

Study on EEMD-LSTM Water Quality Prediction Model with Attentional Mechanism

Wenrui Bai¹, Yiqiang Bai^{1,*}, Qiang Li¹

¹School of Automation and Information Engineering, Sichuan University of Science and Engineering, Yibin 644000, China.

Abstract

In order to improve the accuracy of water quality prediction, the ensemble empirical mode decomposition (EEMD) was used to mine the hidden features of water quality data. Based on the decomposed data, four EEMD-LSTM-AT multi-step water quality prediction models were established by combining the attention mechanism (AT) and short and long duration memory network (LSTM). On the one hand, time step dimension or feature dimension of AT concentration; Second, AT is added before or after the LSTM network. Data from Laokou station in the Pearl River Basin were used to verify the effect of the model. The experimental results show that the LSTM model is better than the traditional MLP and SVR models in dealing with this problem. After the combination of EEMD, LSTM and AT, the effect of the model is further improved and the accuracy of water quality prediction is improved.

Keywords

EEMD Decomposition; LSTM Model; Attention Mechanism; Time Series.

1. Introduction

With the development of modern industrialization, industrial waste water, domestic waste water and so on inevitably lead to the deterioration of water quality, the problem of water pollution has also attracted extensive attention, the country is now emphasizing green development also shows the importance of environmental issues. The accurate prediction of water quality can give early warning to relevant departments in the early stage of pollution, reduce the cost of water quality control and improve the efficiency of water quality control. So there is a lot of research on water quality prediction in recent years, such as support regression machine the BP neural network and grey system, the regression model, fuzzy neural network are applied in the field of water quality prediction, these methods have their own advantages, such as when dealing with small sample data support regression machine performance is good, not only high precision, And there is the advantage of faster processing speed; For example, autoregressive model has the advantage of simple modeling, which can be very convenient to be applied in practical applications. But as nowadays society into the era of big data, automated collection and the amount of data is very large, the traditional method when applied to large sample data scene accuracy is lower than the neural network method, and the traditional neural network method is only mechanical reflect the shortcoming of the mapping relationship between input and output, hidden in the data sequence information cannot be effectively use, That's where the improved network comes in.

The short and Long-term memory network (LSTM) proposed by Hochreiter et al. [6] solves this problem. LSTM network realizes the long-term preservation of information in time series by spreading errors along time. Due to this characteristic, LSTM network has gradually attracted people's attention when dealing with timing issues since it was proposed. For example, LSTM is used to fit market trends in financial transactions [7], to predict pollution levels in

environmental protection [8], and to predict vehicle flow in urban traffic [9]. Wang et al. [10] used LSTM to analyze water quality problems. It is proved that the network is also useful in water quality prediction. At the same time, many people also consider the combination of feature extraction and LSTM model. For example, CONV-LSTM proposed in rainfall prediction [11] achieves success by fully extracting spatial and temporal features. In addition, the power forecast [12] also achieves the prediction target through the combination of CNN and LSTM network. After attention mechanism (AT) is popular in translation [13], it is often combined with LSTM in other places, such as semantic recognition [14] and load prediction [15].

So, through the above inspiration. The EEMD-LSTM-AT model is established to predict water quality through the combination of LSTM network and AT after feature extraction. Finally, the dissolved oxygen data of Laokou station in the Pearl River Basin were used as sample validation model to prove the superiority of EEMD-LSTM-AT model in water quality prediction.

2. Introductionr to model methods

2.1. EEMD decomposition

EEMD decomposition extracts the hidden information in the data by decomposing the data into modal functions. This feature extraction method has obvious effects on timing problems with nonlinear fluctuations. The extraction steps are as follows:

