American Journal of Engineering Research (AJER)
e-ISSN: 2320-0847 p-ISSN : 2320-0936
Volume-11, Issue-05, pp-43-50
www.ajer.org
Research Paper Open Access

Fan Fault Diagnosis Via Long Short-Term Memory

Jiaohui Xu, Xinying Li, Wen Tan, Rong Zhou
(School of Control and Computer Engineering, North China Electric Power University, Beijing, China)
Corresponding Author: Xinying Li

ABSTRACT: *The rapid development of wind power industry poses challenges to the operation and maintenance of wind power. On-line fault diagnosis is the basis to ensure the normal operation of wind turbines. Traditional fault models or signal-driven diagnosis methods cannot solve the problems of threshold design, time series dependence of eigenvalues and applicability under multiple working conditions. With the development of science and technology, wind farms can record a large amount of wind turbine operation process data in real time. Based on these data, this paper proposes a Long short-term Memory network (LSTM) algorithm to solve the problem of fan operation fault diagnosis. LSTM is a special neural network algorithm which can fully consider the timing dependence of running variables. This paper analyzes the correlation of wind turbine operating variables, and using LSTM to establish the cross prediction model and sequential classification model respectively, combined with the feature extraction method of sliding window, which can solve the temporal correlation problem of sensor sampling data and take into account the high-dimensional characteristics of the characteristic values in space. Diagnosis by comparing the actual fan data, this algorithm is better than the traditional machine learning algorithms, including K-nearest Neighbor, Decision Tree and Random Forest, for the benchmark model of multiple classification problem has obvious advantages, because considering the temporal correlation operation variables for the degree of hidden deep fault recognition has prominent advantages.*

KEYWORDS: *Wind turbine fault diagnosis; Long Short-Term Memory; Cross prediction model; Sequential classification model*

Date of Submission: 28-04-2022 Date of acceptance: 09-05-2022

I. INTRODUCTION

Wind energy has become the most promising green energy in the 21st century. In order to maintain the operational safety and reduce the maintenance cost of wind turbines, wind turbine fault diagnosis technology has become a current research hot spot. Scholars at home and abroad have proposed a variety of fault diagnosis methods. Xing-Ze Dai [1] used genetic algorithm to optimize the BP neural network, which solved the problem that the network is easy to fall into local minima, but the convergence speed of the algorithm is slow and the computational efficiency is low. Yolanda Vidal [2] first performed group scaling and feature transformation of SCADA measurement data by multiplexed principal component analysis and then a Support Vector Machine (SVM) classifier is used to complete the classification. The simulation results show that all the faults studied can be detected and classified with an overall accuracy of 98.2%. SVM has better classification ability and computational efficiency than BP neural network, but there are still some problems for handling large amount of data. Chun-Lin Ye et al [3] proposed a fault diagnosis method based on data mining, firstly screening the fault features, then processing the SCADA dataset, and finally classifying the faults, and the experiments show that the model has good diagnostic performance and generalization. The random forest algorithm can handle high-dimensional data, but because wind turbines are an integrated system with coupling between sensors, real-time and accurate diagnosis of the cause of faults cannot be achieved. There are special features in the structure of the Long Short-Term Memory Network itself, so it has been widely used in dealing with highly time-dependent and high-dimensional strong coupling problems. The experimental results of Jing Li et al [4] showed that the LSTM-based model was able to predict the trend of hot events most accurately, followed by SVM, while the BP neural network was relatively poor in prediction. Xing-Yu Qu [5] proposed an RNN-LSTM-based fault diagnosis technique for grinding systems, which batches the datasets used in the input LSTM network, extracts their correlations in the time dimension, and achieves the classification of faults by comparing the characteristics of different moments before and after. The experimental results show that LSTM has strong classification ability for sample data sets with large temporal dependence. In this paper, based on the Long and Short-Term Memory Network, we propose a cross LSTM prediction model and a sequential LSTM fault classification model to achieve the fault diagnosis of wind turbines with full consideration of the time-dependent characteristics in the face of massive data.

