Proceedings of the 33rd European Safety and Reliability Conference (ESREL 2023)
Edited by Mário P. Brito, Terje Aven, Piero Baraldi, Marko Čepin and Enrico Zio

Published by Research Publishing, Singapore.
doi: 10.3850/978-981-18-8071-1_P706-cd

A data-driven failure prediction method for offshore wind turbines using Long Short-Term Memory model

Heping LI

School of Economics and Management, China University of Petroleum (Beijing), China. E-mail: liheping@cup.edu.cn

Li CHEN

School of Economics and Management, China University of Petroleum (Beijing), China. E-mail: 17709575502@163.com

Kejia YANG

School of Economics and Management, China University of Petroleum (Beijing), China. E-mail: 2022211421@student.cup.edu.cn

As operating in harsh marine environment, the offshore wind turbines often lead to high failure rate which affect the efficiency and reliability of wind power generation significantly. To improve the power generation capacity and decrease the breakdown time, this paper proposes an early failure prediction model for offshore wind turbines using Long Short-Term Memory (LSTM) model. With the SCADA data, a main feature is distinguished by the coefficient analysis to each failure mode. LSTM is used to capture the representation between the main feature and other relative features in the normal operation. When a failure occurs, the consistency of the representation should change dramatically. Therefore, a rule is set to distinguish the pattern between the normal operation and the failures using residual value indicator. With the SCADA data set of an offshore wind farm provided by EDP, it is proved that the algorithm can warn the hydraulic group, bearing and transformer failure about 31 hours, 5 hours and 15 hours in advance respectively.

Keywords: failure prediction, offshore wind turbines, Long Short-Term Memory (LSTM).

1. Introduction

Wind energy, one of the renewable and clean energy sources, has been regarded as a promising approach toward energy transmission in the last few decades. Offshore wind power generation has more advantages such as large energy capacity, stable wind speed and small environmental pollution compared to onshore wind power generation. However, the development of offshore wind turbines is limited by high failure rate, harsh operation and maintenance accessibility which leads to high operation and maintenance costs. To improve the power generation capacity and decrease the breakdown time, condition-based maintenance of wind turbines is being investigated. With the monitoring data, a preventive maintenance is executed before predicted failure.

In literature, a lot of research has been done on the failure prediction of offshore wind turbines. Nguyen et al. (2018) used the two-parameter Weibull distribution to model the failure time of offshore wind turbines[1]. It is efficient except that it can’t well define the heterogenous cause by environment. The Gamma process used widely to model the degradation of the component over time then the condition-based maintenance is established to prevent the failure[2]. Recently with the rise of sensor technique and deep learning methods, data driven methods are more developed to predict the failure or the remaining useful lifetime. Based on the condition monitoring information, such as vibration monitoring data, temperature monitoring data and oil detection data, artificial neural network (ANN) is used to establish a prediction model of failure time distribution then the optimal maintenance decision is proposed[3]. A framework based on deep neural network (DNN) is established by modelling the lubricating oil pressure of gearbox for failure monitoring[4]. The results show that the DNN model has more accurate results than K-nearest neighbor, support vector product and artificial neural network. The recurrent neural network (RNN) is better to deal with the time series, it is applied to calculate the bearing failure threshold which improved the accuracy[5]. The long short-term memory network (LSTM) is a specific RNN algorithm to deal with the problem of gradient disappearance or explosion. Sun et al. (2022) used the LSTM network to evaluate the four key components of the offshore wind turbine, predict potential risks and perform preventive maintenance on components thereby reducing maintenance costs and downtime[6].

In this paper, LSTM is used to establish a failure prediction model for offshore wind turbines in order to detect the anomaly and avoid downtime losses. The operation of component of wind turbine is supposed to be divided into two stages. The first period is the normal operation stage and the other is degraded operation stage. When the component begins to degraded, the data pattern is changed and the failure comes up. As LSTM is a supervised model, the main challenge is how to build the label data. In this paper, we proposed two-stage method to figure out the problem. Firstly, we analyze the main feature of failure mode and sign it as the reference value. Secondly, according to the correlation analysis with the main feature, a set of characteristics is selected. LSTM model is used to predict the main feature by the selected set of relevant characteristics under normal operation. The residual value of the actual value and the predicted value of main feature of each component under the healthy state is used to characterize the degree of deviation of the component from the healthy state. The set sliding window failure threshold is used to determine the state of the component and make the failure warning in advance.