- 1) Add white noise with a amplitude coefficient of K to the original data.
- 2) Empirical mode decomposition is performed based on the data after adding white noise, and the process is as follows:
 - ① Based on the original sequence $x(t)$, the upper and lower envelope formed after the cubic spline interpolation function, where, the upper envelope sequence is denoted as $x_{max}(t)$, and the lower envelope sequence is denoted as $x_{min}(t)$.
 - ② The average envelope is obtained by taking the mean values of the upper envelope and the lower envelope, denoted as $m(t)$, and then subtracting the average envelope on the basis of $x(t)$, to obtain a new sequence $h(t) = x(t) - m(t)$.
 - ③ when the $h(t)$ following two conditions are met, is the component of the modal function, denoted as $I_1(t)$: First, the number of extreme points and zero crossing points in the data range should be restricted to be equal or the difference between them should not be greater than 1; And then the mean of each instant has to be zero. When either condition is not met, the appeal process is repeated until both conditions are met.
 - ④ when calculate the first modal function component, on the basis of the original sequence minus modal function component, residual sequence for remember, i.e., then as a new original sequence $r_1(t)$, the new sequence according to the appeals procedure again into a modal function component, finally only a drab trend sequence, namely the residual signal. The original sequence is the superposition of various modal function signals and residual signals.
- 3) Repeat Step 1 and Step 2, adding new normal distributed white noise each time.
- 4) Calculate the mean value of IMF obtained by each decomposition to obtain the final result.

2.2. Attentional mechanism

In deep learning, attention mechanism (AT) is a mechanism that simulates human brain's weight allocation of the concerned affairs. When people are watching television, for example, although can see the whole content of TV, we focus on television in different area weight is obviously different eyes will focus on some details or has obvious characteristics of the character, so as to ignore some important background panel, reduction of this kind is not important. The LSTM-AT model established by this mechanism can not only change the original

way of random initial weight coefficients into the way of probability distribution, but also improve the problem that part of information will be lost when the sequence is too long.

2.3. EEMD decomposition

LSTM network is an algorithm derived from recurrent neural network, which is characterized by not only spreading errors to the next layer but also spreading along time during training, so as to realize long-term preservation and memory of information in time series. A single LSTM neuron is shown in Figure 1:

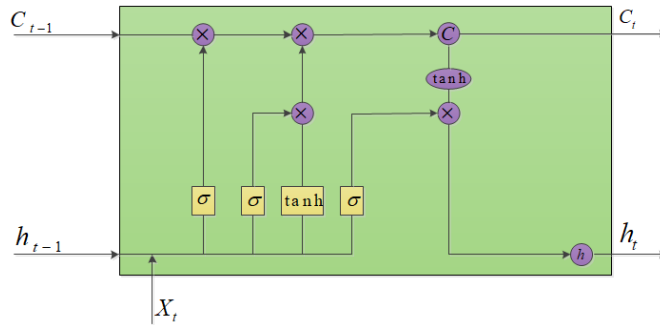


Figure 1. Neural unit structure diagram

In Figure 1, a single neuron receives three messages, The input of time t denoted by $x(t)$;

C_{t-1} is the unit state at $t-1$; Let's call it the hidden state of t minus one. When each neuron receives the information, it can control the update of the information through three corresponding "gates". The forgetting gate controls whether the information is discarded in the cell unit. When each neuron receives the information, it can control the update of the information through three corresponding "gates". The forgetting gate controls whether the information is discarded in the cell unit.

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

Where σ is the activation function sigmoid, It outputs a probability value to determine the proportion of components passing through. w_f and b_f are matrix coefficients of forgetting gate and bias term respectively.

$$i_t = \sigma(w_i[h_{t-1}, x_t]) + b_i \quad (2)$$

Where i_t is the input gate vector. The Tanh layer can create a new candidate value denoted as C'_t , and the current cell state C'_t can be output according to the information lost in the forgetting gate and the C_t updated by the input gate.

$$C'_t = \text{Tanh}(w_c[h_{t-1}, x_t] + b_c) \quad (3)$$

$$C_t = f_t C_{t-1} + i_t C'_t \quad (4)$$

Finally, the Tanh function is used to process the cell states and a value of 1 to -1 is obtained, which is multiplied by the output of the Sigmoid layer to obtain the final output. Through the appeal process, long-term memory can influence the output of the current state under the control of the output gate.

$$O_t = \sigma(w_o[h_{t-1}, x_t] + b_o) \quad (5)$$

$$h_t = O_t \times \text{Tanh}(C_t) \quad (6)$$

3. Model design and evaluation index

3.1. Model design process

EEMD decomposition extracts the hidden information in the data by decomposing the data into modal functions. This feature extraction method has obvious effects on timing problems with nonlinear fluctuations. The extraction steps are as follows:

