Classification of gear faults using cumulants and the radial basis function network

Lai Wuxing\textsuperscript{a,}\textsuperscript{*}, Peter W. Tse\textsuperscript{b}, Zhang Guicai\textsuperscript{a}, Shi Tielin\textsuperscript{a}

\textsuperscript{a}Department of Mechtronics, School of Mechanical Science and Engineering, Huazhong University of Science and Technology, Wuhan 430074, People’s Republic of China

\textsuperscript{b}Department of Manufacturing Engineering and Engineering Management, City University of Hong Kong, Hong Kong, People’s Republic of China

Abstract

Every tooth in a gearbox is alternately meshing and detaching during its operation. Hence, the loading condition of the tooth is alternately changing. Such a condition will make the tooth easily subject to spalling and worn. Moreover, Gaussian type of noise which is always embedded in the measurements makes the signal-to-noise ratio (SNR) of the collected data low and difficult to extract in fault-related features. This paper aims to propose an approach for gear fault classification by using cumulants and the radial basis function (BRF) network. The use of cumulants can minimize Gaussian noise and increase the SNR. The RBF network has proven to be superior to back-propagation networks. The RBF network provides better functions to approximate non-linear inputs and faster in convergence. In this paper, experiments have been conducted on a real gearbox. The cumulants calculated from the vibration signal collected from the inspected gearbox are used as input features. The RBF network is then used as a classifier for various kinds of operating conditions of the gearbox. Results show that the method of classification by combining cumulants and the RBF network is promising and achieved better accuracy.

© 2003 Elsevier Ltd. All rights reserved.

Keywords: Artificial neural networks; Fault diagnosis; Cumulants; High-order statistical analysis

1. Introduction

In recent years, on-line and automatic types of fault detection and diagnostic systems have been gaining considerable amount of business potential. The need for automating industrial processes and optimizing the cost of maintenance have stimulated the research and development of faster and robust fault diagnosis. New methods include dynamic model-based methods, multivariate...
statistical analysis, and fuzzy logic and artificial neural networks based methods [1–5]. Since neural networks is a non-linear dynamics system and capable of mapping non-linear functions, they are widely applied in the field of pattern recognition and classification. Neural networks have been successfully developed for automatic and robust machine fault classification [10]. There are many types of neural networks, among them, the back-propagation (BP) network, which is a kind of multi-layer feedforward networks, is the most popular network used in engineering applications. Although many improved algorithms for BP network have been developed to increase the speed of training and avoid fall into local minimum during training, their effectiveness in solving these problems are not promising. Moreover, the determination of a suitable architecture for the BP network, such as the number of neurons in the hidden layer, is difficult for inexperienced users. The radial basis function (RBF) network offers a new and more effective method for training. Since it can avoid complicated and prolix calculations, the time required for training can be much faster than that of the BP network. Moreover, it has the ability of fast convergence and is able to automatically determine the number of neurons in the hidden layer during training. Hence, an optimized architecture for the RBF network can be obtained.

In the application for machine fault diagnosis, neural network is able to perform complicated non-linear mapping to identify different kinds of features to different types of faults. However, the process, which is used to determine vital features from the measured signal, must be completed prior to the use of neural networks for fault classification. For vibration-based machine fault diagnosis, when a fault starts to occur in a machine, its fault-related signal is usually weak and overwhelmed by noise and other structural vibrations. Moreover, the measured signal often embedded with noise and its signal-to-noise ratio (SNR) is low. Therefore, an effective preprocessing method must be developed to extract essential features. One of the suggested methods is to determine the collected signal falls within Gaussian distribution. If vibration signals are collected from a healthy machine, they will fall within the Gaussian distribution. Whereas, if the inspected machine is faulty, the signals will not be Gaussian distributed. In recent years, a number of tools have been developed based on the theory of high-order statistics to process non-Gaussian signals [6]. These tools can be used as preprocessing methods to determine vital features from measured signals.

The vibration signals generated from a faulty gearbox are usually non-stationary. Non-stationary signals can be analyzed by time–frequency methods [7,8] or time-scale (wavelets) methods [9]. On the other hand, statistical methods, such as high-order statistics, can also provide good representations to various fault signatures. It can be used as an effective preprocessing method to determine vital features from measured signals. This paper introduces the use of cumulants to preprocess the collected signals and extract the vital features. Then the extract features will act as inputs to the RBF network for classifying a gearbox’s operating conditions, which include normal, spalling and worn gear conditions. The results are promising.

2. Radial basis function neural network

The RBF neural network [11,12], is a feedforward network with its architecture as shown in Fig. 1. It consists of three layers: an input layer of \( R \) neurons, a hidden radial basis layer of \( S^1 \) neurons and an output linear layer of \( S^2 \) neurons. The information of the input neurons will
transfer to the neurons in the hidden layer. The BRF in the hidden layer will response to input information, and then generates the outputs in the neurons of the output layer. The advantage of the RBF network is that the hidden neurons will have non-zero response if the minimum of a function is in the pre-defined limited range of the input values, otherwise, the response will be zero. Therefore, the number of active neurons is smaller and the time required in training the network is less. Hence, the RBF network is also referred to as the local range network.

