Early fault detection for power plant fans based on dynamic neural network

Guobin Zhang¹, Ronghua Du¹, Xiaogang Xin¹, Wei Zhao¹, Shaojia Dang¹, Song Zhao¹, Pei Yun²

¹Inner Mongolia Power Research Institute, Hohhot 010020, P.R. China;
²School of Control and Computer Engineering, North China Electric Power University, Beijing 102206, P.R. China

¹E-mail: zgb7230@163.com
²E-mail: yunpei@ncepu.edu.cn

Abstract: Early fault detection is increasingly important for the reliable and secure operation of power plant fans. In this paper, we propose an early fault detection strategy for power plant fans based on dynamic neural network. First, the nonlinear autoregressive (NARX) with exogenous inputs network is utilized to predict the behavior of fans by using normal operating data, and the discrete particle swarm optimization (DPSO) is utilized to optimize the hyper-parameters of NARX network to enhance prediction accuracy. Then, the fault alarm strategy is proposed by the prediction deviations and generalized extreme value (GEV) theory. Finally, the proposed method is applied to a forced draft (FD) fan in a coal-fired power plant. Experiment results show that the proposed model has high prediction accuracy on normal operations and produces large prediction errors when a failure occurs. Furthermore, it can detect early faults accurately and timely before the failure occurs.

1. Introduction

With the rapid development of renewable energy such as wind power and solar power, thermal power plant plays an important role in peak regulation rapidly and flexibly, which put forward higher requirements for safe and reliable operation of thermal power plant equipment [1]. Fans are important auxiliary equipment in thermal power plant, which maintain the air circulation required by power generation. However, fans usually run under adverse environment, which is the leading reason for high failure rate. Fans failure could not only lead to potential safety accidents, but also cause unscheduled shutdown and expensive maintenance cost. Thus, early fault detection is demand to give warning signals before fault occurs and reduce maintenance cost [2].

A number of strategies, including: model-based methods and data-based methods, have been suggested to achieve early fault detection in the past. Due to the complex structure of fans, it is difficult to accurately characterize the degradation model to predict the fault in advance [3]. However, data-based methods do not need accurate mechanism models. By using data mining technology, intelligent monitoring and early fault detection become possible. Therefore, several data-based methods, such as neural network [4] and support vector machine (SVM) [5], have been used in early fault detection. However, the above methods need both normal operating data and fault data. Due to no insufficient fault data in power plant, fault identification for fans could be achieved by modeling the normal operation behavior. Any different from normal operation is regarded as a failure. Nonlinear state estimation (NSET) [6] has been proven to have good performance in
normal state modeling. These methods mainly identify faults by extracting the cross-relationships between variables. Considering that operating data of fans are nonlinear dynamic time series, this paper utilizes a dynamic neural network to model normal operation behavior. Nonlinear autoregressive (NARX) network with exogenous input performs well in modeling nonlinear dynamic system, which is widely used for wind speed prediction and fault detection\cite{7}.

In this paper, NARX network is used to extract the nonlinear dynamic features hidden in historical normal data and predict the behavior of fans, and the hyper-parameters of NARX are determined by using discrete particle swarm optimization (DPSO) to improve the prediction accuracy. Then, a novel threshold-based alarm strategy based on generalized extreme value (GEV) theory of prediction deviation is proposed and applied to a real coal-fired power plant for early fault detection.

2. Fans description

In this paper, we focus on a forced draft (FD) fan in a real coal fired power plant. A large lot of parameters, including current, pressure, temperature, and vibration, are monitored and saved in the Supervisory Information System (SIS) in plant. Table 1 shows the variables used for early fault detection in this paper. The motor current and motor power represent the load of the fan. Due to the FD fan is a large rotating equipment, bearing is an important component. Thus, bearing temperature and bearing vibration are also important parameters to reflect the running status of the FD fan.