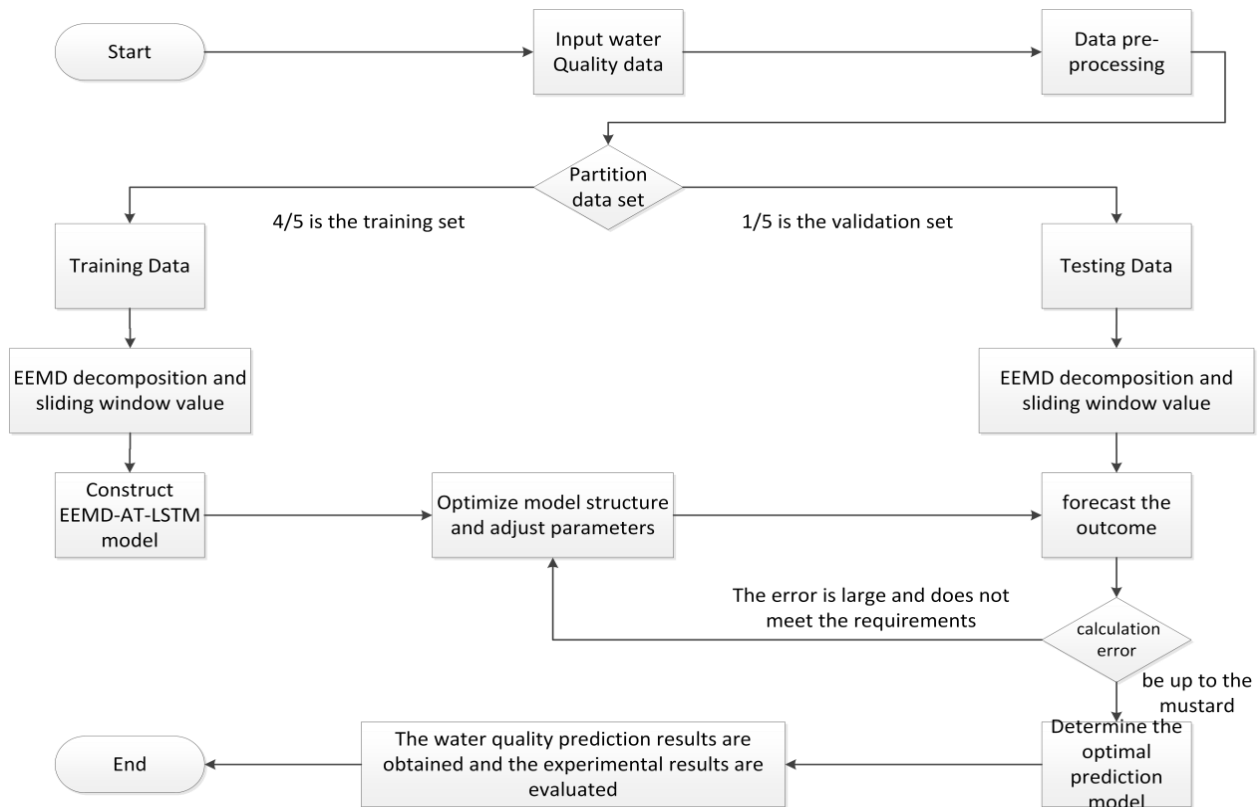


Figure 2: Flow chart of water quality prediction

3.2. Evaluation Indicators

In this paper, absolute error (MAE), mean square error (MSE) and root mean square error (RMSE) are adopted as evaluation indexes of the model. The smaller the value is, the smaller the error between the predicted results and the real results will be, and the final fitting degree will be higher. The specific formula is as follows:

$$E_{MAE} = \frac{1}{N} \sum_{i=1}^N |f(x_i) - y_i| \quad (7)$$

$$E_{MSE} = \frac{1}{N} \sum_{i=1}^N |f(x_i) - y_i|^2 \quad (8)$$

$$E_{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N |f(x_i) - y_i|^2} \quad (9)$$

Where, N is the total number of samples, $f(x_i)$ is the actual value of the sample, and y_i is the predicted value of the sample.