The rest of the paper is organized as follows: Section 2 introduces the failure prediction model of offshore wind turbine. In Section 3, a numerical example is provided, and Section 4 deals with the conclusions and future directions.

2. Failure prediction model of offshore wind turbine

This section uses the SCADA system monitoring signals from a certain offshore wind farm in the Gulf of Guinea in Africa provided by EDP company as the dataset for the model from 2016 to 2017. It calculates the failure threshold and early warning time for bearings, transformers, and hydraulic components, respectively. In this paper, we only do the case of failure which is alarmed by temperature, for the other failures, the main feature should be reselected instead of temperature.

2.1. *Wind turbine failure prediction process*

During wind turbine health monitoring, the SCADA system detects a stable relationship between certain signals monitored by sensors. By using an appropriate model, the temporal changes of a certain monitoring signal can be fitted. As the wind turbine approaches a failure state, the values fitted by the model based on the healthy state will have significant deviations from the actual values monitored by the SCADA system. As the wind turbine degrades further, these deviations will increase. Once the deviation exceeds a set threshold, the system will identify an impending failure and issue a failure warning, enabling maintenance personnel to respond in a timely manner.

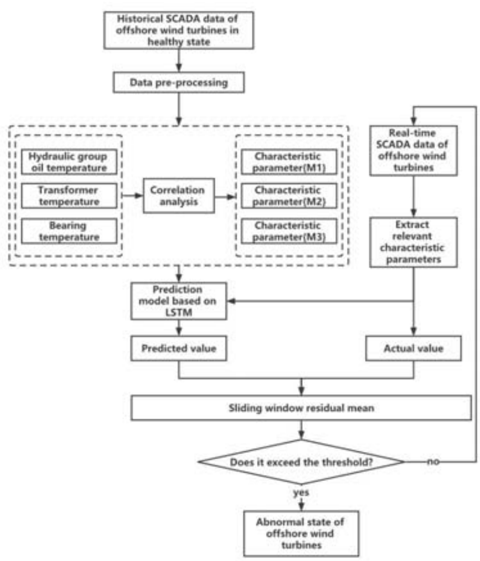

Fig. 1. Flow chart of wind motor failure warning

We have developed a wind turbine failure prediction model, which is illustrated in Fig. 1, with the specific steps of the model as follows:

Step 1: The data will be preprocessed, including removing invalid values, filling in missing values, and normalizing the data.
Step 2: Determine the state parameters and relevant features that characterize the components of offshore wind turbines. In this study, we chose to predict the faults by studying the monitoring parameters of transformer, hydraulic group, and bearing temperature signals. The correlation coefficients between transformer temperature, hydraulic group oil temperature, bearing temperature data, and other feature data were calculated. Based on the magnitude of the correlation coefficients, the features that exhibit high correlation with transformer temperature, hydraulic group oil temperature, and bearing temperature were identified.
Step 3: Build temperature prediction models based on LSTM for the three components: hydraulic group, transformer, and bearing. The inputs of the models are the features with high correlation to their respective temperatures identified in Step 2, and the outputs are the temperatures of each component.
Step 4: Based on the established LSTM-based temperature prediction models for each component, obtain real-time temperature prediction values for each component, and calculate the residuals between the temperature prediction values obtained using the prediction model and the actual values.
Step 5: The failure thresholds for the components are determined based on the residual values of the component's temperature during normal operation. The thresholds for distinguishing between the healthy and failure states are obtained by calculating the average residual value within a sliding window. This approach effectively reduces the fluctuation range of the residuals and minimizes the interference of false alarm points.

