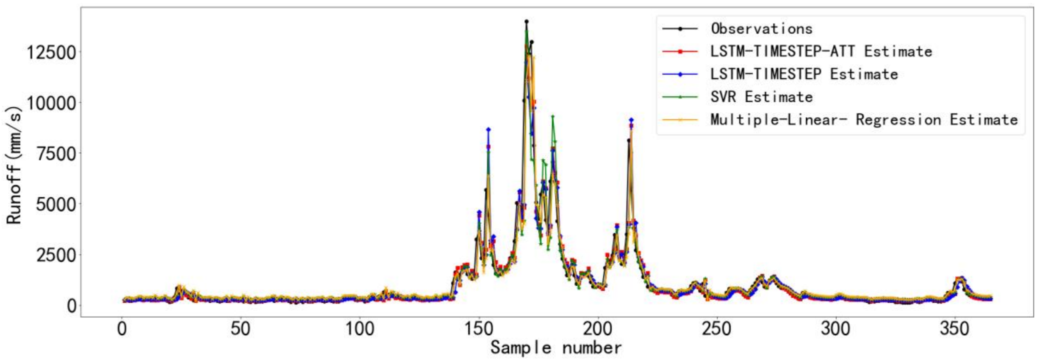

Fig. 7. Comparison of Prediction Results of Four Models

From Figure 7, it can be seen that in the 365 day prediction, the four models basically overlap in the case of flatter values, with little difference. However, in the peak area where there is a sudden change in water level, there is a slightly greater difference.

Table 1 compares prediction outcomes across four models. It is obviously discovered that prediction results of LSTM-TIMESTEP-ATT model have the smallest values of the three evaluation indexes of MAPE, MAE, and RMSE, and the value of the coefficient of determination is closer to 1. This suggests LSTM-TIMESTEP-ATT can focus on more important timing steps more efficiently, and the model's generalization performance is the best.

Table 1

Comparison of runoff prediction results between LSTM-TIMESTEP-ATT and various models.

Forecasting model	MAPE (%)	MAE (m3/s)	RMSE) (m3/s)	R2
Multiple-Linear- Regression	37.4	267.5	655.5	0.86
SVR (Radial basis kernel)	32.1	252.7	671.0	0.85
LSTM	23.9	245.8	650.1	0.86
LSTM-TIMESTEP-ATT	**18.8**	**223.3**	**596.5**	**0.88**

To improve analysis of prediction accuracy, focusing on divergence between forecasted and observations across models., this paper also compares the relative error percentage (REP). Table 2 gives the number of days for each of the three scenarios: |REP| greater than or equal to 20%, |REP| between 20% and 50%, and |REP| greater than 50% for the 365-day runoff forecasts of the four models.

Table 2

Comparison of different REP value ranges and days for 365 day runoff prediction results of four models.

Absolute range of REP values	Multiple-Linear-Regression	SVR (Radial basis kernel)	LSTM	LSTM-TIMESTEP-ATT
|REP|<=20%	130	154	196	**233**
|REP|>20% and |REP|<=50%	136	136	131	109
|REP|>50%	99	75	38	**23**
Total days	365	365	365	365

Table 2 shown that the number of days in which the relative error of the LSTM-TIMESTEP-ATT model is within 20% reaches 233 days in the 365-day runoff prediction, which is the highest among the four models. And the number of days with relative errors exceeding 50% is only 23 days, which is the smallest of the four models. Compared with the other three models, the LSTM-TIMESTEP-ATT model shows better prediction results.

5. Conclusion

In this paper, the LSTM model is optimized by using multiple time-step attention mechanism, and LSTM model with multiple time-step attention mechanism is established to be applied to day-by-day runoff forecasting of Liujiang River in Guangxi, and experimental comparisons are carried out with the LSTM without attention mechanism, multiple linear regression, and Support Vector Machine SVR (Radial Basis Kernel), etc., and the primary conclusions can be summarized as follows:

The LSTM-TIMESTEP-ATT model has effectively enhanced the prediction accuracy of runoff time series by incorporating a temporal attention mechanism. This model is capable of assigning different weight coefficients to various time steps, thereby more flexibly capturing the dynamic changes and trends within the time series. In the runoff prediction of the Liujiang River Basin in Guangxi, the model has demonstrated a high level of predictive accuracy, providing an important reference for water resource management.

Accurate runoff prediction holds significant practical importance for water resource management, including flood control and drought prevention. By forecasting the timing and peak of floods, it is possible to arrange defensive measures in advance, reducing the losses caused by floods. Moreover, Accurate runoff prediction also contributes to the comprehensive development and management of water resources, scientific management, and optimized scheduling, playing a pivotal role in guaranteeing the sustainable utilization of regional water resources.