II. LONG SHORT-TERM MEMORY(LSTM)

The Long Short-Term Memory (LSTM) was proposed by Hochreiter and Schmidhuber [6] in 1997 and was improved and generalized in 2000[7]. It is a special temporal Recurrent Neural Network(RNN) designed with input gates, forgetting gates, output gates and cell cells to solve the gradient disappearance problem that exists in general Recurrent Neural Network , so LSTM is well suited to handle events with long time series intervals. LSTM has been widely used in image semantics and speech recognition [8~10]. In view of its superior timing signal processing capability, LSTM has also been gradually tried to be applied to industrial processes in recent years. In this paper, we have proposed a fault diagnosis method based on the prediction and classification of Long and Short-Term Memory Networks, which can accurately identify faults in wind turbines by fully considering the time-dependent characteristics of variables on the basis of using the correlation of variables in wind turbines.

In a cell unit of the LSTM, each sigmoid layer (σfunction) is performing the computation of the weights, and the forgetting gate f_t indicates the degree of forgetting of the information at the previous moment, the input gate i_t indicates the degree of updating of the current information, and the output gate o_t indicates the degree of output after cell state excitation. The representation of the current information $\tilde{c}_t$ can be obtained after excitation through the tanh layer, which is multiplied with the input gate to determine how much new information stays in the current cell state, while the forgetting gate is multiplied with the cell state c_{t-1} of the previous moment to determine how much information is left behind. The cell state ct at moment t is obtained by adding the two, and then multiplied with the output gate after excitation through the tanh layer to obtain the output h_t at moment t. This operation can be understood as compressing or updating some information that existed at the previous moment, and later adding information from this moment, using this way to remember long-term information (where $\tilde{c}_t$, c_t, h_t, and x_t denote current information, cell state, output and input, respectively). Figure 1 represents the structure, data flow inside the neuron.

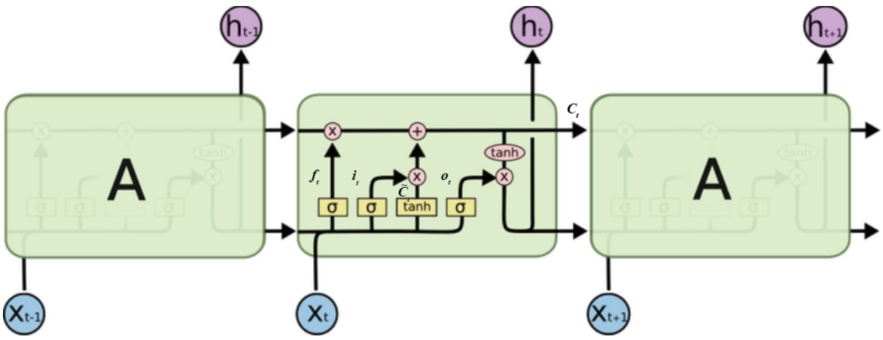

Fig.1. Structure diagram of LSTM unit

The intra-neuronal formula is:

$$i_t = \sigma\left(U_i h_{t-1} + W_i x_t + b_t\right) \quad (1)$$

$$f_t = \sigma\left(U_f h_{t-1} + W_f x_t + b_f\right) \quad (2)$$

$$o_t = \sigma\left(U_o h_{t-1} + W_o x_t + b_o\right) \quad (3)$$

$$\tilde{c}_t = \tan h(U_c h_{t-1} + W_c x_t + b_c) \quad (4)$$

$$c_t = f_t \otimes c_{t-1} + i_t \otimes \tilde{c}_t \quad (5)$$

$$h_t = o_t \otimes \tan h(c_t) \quad (6)$$

Where, U, W, and b are the weights to be learned.

III. USING LSTM TO PREDICT AND CLASSIFY

Using LSTM network can perform timing prediction and fault classification. The basic idea of LSTM timing prediction is a model that uses the previous value of a sensor to predict the present value of that sensor, but wind turbines are an integrated system where the sampled values of all sensors are interrelated and the value of a sensor also depends on the values of other associated sensors. Based on this feature, a cross LSTM prediction model was proposed in the literature [11], which can predict the future value of one sensor using the past values of all the related sensors. Fault classification generally establishes the mapping relationship between feature values and corresponding fault categories, but in order to achieve fault diagnosis on the basis of fully considering the time-dependent characteristics, this paper also proposes a sequential LSTM fault classification model, and the specific structures of the two models are described as follows.

