Research on Condition Evaluation of Gearbox Based on PCA-FCM

Haitao Wang¹, Sixia Fan¹, Bin Wu¹*

¹ School of Business, Shanghai Dianji University, Shanghai, China
*Corresponding author’s e-mail: wub@sdju.edu.cn

Abstract. In order to evaluate the state of the gearbox, this paper proposes to use PCA-FCM algorithm to divide the state of the gearbox. By decomposing the vibration signal of the gearbox and taking the principal component, the state of the gearbox is divided. At the same time, this paper considers the fuzzy partition and geometric structure partition of the data set to determine the classification, which is verified by the wind turbine gearbox simulation experiment platform. The experimental results show that the gearbox state assessment based on PCA-FCM algorithm can effectively divide and identify the operation state of the gearbox, which makes the operation state division of the gearbox more scientific, and provides a theoretical basis for subsequent fault diagnosis and maintenance strategy.

1. Introduction

China started to develop wind power as early as the sixth five year plan (1981-1985). After 40 years of development, wind power has gradually become one of the important energy sources in China. According to statistics, gearbox is the most vulnerable component in the operation of wind farm, and its failure rate is as high as 60%[1]. The state assessment of gearbox can understand the current operation state of gearbox, and provide the basis for fault diagnosis and maintenance strategy formulation of gearbox, so it is of great significance to carry out state assessment. Xing[2] divides the degree of deterioration of gearboxes into three categories, and builds a gearbox deterioration model through PCA-NAR for state assessment to predict the degree of deterioration and change trend of the gearbox. Zhai[3] uses AE technology to collect signals on the gearbox, and uses SVM to identify the normal and fault states of the gearbox. Dong[4] uses k-means to identify the operating conditions of wind turbines, and uses Gaussian mixture models to evaluate the health status of different operating conditions in real time. Zhang[5] composes an observable sequence of signals such as vibration and temperature, and divides the health status of the fan into four states by sensing the sequence. It can be seen that the current state evaluation research is generally to judge whether the gearbox is in fault state directly, but the research on the division of operation state is still a little insufficient, and many of them are divided according to engineering experience, which has no scientific basis and strong subjectivity.

In this paper, the vibration signal of gearbox is processed in time and frequency domain, and then the corresponding feature quantity is extracted, and the principal component analysis (PCA) is used to reduce the dimension. The feature quantity which can represent the vibration signal is selected, and the fuzzy C clustering (FCM) is used for the selected feature quantity, and the evaluation index is constructed to evaluate the clustering results. Finally, the simulation experiment platform of fan gearbox is built the data are verified by experiments.
2. Fuzzy clustering model of wind turbine gearbox

2.1. Vibration signal processing and PCA dimension reduction

The gearbox of wind turbine will produce a large number of vibration signals during operation, which hide a lot of information. The most commonly used feature extraction methods are time domain analysis and frequency domain analysis. For a group of vibration signals, the commonly used time-domain processing methods include maximum value, minimum value, average value, peak to peak value, absolute average value, variance, standard deviation, kurtosis, root mean square, waveform factor, peak factor, pulse factor, etc. Common frequency domain processing methods include barycenter frequency, frequency variance, frequency standard deviation, mean square frequency, etc. Since a signal is decomposed into several data during signal processing and feature decomposition of the vibration signal, there are too many input signals that need to be processed in the cluster analysis of the gearbox, and too many input signals will cause the FCM to fall into the local extreme. So this paper proposes to use PCA to reduce the dimensionality of the vibration signal after the feature is extracted to reduce the input signal of the cluster. The process is shown in Figure 1.

![Diagram of Gearbox condition assessment process based on PCA-FCM](image-url)

Figure 1. Gearbox condition assessment process based on PCA-FCM
2.2. Cluster evaluation index

At present, the evaluation indicators of clustering can generally be divided into two categories: fuzzy partition based on data set and geometric structure based on data set. The representatives of the two categories are partition entropy and XB coefficient. The partition entropy $V_{pe}$ indicates the amount of information carried by the clustering result. The smaller the value of $V_{pe}$, the better the result of clustering; the numerator of the XB coefficient is the degree of closeness between each type of data in the data set, and the denominator represents the data set class and class The degree of separation between the two, the smaller the value of $V_{xb}$, the better the result of clustering.

$$V_{pe} = -\sum_{i=1}^{c} \sum_{k=1}^{K} u_{ki} \log(u_{ki}) / n$$

$$V_{xb} = \frac{\sum_{i=1}^{c} \sum_{j=1}^{J} U_{ij}^{m} \|x_{ij} - v_{i}\|^2}{n \cdot \min_{j} \|v_{i} - v_{j}\|^2}$$

When evaluating the effectiveness of clustering results, it is difficult to reflect the effectiveness of clustering by using only a certain index, so the evaluation is often performed by considering both. In order to comprehensively consider fuzzy partition and geometric structure, this paper combines partition entropy and XB coefficient to obtain a new validity evaluation function $V_{eb}$. The first part of $V_{eb}$ considers the fuzzy division between the data sets for evaluation, and the second part considers the geometric structure of the data sets for evaluation. When the value of $V_{eb}$ is smaller, the clustering effect is better. According to the verification of literature [6], Exponentiation can reduce the influence of noise data and offset data on the evaluation function, making the evaluation function more robust.

$$V_{eb} = -\sum_{i=1}^{c} \sum_{j=1}^{J} u_{ij} \log(u_{ij}) / n - \sum_{i=1}^{c} \sum_{j=1}^{J} U_{ij}^{m} \exp \left( -\frac{\|v_{i} - x_{ij}\|^2}{\min_{j} \|v_{i} - v_{j}\|^2} \right)$$

3. Experimental results and verification

3.1. Experimental platform construction

In order to collect the operating data of the fan gearbox to classify and evaluate its status, this paper builds a fan gearbox simulation experiment platform. The experiment platform is mainly composed of electromagnetic speed-regulating three-phase motors, frequency converters, vibration signal acquisition cards, temperature signal acquisition cards, gear box and magnetic powder brake.

