

Fan Fault Analysis Based on Time Domain Features and Improved *k*-means Clustering Algorithm

C.L.Shao & W.Lv

Hangzhou cigarette factory of China tobacco Zhejiang industrial Co., Ltd Hangzhou 310024, China.

Keywords: centrifugal fan, time domain features, improved *k*-means clustering algorithm.

Abstract. Aiming at the non-stationary and nonlinear characteristics of fan vibration signals, a method based on time domain signal analysis combined with the improved *k*-means clustering algorithm is proposed. In order to estimate the fault types, peak to peak values of several typical fan fault signals, Hurst exponent and approximate entropy have been extracted and put into the improved *k*-means clustering classifier as feature vectors. The experiments on the centrifugal fan show that: the selected three kinds of time-domain characteristics can reflect the difference between faults and the effect. Besides, the improved *k*-means clustering algorithm whose average recognition rate may come up to 88.67% has a better classification performance compared with the original *k*-means, and runs much more stably.

1. Introduction

The essence of equipment fault diagnosis is pattern recognition, which includes feature extraction and fault recognition. At present, most intelligent diagnostic methods are characterized by spectral band energy, using neural networks to recognize fault. This mode has the following problems. (1) Since the Fourier spectrum does not reflect any time-domain information, it is only valid for stationary signal analysis, However, due to the fluctuation of the grid voltage and the non-linearity of the equipment itself, the wind buffeting signal usually shows non-stationary. (2) Using neural network to diagnose the bearing fault has been a good result, however, neural network for fault identification requires a large number of data samples, which in reality is often difficult to meet, and the generalization of neural network capacity is not good^[1]. In this paper, an intelligent diagnosis method based on time-domain characteristic parameters and improved *k*-means clustering is proposed, which makes full use of the most direct time-domain information, pick-up the peak-to-peak value, Hurst exponent and approximate entropy of the chattering signal are extracted to characterize the characteristics of the non-stationary signal. *K*-means clustering is an unsupervised learning method, which can be used to classify samples according to the feature vectors of samples without knowledge of the sample category, and presents unique advantages and good application prospects in solving small sample problems, it has excellent generalization ability.

2. Experimental Apparatus And Test Methods

Test device using Siemens Y90S-2 fan, the maximum speed is 2900r / min, The wind pressure is 803 Pa, Air volume of 1830 m³ / h, power is 1.2kW, The voltage is 380V, Current is 3.4A. In order to facilitate the buffeting signal, compared with the actual fan system, the test device add two bearings between the fan and the motor, the rigid coupling is connected between the fan and the motor, centrifugal fan shaft vertical and horizontal directions were installed non-contact eddy

current displacement sensor to measure the radial displacement, plane - mounted acceleration sensor on fan housing. The vertical surface of the fan coupling is used as the test surface, horizontal installation of non - contact eddy current displacement sensor to measure axial displacement, the system measured the rotor coupler misalignment buffeting acceleration signal. During the test, at the fan rated speed, keep the centrifugal fan inlet to adjust the threshold opening, So that the fan load is maintained at 80%, sampling frequency of 800Hz, the overall structure of the test system shown in Fig. 1. The first signal in Figure 2 for the normal operation of the fan signal, followed by the first 2 to 6 are unbalanced, misalignment, pedestal loosening, friction and bearing damage fault signal.

3. Feature Extraction of Buffeting Signal Based On Time Domain Analysis

3.1 Peak to peak

The peak-to-peak analysis of the waveform reflects the change in the local amplitude of the buffeting signal, in the experiment, the acceleration signal was used as the analysis signal, and acceleration is the rate of change in velocity per unit time, when the fan is running, the stress of the rotor under different fault conditions is different, its acceleration will be different, the peak-to-peak values reflected on the waveform will also differ, since the actual sampling of the original signal is not clear starting point, fan rotation cycle is not an integer multiple, it will cause poor comparability between the signal, and it is not conducive to the next step of fault diagnosis, so it is necessary to analyze the n-cycles to intercept the original signal of fan rotation, so it will reduce the influence of error on feature extraction. Define the chirp signal peak-to-peak value:

$$x_{pv} = \max(x_i) - \min(x_i) \quad (1)$$

Thereinto: x_{pv} is the buffeting signal peak-to-peak value, $\max(x_i)$ is the peak of the buffeting signal, $\min(x_i)$ is the buffeting signal trough value, the average peak-to-peak value of n-cycles is:

$$\overline{x_{pv}} = \sum_{i=1}^n x_{pv} / n \quad (2)$$

3.2 Hurst analysis calculation method

The fractal Brownian motion proposed by Hurst, a British hydraulic scientist, it is an analytical model that can reflect the nature of irregular motion on a wide range for natural objects, its numerical changes are very complex, continuous but not derivative, it is a non-stationary process, and it has similarity to the change of time and scale^[2]. A large number of experiments show that the fan fault buffeting signal has a non-stationary, so we can use the fractal Brownian motion to describe this signal. Fractal Brownian Motion (FBM) Incremental variance is:

$$M(B_H(t) - B_H(t_0)) = \sigma_0^2 |t - t_0|^{2H} \quad (3)$$

Thereinto: B_H is a function of fractional Brown, t is time, σ_0^2 is the sample variance at time t_0 , H is the Hurst index. The Hurst index determines the degree of irregularity of an FBM, and it describes the long-term correlation of stochastic processes. The author uses the R / S analysis method to calculate the Hurst index of the wind buffeting signal, the calculation process is as follows:

