Vibration Signal Analysis of Rotating Machinery Based on Manifold Learning

Hao Wang¹, Yuyun Zeng¹, Guangyao Xie¹, Xiaoshang Wang² and Jingquan Liu¹*

¹Department of Engineering Physics, Tsinghua University, Beijing, 100084, China
²Qingdao LNG Co. Ltd., Sinopec, Qingdao, 266400, China
*Corresponding author’s e-mail: jingquan@tsinghua.edu.cn

Abstract. Vibration signal contains the fault information of rotating machinery, so it can be selected to monitor the health status of rotating machinery and provide an early warning. But vibration data of rotating machinery usually has the characteristics of large amount of data and non-linearity, and traditional data processing methods are difficult to analyze effectively. Therefore, in this paper, based on the statistics and wavelet packet decomposition, the time-frequency domain characteristics of the signal are extracted. And then the feature set that can characterize the running state of the rotating device is selected according to the Pearson correlation coefficient. At last, the manifold learning method is used to calculate the feature set. The result shows that this method can effectively monitor the running state of the rotating machinery.

1. Introduction
In today's society, rotating machinery have a large number of applications in industrial production. At the same time, with the development of science technology, rotating machinery is more high-speed, high-efficiency and more intellective, so it is necessary to carry out monitor and diagnosis to ensure their healthy and stable operation. Thanks to the progress of sensor technology and the development of big data technology, the types and quantity of data that can be used to analyse the status of equipment are increasing, such as temperature, speed, vibration and flow, etc [1]. Among them, vibration signal processing method has always been a research hotspot, because it is very important for equipment state monitoring and fault diagnosis. The amplitude range analysis method and the wavelet analysis method can effectively process the vibration signal [2]. Wavelet packet decomposition technology can divide the signal frequency band and calculate the energy distribution. This is an effective method to extract signal features [3]. A model based on wavelet packet sample entropy and manifold learning can effectively extract fault features of nonlinear vibration signals [4].

In this paper, a fault feature processing method based on the local linear embedding method in manifold learning is applied to the running state monitoring of rotating machinery. And the validity of this method is verified in the practical application of high-pressure pump.

2. Time-frequency domain feature calculation of vibration signals
The fault feature set of rotating equipment is the key step of state evaluation. The quality of feature set is directly related to the accuracy of evaluation results. The vibration signal of rotating equipment is a kind of non-stationary signal. When there is a fault, the time-domain and frequency-domain
characteristic statistical parameters of the vibration signal will change to some extent. Therefore, this paper selects the time-domain and frequency-domain characteristics to form the fault feature set.

2.1. Data processing
The data used in this paper is the vibration data of high-pressure pump in Qingdao LNG receiving terminal of Sinopec. There are four high-pressure pumps in this receiving terminal. The type of failure is the gradual wear of bearings. The system records every 10 minutes and the sampling frequency is 10 kHz. The vibration data is stored in hexadecimal form, so it cannot be analysed directly. Therefore, the first step is to process the data and convert it to decimal for subsequent analysis.

2.2. Feature extraction in time domain
Time domain features can reflect the overall operation of rotating machinery. In general, the commonly used time-domain features include mean value, standard deviation, peak value, root mean square value, waveform factor, peak factor, pulse factor, margin factor, skewness and kurtosis. Different features can represent different fault forms. Time domain features and their interpretations are shown in Table 1.

| Feature   | Interpretation                                      |
|-----------|----------------------------------------------------|
| Mean      | Balance point position of vibration                |
| Std       | Degree of data dispersion                          |
| Skewness  | Asymmetry of vibration signal                      |
| Peak      | Degree of deviation from average                   |
| Kurtosis  | Regularity of amplitude                            |
| Rms       | Bearing surface corrugation and other defects      |
| Crest     | Whether the waveform has impact                     |
| Shape     | Damage type of rolling bearing                     |
| Impulse   | Frequency of vibration pulse                       |
| Margin    | Wear of mechanical equipment.                      |

2.3. Feature extraction in frequency domain
There are many vibration sources in the rotating equipment, and the vibration signals recorded by the sensors are disordered. Here we choose wavelet packet transform to analyse the signal in frequency domain. And the energy of each frequency range after signal decomposition is calculated as the frequency domain feature.