3. Early fault detection model

3.1 NARX network

NARX network is a kind of dynamic artificial neural network, which could represent the nonlinear dynamic characteristics between input data and output data\cite{8}. Its dynamic behavior could be expressed as:

\[ y(t) = f(y(t - 1),..., y(t - n_y), u_1(t), u_1(t - 1),..., u_1(t - n_u),..., u_m(t), u_m(t - 1),..., u_m(t - n_m)) \]  

Where \( y(t) \) is predict value, \( f \) denotes a nonlinear mapping function, \( u_1(t), u_1(t - 1), u_1(t - 2),..., u_1(t - n_u) \) are actual values of variable \( i \) at time \( t \) and its historical moment. There are total of \( m \) variables. \( n_{u_1}, n_{u_2},..., n_{u_m} \) are input delays of variable \( u_1, u_2,.., u_m \) respectively and \( n_y \) is output delay.

NARX network consists of two structures, including closed-loop structure and open-loop structure. Closed-loop structure uses the predict value to provide feedback, which is suitable for the situation where the target value is unknown. The open-loop structure uses the actual measurement to provide feedback. The operating parameters of FD fan in this paper are known, so this paper adopts open-loop structure to predict the behavior of FD fan.

The accuracy of NARX network is the most important premise of the effectiveness of anomaly identification. Selecting of hyper-parameters, including input delay, output delay, and number of hidden neurons, directly affects the topological structure and prediction accuracy of NARX network. In this paper, DPSO is utilized to find the optimal hyper-parameters. To simplify the optimization algorithm, the input variables are divided into three types, including current, temperature, and vibration. Each type of input variables has the same input delay. Therefore, there are 5 parameters to be optimized, including output delay \( n_y \), input delay \( n_i, n_1, n_2 \), and number of hidden nodes \( n_h \).

3.2 DPSO

PSO is a population-based global optimization technique inspired by social behavior of bird flocks looking for corn\cite{9}. In PSO, each potential solution of optimization problem is known as a particle. Considering a d-dimensional search space, a population consists of n particles represented by \( X = \{x_1, x_2,..., x_n\} \), where \( x_i = [x_{i1}, x_{i2},..., x_{id}] \). Considering that the hyper-parameters of NARX network is discrete, discrete PSO (DPSO) is utilized to optimize the hyper-parameters. In DPSO, each particle is coded by a binary bit 0 or 1.
At the $t^{\text{th}}$ iteration of the DPSO, the position and velocity of the $i^{\text{th}}$ particle are represented by $x_{i,t}$ and $v_{i,t}$, respectively, the personal best and global best position are $p_{best_{i,t}}$ and $g_{best_{t}}$, respectively. Velocity and position of the $i^{\text{th}}$ particle at the $(t+1)^{\text{th}}$ iteration are updated as follows:

$$v_{i,t+1} = w v_{i,t} + c_1 r (p_{best_{i,t}} - x_{i,t}) + c_2 r (g_{best_{t}} - x_{i,t}),$$

where $w$ is the inertia weight for controlling the effective balance between personal and global optimization capabilities. $c_1$ and $c_2$ are acceleration coefficients, determining the influence of particle itself and other particle experiences on position and velocity. $r$ denotes independent and uniformly distributed random vector within the range of $[0, 1]$.

The fitness function is defined as the mean square error of actual value and predict value from NARX network with validation data, shown as follows,

$$fit = \frac{1}{N} \sum_{n=1}^{N} (\hat{y}_n - y_n)^2$$

where $\hat{y}_n$ is the predict value and $y_n$ is the actual value.

### 3.3 Alarm strategy

The residual between actual value and predict value from NARX network can reflect the change of a single variable. In order to represent the overall operating state of FD fan system, the predicted deviation of the system is defined as:

$$D_{t} = \sqrt{\frac{\sum_{n=1}^{m} (X_{i,n} - Y_{i,n})^2}{m}}$$

where $m$ is the total number of samples, $X$ and $Y$ represent the predict value and actual value, respectively.

Prediction deviation obtained under normal states are considered as normal and healthy, which could be used as a reference for early fault detection. Due to prediction deviations usually have an extreme distribution that may cause false alarms, the threshold based on generalized extreme value (GEV) theory is proposed and utilized as alarm rules.

