Wind turbine fault prediction based on seq2seq model

Haixing Huang[a,b], Zhonghu Li[a,b,*], Jinming Wang[a,b], Jihong Zhang[a,b]

[a]School of Information Engineering,Inner Mongolia University of Science and Technology,Baotou 014010,China; [b]Key Laboratory of Solar Thermal and Wind Power Generation of Inner Mongolia Autonomous Region, Baotou 014010,China.

ABSTRACT

Aiming at the advantages of processing time series data, an excellent variant network seq2seq of recurrent neural network is constructed, and the basic unit of the network adopts LSTM, and a wind turbine fault prediction method based on SCADA data is proposed. The method first reduces the dimensionality of a certain sequence length through the encoder and decoder of SEQ2SEQ, and then predicts the magnitude of the active power through the fully connected layer, and then calculates the residual size between the predicted value of the active power and the actual value to analyze the operating state of the wind turbine, and finally verifies the method with obvious fault data. The results show that this method detects the abnormal occurrence of the wind turbine 6 days earlier than the alarm time of the SCADA system, which provides a technical guarantee to avoid the further deterioration of the wind turbine failure.

Keywords: Wind turbines; SCADA system; failure prediction; LSTM; seq2seq

1. INTRODUCTION

As an economical and environmentally friendly renewable and clean energy, wind energy is favored by countries around the world. By the end of 2021, the installed capacity of onshore wind power in China has exceeded 300 million kilowatts[1]. Annual turbine maintenance costs account for about 10% to 20% of turbine operating revenue[2]. Because the operating environment of wind turbines is relatively harsh, resulting in too high a frequency of failures, Therefore, it is important to evaluate the operating performance of wind turbines and predict potential failures.

At present, most of China's research on wind turbine fault prediction is based on data-driven machine learning methods. For example, Jin Xiaohang[3] and others use sparse self-encoders to encode and decode feature data for dimensionality reduction, and then predict power through deep neural networks. Wang Chao[4] et al. adopted the LSTM neural network fault method, and analyzed and processed the prediction residuals through a sliding window. Li Senjuan[5] et al. used the method based on SVM to classify and predict each fault and normal operation data. Liu Jiarui[6] et al. combined the automatic encoder (AE) with the convolutional neural network (CNN) to propose a wind turbine fault warning method based on deep convolutional self-coding (DCAE). The above research mainly uses relevant algorithms to predict some important characteristics of wind turbines, Determine the threshold for abnormal alarms by comparing the residuals, and finally uses the threshold to analyze the system fault, and the experimental results are ideal. However, there are problems such as dimensional differences caused by excessive data volume of the network model, which leads to some errors in the model, and there are certain defects in data drive.

*Email: lizhonghu@imust.edu.cn

The output active power of wind turbines is a direct reflection of the performance of wind turbines, and the use of operating data to model and analyze them has become a research hotspot for wind turbine performance evaluation[7]. For example, Huang Lingling[8] et al. took the prediction error of the long short-term memory neural network as the dynamic deterioration of the monitoring index, and then used the fuzzy comprehensive evaluation method to evaluate the operating state of the wind turbine. Wang Yuhong[9] et al. proposed an ultra-short-term power prediction method for multi-wind turbines based on the BiLSTM network based on TPA mechanism. In view of the above research on power prediction based on LSTM and its improved algorithm, retaining its advantages, this paper mainly uses the relevant theory of residual analysis to use the seq2seq (Sequence-to-sequence) neural network based on the LSTM model unit to perform power prediction analysis on the SCADA system data of wind turbines.2. Wind turbine SCADA system operation data processing

2. THE WIND TURBINE SCADA SYSTEM RUNS DATA PROCESSING

2.1 Data preprocessing

In this paper, the SCADA system operation data of a wind farm in Inner Mongolia is used, and the data mainly includes fault data and normal operation data, and the data characteristics are mainly running time, fan wind speed, active power and engine speed. Table 1 lists some of the important feature parameters used in this document

Table 1. Characteristic data used in wind turbine fault prediction

serial number	Active power	30s average wind velocity	dynamo rotate speed	dynamo winding temperature	dynamo Drive side Bearing temperature	dynamo Non-driver side Bearing temperature	Cabin temperature
1	1053.9	8.89	1839.2	81.4	41.5	53.5	13.5
2	0	2.99	0	59.9	51.3	51.9	20.3

There is a part of the data of 0 in the data, this kind of data is generally the data when the wind turbine is stopped and the manufacturer will set part of the data to 0 data, which often produces interference in modeling training, so it needs to be eliminated, and the principle of exclusion is mainly as follows: speed is 0, power is 0, The wind speed does not meet the data between the cut in and cut out wind speeds.