3.3. Accuracy rate of predicting water quality categories (ACC)

According to the national Surface water quality standard GB3838-2002, the water quality is divided into 5 categories, and some of the water quality standards are shown in Table 1:

Table 1. Water quality standard tabl

Project	I) Category	II) Category	III) Category	IV) Category	V) Category
AN $mg.L^{-1}$	≥ 7.5	≥ 6	≥ 5	≥ 3	≥ 2
P $mg.L^{-1}$	≤ 0.15	≤ 0.5	≤ 1.0	≤ 1.5	≤ 2.0
DO $mg.L^{-1}$	≤ 0.02	≤ 0.1	≤ 0.2	≤ 0.3	≤ 0.4

The accuracy rate of predicting water quality is as follows:

$$ACC = \frac{n_i}{n} \times 100\% \quad (10)$$

Where n_i is the number of samples whose predicted results are consistent with the real pollution situation, and n is the total number of predicted samples.

4. Experiment and analysis

4.1. Data Sources

Old data from the zhujiang river mouth stood open data, contains a day between 2006 and 2018 samples of PH, temperature, water quality indexes such as dissolved oxygen, including dissolved oxygen water quality data as a follow-up data sets, then four 5 of the data set as the training data set, the rest of the data set as a validation set, data details are shown in table 2:

Table 2. Water quality data detail table

Category	Training Data	Testing Set	Total
I	628	71	699
II	2427	375	2802
III	504	246	750
IV	7	42	49
Total	3566	734	4300

4.2. Experimental model design

Four EEMD-LSTM-AT multi-step water quality prediction models were established by considering the following two aspects: on the one hand, the time step dimension of AT concentration or the characteristic dimension; Second, AT is added before or after the LSTM network. Then, the most suitable model for current water quality prediction is selected from the four EEMD-LSTM-AT models according to the evaluation indicators, and the eEMD-LSTM-AT model with the best performance is selected by comparing the four models with the LSTM model and EEMD-LSTM model before improvement. Finally, the optimal EEMD-LSTM-AT model and LSTM model as well as the traditional SVR and MLP methods are compared. The experimental models involved are shown in Table 3:

Table 3. Experimental model table

Number	Model	describe
1	EEMD-LSTM-AT-DIM	First, EEMD decomposition is performed. After AT is added to LSTM network, AT focuses on features
2	EEMD-LSTM-AT_TIME	EEMD decomposition is carried out first, and AT is added to LSTM network, and AT pays attention to time step
3	EEMD-AT-LSTM-DIM	EEMD decomposition is carried out first. AT is added before LSTM network, and AT focuses on features
4	EEMD-AT-LSTM-TIME	EEMD decomposition is carried out first. AT is added before LSTM network, and AT focuses on features
5	LSTM	LSTM network
7	SVR	Support regression machine
8	MLP	Multilayer sensor

4.3. EEMD decomposition

First of all, it is necessary to preprocess the missing water quality data caused by instrument damage, table missing and other problems. In this paper, the mean value is adopted to fill the missing data, and the mean value of the time before and after the missing value is used to supplement the missing data. Then, EEMD is used to extract the characteristics of dissolved oxygen data. The data of water quality data before decomposition is shown in Figure 3, and the data after decomposition is shown in Figure 4.