3.1 Cross LSTM prediction model

(1) Suppose that a system collects measurements from *t* sensors $(\alpha \;, \beta \quad \gamma \;\ldots)$, sampling a total of *n* observations, which are filtered and normalized to form a data set *x* with dimension *t*n*.

$$x = \begin{bmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_n \\ \beta_1 & \beta_2 & \cdots & \beta_n \\ \gamma_1 & \gamma_2 & \cdots & \gamma_n \end{bmatrix}_{t\times n} \tag{7}$$

The filtering is used to remove high-frequency noise from the sensor measurements, and normalizing the data allows the data to have zero mean and unit variance for a better fit. When making predictions, care needs to be taken to normalize the test set using the same normalization parameters as the training set.

(2) The training prediction model ψ. Dividing x into train set and test set, the cross LSTM prediction model is shown in Fig. 2. The input *x* is the sampled values of *n* sensors at moments *t-15* to *t-1*, and the regression value *y* is the value α_t of sensor *α* at moment *t*. The network has 15 loop steps and the dimension of the input data is *n*. The network is trained using the Adam optimization algorithm and has four layers: the input layer, the LSTM layer, the fully connected layer and the regression layer.

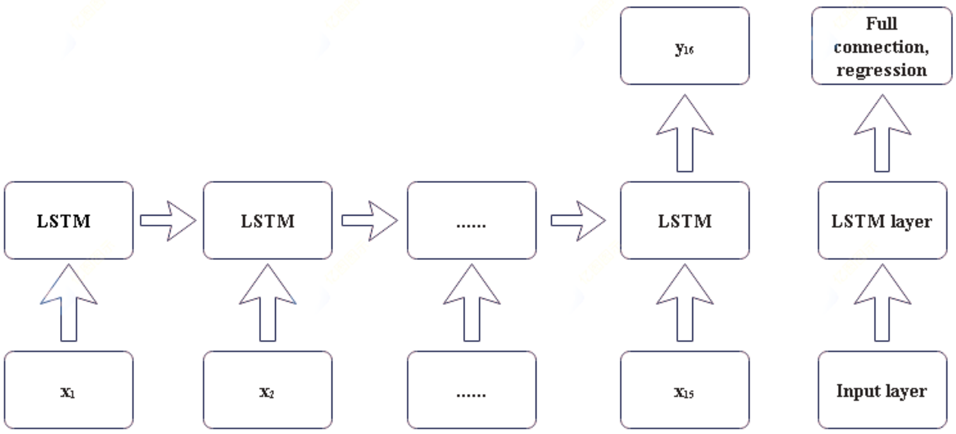

Fig.2. Structure diagram of cross LSTM prediction

(3) Generate the residuals of the test set. As shown in equation 8, the prediction model Ψ for sensor *α* is constructed by using the training set and then input to the test set x_{test}, the predicted value $\hat{x}_{test}$ of the corresponding sensor can be obtained. It should be noted that the test sets have been normalized, so it is also necessary to recover their original size. Finally, the difference between x_{test} and $\hat{x}_{test}$ constitutes the residual signal R_α, which contains information about various faults.

$$\hat{x}_{test} = \Psi(x_{test}) \tag{8}$$

3.2 Sequential LSTM classification model

In order to fully consider the time-dependent characteristics of variables and achieve real-time fault diagnosis, this paper designs a sequential LSTM classification network, as shown in Fig. 3, which has four layers, namely, input layer, LSTM layer, fully connected layer and classification layer, the number of cyclic steps is the length of the whole sequence 8000, the gradient at each update of the weights is the average gradient of 8000 steps, the output corresponding to each cyclic step is the fault category at that moment, the input dimension is the characteristic dimension of the residuals, and the output dimension is the number of categories.