In the RBF neural network, the transfer function of the hidden layer is a Gaussian function

\[
A_j = \exp \left( -\frac{(p - c_j)^T(p - c_j)}{2\sigma_j^2} \right), \quad j = 1, 2, \ldots, S^1,
\]

where \(A_j\) is the output of the \(j\)th neuron in the hidden layer, \(p\) is the input mode, \(c_j\) is the center of the \(j\)th neuron Gaussian function. \(\sigma_j^2\) is the unitary parameter, and \(S^1\) is the neurons in hidden layer 1.

There are two steps in the training of the RBF network. In the first step, depending on the information contained in the input samples, the neurons in the hidden layer \(S^1\), the center of Gaussian function \(c_j\) and the unitary parameter \(\sigma_j^2\) will be determined. The most frequently used method to determine the Gaussian function is the K-means aggregation algorithm. Assume that the K-means aggregation algorithm aggregates the input samples, \(\theta_j\) are all the samples of the \(j\)th group, then the center of the \(j\)th aggregation \(c_j\) is given by the following equation:

\[
c_j = \frac{1}{M_j} \sum_{x \in \theta_j} x.
\]

And the \(j\)th unitary parameter \(\sigma_j^2\) is given by the following equation:

\[
\sigma_j^2 = \frac{1}{M_j} \sum_{x \in \theta_j} (x - c_j)^T(x - c_j).
\]

In Eq. (3), \(M_j\) is the mode of \(\theta_j\), and \(\sigma_j^2\) is the estimate of samples that distributed the center of the \(j\)th aggregation, \(c_j\).
In the second step, according to the parameter in the hidden layer, input samples and the target values, the weight $w_j$ will be determined and adjusted by the principle of least squares.

In comparison, both the RBF and BP networks can approximate any non-linear function in reasonable precision. However, because the RBF uses a different transfer function in the hidden layer than that of the BP network, the accuracy in approximating the input function is also different. Poggio and Girosi [13] have proven that the RBF network is superior in approximating continuous functions. Although sigmoid function commonly used in the BP network has better ability in generalization, its results will affect much more neurons in adjusting their weights. Moreover, the transfer function overlaps in a large range of the input values, so the outputs interreference each other. The RBF network increases the speed of training by using a local transfer function so that only a few neurons have a non-zero response and become active to each input value. Therefore when new data are adopted, only the weights of a few active neurons are required to be modified. In summary, the training of the BP network is slower and more difficult to be convergent than the RBF network. Moreover, the RBF network can avoid falling into local minimum when the training is in progress [14].

3. Fault feature extraction method

The vibration signal collected from a gearbox often has low SNR, especially when a fault is occurring in the gearbox and starts to propagate. Since the signal is weak and overwhelmed by noise, its fault-related features are difficult to be identified. Preprocessing must be performed to minimize the effect of noise and increase the SNR for better fault diagnosis performed by the post-processing RBF network. Conventional fast Fourier transforms (FFT) can reveal the average energy or power of the signal using the frequency spectrum. However, vibrations generated by large structural components and noises often cover the faulty vibration signal generated by the smaller gears. Therefore, the spectrum analysis provided by the FFT is not useful when the fault is at its early stage of propagation. Digital filter is another popular method to minimize noise. However, one must have prior knowledge of the existing noise and the interested range of frequency before the selection of frequency bands of the digital filter. Such prior knowledge is usually unknown making the use of digital filter unsatisfactory.

Another approach is to measure the Gaussian distribution of the collected signals. The signals may be disturbed according to Gaussian white and color noises with unknown frequencies. The development of high-order statistics in recent years has offered a promising method to minimize these noises. High-order statistics have been proven to be effective in detecting whether the signal has embedded Gaussian distributed noise, and muffling of both white and color Gaussian noises with unknown frequencies. Assuming that the interested signal has zero mean and is discrete in time, the second-, third- and fourth-order of moments of a zero-mean signal generated from a random process $x(n)$ can be found by using the following equations [15]:

$$m_{2x}(i) = E[x(n)x(n + i)],$$

(4)

$$m_{3x}(i,j) = E[x(n)x(n + i)x(n + j)],$$

(5)
and
\[ m_{4x}(i,j,k) = E[x(n)x(n+i)x(n+j)x(n+k)]. \] (6)