Wavelet packet decomposition is developed from wavelet transform. Wavelet transform only decomposes the low-frequency part of the signal, but no longer decomposes the high-frequency part, that is, the detail part of the signal. Therefore, wavelet transform can well represent a large class of signals with low-frequency information as the main component, and cannot well decompose and represent the signals containing a large number of small edges or textures, such as the non-stationary mechanical vibration signals used in this paper. The wavelet packet transform can divide the frequency band into several levels, decompose the high frequency part which is not subdivided by wavelet transform, and select the corresponding frequency band adaptively according to the characteristics of the signal, so as to make better time-frequency localization analysis for the signal which contains a lot of medium and high frequency information. After decomposing the signal by wavelet packet and sorting by frequency band, the frequency range of each frequency band is as shown in Formula 1.
\[ \left[ \frac{k-1}{2^n}, \frac{k}{2^n} \right) F_N \]  

\( n \) is the number of wavelet packet decomposition layers, \( k \) is the node, \( F_N \) is the signal frequency. Assuming that \( f \) is the original signal, formula 2 can be obtained according to the law of conservation of energy.

\[ \langle f(t), f(t) \rangle = \int f(t)^2 dt = \int a^{-2} da \int W_f(a,b)^2 db \]  

(2)

According to the limitation of Heisenberg uncertainty principle, \( W_f(a,b)^2 / a^2 \) cannot be regarded as instantaneous energy density, but \( W_f(a,b)^2 / a^2 \) can be regarded as density function on \( (a, b) \) plane, that is, \( W_f(a,b)^2 \Delta a \Delta b / a^2 \) gives the energy of scale interval \( \Delta a \) and time interval \( \Delta b \) centered on scale \( a \) and time \( b \). We can get formula 3 and formula 4.

\[ \int f(t)^2 dt = \int E(b) db \]  

(3)

\[ E(b) = \int W_f(a,b)^2 / a^2 da \]  

(4)

After wavelet packet decomposition, the width of each band is constant. According to formula 4, the energy of each band can be obtained, as shown in formula 5. \( W_k(i) \) is the reconstruction coefficient of each node.

\[ E_k = \sum_{i=1}^{2^k} |W_k(i)|^2 \]  

(5)

According to the characteristics of vibration signal, this paper decomposes the signal into three layers of wavelet packet, and the wavelet function is dmey. The frequency bands and their Fourier transform of the vibration signal after wavelet packet decomposition are shown in Figure 1.

![Wavelet packet decomposition of vibration signal](image)

Figure 1. Wavelet packet decomposition of vibration signal

Because the energy difference between different frequency bands is too large, it is necessary to normalize the energy. The ratio of the energy of each frequency band divided by the total energy is taken as the frequency-domain feature of the signal. Thus, the feature is normalized to (0,1).
3. Feature select based on Pearson correlation coefficient

We have calculated many characteristics of vibration signal in time and frequency domain. However, some features have little to do with the running state of the equipment, so we need to select some features that are most relevant to the running state to construct feature sets. If the feature is not selected, on the one hand, the amount of calculation will be large, on the other hand, the irrelevant features will also cause interference to the calculation results.

In mathematics, covariance is usually used to calculate the relationship between two variables. However, because the values of different features are quite different, the size of covariance cannot accurately compare the size of correlation. Therefore, Pearson correlation coefficient is used here to compare the correlation between different characteristics and equipment operation status. Here we use the magnitude of vibration amplitude to judge the operation of the equipment. The amplitude increases, indicating that the equipment condition becomes worse. The correlation coefficient of each feature and vibration amplitude is shown in Table 2.

| Feature     | Correlation coefficient |
|-------------|-------------------------|
| Mean        | 0.2582                  |
| Std         | 0.9867                  |
| Skewness    | 0.1208                  |
| Peak        | 1.0000                  |
| Kurtosis    | 0.1427                  |
| Rms         | 0.9867                  |
| Crest       | 0.5423                  |
| Shape       | 0.3772                  |
| Impulse     | 0.4574                  |
| Margin      | 0.5062                  |
| Energy0     | 0.5658                  |
| Energy1     | 0.7003                  |
| Energy2     | 0.7213                  |
| Energy3     | 0.5042                  |
| Energy4     | 0.5997                  |
| Energy5     | 0.6956                  |
| Energy6     | 0.0102                  |
| Energy7     | 0.0961                  |

According to the Pearson correlation coefficient and the actual size of each feature, we choose peak value, root mean square value, standard deviation and crest factor as the time-domain features, and choose the energy ratio of the 1st, 2nd, 4th and 5th frequency bands as the frequency-domain features. Four time-domain features and four frequency-domain features together constitute the eigenvector set of the high-pressure pump.