The deviation sequence is divided into several minimum intervals with the same data points, and 5 mins is selected as the minimum interval. $M_n$ represents the maximum value within each minimum interval, the relationship between the distribution of $M_n$ and the distribution $F$ is:

$$P_n\{M_n \leq z\} = P\{x_1 \leq z, ..., x_n \leq z\} = \{F(z)\}^n$$

However, the distribution $F$ is unknown. Assuming that $\mu$ and $\sigma$ satisfy the follow function,

$$P_n\{(M_n - \mu) / \sigma\} \rightarrow G(z)$$

where $\mu$ is the location parameter, $\sigma$ is the scale parameter, and $G$ is the generalized extreme value distribution function\(^{10}\), which can be described as follows:

$$G(z) = \exp\{-1 + (z - \mu) / \sigma\}^{-\xi}$$

We define the set: $\{z: 1 + (z - \mu) / \sigma > 0\}$, where $-\infty < \mu < \infty$, $\sigma > 0$, $-\infty < \xi < \infty$, and $\xi$ is the shape parameter. All the three parameters, including location parameter, scale parameter and shape parameter, are obtained by using maximum likelihood estimation approach. The alarm threshold could be calculated as an inverse function of the cumulative distribution, which is calculated as:

$$D_{\alpha} = \mu - \sigma \xi \left[1 - \ln(1 - \alpha)^{-\xi}\right]$$

where $\alpha$ is confidence limit.

The warning threshold $D_D$ is set as when the confidence limit is 95%, and alarm threshold $D_H$ is set as when the confidence limit is 99%. The alarm rule is:
\[ S = \begin{cases} 
0 & D < D_k \\
1 & D_k \leq D < D_n \\
2 & D \geq D_n 
\end{cases} \quad (9) \]

3.4 NARX network training and validation
In this paper, a total of 11520 samples before the failure were acquired for research. Among them, first 8000 samples are considered as normal samples, while the last 3520 samples are considered as fault samples, which is named as Fault set for early fault detection. Normal samples are divided into Training set, Validation set and Test set. The Training set is used to train the NARX network, Validation set is used to optimize the hyper-parameters of NARX network, and Test set is used to test the prediction accuracy of the model.

The parameters of NARX network need to be determined before the model training, including output delay \( n_y \), current variable input delay \( n_i \), temperature variable input delay \( n_t \), vibration variable input delay \( n_v \), and number of hidden nodes \( n_h \). In this paper, the vary range of the four hyper-parameters are \([1, 20]\), \([1, 20]\), \([1, 20]\), and \([1, 50]\), respectively. In DPSO, the population size is set as 10, the maximum number of iterations is 100. Fig. 1 shows the iterative process. It can be seen that the fitness value converges to 0.017 after the 39th iterations, indicating that at the 39th iterations, the optimal hyper-parameters are found, where \( n_y = 8 \), \( n_b = 10 \), and \( n_h = n_i = n_v = 6 \).

![Figure 1. Iterative process](image)

3.5 Prediction accuracy analysis
The most commonly performance indicators such as mean absolute error (MAE) and root mean squared error (RMSE) are used to analyze the performance of NARX network. Tab. 1 shows the prediction accuracy of feature variables with Test set. It can be observed that although the prediction accuracy of each variables is different, the maximum RMSE of temperature variables is lower than 0.03\( ^\circ \)C, and the maximum RMSE of vibration variables is lower than 0.05mm, which indicating that NARX network has high prediction accuracy for normal operating data and could accurately map the nonlinear dynamic relationship between the operating parameters of power plant fans.

| Variables                  | RMSE  | MAE  | Unit |
|----------------------------|-------|------|------|
| Motor winding temperature   | 0.0070| 0.0044| \(^\circ\)C |
| MNDE bearing temperature    | 0.0109| 0.0078| \(^\circ\)C |
| MDE bearing temperature     | 0.0177| 0.0106| \(^\circ\)C |
| FNDE bearing temperature    | 0.0236| 0.0136| \(^\circ\)C |
| Main bearing temperature    | 0.0248| 0.0144| \(^\circ\)C |
| FDE bearing temperature     | 0.0266| 0.0180| \(^\circ\)C |
| Bearing horizontal vibration| 0.0475| 0.0359| mm   |
| Bearing vertical vibration  | 0.0167| 0.0119| mm   |
3.6 Fault detection

To demonstrate that NARX network could distinguish normal data and fault data accurately, the prediction errors in Test set and Fault set are calculated and compared. Fig. 2 shows the prediction error of FNDE bearing temperature, FDE bearing temperature. It is obviously that the prediction error in Test set is lower than that in Fault set. And at the last of Fault set, the prediction errors of temperature variables gradually increase until they reach the maximum value. Therefore, NARX network could not only accurately extract normal operation behavior of FD fans, but also provide the possibility of early warning because of the large prediction error of fault data.