Although LSTM-TIMESTEP-ATT model has shown high predictive accuracy, its generalization capability still requires further validation. Future research could consider integrating this model with other machine learning models to enhance the robustness and accuracy of predictions. Additionally, optimizing model and adjusting its parameters are also important avenues for enhancing predictive performance.

Acknowledgments

This work was Funded by National Natural Science Foundation of China (Grant 42065004）, Guangxi Innovation Driven Development Special Project (Major Science and Technology Special Project) (Guike AA21077018, Guike AA21077018-2), Guangxi Normal University of Science and Technology College Student Research Fund Project (GXKS2022DXSA009,GXKS2023DXSA004)

REFERENCES

[1]. *T. -T. -H. Nguyen, D. -Q. Vu, S. T. Mai and T. D. Dang*, "Streamflow Prediction in the Mekong River Basin Using Deep Neural Networks," in IEEE Access, vol. 11, pp. 97930-97943, 2023.

[2]. *Jiang LL, Li J, Wu JS et al*. Research on runoff forecasting model using PSO-SA hybrid optimization support vector regression. Computer Application Research, 2019, 36(09): 2599-2603(in Chinese).
[3]. *M. Moishin, R. C. Deo, R. Prasad, N. Raj and S. Abdulla*, "Designing Deep-Based Learning Flood Forecast Model With ConvLSTM Hybrid Algorithm," in IEEE Access, vol. 9, pp. 50982-50993, 2021.
[4]. *Longyang Q, Zeng R*. A Hierarchical Temporal Scale Framework for Data-Driven Reservoir Release Modeling. Water Resources Research, 2023(6): 59.
[5]. *H. Liu, W. Liu, J. Luo and J. Li*, "Comparative Study of Multi-Combination Models for Medium-and Long-Term Runoff Prediction in Weihe River," in IEEE Access, vol. 11, pp. 97099-97106, 2023.
[6]. *Deng H, Chen W, Huang G*. Deep insight into daily runoff forecasting based on a CNN-LSTM model. Natural Hazards, 2022, 113(3): 1675-1696.
[7]. *Yuan X, Chen C, Lei X, et al*. Monthly runoff forecasting based on LSTM–ALO model. Stochastic Environmental Research and Risk Assessment, 2018, 32(1).
[8]. *S. Yao, H. Chen, E. J. Thompson and R. Cifelli*, "An Improved Deep Learning Model for High-Impact Weather Nowcasting," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 15, pp. 7400-7413, 2022.
[9]. *Tao S. M., Liang Z. M., Chen Z. N. et al*. Application of long- and short-term memory networks in medium- and long-term runoff forecasting. Journal of Wuhan University (Engineering Edition), 2021, 54(1): 21-27(in Chinese).
[10]. *Ma Y, Zhong P, Xu B, et al*. Stochastic generation of runoff series for multiple reservoirs based on generative adversarial networks. Journal of Hydrology, 2022, 605: 127326.
[11]. *LIANG Z, TANG T, LI B, et al*. Long-term streamflow forecasting using SWAT through the integration of the random forests precipitation generator:case study of Danjiangkou Reservoir. Hydrology Re-search, 2017, 49(5): 1513-1527.
[12]. *Wei, Shouke, J. Zhao, and L. M. Yuan*. "Seq2seq model for human action recognition based on skeleton and two-layer bidirectional LSTM." Journal of ambient intelligence and smart environments 15. 4(2023): 315-331.
[13]. *Wang H, Qin H, Liu G, et al*. Hierarchical attention network for short-term runoff forecasting. Journal of Hydrology, 2024, 638.
[14]. *Zhou K, Wang W Y, Hu T, et al*. Comparison of time series forecasting based on statistical ARIMA model and LSTM with attention mechanism//Journal of physics: conference series. IOP Publishing, 2020, 1631(1): 012141.
[15]. *M.A.R. Suleman and S. Shridevi*, "Short-Term Weather Forecasting Using Spatial Feature Attention Based LSTM Model," in IEEE Access, vol. 10, pp. 82456-82468, 2022.
[16]. *Mustaqeem, M. Ishaq, and S. Kwon*. "Short-Term Energy Forecasting Framework Using an Ensemble Deep Learning Approach." IEEE Access PP.99(2021): 1-1.
[17]. *Yang S, Yi X, Dong R, et al.* Long-term autonomous time-keeping of navigation constellations based on sparse sampling LSTM algorithm. Satellite Navigation, 2024, 5(1).
[18]. *Majnooni S, Nikoo M R*, Banafsheh NematollahiMahmood FooladiNasrin AlamdariGhazi Al-RawasAmir H. Gandomi. Long-term precipitation prediction in different climate divisions of California using remotely sensed data and machine learning. Hydrological sciences journal, 2023, 68(13/16): 1984-2008.
[19]. *Long Y, Lv Q, Wen X, et al*. Bayesian logistic regression in providing categorical streamflow forecasts using precipitation output from climate models. Stochastic environmental research and risk assessment, 2023.