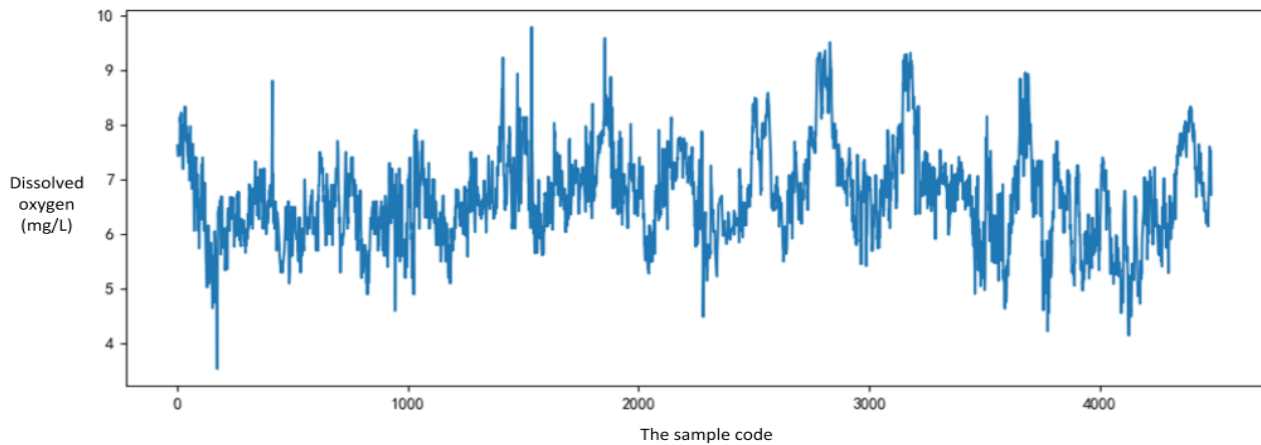


Figure 3. Original dissolved oxygen data

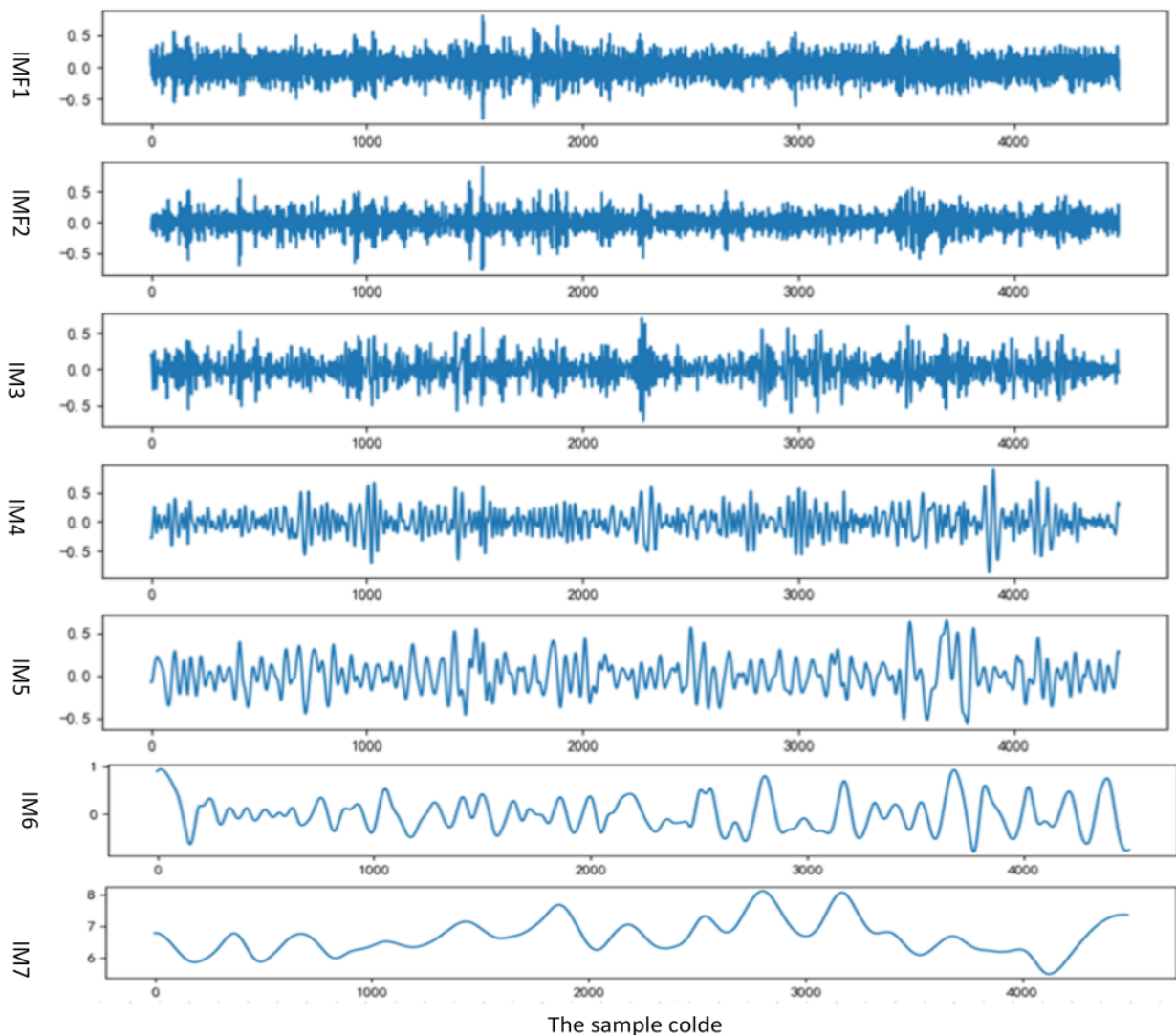


Figure 4. Dissolved oxygen data after decomposition

4.4. EEMD LSTM - AT model

The decomposed data of EEMD are converted into a data set that can be trained by EEMD-LSTM-AT model through sliding window through water quality data set. Taking sliding window size of 2 and predicted step number of 1 as an example.

Based on the decomposed data, a model was established to predict water quality after 14 days. The model was built for two aspects: on the one hand, the AT layer was added before the GRU network in terms of time dimension and characteristic dimension respectively, and the dense layer was linked to output the prediction results. On the one hand, AT layer was added to LSTM network for time dimension and characteristic dimension respectively, and