2.2. Correlation analysis

The original SCADA dataset contains multiple features, but not all of them are closely related to wind turbine failures. Therefore, feature dimensionality reduction is necessary to remove redundant or low correlation features.
Based on wind power experts' experience, temperature information of wind turbine components can be used to evaluate their operating status. Therefore, this paper mainly selects state features from the SCADA system that have a strong correlation with the temperature of each component. Therefore, Spearman correlation coefficient is used to analyze the correlation between each feature, and indicators with $P_{sr} > 0.6$ are selected according to the correlation strength standard. Taking transformer temperature as an example, features with a higher correlation are selected. For instance, the average temperature of the first phase of the stator winding in the generator, which has a Spearman correlation coefficient of 0.7995 with the transformer temperature, is chosen to be included in the construction of the transformer temperature prediction model.

2.3. *Health state model based on LSTM*

Temperature is used as an indicator to characterize the health status of various components of wind turbines. Health status models for three monitoring variables, namely hydraulic group oil temperature, bearing temperature, and transformer temperature, are established separately. The specific process is shown in Fig. 2.

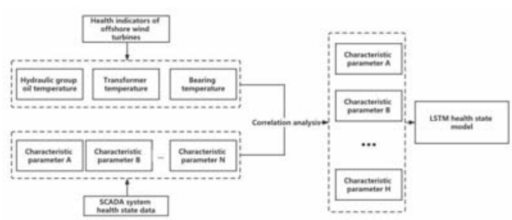

Fig. 2. Flow chart of health status modeling

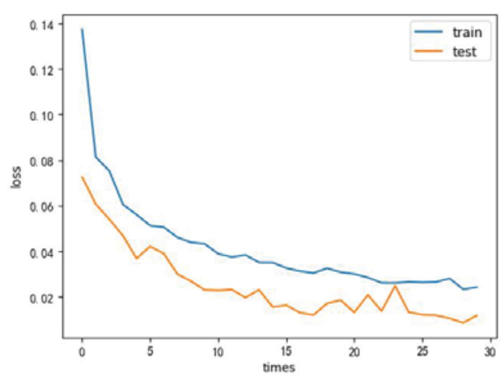

Fig. 3. Oil temperature loss function value curve of hydraulic group

We establish an LSTM prediction model with the ReLU activation function, whose expression is:

$$ReLU(x) = max\{0, x\} \tag{1}$$

Taking the LSTM health status model of the hydraulic group oil temperature as an example, the loss values of its training set and test set are shown in Fig. 3.

2.4. *Failure threshold determination*

The SCADA dataset of a healthy wind turbine is input into the established temperature prediction models for each component, and the predicted temperature values for each component are obtained. The difference between the actual and predicted temperature values for each component represents the degree to which the current state deviates from the normal state. Therefore, using temperature residuals can capture early features of component failures in a timely manner.

By calculating the residuals between the actual and predicted values, the residual distributions of normal and failure states are compared. This paper determines the failure warning threshold by calculating the average residual value within a sliding window. After inputting the actual operational SCADA data of the wind turbine into the model to generate temperature prediction residuals, if the residual exceeds the set threshold, it indicates that the wind turbine component has entered the early failure stage, and a warning can be issued for the component failure in advance.

3. Example analysis

The data is sourced from an open dataset provided by EDP (Energias de Portugal). The offshore wind farm is near the Gulf of Guinea in Africa, consisting of 16 wind turbines with a rated power of 2000 kW and cut-in wind speed of 4 m/s and cut-out wind speed of 25 m/s. The dataset contains a total of 410,000 SCADA signals of four wind turbines, namely T01, T06, T07, T11, including 21 failure records. The monitored components include the gearbox, generator, generator bearing, transformer, and hydraulic group, with a monitoring frequency of six readings per hour. The data mainly comes from the 82 sensors installed on the offshore wind turbines, including environmental parameters such as wind speed, and wind direction, and turbine condition parameters such as gearbox oil temperature, transformer temperature, and hydraulic group oil temperature.

3.1. *Data pre-processing*

Data pre-processing involves steps such as deleting invalid data, filling missing values, and normalizing data.