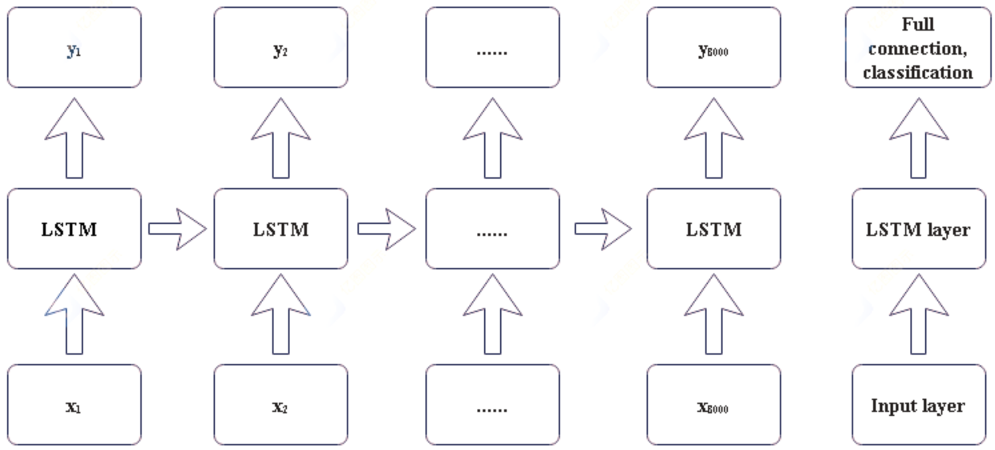

Fig.3. Structure diagram of sequential LSTM classification

The sequential LSTM classification network uses a stochastic momentum gradient descent optimizer for iteration. Compared with the traditional stochastic gradient descent algorithm, the former adds a momentum term to the parameter update to reduce the oscillation phenomenon along the most rapid descent path as the latter moves toward the optimal direction. The process of its update is as follows:

$$\theta_{l+1} \approx \theta_l - \alpha \nabla E(\theta_l) + r(\theta_l - \theta_{l-1}) \qquad (9)$$

Where θ, l, and α are the weight vector, the number of iterations, and the learning rate, respectively; E(θ) and ∇E(θ) are the loss function and the gradients of the loss function , respectively; and r is the contribution of the previous step's gradient to the current step.

In addition, a regularization term (10) is added to the loss function E(θ) to reduce overfitting, and the regularization function Ω(w) is shown in equation 11.

$$E_R(\theta) = E(\theta) + \lambda\Omega(w) \qquad (10)$$

$$\Omega(w) = \frac{1}{2} w^T w \qquad (11)$$

Where λ and w are the regularization factor and the weight vector, respectively.

IV. USING LSTM FOR FAULT DIAGNOSIS

The first step is the timing prediction. A model is trained to predict the output of a wind turbine at the next moment under normal conditions. If a certain fault occurs in the turbine, the predicted value will differ from the measured value and show a certain regularity, so that the fault can be judged to have occurred. The cross LSTM prediction network performs well in mining the intrinsic connections of variables [12], connected to this paper, the cross LSTM model can be used to predict the output of a particular sensor to obtain the difference between the true value and the predicted value, as described in Section 3 of this paper.

The second step is to extract the residual features. The fault classification by using the residuals directly often does not yield the desired results because this does not take full advantage of the hidden features in the multi-sensor data, so the time and frequency domain features of the residual signals can be extracted to improve the performance of fault diagnosis. This chapter utilizes a sliding window feature extraction method that selects five common time domain features, mean, peak, peak-to-peak, energy, and root mean square amplitude, using this method to preserve the time-series dependence of the signal and facilitate real-time fault diagnosis. For the residual signal $R\alpha = \{r\alpha 1 \ , r\alpha 2 \ldots \quad r\alpha n\}$ of sensor α , n is the length of the signal, defining Φ as the feature extraction function and y as a certain time-domain feature to be extracted, the length of the signal y is n-m+1 and the value of y at moment t is:

$$y_t = \Phi(r_{t-m} : r_t) \qquad (12)$$

Where, m is the width of the sliding window.

Table.1. Five feature extraction functions

Time domain features	feature extraction function Φ

mean	
peak	$x_p = \max(X)$
peak-to-peak	$x_{p-p} = \max(X) - \min(X)$
energy	$E = \sum_{i=1}^{m} x_i^2$
root mean square amplitude	$x_r = \left(\frac{1}{m} \sum_{i=1}^{m} \lvert x_i \rvert^{1/2} \right)^2$

Where the left side of the equation is the time domain characteristic of the output and the input is the residual signal $X = \{x1 , x2 ... \quad xm\}$, with length m.