the dense layer was linked to output prediction results. The loss function is MSE, and the optimization algorithm is Adam optimization. After the model training, the validation set data was input into the model to obtain the prediction results, and then the prediction results were visualized. Finally, the comparison results between the predicted values and the real values of the four EEMD-LSTM-AT models were shown in Figure 5:

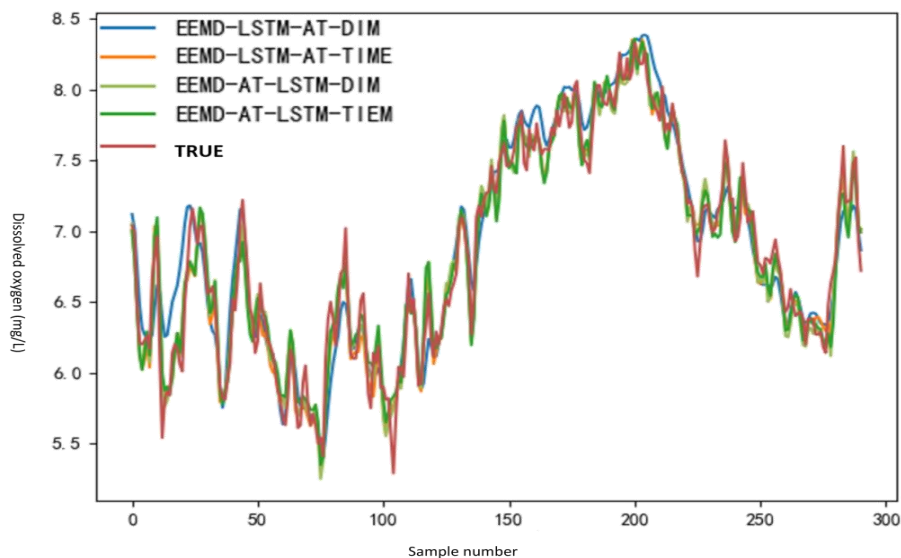


Figure 5. Comparison between actual and predicted values of dissolved oxygen

The trend of the predicted results of the four models in the figure is consistent with the real curve, but the fitting degree of the blue curve is significantly worse than the other three curves. Finally, MSE, MAE, ACC and other evaluation indicators are used to measure the effect of the model. The results are shown in Table 4.