We assume that the chattering time sequence is $X(t)(t = 1, 2, \dots, T)$, use $X^*(t) = \sum_{u=1}^t X(u)$, and

then $R(t, f)$ is:

$$R(t, f) = \max \left| X^*(t+u) - X^*(t) - \frac{u}{f} [X^*(t+u) - X^*(t)] \right| \quad (4)$$

$$- \min \left| X^*(t+u) - X^*(t) - \frac{u}{f} [X^*(t+u) - X^*(t)] \right|$$

Thereinto: f is the delay time. $R(t, f) / S(t, f)$ and the delay time(f) have the following relationship:

$$E(R(t, f) / S(t, f)) \propto f^H \quad (5)$$

$E()$ denotes the averaging of different initial instants t at the same delay time (f), this eliminates the effect of end effects at different initial instants on the statistical calculations. The Hurst exponent can be obtained by plotting the A-relation and regressed the slope of the straight-line relationship.

3.3 Definition and Properties of Approximate Entropy

Approximate entropy uses a nonnegative number to represent the complexity of a time series, the more complex time series corresponding to the larger the approximate entropy^[3]. Assume that the raw data collected is $\{u(i), i = 0, 1, \dots, n\}$, the value of the mode dimension(m) and the similar tolerance(r) is given in advance, then the approximate entropy can be calculated by the following steps:

1) The sequence $\{u(i)\}$ is composed in order of the m -dimensional vector($X(i)$), namely,

$$X(i) = [u(i), u(i+1), \dots, u(i+m-1)], (i = 1 \sim n - m + 1) \cdot$$

2) The distance between the vector $X(i)$ and the remaining vector $X(j)$ is calculated for each i value:

$$d[X(i), X(j)] = \max |u(i+k) - u(j+k)| \cdot$$

3) According to the given threshold r ($r > 0$), counted the number of $d[X(i), X(j)] < r$ and the ratio of this number to the total number of vectors for each i value, it is $C_i^m(r)$, namely,

$$C_i^m(r) = \{d[X(i), X(j)] < r \text{ 的数目} \} / (n - m + 1).$$

4) Let $C_i^m(r)$ take the logarithm and find its average for all i , it is $H^m(r)$, namely,

$$H^m(r) = \frac{1}{n - m + 1} \sum_{i=1}^{n-m+1} \ln C_i^m(r) \quad (6)$$

5) Then $m + 1$, repeat the process of 1 to 4, obtained $H^{m+1}(r)$.

6) the approximate entropy of this sequence is $ApEn(m, r) = \lim_{N \rightarrow \infty} [H^m(r) - H^{m+1}(r)] \quad (7)$

In general, this limit exists at probability 1. But in practical work n cannot be ∞ , when n is a finite value, according to the above steps is the sequence length is n when the estimated value is:

$$ApEn(m, r, n) = H^m(r) - H^{m+1}(r) \quad (8)$$

The value of $ApEn$ is related to the value of m, r, n . According to engineering experience, usually take $m = 2$, $r = 0.2SD(u)$ (SD represents the standard deviation of the sequence $\{u(i)\}$), select n is

1600, the length of time is 2s, the fan turns 10 cycles of data points.

The approximate entropy calculation is actually the size of the probability of generating a new pattern in the time series when measuring the change in dimension. The greater the probability of generating a new pattern, the more complex the sequence, so in theory, approximate entropy can represent the irregularity (complexity) of the signal, the more complex the signal, the larger the approximate entropy. At the same time, the approximate entropy has a good ability of anti-noise, anti-wild^[4].

4. Improved K-means Clustering Algorithm

Suppose there are n samples with unknown labels (x_1, x_2, \dots, x_n) , how to eigenvector according to the sample, The samples were divided into k -classes, T_1, T_2, \dots, T_k . Assume that the number of samples

for class k is nk , then $n = \sum_{i=1}^k n_i$, the means of each T_k is m_1, m_2, \dots, m_k , then $m_k = \frac{1}{nk} \sum_{i=1}^{nk} x_i, k = 1, 2, \dots, k$.

k -means clustering is based on the sum of squares of errors, namely, the objective function for k -means clustering minimization is:

$$J = \sum_{k=1}^k \sum_{i=1}^{nk} \|x_i - m_k\|^2 \quad (9)$$

3.1 Theory of k-means algorithm

- 1) The k objects in the random sample are the initial cluster centers.
- 2) For other objects, they are classified into the most similar clusters according to their distance (similarity) with the selected k cluster centers, and the center of the clustering is recalculated.
- 3) If the clustering minimization of the objective function to achieve accuracy requirements, the cluster center does not move, the algorithms terminated, or go to step 2.