![Figure 2. Simulation experiment platform for fan gearbox](image)
3.2. PCA-FCM algorithm implementation

3.2.1. Signal noise reduction

After sorting out the data collected during the experiment, 6400 valid data were obtained after screening them. Considering that the vibration signal will inevitably be interfered by the surrounding noise during the acquisition process, it is necessary to reduce the noise signal before using the noise signal. In this paper, the wavelet soft threshold noise reduction is used to reduce the noise of the vibration signal. The noisy signals are shown in Figure 3 (a) and (b) respectively. Comparing the two images, it can be observed that the data after wavelet soft threshold denoising basically retains the previous amplitude, but it is smoother than the original signal, which can achieve a good noise reduction effect.

![Figure 3. Vibration data](image)

3.2.2. Pivot selection

After performing wavelet soft threshold denoising on the vibration signals in the normal state, 400 vibration signals are selected for principal component analysis and the principal components of the cluster are selected. Calculate 12 time domain signals and 4 frequency domain signals to form a sample matrix $A_{10*16}$ under normal conditions. According to the principal component analysis method, the variance contribution rate (PV) and the cumulative principal component contribution rate (CPV) are shown in Table 1. It can be seen from the table that the first five main elements can make the cumulative contribution rate reach more than 95%, so the first five main elements can be selected as the input signal of fuzzy C clustering.

| Number of pivots | 1   | 2   | 3    | 4     | 5     | 6     |
|------------------|-----|-----|------|-------|-------|-------|
| PV(%)            | 41.96 | 25.03 | 12.75 | 11.31 | 6.06  | 1.41  |
| CPV(%)           | 41.96 | 66.99 | 79.74 | 91.05 | 97.11 | 98.52 |

3.2.3. FCM results and evaluation

From the collected gearbox vibration signals, select 4800 normal status and 1600 fault status, a total of 6400 vibration data, and calculate the input matrix $A_{320*5}$ from the 5 principal elements obtained in the previous step of dimensionality reduction. Perform clustering according to the PCA-FCM flow chart and obtain the evaluation index $V_{eb}$, as shown in Table 2.
Table 2. Clustering results and evaluation indicators.

| Number of clusters | 3   | 4   | 5   | 6   | 7   |
|--------------------|-----|-----|-----|-----|-----|
| Number of samples  |     |     |     |     |     |
| category 1         | 190 | 55  | 49  | 50  | 105 |
| category 2         | 84  | 38  | 27  | 76  | 14  |
| category 3         | 46  | 153 | 79  | 37  | 25  |
| category 4         | -   | 74  | 120 | 26  | 34  |
| category 5         | -   | -   | 45  | 15  | 34  |
| category 6         | -   | -   | -   | 116 | 38  |
| category 7         | -   | -   | -   | -   | 70  |
| $V_{eb}$           | -0.4884 | -0.3068 | -0.2107 | -0.2124 | -0.1884 |

It can be seen from the table that when the number of clusters is 3, the evaluation index $V_{eb}$ takes the minimum value. At this time, the number of various types of data is basically consistent with the collected data, so when the running state of the gearbox is divided into three, it is most appropriate to define these three states as "normal", "deteriorated", and "faulty" respectively. This classification is similar to that of "good", "qualified" and "alarm"[3], which is more reasonable than that of "normal" and "fault" in most cases, and more concise than that of "normal", "abnormal", "deterioration" and "fault"[5]. It reduces the possibility that too many state categories lead to cumbersome evaluation criteria, and it is difficult to judge the specific state.

4. Conclusion

In this paper, PCA is used to reduce the dimensionality of the time-frequency domain feature extraction of the gearbox vibration signal, which effectively reduces the dimensionality of the input data used for FCM and reduces the complexity of clustering. At the same time, the operation status of the gearbox was successfully divided through the PCA-FCM algorithm and the evaluation index $V_{eb}$. Finally, the effectiveness is verified by experiments, and the classification is more scientific and reasonable, which provides a theoretical basis for subsequent fault diagnosis and maintenance strategies.

Acknowledgments

Humanity and Social Science Youth foundation of Ministry of Education of China (Project No.20YJCZ027)

References

[1] Wang C, Li S, Zhang T.Y, Lu W. Research on Diagnosis Method and Early Warning Strategy of Abnormal Oil Temperature in Gearbox of Wind Turbine [J]. Ship Engineering,2020,42(S1):542-546+549.
[2] Xing Y.S. Evaluation and Prediction of Deterioration Trend of Wind Turbine Gearbox Based on PCA-NAR Neural Network [D]. Southwest Jiaotong University, 2019.

[3] Zhai Y.M. Research on Degradation State Evaluation of Key Components of Wind Turbine Transmission System [D]. Southeast University, 2019.

[4] Dong Y.L, Li Y.Q, Cao H.B, He C.B, Gu Y.J. Real-time evaluation method of wind turbine health status based on operating condition identification [J]. Chinese Society for Electrical Engineering, 2013, 33(11): 88-95+15.

[5] Zhang X. Wind turbine state reliability assessment and maintenance strategy optimization [D]. Huazhong University of Science and Technology, 2019.

[6] Zhang Y.X, Liu T, Dong X, Li S. Fuzzy clustering validity function based on improved partition coefficient [J]. Journal of Shenyang University of Technology, 2014, 36(04): 431-435.