3.1.1. *Delete invalid data*

Before removing invalid data, a scatter plot of wind speed and power was plotted using matplotlib, as shown in Fig. 4. It can be observed that there are data points with wind speed greater than 0 and power equal to 0, as well as data points with wind speed less than the cut-in wind speed or greater than the cut-out wind speed. It is known that the wind turbines only start to generate electricity when the wind speed is greater than the cut-in wind speed, and the wind turbines will be disconnected from the grid and stop running when the wind speed is greater than the cut-out wind speed. Therefore, the data points related to these conditions were removed.

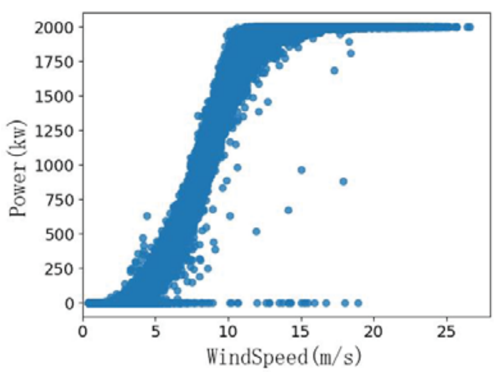

Fig. 4. Wind speed-power scatter diagram (Before deleting invalid data)

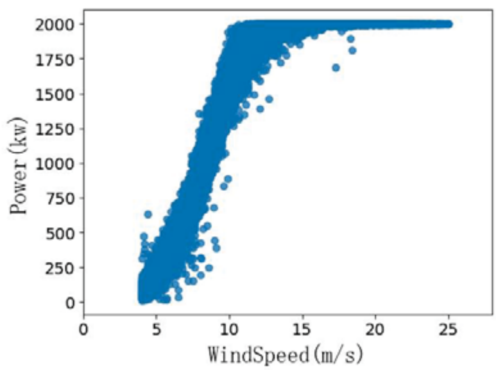

Fig. 5. Wind speed-power scatter diagram (After deleting invalid data)

After processing the data points as described above, the wind speed-power diagram is shown in Fig. 5.

3.1.2. *Filling the missing value*

By visualizing the wind speed characteristics around the T06 wind turbine (as shown in Fig. 6), it is observed that there are missing data during the period of June to July 2016.

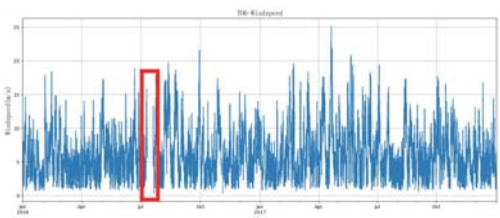

Fig. 6. Wind speed sequence diagram around the T06 wind turbine

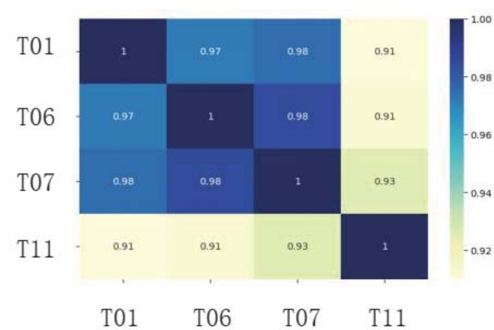

Fig. 7. Wind speed correlation heat map around each wind turbine

Analyzing the correlation of wind speeds among each wind turbine and depicting a correlation heatmap (as shown in Fig. 7), it is evident that the wind speed around the T06 wind turbine has the highest correlation with the T07 wind turbine. Consequently, the wind speed data from T07 is utilized as the basis to fill in the missing values of the wind speed data for the T06 wind turbine.

3.1.3. *Data normalization*

The numerical range and units of each feature in the original SCADA data are different. In this paper, the Min-Max normalization method is used to linearly transform the original data, mapping the data to a range of 0-1, in order to eliminate the dimensional influence of each variable. The formula is as follows:

$$x' = \frac{x - min(x)}{max(x) - min(x)} \quad (2)$$

Where $min(x)$ represents the minimum value of the data set, and $max(x)$ represents the maximum value of the data set.