The third step is using a sequential LSTM for fault separation. The specific procedure of this network training can be seen in Section 3, where the cross-entropy loss is chosen as the loss function. All the feature signals obtained from the previous step form the matrix $Y = \{y1 , y2 \ldots \quad yn\}$, corresponding to the fault label $T = \{t1 , t2 \ldots \quad tn\}$. The above signals are divided into training and test sets, and after several iterations, a classification model Y between T and Y can be built, and then the fault class of the test set can be predicted. The flow of this fault diagnosis strategy is shown in Figure 4:

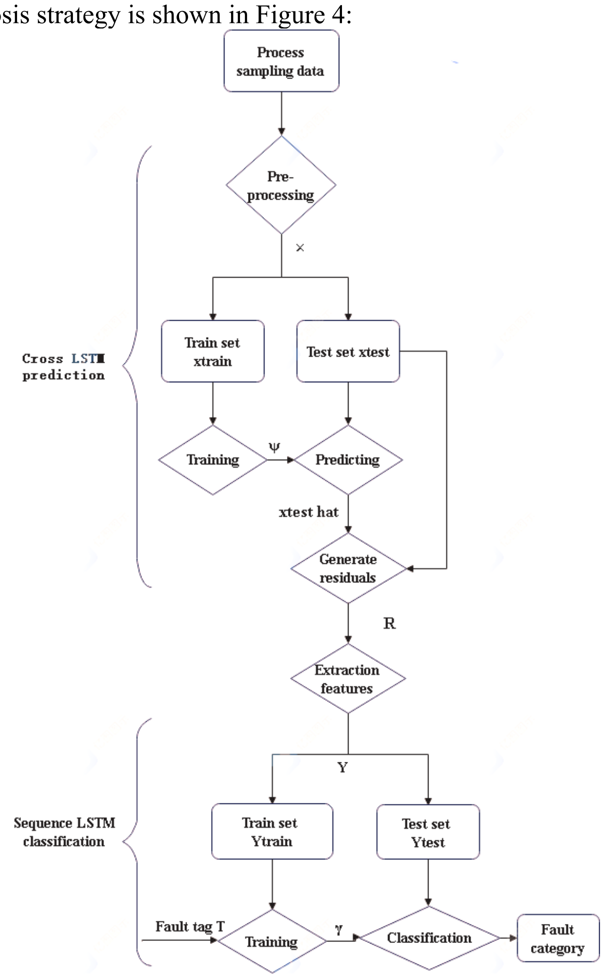

Fig.4. Flowchart of fault diagnosis strategy

V. SIMULATION RESULTS AND COMPARISON

In this paper, three traditional machine learning algorithms, k-nearest Neighbor (KNN) , Decision Tree (DT) and Random Forest, are compared with the fault diagnosis strategy designed in this paper. In order to measure the performance of the method on different datasets and to verify the generality of the method on a real wind turbine data platform, three common metrics, Prec, MDR, and F_1-score, are used here to compare the performance of each method.

$$\Pr ec = \frac{TP}{TP + FP} \tag{13}$$

$$\text{MDR} = \frac{TP}{TP + FN} \tag{14}$$

$$F_1\text{-}score = \frac{2TP}{2TP + FP + FN} \tag{15}$$

Where TP, TN stand for True positive, True negative; FP, FN stand for False positive, False negative respectively.

In this paper, we use the operational data of a wind turbine blade icing prediction contest, which is the data of a certain unit from November 1, 2015 to January 1, 2016 after collecting 26 sensors and normalizing them for two months, some of which have uncertain blade status. Some of these data blade state uncertainty. The data in the blade state determination of the observation points have 374147, of which the state of normal has 26 segments a total of 350255 observations, blade icing has 29 segments 23892 observations, segments and segments are not continuous between, and the number of observations included are not the same. Generally, the fault duration is less and the non-fault duration is longer. To facilitate diagnosis, connect the time segments of the same state. There are too many normal state data points, so only the first 6 segments are taken for a total of about 120,000 data points, the fault time segments are linked together for more than 20,000, and these two segments are named normal and fault.