3.2 Improvement of k-means algorithm

Since the k -means algorithm is random for the initial clustering center, it is easy to fall into the local optimal value, which leads to the classification error, it is necessary to move the local clustering center to a position more favorable for classification. So we define the deformation error formula is:

$$I = S - N[d(\omega, x_0)]^2 \quad (10)$$

Thereinto, S is the sum of squares of the distances between all the objects in a cluster and the European space center, N is the number of objects belonging to this cluster, $d(\omega, x_0)$ is the distance from the center of this cluster to the center of the European Space x_0 .

We define $\Delta M = \Delta I - \Delta D$ as the cluster center movement criterion, thereinto: ΔI is the overall deformation error caused by moving out of the cluster center, and ΔD is the deformation error caused by the insertion of the new clustering center. When $\Delta M < 0$, clustering center movement can reduce the overall deformation error.

- 1) The k objects in the random sample are the initial cluster centers.
- 2) Each object in the training sample is classified into the nearest cluster and the cluster center is recalculated.
- 3) If the clustering minimization objective function meets the precision requirement, then the cluster center does not move, go to step 4.
- 4) According to the cluster center moving rule, If a cluster center can be moved to a better position

to reduce the overall deformation error, then it is moved to a better position, then go to step 2, otherwise stop.

5. Test Results And Analysis

The buffeting signals of the wind turbines collected in the experiment under different operating conditions, the peak-to-peak values of the time-domain signals, Hurst exponents of the chaotic properties and approximate entropy data as shown Tab.1. In the experiment, six kinds of working conditions appeared in the operation of the fan, we extracted 300 samples, each condition is 50 samples, in each buffeting signal selected 30 groups, a total of 180 groups as a learning sample, the remaining 120 groups were used as test samples and classified by improved k-means clustering algorithm, when the total deformation error of the last two iterations is less than X , then the algorithm terminates. Where k is 6 and X is 10^{-4} , the data comparison of the k-means clustering algorithm before and after the improvement as shown Tab.2. It can be seen from the experimental results, the original k-means clustering algorithm is randomly selected cluster centers, it is easy to fall into local minimum, so the average recognition rate is not high, the improved k-means clustering algorithm improves the classification performance and improves the stability due to the steps of moving the local optimal clustering center, however, the complexity of the algorithm is higher than that of the original algorithm, so the recognition time is longer than before.

Tab.1 Characteristic Data of Partial Fan Vibration Signals

number	vibration type	peak to peak	Hurst index	Approximate entropy
1	normal	11.223	0.1533	0.5549
1	normal	13.152	0.1537	0.5632
2	unbalanced	35.014	0.1237	1.5520
2	unbalanced	34.112	0.1209	1.5504
3	misalignment	89.123	0.1513	4.6599
3	misalignment	90.763	0.1520	4.7643
4	base loosed	62.374	0.0843	2.8430
4	base loosed	58.323	0.0811	2.7455
5	friction	68.434	0.0651	3.3598
5	friction	67.076	0.0642	3.3223
6	bearing inner wear	37.764	0.1185	2.2376
6	bearing inner wear	37.534	0.0943	2.2143

Tab.2 Performance Comparison between Improved k - Means Clustering Algorithm and Original Algorithm

algorithm	classification time-consuming (s)	average recognition accuracy (%)
<i>k</i> -means clustering algorithm	3.5	79.56
improved <i>k</i> -means clustering algorithm	1.6	89.71

6. Conclusion

- 1) The peak-to-peak value, Hurst exponent and approximate entropy of the wind buffeting signal well reflect the non-stationary and complexity, it is a valid time-domain signal recognition metric.
- 2) The improved k-means clustering algorithm is simple to implement and can overcome the problem of local minimization caused by randomly selecting initial clustering centers without long

training, however, because of its higher complexity, the classification time will be longer.

3) Experiments show that the fault diagnosis method based on the time-domain hybrid feature and the improved k-means clustering algorithm is reliable. It must be pointed out that the above test is obtained in the case of small samples, and how to improve its classification stability and accuracy in large sample cases is the key to future research.

Reference

- [1] Li Meng, Lu Shuang, Ma Wenxing . Fractal characteristics of the rolling bearing fault diagnosis [J]. Transactions of the Chinese Society of Agricultural Machinery, 2005, 36(12):162-164.
- [2] Zhou Yunlong , Liu Chang xin, Zhao Peng, et al. Fault diagnosis methods for centrifugal pump based on autoregressive and continuous hidden Markov model[J]. Proceedings of the CSEE, 2008, 28 (20):88-93.
- [3] Qi Ziyuan, Xu Zhang sui. Application of approximate entropy in fault diagnosis of engine [J]. Journal of Ordnance Engineering College, 2008, 20 (2):39-42.
- [4] Xu Yonggang , Li Ling jun, He Zheng jia. Approximate entropy and its applications in mechanical fault diagnosis [J]. Information and Control, 2002, 31(6):547-551.