3.2. *Hydraulic group failure warning*

To establish a hydraulic group failure warning model, the first task is to determine the input indicators of the model. Based on the results of correlation analysis, features with Spearman correlation coefficients greater than 0.6 with the hydraulic group oil temperature are selected as the input of the model, and the hydraulic group oil temperature is selected as the output of the model, constructing the experimental dataset.

A total of 14,265 SCADA signals from December 1, 2016 to June 30, 2017 were selected as the dataset for building a healthy operating hydraulic oil temperature prediction model. The dataset was divided into training and testing sets in a 7:3 ratio, and a hydraulic oil temperature prediction model was established. With the parameters of the prediction model fixed, the regression prediction model for the generator bearing temperature of the wind turbine was established, and the predicted results for normal operation of the generator bearing temperature are shown in Fig. 8.

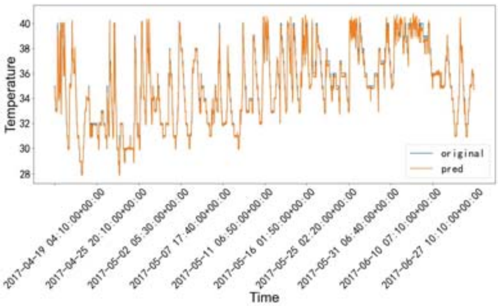

Fig. 8. Prediction results of oil temperature of hydraulic group in healthy state

The blue curve in Fig. 8 represents the actual oil temperature curve of the hydraulic group under healthy conditions, while the orange curve represents its predicted curve.

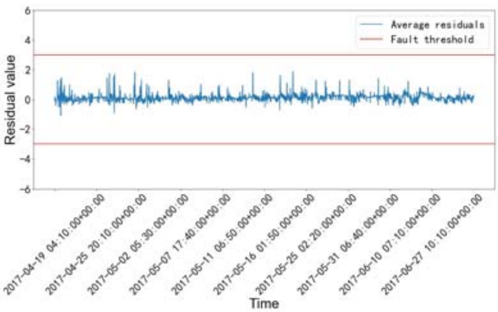

Fig. 9. Health status hydraulic group oil temperature prediction sliding window residual mean value results

We calculate the residual between the predicted hydraulic oil temperature values and the actual values output by the model, and use sliding window to calculate the average of residuals and obtain the residual mean of the hydraulic group prediction model, as shown in Fig. 9.

The blue curve in Fig. 9 represents the distribution of sliding window residual mean values for the prediction model. All residual mean values are within the range of [-3,3]. Therefore, this paper uses a sliding window residual mean value of ±3 as the threshold for hydraulic group failure detection, as indicated by the red line in Fig. 9.

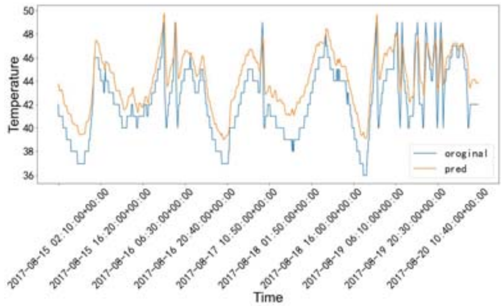

Fig. 10. Failure state hydraulic group oil temperature prediction results

From the historical data of the SCADA system of the wind turbine, data records where hydraulic group failures occurred were extracted. In this section, the SCADA signals of wind turbine T06 from August 15, 2017 to August 21, 2017 were used as inputs to the established hydraulic group oil temperature prediction model.

In Fig. 10, the blue curve represents the actual hydraulic oil temperature curve of the hydraulic group, while the orange curve represents the predicted curve. Using the same method, the residual values of the predicted and actual hydraulic oil temperatures under the failure state were calculated, and the sliding window was used to calculate the mean value of the residual values.

In Fig. 11, the blue curve represents the residual curve, and the red line represents the failure threshold. The sliding window residual mean value exceeded the threshold for the first time at 14:30 on August 16, 2017, while the actual hydraulic group failure occurred at 09:47 on August 19, 2017. Therefore, the model proposed in this paper for predicting component failures using the predicted oil temperature of the hydraulic group through the wind turbine SCADA system could detect hydraulic group failure about 31 hours in advance.