4. State monitoring based on Manifold Learning

Eight kinds of time-frequency features are included in the feature vector set. In the face of an 8-dimensional data set, we cannot directly observe the changes of data to determine the operation status of the equipment. Therefore, further processing of 8-dimensional data is needed. Manifold learning is a good way to deal with high dimensional data.
Manifold learning believes that the high-dimensional data we observe is actually the mapping of low-dimensional data. Analysing the structural characteristics of data, high-dimensional data is redundant in dimension, that is to say, lower dimensional data can represent the characteristics of high-dimensional data. Manifold learning has the properties of Euclidean space locally, and can calculate the distance with Euclidean space. If the low dimensional manifold can be embedded into high-dimensional space, although the data is complex in the high-dimensional space, it has the properties of Euclidean space locally. Therefore, the mapping relation can be established locally, and then the local mapping relation can be generalized to the global situation. When the data is reduced to the familiar low dimensional space, it can be visualized. There are two methods of manifold learning: Isometric Mapping (Isomap) and local linear embedding (LLE). Considering the huge amount of data, we choose the LLE method with small amount of calculation.

Local linear embedding is just trying to maintain the relationship between the data in the neighbourhood. When the data is mapped from high-dimensional space to low-dimensional space, the linear relationship between the data in the neighbourhood remains unchanged. Here are the steps of the LLE method.

First of all, the data of feature vector set \( X \) should be standardized. This paper chooses the Z-score method. After processing, the mean value of the data is 0 and the standard deviation is 1.

Then the domain reconstruction coefficient \( \omega \) of all data is calculated according to the neighbourhood relationship, that is to find out the linear relationship between each data and other data in the neighbourhood. As shown in formula 6.

\[
\min \sum_{i=1}^{m} \left\| x_i - \sum_{j \in Q} \omega_{ij} x_j \right\|_2^2
\]

\( x_i \) and \( x_j \) are known. \( C_{jk} = (x_i - x_j)^T (x_i - x_k) \). We can get formula 7.

\[
\omega_{ij} = \frac{\sum_{k \in Q} C_{jk} c_{ik}^{-1}}{\sum_{l \in Q} C_{il} c_{ls}^{-1}}
\]

According to the constant coefficient of neighbourhood reconstruction, the coordinates of each sample in low dimensional space can be obtained. As shown in formula 8.

\[
\min \sum_{i=1}^{m} \left\| z_i - \sum_{j \in Q} \omega_{ij} z_j \right\|_2^2
\]

Let \( Z = (z_1, z_2, \cdots, z_m) \in \mathbb{R}^{d' \times m}, (W)_{ij} = \omega_{ij} \), we can get formula 9.

\[
M = (I - W)^T (I - W)
\]

After the M matrix is obtained, the problem becomes the decomposition of M. Then the feature vector corresponding to the minimum feature value is taken to form the low-dimensional coordinate of the data.

In this paper, the 8-dimensional eigenvector set are processed by manifold learning, and 1-dimensional output is obtained. In order to show the running state of mechanical equipment more clearly, the output results are zoomed to (0,1). The result shows that the output can well show the running state of the equipment. As shown in Figure 2. The larger the value is, the better the health status of the equipment is, and the smaller the value is, the greater the possibility of equipment failure.

According to the equipment maintenance records of the factory, the high-pressure pump appeared an obvious fault on April 4, and then was shut down for overhaul. Due to the lack of preparation for response measures, the unplanned shutdown caused considerable losses. From the condition monitoring curve, we can see that in the first few months of shutdown, the health status of high-pressure pump has a significant decline. According to the operation state curve obtained by the method in this paper, preventive maintenance strategy can be adopted to reduce the loss. Therefore, the method proposed in this paper can be used to evaluate the health status of the equipment and give the indication information before the equipment fails.
5. Conclusion
In this paper, the vibration signal of high-pressure pump of typical rotating mechanical equipment is extracted. Then the time-frequency characteristics of the vibration signal are calculated. Through Pearson correlation coefficient, the feature set of the high-pressure pump which is closely related to the operation state of equipment is constructed. The feature set is processed by the local linear embedding method in manifold learning, and the intuitive one-dimensional data is obtained. The results show that this method has a small amount of calculation and can monitor the running state of equipment in time and reliably.

Acknowledgments
This study is sponsored by the project “Research on health assessment and maintenance strategy for critical equipment in LNG terminals” of Qingdao LNG Co. Ltd., Sinopec.

References
[1] Chu, D. (2017) Research on Fault Signal Analysis and Diagnosis Technology for Rotating Machinery. D. North China Electric Power University. Beijing, China.
[2] Li, S., Guo, H., Li, D. (2013) Review of vibration signal processing methods. J. Chinese Journal of Scientific Instrument., 34: 1907-1915.
[3] Chen, W., Xie, B., Long, Z., et al. (2016) Stage Identification in Air-gap Discharge of Oil-impregnated Paper Insulation Based on Wavelet Packet Energy Entropy. J. Proceedings of the CSEE., 36: 563-569.
[4] Xiang, D., Ge, S. (2014) A model of fault feature extraction based on wavelet packet sample entropy and manifold learning. J. Journal of Vibration and Shock., 33: 1-5.