Diesel Engine Bearing Fault Diagnosis Based on Underdetermined Blind Source Separation

Hong-bo FAN*, Yi-quan SUN, Ming LIU and Jin-peng LI

Seventh Department, Shijiazhuang Campus of Army Engineering University, Shijiazhuang, China

*Corresponding author

Keywords: Vibration signal, Underdetermined blind source separation, Phase space reconstruction, Singular value decomposition, Bearing fault diagnosis.

Abstract. The method of diesel engine bearing fault diagnosis based on underdetermined blind source separation was proposed. The phase space reconstruction matrix of the vibration signal was decomposed with singular value decomposition, and reconstruction order number was determined by the method of dynamic clustering in order to construct virtual observed signal, so the blind source separation was changed from underdetermined to determined or over-determined. Then, the virtual observed signal and the original signal were composed into propagation paths, and the signal containing bearing fault characteristics was separated from mixed signal by the method of blind source separation which was based on the algorithm of adaptive PARAFAC. The simulation and test results show that this method effectively isolates the source signal.

Introduction

Due to various sources of diesel engine vibration signal, the vibration signals which reflecting the failure of bearing are usually drowned among other vibration signals in the diesel engine bearing fault diagnosis. A new method to extract the bearing fault feature information from the mixed signals is provided by blind source separation technique. In the blind source separation, it is often assumed that the number of source signals is not less than the number of observed signals. But in fact, since the sensor installation is not so convenient, or source signal is seriously interference, this assumption is usually not true. For underdetermined blind source separation, existing research is basically based on sparse representation of the source signal, with the aid of clustering algorithm [1].

However, due to the diesel engine vibration signal is consisted of more than one local oscillator signal in an unknown form of aliasing, for the diesel engine vibration signal, the assumption that the source signal is sparse is not necessarily true [2].

To solve those problems, the method of diesel engine bearing fault diagnosis based on underdetermined blind source separation is proposed. The virtual observed signals were constructed by using the method of phase space reconstruction and singular value decomposition, so the blind source separation is changed from underdetermined to determined or over-determined.

Approach of Getting the Virtual Observation Signals

Principle of Phase Space Reconstruction

\( x(i), i = 1, 2, \ldots, N \) is univariate time series, and an attractor trajectory matrix is obtained after the series’ phase space reconstruction. The embedding dimension and time delay of the matrix is \( p \) and \( \tau \) [3]:

\[
X = \begin{bmatrix}
x_1 & x_{1+r} & \cdots & x_{1+(q-1)r} \\
x_{1+r} & x_{1+2r} & \cdots & x_{1+qr} \\
\vdots & \vdots & & \vdots \\
x_{1+(p-1)r} & x_{1+pr} & \cdots & x_N 
\end{bmatrix}
\]

(1)

In the formula, \( q = N - (p - 1) \cdot \tau \), and \( p > q \).
In order to obtain a suitable embedding dimension, F. Takens proposed that the system constituted by the track of reconstruction is topologically equivalent to the original driving force system under the assumption that the data is unaffected by noise and its length is infinite, as long as the embedding dimension meet the conditions: \( p \geq 2u + 1 \), in which \( u \) is attractor dimension of the original driving force system. This theory is called the delay embedding theorem\(^4\).

**Singular Value Decomposition and Reconstruction Based on Dynamic Clustering**

**Principle of Singular Value Decomposition.** According to matrix theory, the singular value decomposition of the attractor trajectory matrix \( X \) can be calculated as the ninth formula.

\[
X = U \Lambda V^T
\]

In the formula, \( U \) is the characteristic matrix of \( X \) whose order is \( p \times p \), \( V \) is the characteristic matrix of \( X \) whose order is \( q \times q \), and \( \Lambda \) is a matrix whose order is \( p \times q \). \( \Lambda \) can be denoted as follows: \( \Lambda = (\text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_r)) \), \( \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > 0 \), \( \lambda_1, \lambda_2, \ldots, \lambda_r \) are all the singular values of \( X \).

Large singular values reflect the change rule of the signal. The reconstruction order \( k \) can be determined by selecting the values at the top of the singular value sequence. The matrix \( \hat{X} \) which is the estimate of the matrix \( X \) can be obtained by inverse transform of SVD. And a virtual observation signal can be obtained by calculating the average of all the elements of the matrix \( \hat{X} \).