2.2 Correlation analysis of data

There is a strong correlation between SCADA data and data, such as state variables such as wind speed, power and generator rotor speed. This is a fixed parameter of learning for model prediction, which has little impact on the training of the model, the lower the correlation of the data in the small change often has a great impact on other data, we often need some relatively low data as an important data for model learning optimization, so the correlation of the data needs to be analyzed.

The Pearson correlation coefficient is a linear correlation coefficient that is mainly used for the analysis of relationships between data. For a set of variables, the correlation coefficient r is calculated as:

$$r = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}} \tag{1}$$

In this paper, the SCADA data used will be used for correlation experiments, and the resulting data correlation curve chart is shown in Figure 1.
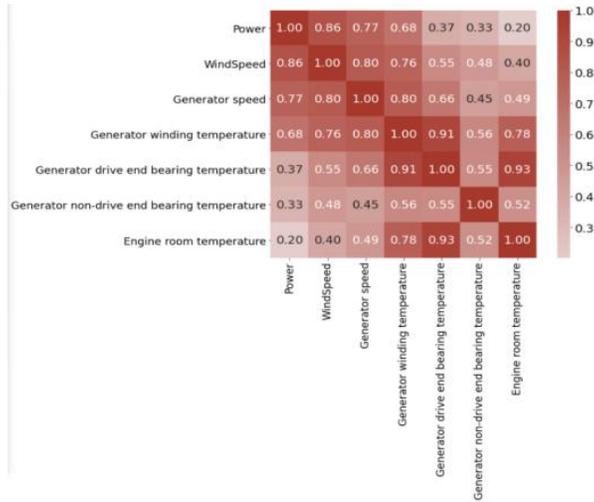
Figure 1. Histogram of SCADA data correlation

It can be seen from Figure 1 that the correlation between the front power, wind speed, speed and generator winding temperature is very strong, while the correlation between the bearing temperature of the generator drive end, the bearing temperature of the generator non-drive end, and the cabin temperature and the previous 4 sets of variables is relatively weak, and the subsequent training will have a greater impact on the model, so in the subsequent failure prediction time, we tend to use the latter 3 sets of faulty data to verify the experiment to ensure the accuracy of the experiment.

3. WIND TURBINE FAULT PREDICTION ALGORITHM BASED ON SEQ2SEQ

3.1 LSTM neural network

LSTM networkIt is a special recurrent neural network[10]. With the advantages of RNN neural network for sequence processing, it has many more gating switches than RNNs, which has a screening effect on information. It can solve a series of problems such as gradient damage and extremely large or minimal overfitting during training due to slow update of the weight relationship. The network has three control gate structures of output gate, input gate and forget gate, the work of the forget gate is to determine whether the memory unit at the previous moment enters the network for calculation, the input gate determines whether the candidate memory unit is used, and the output gate determines whether the hidden state is used. It is effectively controlled by the activation function. The internal structure of the LSTM network as a whole is shown in Figure 2.
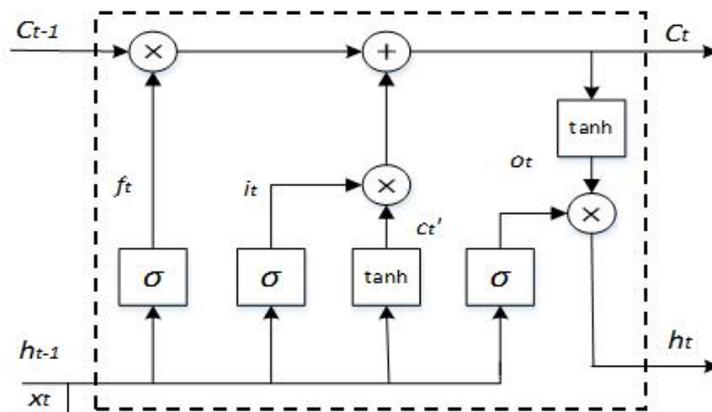