Table.2. Characteristics of data points

Number of continuous observation points under fault conditions	0-250	250-500	500-750	750-1000	1000-1300	Total
Number of segments	2	3	3	17	4	29
Number of continuous observation points under normal conditions	0-8000	8k-16k	16k-24k	24k-32k	32k-48k	Total
Number of segments	13	5	3	1	4	36

The process of fault diagnosis is as follows:

(1) Pre-processing. Noise is inevitably mixed into the sensor measurements, so the first step is using a low-pass filter to filter out the high-frequency noise from both data segments.

(2) Build predictive models. We divided the pre-processed data into training set and test set. As mentioned before in paragraph 4, the training set for building the prediction model should be the data in normal state, so we divided normal into two segments, the first 80,000 observations as the training set, and the second 40,000 or more connected with failure as the test set, with more than 70,000 observations. After the prediction model is established, the test set is used as input to generate the predicted values, which are made to differ from the actual output values, and then the residual signal is generated. For the convenience of representation, the normal state is numbered as 0 and the icing state is numbered as 1. Randomly select 6000 normal state and 6000 icing state data points forming 3000(0)+3000(1)+3000(0)+3000(1) totaling 12,000 observations to form a signal for the next step. Figures 5 shows the trends of the residual signals for the two variables int_tmp, environment_tmp in this segment of data.

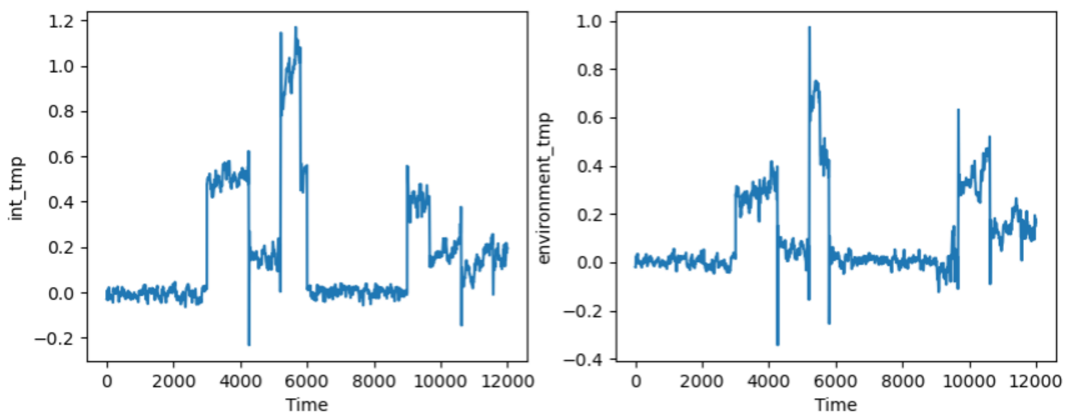

Fig.5. Residual signal of int_tmp and environment_tmp

(3) Extracting features. Selected the residual signals of 2 sensors int_tmp, environment_tmp, and extracted 5 features for each residual signal, so the size of the data after extracting features is 10*12000.

(4) Fault classification. Divide the above data into training and test sets, which are both 10*6000 in size. The trained classification model is used for fault identification in the test set, and the performance of the method is evaluated by comparing the true labels of the test set after obtaining the classification labels. The following figure shows the performance comparison of four fault diagnosis methods: K-nearest Neighbor, Decision Tree, Random Forest, and LSTM predictive classification.

Table.3. Comparison of the four methods

State	Indicator	KNN	DT	Forest	LSTM
Normal	Prec	99.7	99.7	99.7	**99.8**
	MDR	8.0	6.6	6.7	**6.2**
	F_1-score	95.7	96.4	96.4	**96.7**
Icing	Prec	91.3	93.0	92.8	**93.7**
	MDR	0.3	0.3	0.3	**0.2**
	F_1-score	95.3	96.2	96.1	**96.5**
Average	Accuracy	95.5	96.3	96.3	**96.6**

Here using the three metrics mentioned above to evaluate the performance of the four fault diagnosis methods. As shown in Table 3, the diagnostic method corresponding to the bolded numbers is the best performer, and it is clear that the advantage of LSTM predictive classification is dominant. However, taken together, the differences between the other three methods and the method proposed in this chapter are not obvious, and the method proposed in this paper does not have particular advantages in the dichotomous problem. The comprehensive simulation results show that LSTM prediction classification is more effective if the problem faced is a multi-classification problem. If faced with a dichotomous classification problem, traditional machine learning methods can be chosen, but the cross LSTM prediction network performs well in mining the intrinsic association of variables, and the residual signal generated after its prediction contains fault information, which helps to identify fault classes.