**Method for Determining the Reconstruction Order.** Singular values reflecting the characteristics of attractor trajectory matrix \( X \) are a number of large amplitude ones, and the amplitude of singular values of the noise is relatively small. If the singular values are projected onto a coordinate axis, those reflecting the signal characteristics will be scattered on the coordinate axes, and the others which are relative to the noise signal will be concentrated. If clustering analysis is applied to all the singular values, the bigger ones will be clustering, and on the contrary, the smaller will be gathered together. So, it can be considered that the number of singular values which are separated is the best reconstruction order\(^5\). The method of k-means which is called hard clustering can be applied to clustering singular values dynamically.

The method of k-means can classify the samples of the sample space into \( K \) classes, \( K \) is a positive integer which is greater than or equal to one. It is assumed that \( N_i \) is the number of samples in \( \omega_i \) whose series number in the category space is \( i \), \( m_i \) is the average of all the samples, and it can be calculated by the following formula.

\[
m_i = \frac{1}{N_i} \sum_{x \in \omega_i} x
\]

Clustering criterion function is defined as follows:

\[
J_e = \sum_{i=1}^{K} \sum_{x \in \omega_i} \|x - m_i\|^2
\]

The value of \( J_e \) changes when the clustering is different, and the clustering result is the best one under the principle of the error sum of squares when the value of \( J_e \) becomes the minimum. The process of the method is as follows:

1. Divide the sample into \( k \) clustering regions, and compute \( m_1, m_2, \ldots, m_k, J_e \).
2. Choose any sample in the regions optionally which can be represented as \( x \in \omega_j \).
3. If \( N_i = 1 \), then recycle to (2); if \( N_i \neq 1 \), then recycle to (4).
4. Calculate the values of the following formulas.
\[ \rho_j = \begin{cases} \frac{N_j}{N_j + 1} \| x - m_j \|^2 & j \neq i \\ \frac{N_i}{N_i + 1} \| x - m_i \|^2 & j = i \end{cases} \]  \tag{5} 

(5) Calculate the value of \( \rho_j \) for each sample. If \( \rho_m \leq \rho_j \), \( x \) will be moved from \( \omega_i \) to \( \omega_m \).

(6) Calculate the values of \( m_i \) and \( m_M \) again, and modify \( J_e \).

(7) If the value of \( J_e \) remains unchanged after loop iteration, the clustering will stop, otherwise, recycle to (2).

**Underdetermined Blind Source Separation of Simulated Signals**

Introduce three simulated source signals as follows:

\[ s_1 = \sin(20\pi t^2), \ s_2 = \sin(20\pi t) \]  \tag{6}

\( s_3 \) is white Gaussian noise and \( t \in [0 \ 1s] \).

\( A \) is a mixed matrix:

\[
A = \begin{bmatrix} 0.23 & 0.65 & 0.39 \\ 0.51 & 0.58 & 0.35 \\ 0.15 & 0.36 & 0.95 \end{bmatrix}
\]  \tag{7}

\( Z \) is the observation signal:

\[
Z = A \begin{bmatrix} s_1 \\ s_2 \\ s_3 \end{bmatrix}^T
\]  \tag{8}

Fig.1 shows the time-domain distribution of each simulated source signal. Fig.2 shows the singular value distribution of the first observed signal. Fig.3 is the time-domain distribution of the observation signals; the first two are actual, and the last one is virtual. So the blind source separation is changed from underdetermined to determined or over-determined. The basic requirement that the number of observation signals is not less than the number of source signals is satisfied.

![Figure 1. Time-domain distribution of source signals.](image1)

![Figure 2. Singular value decomposition distribution of observation signal’s trajectory matrix.](image2)
Blind source separation is applied to the observation signals by using the algorithm of FastICA, and Fig.4 shows the separation signals. The result comparing Fig.4 and Fig.1 shows that the source signals separated from the observation signals effectively. The simulation result shows that it is feasible to obtain the virtual observation signal by using the method of phase space reconstruction based on instantaneous hybrid model.

![Figure 3. Instantaneous mixture observation signals.](image)

![Figure 4. Separation signals.](image)

Test Analyses

**Signal Acquisition Program of the Bearing Fault Vibration Signals**

Experiments are carried out on the EQ6BT diesel engine. The vibration sensors are installed at the junction of the oil pain and the cylinder which is just aligned at the fourth main bearing and at the oil pain. The working states of the crankshaft bearings are set as the table 1 in this test. The four installed sensors are nearest to the vibration source, so the measurement signal is sensitive to fit clearance changes.