Table 4. Model evaluation table

Model	MAE	MSE	RMSE
EEMD-LSTM-AT-DIM	0.2359	0.1037	0.3220
EEMD-LSTM-AT-TIME	0.1256	0.0251	0.1584
EEMD-AT-LSTM-DIM	0.1212	0.0233	0.1527
EEMD-AT-LSTM-TIME	0.1355	0.0292	0.1710

As can be seen from the experimental results, in the current water quality prediction, when AT is added after LSTM network and focuses on the characteristic dimension, the effect is the worst, MAE is close to 2 times of the other three models, MSE is close to 4 times of the other models, and the accuracy is also about 10% lower. However, when AT focuses on the time dimension, the errors and accuracy are almost the same whether it is added before or after LSTM network. However, when AT focuses on the feature dimension and is added before LSTM network, it achieves the best effect. Compared with the two models of attention plus time dimension, the error is slightly reduced and the accuracy is slightly improved. (For the convenience of comparing with other models, eEEMD-AT-LSTM-DIM with the best performance is denoted as EEMD-LSTM-AT model, and no specific distinction is made.)

4.5. Comprehensive comparison of models

In order to verify eEMD-LSTM-AT does have advantages in water quality prediction, comparison experiments were conducted between EEMD-LSTM-AT model and LSTM, MLP and SVR. The comparison results of predicted values and real values of the four models are shown in the figure.

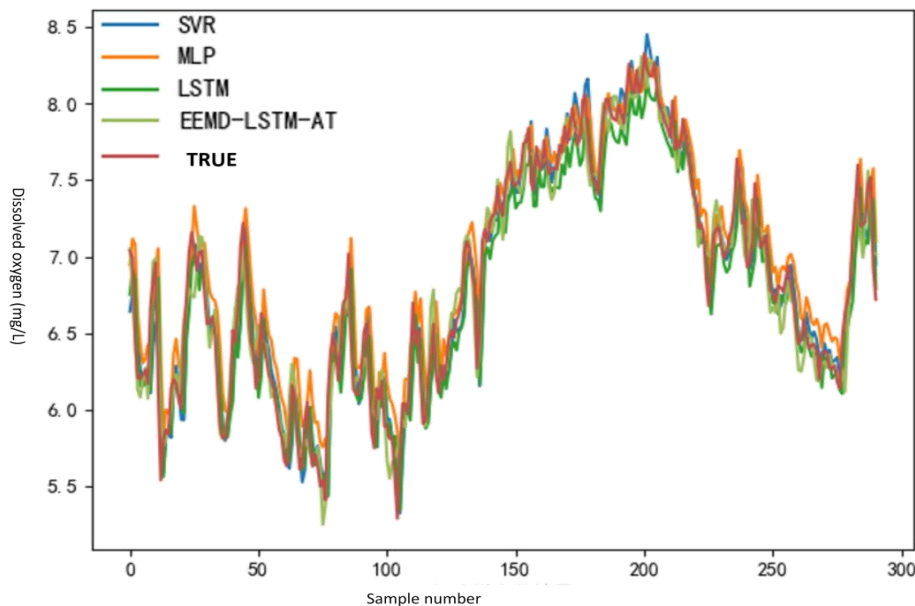


Figure 6. Comparison between actual and predicted values of dissolved oxygen

Finally, MSE, MAE, ACC and other evaluation indicators were used to measure the effect of the model. The results are shown in Table 5.

Table 5. Model evaluation table

Model	MAE	MSE	RMSE
EEMD-LSTM-AT	0.1212	0.0233	0.1527
SVR	0.2216	0.0857	0.2928
MLP	0.2323	0.0864	0.2940
LSTM	0.1871	0.0594	0.2437

that of MLP model. Compared with SVR model, the RMSE evaluation index obtained by LSTM model decreased by 17.1% compared with MLP model and 16.7% compared with SVR model, and the accuracy of category prediction improved by 8% compared with MLP model and 6% compared with SVR model. By combining EEMD decomposition and AT mechanism, the hidden information of time series is further mined, and then the prediction is made on the basis of decomposed sequence. The MAE index of eEMD-LSTM-AT with the best performance is reduced by 35.2% and THE MSE index by 60.7% compared with LSTM model. Compared with LSTM model, RMSE evaluation index is reduced by 37.3%, and the accuracy of category prediction is improved by 4%.