Figure 2. Internal structure diagram of LSTM network

The LSTM is calculated as follows:

$$f_t = \sigma\left(x_t w_{xf} + h w_{hf} + b_f\right) \tag{2}$$

$$i_t = \sigma(x_t w_{xi} + h_{t-1} w_{hi} + b_i) \tag{3}$$

$$c_t' = tanh(x_t w_{xc} + h_{t-1} w_{hc} + b_c) \tag{4}$$

$$o_t = \sigma(x_t w_{xo} + h w_{ho} + b_o) \tag{5}$$

$$c_t = f_t * c_{t-1} + i_t * c_t' \tag{6}$$

$$h_t = o_t * tanh(c_t) \tag{7}$$

Formula: f_t for the Forgotten Door; σ is the sigmoid activation function; x_t Enter for the vector at the current moment; w_{xf} is the weight between the input vector and the forget gate; h_{t-1} is the hidden layer state of the previous moment; w_{hf} is the weight between the hidden layer and the forgetting gate; b_f for the bias of the Forgotten Gate; i_t is the input door; w_{xi} is the weight between the input vector and the input gate; w_{hi} is the weight size between the hidden layer and the input gate; b_i is the bias of the input gate; c_t' is a candidate memory unit; w_{xc} is the weight size between the input vector and the candidate memory cells; w_{hc} is the weight size between the hidden layer and the candidate memory cells; b_c is the bias of the candidate memory cells; o_t is the output gate; w_{xo} is the weight between the input vector and the output gate; w_{ho} is the weight size between the hidden layer and the output gate; b_o is the bias of the output gate; c_t is the memory unit; c_{t-1} is the memory unit of the previous moment; h_t is the hidden state of the current moment.

3.2 seq2seq neural network

the sequence-to-sequence model was mainly used in natural language processing tasks such as machine translation and speech and text recognition[11], and later studies applied the model to time series forecasting tasks and achieved good prediction results. SEQ2SEQ can be divided into encoder and decoder as a whole, the basic unit uses the LSTM model[12], and the encoder and decoder expansion diagram is shown in Figure 3.
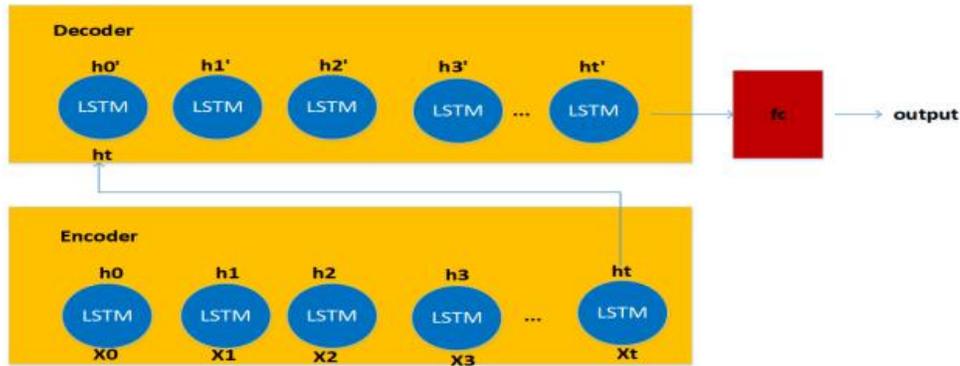

Figure 3.seq2seq encoder and decoder unfolded

In the figure, the encoder input sequence length is t, the decoder output sequence length is t', the encoder obtains the final hidden layer state ht as the input of the decoder, and finally the output of the decoder transforms the dimension through the fully connected layer to obtain the final output output.