VI. SUMMARY

This paper focuses on a fault diagnosis strategy based on Long Short-Term Memory Network, which modifies the training process of LSTM , and proposes two networks with different functions and different training methods, namely cross prediction and sequential classification, to diagnose the faults of wind turbines. Selected data provided by a wind turbine blade icing prediction, the pre-processed data is divided into training set and test set, the test set is used as input to generate the predicted values, which are made to differ from the actual output values, and then the residual signal is generatedand , using the extracted features of the residual signal for fault classification and comparison with other fault diagnosis methods.

From the simulation results, this approach outperforms traditional machine learning algorithms, including K-nearest Neighbor, Decision Tree and Random Forest, in all diagnostic performance on real data sets. It has obvious advantages for the multi-classification problem of faults, especially for actuator faults that are relatively difficult to diagnose; while in the binary classification problem of whether the blades are iced, traditional machine learning algorithms have performed well, so the advantages of the method are not particularly obvious, but the performance of traditional machine learning also benefits from the ability of the LSTM prediction network to extract fault information. Collectively, it seems that this approach, which considers the spatio-temporal correlation of variables, has outstanding advantages for faults with a deeper degree of concealment.

REFERENCES

[1] Dai Xingze, Dou Huiyang. Fault warning algorithm of fan gearbox based on GA-BP neural network [J]. Electronic Technology & Software Engineering,2020(09):160-163.

[2] Vidal Y, Pozo F, Tutivén C. Wind Turbine Multi-Fault Detection and Classification Based on SCADA Data[J]. Energies. 2018, 11(11): 3018.

[3] Ye Chunlin, Qiu Yingning, Feng Yanhui. Faults Diagnosis of Wind Turbine Blade Icing based on Data Mining [J]. Noise and Vibration Control, 2018,38(S2):643-647.

[4] Li Jing, Xu Lulu. Comparison and Analysis of Research Trend Prediction Models Based on Machine Learning Algorithm——BP

Neural Network Support Vector Machine and LSTM Model [J]. Journal of Modern Information, 2019,39(04):23-33.

[5] Qu Xingyu, Zeng Peng, Li Junpeng. Fault Diagnosis Technology of Grinding System Based on RNN-LSTM [J]. Information and Control,2019,48(02): 179-186.

[6] Hochreiter S, Schmidhuber J. Long Short-term memory[J]. Neural Computation, 1997,9(8):1735-1780.

[7] Gers F A, Schmidhuber J, Cummins F. Learning to Forget: continual prediction with LSTM[M]. Switzerland: Istituto Dalle Molle Di Studi Sull Intelligenza Artificiale, 1999:2451–2471.

[8] Auli M, Galley M. Joint language and translation modeling with recurrent neural networks[A]. Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing[C], Association for Computational Linguistics, 2013:1044-1054.

[9] Jia X, Gavves E. Guiding the long short term memory model for image caption generation[J] International Conference On Computer Vision, 2015: 2407-2415.

[10] Sak H, Senior A. Long short-term memory based recurrent neural network architectures for large vocabulary speech recognition[J]. Comput. Sci, 2014, 338–342.

[11] Yang J, Guo Y, Zhao W . Long short-term memory neural network based fault detection and isolation for electro-mechanical actuators[J]. Neurocomputing, 2019, V360(Sep.30):85-96.

[12] Lu J, Liu F. Statistical Modeling of Dynamic Multivariate Process Using Canonical Variate Analysis[A]. International Conference on Information & Automation[C]. IEEE, 2006:218-221.

Xinying Li, et. al. "Fan Fault Diagnosis Via Long Short-Term Memory."*American Journal of Engineering Research (AJER),* vol. 11(05), 2022, pp. 43-50.