5. Conclusion

In order to improve the accuracy of water quality prediction, this paper adopts an EEMD-LSTM-AT prediction model combining EEMD decomposition, AT mechanism and LSTM network. Then through specific experimental analysis, this method is compared with traditional MLP and SVR methods, and the following conclusions are drawn: Through EEMD decomposition of water quality data, multi-feature water quality data components can be obtained, which is conducive

to fully mining data information and facilitating multi-dimensional research on water quality change rules. The prediction effect of input data after EEMD decomposition into LSEM-AT model is better than that of traditional MLP and SVR models.

(1) Through EEMD decomposition of water quality data, multi-feature water quality data components can be obtained, which is conducive to fully mining data information and facilitating multi-dimensional research on water quality change rules. Through the attention mechanism, AT-LSTM can effectively focus on important features and improve the accuracy of prediction.

(2) Compared with traditional algorithms, LSTM model has a stronger ability to deal with time series problems than traditional MLP and SVR models. Through the combination of EEMD decomposition and AT mechanism, LSTM model has a better ability to learn information and better fit the relationship between input and output. The MAE error, MSE error and RMSE error of dissolved oxygen in water quality are only 0.1212, 0.0233 and 0.157 respectively. The prediction accuracy of water quality category is 92%.

(3) The algorithm provides a more scientific and effective basis for water quality protection and treatment, which is of great significance to real life and practical application.

References

- [1] Da Yingyun, WANG Xiaodong, ZHAO Yonggang, JIANG Minlan, Ye Meiying. Water quality prediction model based on correlation vector machine regression [J]. *Acta scientiae sinica*, 2015, 35(11): 3730-3735. (in Chinese)
- [2] Gao Feng, Feng Min-min, Teng Su-fen. Water quality prediction based on PSO optimized BP neural network [J]. *Journal of safety and environment*, 2015, 15(04): 338-341.
- [3] Zhang Y, Gao Q Q. Research on comprehensive water quality prediction model based on grey model and fuzzy neural network [J]. *Chinese journal of environmental engineering*, 2015, 9(02): 537-545.
- [4] Shi Bo, Zou Zhihong. *Chinese journal of environmental engineering*, 2014, 8(10): 4550-4554. (in Chinese).
- [5] Zhu Y L. Application research of grey system in surface water quality evaluation and prediction [D]. Tianjin University, 2008.
- [6] Long Short-Term Memory [J]. Sepp Hochreiter, Jürgen Schmidhuber. *Neural Computation*. 1997(8)
- [7] TROIANO L, VILLA E M, LOIA V. Replicating a trading strategy by means of LSTM for financial industry applications [J]. *IEEE Transactions on Industrial Informatics*, 2018(99): 1-11
- [8] Zhou Y S. PM2.5 prediction based on LSTM neural network [D]. Hunan University, 2018.
- [9] Discussion MIAO Miao. Short-term Traffic Flow Prediction Based on LSTM and Grey Model Integration [D]. Nanjing University of Posts and Telecommunications, 2017.
- [10] WANG Y, ZHOU J, CHEN K, et al. Water quality prediction method based on LSTM neural network [C]// LI Tianrui, LUIS M Z, LI Yun. *Proceedings of 2017 12th International Conference on Intelligent Systems and Knowledge Engineering (ISKE)*. Washington: IEEE Computer Society, 2017: 1-5
- [11] Shi X, Chen Z, Wang H, et al. *Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting* [J]. MIT Press, 2015.
- [12] Lu Jixiang, ZHANG Qipei, Yang Zhihong, Tu Mengfu, Lu Jinjun, Peng Hui. Short-term Load Forecasting Method Based on CNN-LSTM Hybrid Neural Network Model [J]. *Automation of electric power systems*, 2019, 43(08): 131-137.
- [13] Mnih V, Heess N, Graves A. Recurrent models of visual attention [C]// *Advances in Neural Information Processing Systems*. 2014: 2204-2212
- [14] Wang Hong, Shi Jinchuan, Zhang Zhiwei. Semantic relation extraction of LSTM based on attention mechanism [J]. *Application research of computers*, 2018, 35(05): 1417-1420+1440.
- [15] Li Zhao-yu, AI Qian, ZHANG Yu-fan, XIAO Fei. An Approach to Ultra-short-term Load Prediction using LSTM Neural Network based on Attentional Mechanism [J]. *Power supply*, 2019, 4(01): 17-22.