The hidden layer state in the encoder at the current moment is calculated as follows:

$$h_t = \text{LSTM}(x_t, h_{t-1}) \tag{8}$$

The hidden layer state at the current moment in the decoder is calculated as follows:

$$h_t' = \text{LSTM}\ (f_{t-1}, h_{t-1}') \tag{9}$$

Formula: h_t is the hidden layer state under the encoder at the current moment; x_t is the input vector at the current moment; h_{t-1} is the hidden layer state of the previous moment in the encoder; LSTM（） is the internal calculation function of the

LSTM model h_t represents the implied layer state at the current moment in the decoder; f_{t-1} is the output vector at the previous moment; h'_{t-1} is the implied layer state at the last moment in the decoder.

4. EXPERIMENTAL RESEARCH AND RESULT ANALYSIS

4.1 Data analysis of wind turbine operation

The wind turbine had caused the SCADA system to malfunction due to generator bearing problems, and then shut down the wind turbine for maintenance. In order to verify the effectiveness of the proposed algorithm, the data of 3 months of the corresponding time period in the year before the failure is used as the training set to train the model, and the data of 13 months, including the fault time point, after the period of the failure, is used as the test data of the test set to verify the time of failure.Since there are abnormal data such as downtime data and fault data, data visualization is first used to preprocess the data. The variables used in this prediction experiment are plotted sequentially from A to G using the features in Table 1, as shown in Figure 4.
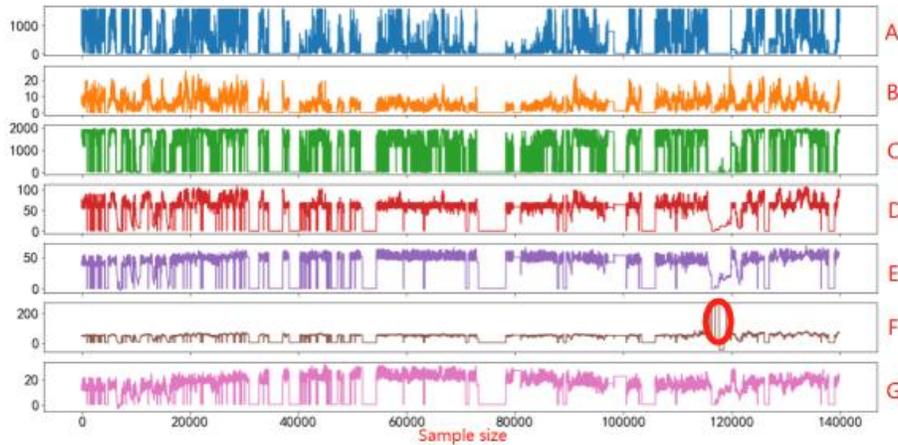

Figure 4. Timing diagram of wind turbine generator related characteristics

As can be seen from the figure, there is a lot of data with 0, which is recorded by the SCADA system when the wind turbine is in a shutdown state. In the F sub-diagram, it can be seen that the temperature of the non-drive bearing of the generator is in a stable state as a whole, but it fluctuates significantly in the red elliptical area, and the temperature rises significantly due to the failure of the non-drive bearing of the generator.4.2 Model parameter optimization

4.2 Model parameter optimization

The data of 3 months of the corresponding time period in the year before the above generator bearing failure was extracted, and 19464 sample data were obtained after processing for SEQ2SEQ network training. Set the batch size to 64, optimize the initial learning rate of 0.001, the sequence length seq_len is 12, and the number of training times is 50. By adjusting the number of neurons in the hidden_dim hidden layer for prediction, the number of neurons in the hidden_dim was 8, 16, 32, 64, 128 for comparative prediction experiments, and the MAE and RMSE were used for evaluation. The specific calculation method is as follows:

MAE calculation formula:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i| \tag{10}$$

RMSE calculation formula:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \tag{11}$$

where n is the number of data samples; y_i is the actual value; $\hat{y}_i$ is the predicted value.

The specific experimental results are shown in Table 2.

Table 2. Model prediction effect under different hidden_dim

hidden_dim	train_MAE	train_RMSE	test_MAE	test_RMSE
8	0.1052	0.1733	0.1103	0.2094
16	0.1003	0.1746	0.1129	0.2110
32	0.1008	0.1739	0.1124	0.2097
64	0.1052	0.1724	0.1086	0.2087
128	0.1007	0.1721	0.1108	0.2088

With the increase of neurons, the model effect will be better and better, but when the number of neurons is not as good as the number of neurons at 128, this can indicate that the number of neurons is not as much as possible, the curve is concave curve, and the number of neurons used here is 32.