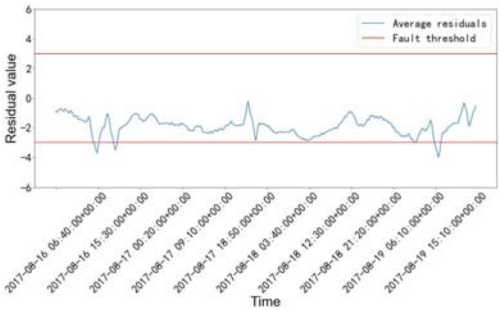

Fig. 11. Failure state hydraulic group oil temperature prediction sliding window residual mean results

3.3. *Bearing failure warning*

By using the same method described in section 3.2, we select a total of 16388 SCADA signals from January 1, 2017 to June 30, 2017 as the data to construct the healthy operation bearing temperature prediction model. The predicted results of the bearing temperature during normal operation are shown in Fig. 12.

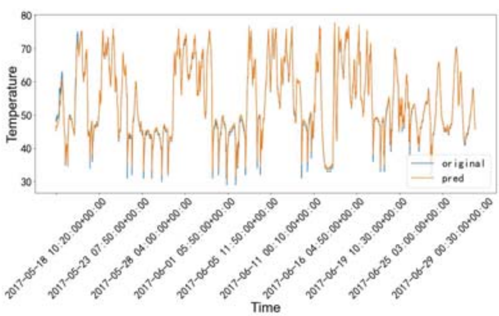

Fig. 12. Health bearing temperature prediction results

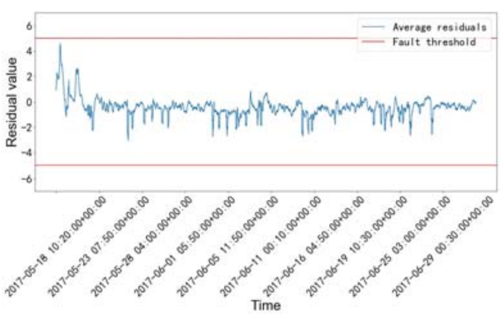

Fig. 13. Sliding window residual mean distribution of bearing temperature prediction in healthy state

In Fig. 13, this paper sets the threshold for bearing failure as the residual mean value of ±5.

In this section, SCADA signals from wind turbine T07 from August 17th to August 23rd, 2017 were used as input to the established bearing temperature prediction model.

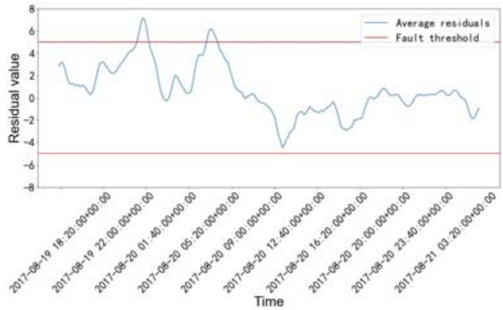

Fig. 14. Failure state bearing temperature prediction sliding window residual mean local graph

In Fig. 14, it can be seen that the first time the sliding window residual mean exceeds the threshold is at 01:00 on August 20, 2017, while the actual failure occurred at 06:08 on the same day. Therefore, this model can detect bearing failures about 5 hours in advance.

3.4. *Transformer failure prediction*

The same method, 17368 SCADA signals from October 1, 2016 to April 28, 2017 were selected. The predicted results of the normal operating transformer temperature were shown in Fig. 15.

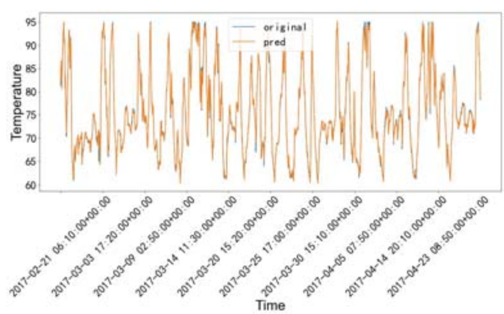

Fig. 15. Health state transformer temperature prediction results

As shown in Fig. 16, this paper sets the bearing failure threshold as the residual mean value of ±3. In this section, the SCADA signals of wind turbine T07 from July 7, 2017, to July 12, 2017, were used as input to the established transformer temperature prediction model.