![Figure 2. Prediction error: (a) FNDE bearing temperature, (b) FDE bearing temperature.](image)

The prediction deviations of Validation set are calculated by using Eq. (5), which is regarded as normal and healthy, as shown in Fig. 3. The deviation sequence is divided into several intervals with the same data points, and the maximum Mn sequence in each interval is formed. The maximum likelihood estimation method is used to estimate the parameter value of the generalized extreme value (GEV) distribution, which is shown in Fig. 4. The warning threshold and alarm threshold are determined by using formula (8).

![Figure 3. Prediction deviations in Validation set](image)

![Figure 4. Frequency distribution of Mn sequence](image)

The deviation values of Test set and Fault set are shown in Fig. 5. It is obvious that all the deviations in Test set and the beginning of Fault set are below the warning threshold and alarm threshold. With time goes by, the deviations in the last stage of Fault set increase gradually and then exceed the warning threshold and alarm threshold. Finally, the deviations reach the maximum value. Fig. 6 shows the alarm situation visually. The warning signal is first produced at the 5016th sample, which alerts plant operators of a possible malfunction. When the alarm signal is first given at 5073th sample, it indicates that there is a fault about to occur. Therefore, the proposed method in this paper could achieve early fault detection and given operators enough time to take a maintenance, which is useful for avoiding downtime and reducing maintenance costs.
4. Conclusion
In this paper, NARX network is used to extract the nonlinear dynamic features hidden in historical normal data and predict normal behavior of fans, and DPSO is used to optimize the hyper-parameters of NARX network to improving prediction accuracy. Then, a threshold-based alarm criterion based on GEV theory of prediction deviation is proposed and applied to a real coal-fired power plant for early fault detection. Experiment results show that the proposed model has high accuracy of normal data and low accuracy of fault data. Furthermore, it could give a warning signal before the fault occurs, which gives operators enough time to take a maintenance.

Acknowledgments
This work is supported by National Natural Science Foundation of China under Grant 52061635102.

References
[1] Y. Sun, C. Xu, T. Xin, G. Xu, Y. Yang, A comprehensive analysis of a thermal energy storage concept based on low-bank coal pre-drying for reducing the minimum load of coal-fired power plants, Appl. Therm. Eng., 156: 77-99, 2019.
[2] G. Muruti, F. Rahim, Z.B. Ibrahim, A survey on anomalies detection techniques and measurement methods, in Proceeding of IEEE Conference on Applications, Information and Network Security (ANIS), 2018: 81-86.
[3] I. Hwang, Y. Sim, Y. Sim, C.E. Seah, A survey of fault detection, isolation, and reconfiguration method, IEEE Trans. on Control System Technology, 18(3): 636-653, 2010.
[4] G. Bolt, Fault models for artificial neural networks, in Proceeding of IEEE Internal Joint Conference on Neural Network, 1991: 18-21.
[5] S. Li, Z. Liu, Application of improved wavelet packet energy entropy and GA-SVM in rolling bearing fault diagnosis, in Proceeding of IEEE Internal Conference on Signal Processing, Communications and Computing (ICSPCC), 2018: 14-16.
[6] Y. Lv, F. Fang, T. Yang, C. Romero, An early fault detection method for induced draft fan based on MSET with informative memory matrix selection, ISA Trans. 102:325-334, 2020.
[7] X. Zhao, J. Liu, D. Yu, J. Chang, One-day-ahead probabilistic wind speed forecast based on optimized numerical weather prediction data, Energy Conversion and Management, 164:560-569, 2018.
[8] P. Bangalore, L.B. Tjernberg, An Artificial neural network approach for early fault detection of gearbox bearing, IEEE Trans. on Smart Grid, 6(2): 980-987, 2015.
[9] J. Kennedy, R. Eberhart, Particle swarm optimization, in Proceeding of IEEE International Conference on Neural Network, 1995: 1942-1948.
[10] D. Toshkova, M. Asher, P. Hutchinson, N. Lieven, Automatic alarm setup using extreme value theory, Mechanical System and Signal Processing, 139:106417, 2020.