4.3 Alarm threshold determination

The above 3 months of training data were used for failure prediction, and then the residual between the predicted and actual values of the active power was calculated, as shown in Figure 5(a). The residual data are normally fitted, and the fitted normal distribution plot is shown in Figure 5(b).

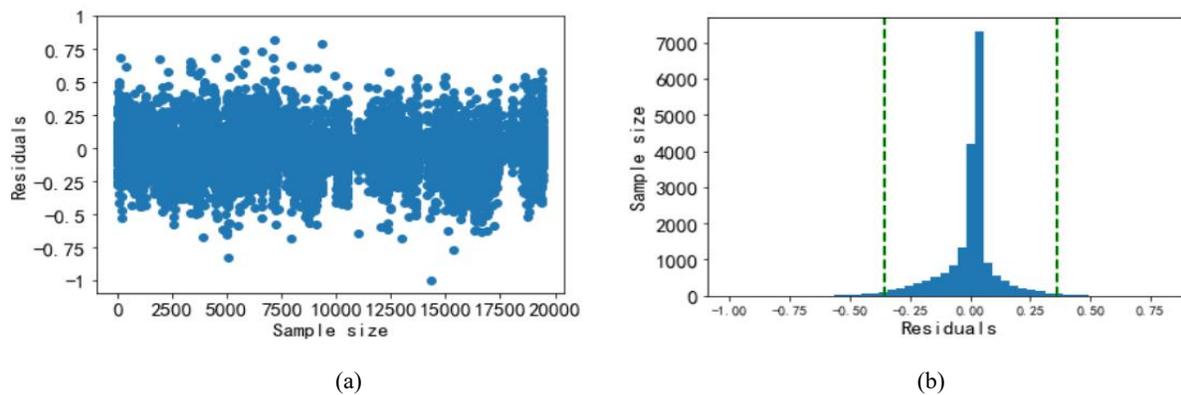(a)   (b)

Figure 5. (a)Predict the residual between active power and actual power, (b)Distribution of residuals in a healthy state

The distribution of residuals is mainly concentrated between ±0.5, and the main part still tends to 0. Using the nature of the normal distribution to set a suitable threshold a, so that the interval [-a,a] contains more than 99.7% of the data, the data between the two green dashed lines can be determined as normal data, and the data outside the dashed line is partial abnormal data, From this we can determine that the alarm threshold is ±0.3584.

4.4 Verification of failure prediction methods

Using the above trained model to predict and verify the test data in the next 13 months, using the number of hidden layer neurons 32 as the subsequent experimental parameters, the distribution curve of the predicted value of the active power output of the test set model and the actual value is shown in Figure 6(a), in order to verify the accuracy of the model prediction, the difference is calculated here, and the residual distribution plot is shown in Figure 6(b).
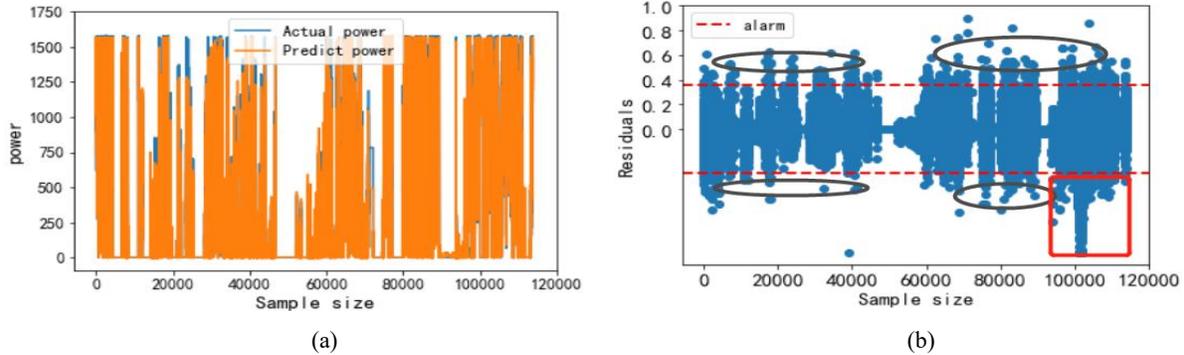
(a)   (b)