<table>
<thead>
<tr>
<th>Fault settings</th>
<th>Normal</th>
<th>Minor fault</th>
<th>Moderate fault</th>
<th>Serious fault</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fit clearance between the third bearing and shaft neck</td>
<td>0.08-0.1</td>
<td>0.08-0.1</td>
<td>0.08-0.1</td>
<td>0.08-0.1</td>
</tr>
<tr>
<td>Fit clearance between the fourth bearing and shaft neck</td>
<td>0.08-0.1</td>
<td>0.24-0.26</td>
<td>0.38-0.4</td>
<td>0.53-0.55</td>
</tr>
<tr>
<td>Fit clearance between the fifth bearing and shaft neck</td>
<td>0.08-0.1</td>
<td>0.08-0.1</td>
<td>0.08-0.1</td>
<td>0.08-0.1</td>
</tr>
</tbody>
</table>

Bearing Fault Diagnosis

Fig.5 shows the observation signals acquired from that four measurement points when the crankshaft bearing stays normal, and Fig.6 shows their frequency spectrums. Fig.7 shows the change of the singular values of the observation signal. It is assumed that there are three cluster centers which divide the singular values into four sections which are respectively represented as 1, 2, 3 and 4 from high to low. The first reconstruction signal is reconstructed by the first section, the others and so on. Fig.8 is the frequency spectrum of each reconstruction signal.

Fig.8 shows that the energy of the normal observation signal mainly concentrates in the first and second reconstruction signals. Cut the timeline into slices to track the signal frequency changes at a time on the time frequency spectrum. Fig.9 shows the slice spectrums with maximum energy, and it demonstrates that the energy change of the first reconstruction signal is closer to the original signal on the frequency axis.

Fig.10 shows the slice spectrums with maximum energy when the crankshaft bearing stays minor fault status. And the energy change of the first reconstruction signal is closer to the original signal on the frequency axis.

A singular value decomposition matrix consisting of the first and second reconstruction signals is
decomposed, and the average and pulse index of the singular values make up the feature vector. The spatial distribution of the vector shows in Fig.11.

Figure 5. Observation signals in normal status of cranked bearing.

Figure 6. Time frequency spectrum of observation signals in normal status of cranked bearing.

Figure 7. Singular value decomposition distribution of normal status observation signal’s trajectory matrix.

Figure 8. Time frequency spectrum of the reconstruction signals in normal status of cranked bearing.
Figure 9. The slice spectrum of cranked bearing in normal status.

Figure 10. The slice spectrum of cranked bearing in slight fault status.

Figure 11 shows that the first and second reconstruction signals contain the information which is the most relevant to the source signal, but the feature values are still have some overlap, and this can be caused by the delay filter during signal transmission. Therefore, the virtual observation signal is reconstructed by using the reconstruction order of the first and second reconstruction signals, so that the blind source separation is changed from underdetermined to determined or over-determined. Then the delay filter on the transmission path is eliminated by using the method of PARAFAC, and the feature vectors of the separation signals in different status are shown in Fig.12.

Figure 11. Spatial distribution of singular values feature vectors of first and second reconstruction signals.
The diagnostic error $E_i$ is defined as follows in order to quantitatively describe the fault diagnosis accuracy.

$$E_i = \frac{m_i}{n_i}, i = 1, 2, 3, 4$$

The values of $i$ are respectively represented the normal, minor, moderate and serious fault status. $m_i$ is the number of error clustering points, and $n_i$ is the sample points. The overall diagnostic accuracy calculated by the first and second reconstruction signals is 81.6%, and the result calculated from the separation signal is 100%. So the diagnostic accuracy is increased by 18.4% after blind source separation.

Conclusions

(1) The reconstruction matrix is decomposed by the method of singular decomposition, and the virtue observation signal is constructed by using the method of dynamic clustering. So that the blind source separation has been changed from underdetermined to determined or over-determined.

(2) The virtual observed signal and the original signal are composed into propagation paths, and the blind source separation is applied to the simulation signal in which the number of observation signals is less than that of the source signals by using the method of adaptive PARAFAC. The simulation result shows that this method can effectively isolated the source signal, and when it is applied to the fault diagnosis of diesel engine crankshaft bearings, the diagnostic accuracy has been increased by 18.4%.

Reference