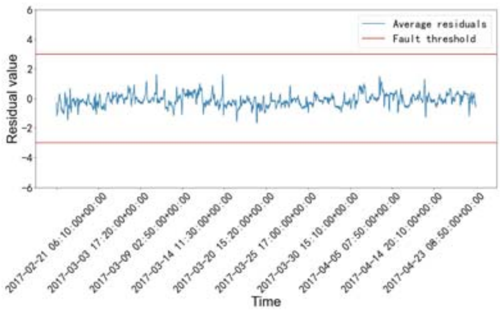

Fig. 16. The mean distribution of sliding window residual of transformer temperature prediction in health state

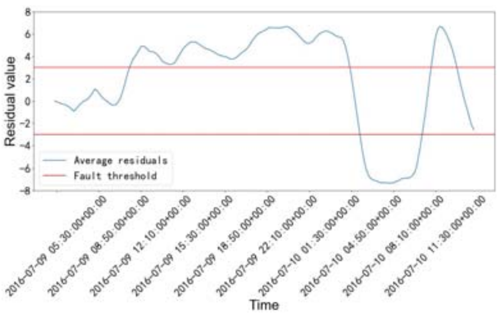

Fig. 17. Failure state transformer temperature prediction sliding window residual mean local graph

In Fig. 17, the sliding window residual mean value first exceeded the threshold on July 19, 2016, at 11:20. The actual time of the failure occurred was on July 10, 2016, at 03:46. Therefore, this model can detect transformer failures about 15 hours and 40 minutes in advance.

4. Conclusion and future work

This paper first explains the process of establishing an early warning model for key components of wind turbines, builds a failure prediction model based on LSTM, and uses the average value of sliding window residual to carry out early warning for failures. The model is then applied to actual windfarm failure early warning. The experiments show that the algorithm can predict failures of the hydraulic group, bearing, and transformer about 31 hours, 5 hours, and 15 hours in advance, respectively, and verify that the algorithm has high prediction accuracy.

Even though SCADA data is installed in most wind turbines, we have found that the data is too noisy and it can only predict with limit kinds of failure. For more failure modes, the diverse kinds of inspection technique should be involved to provide more feature data. Furthermore, due to the limited number of failure cases, this paper lacks a statistical perspective to validate the effectiveness of the proposed algorithm. Therefore, future work will focus on augmenting failure data.

With the failure prediction of component on wind turbine, we then develop predictive maintenance policies for wind turbine and wind farm but it will be presented in the future work.

Acknowledgement

This work was supported by the CNOOC Energy Development Limited Company-China University of Petroleum (Beijing) joint innovation fund(No. GD2021ZCAF0021).

References

[1] Thi-Anh-Tuyet Nguyen, Chou Shuo-Yan. Maintenance strategy selection for improving cost-effectiveness of offshore wind systems[J]. Energy Conversion and Management, 2018, 15786-95.

[2] Bei Wu, Cui Lirong, Yin Juan. Reliability and maintenance of systems subject to Gamma degradation and shocks in dynamic environments[J]. Applied Mathematical Modelling, 2021, 96367-381.

[3] Yang Lu, Sun Liping, Zhang Xinyue, et al. Condition based maintenance optimization for offshore wind turbine considering opportunities based on neural network approach[J]. Applied Ocean Research, 2018, 7469-79.

[4] Wang L., Z. Zhang, H. Long, et al. Wind Turbine Gearbox Failure Identification With Deep Neural Networks[J]. IEEE Transactions on Industrial Informatics, 2017, 13(3): 1360-1368.

[5] Liang Guo, Li Naipeng, Jia Feng, et al. A recurrent neural network based health indicator for remaining useful life prediction of bearings[J]. Neurocomputing, 2017, 24098-109.

[6] Yu Sun, Kang Jichuan, Sun Liping, et al. Condition-based maintenance for the offshore wind turbine based on long short-term memory network[J]. Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability, 2022, 236(4): 542-553.