Figure 6.(a)Test set predicted power and actual power, (b)The test set predicts the residual power from the actual power

As can be seen from the figure, the predicted value output by the model basically matches the actual value. Only some of the data have obvious deviations. Through the determination of the alarm threshold, the red dotted line is determined as the alarm line, and only part of the data in front of more than 90,000 data points intermittently exceeds the alarm line, as shown in the black elliptical area in the figure, but the residual in the red rectangular area has exceeded the alarm threshold of the red dotted line a lot, at this time it has been possible to judge that the wind turbine is in an abnormal state until the subsequent complete exceeding of the alarm threshold. The point where the threshold is exceeded three times in a row for the first time is at 23476, and it should be discarded because it does not meet the reality, and the point where the threshold is exceeded for the second three consecutive times is at 88890 points, at which point it can be determined that the wind turbine has been abnormal, and then the residual exceeds the threshold over time, until the subsequent residuals continue to exceed the threshold. According to the correspondence between data points and time and combined with the alarm records provided by the SCADA system, the wind turbine failure prediction through the seq2seq neural network can know that the wind turbine is abnormal about 6 days in advance, which can replace or repair the relevant components early to avoid unnecessary losses.

5.   CONCLUSION

In this paper, the seq2seq neural network is used to carry out power prediction experiments on wind turbines, calculate the residual difference between the predicted value of active power and the actual value, and determine the alarm threshold. Using the fault data for verification, it is found that the time of exceeding the threshold three times in a row is 6 days earlier than the time of the alarm of the SCADA system, indicating that the use of seq2seq neural network will effectively avoid the deterioration of the fault, provide technical support for optimizing the maintenance strategy of the unit, and improve the reliability of the operation of the wind turbine.

ACKNOWLEDGMENTS

REFERENCES

[1] Lin,C., Multiple measures to promote the high-quality development of distributed wind power, Machine E-commerce News, 2022-08-22(A07) (2022).
[2] Jin, X, H., Sun, Y., Shan, J, H. et al., Review of fault diagnosis and prediction technology of wind turbines, Chinese Journal of Scientific Instrument, 38(05):1041-1053 (2017).

[3] Jin, X. H., Xu, Z. W., Sun, Y, et al., Online operation status monitoring of wind turbines based on SCADA data analysis and sparse self-coding neural network, Journal of Solar Energy, 42(06):321-328 (2021).

[4] Wang, C., Li, Z. D., Wind turbine gearbox bearing fault warning based on LSTM network, Electric Power Science and Engineering, 36(09):40-45 (2020).

[5] Li, S. J., Zhang, P., Yue, D. W., et al., Fault prediction of wind turbine based on support vector machine, Computer Simulation, 39(05):84-88+180 (2022).

[6] Liu, J. R., Yang, G. T., Yang, X. Y., Research on fault warning method of wind turbine based on deep convolutional autoencoder, Journal of Solar Energy, 43(11):215-223 (2022).

[7] Ma, T. S., Modeling and performance evaluation method of wind turbine based on improved LSTM, Shenyang:Shenyang University of Technology (2021).

[8] Huang, L. L., Li, S., Fu, Y., et al., Ultra-short-term offshore wind power prediction based on wind turbine status, Acta Solar Sinica, 43(08):391-398 (2022).

[9] Wang, Y. H., Shi, Y. X., Zhou, X., et al., Ultra-short-term power prediction of BiLSTM multiwind turbine based on time mode attention mechanism, High Voltage Engineering, 48(05):1884-1892 (2022).

[10] Chen, R., Research on English Machine Translation Based on LSTM Attention Embedding, Automation and Instrumentation, 264(10):140-143 (2021).

[11] Men, D., Chen, L., Text abstract generation method based on improved Seq2Seq-Attention model, Electronic Design Engineering, 30(23):6-10 (2022).

[12] Chen, Y. F., Zhang, D. H., Yu, H., Wang, Y. Q., Multi-feature short-term bus load prediction based on Seq2seq model, Transactions of Electric Power System and Automation, 35(01):1-6+35 (2023).