The second-, third- and fourth-order cumulants of the stationary process \( x(n) \) can also be defined as [16]
\[ c_{2x}(i) = m_{2x}(i) = E[x(n)x(n+i)], \] (7)
\[ c_{3x}(i,j) = m_{3x}(i,j) = E[x(n)x(n+i)x(n+j)], \] (8)
\[ c_{4x}(i,j,k) = m_{4x}(i,j,k) - m_{2x}(i)m_{2x}(j-k) - m_{2x}(j)m_{2x}(k-i) - m_{2x}(k)m_{2x}(i-j), \] (9)

where \( m_{2x}(i) = E[x(n)x(n+i)] \), and equals \( C_{2x}(i) \), for a real-valued process. The first-order cumulant is the mean of the process and the second-order cumulant is the autocovariance sequence. Note that for complex processes, there are several ways of defining cumulants depending upon which terms are conjugated. The zero-lag cumulants have special names: \( C_{2x}(0) \) is the variance and is usually denoted by \( \sigma^2 \); \( C_{3x}(0,0) \) and \( C_{4x}(0,0,0) \) are usually denoted by \( r_{3x} \) and \( r_{4x} \). These terms can be referred to as normalized quantities—the skewness and the kurtosis. These normalized quantities are both time-shifted and scale invariant. If \( x(n) \) is symmetrically distributed, its skewness is zero. If \( x(n) \) is Gaussian distributed, its kurtosis is zero. Often, the terms of skewness and kurtosis are used to refer to the unnormalized quantities, \( r_{3x} \) and \( r_{4x} \). If \( x(n) \) is an i.i.d. (higher-order white) process, its cumulants are non-zero only at the origin. If \( x(n) \) is statistically independent of \( y(n) \), and \( z(n) = x(n) + y(n) \), then
\[ c_{4z}(i,j,k) = c_{4x}(i,j,k) + c_{4y}(i,j,k) \] (10)

with similar relationships holding for cumulants of all orders. Eq. (10) provides a simplification in the cumulant-based analyses. With the help of these analyses, the effect of noise can be minimized by high-order cumulants. Therefore, the fault-related features are easier to be identified and extracted for further processes of fault classification using the RBF network.

4. Gear fault feature classification

Extensive experiments were conducted on a helical gear train having a gear ratio of 41:37 and a module of 5 mm as shown in Fig. 2. An accelerometer was installed to measure the vibration.

Fig. 2. Helical gear train.
signals generated on the axes of the 41 teeth pinion. Two phase references and appropriate pick-ups were used to ensure that the vibration signal was acquired when the same pairs of teeth were meshing. Different vibration data sets were collected when the helical gear train was working at normal, spalling, one worn tooth, and two worn teeth conditions. A total of 30 groups of data were collected for each operating condition with a sampling rate of 10 kHz. Fig. 3 shows a typical sample of the temporal raw signal when the gear train is in spalling condition.

Before calculating the cumulants of the temporal raw vibration signal \( x(t) \), the signal will be pretreated to have zero-mean and subjected to Hilbert transfer. The resultant signal \( z(t) \) is obtained by

\[
z(t) = x(t) + jH[x(t)],
\]

where \( H[x(t)] \) is the Hilbert transfer of \( x(t) \). Since modulation of signal often occurs in the vibration signals generated by gears and bearings, demodulation is required to reveal the true signals. Envelope analysis is a widely adopted method for signal demodulation. The enveloped signals \( Z(t) \) can be obtained by

\[
Z(t) = |z(t)| = \sqrt{x^2(t) + (H[x(t)])^2}.
\]

Experiments show that the third cumulant, the value of skewness, is a vital feature for determining the health of the gear train. Therefore, in this experiment, the third cumulants of the 30 groups of data collected for each operating condition of the gear train were calculated and compared. The results are shown in Fig. 4. The values of the third cumulants for the conditions of normal and spalling are close to each other, whilst, the cumulants for one worn tooth and two worn teeth are relatively distinctive.

In the process of fault classification using the RBF network, for each condition, 25 groups of data were used as training samples to train the RBF network. Another independently collected 5 groups of data were used as test samples. Table 1 shows the training input samples and the arranged outputs. The neurons 1, 2, 3, and 4 in the output layer correspond to the operating conditions of normal, spalling, one worn tooth, and two worn teeth, respectively. Table 2 shows the values of each output neuron for the 5 groups of data used for testing. For instance, in the test sample collected under normal condition, the output that corresponded to the normal condition is output neuron 1 as defined in Table 2. Hence, the values listed in the column for output neuron 1 are the highest as compared to the other three output neurons. All the values of the output neurons are matching with their corresponding operating conditions of the gear train. Hence, the accuracy in classification is satisfactory. Table 3 shows the performance comparison of the RBF

![Fig. 3. Temporal raw signal of spalling gear.](image-url)
network and the BP network with 3 hidden layers. It shows that the RBF network employing the K-means aggregation algorithm used only 1.1 s and 6 steps in training, whilst the BP network used 100 times more in time and a lot of steps in training, but still could not converge. Hence, the performance of the RBF network is significantly better than the BP network, particularly in the computational resource for training.

5. Conclusions

This paper presents an effective method for classifying machine faults that exhibits non-linear and noisy signals. By using the high-order statistics, various cumulants can be calculated from the raw signals and used as vital features for robust fault diagnosis. By using the RBF network, the accuracy in classification and the performance in training can be significantly enhanced. A real helical gear train was used to verify the effectiveness of the combined method of cumulant and the RBF network. The operating conditions of the gear train include normal, spalling, one worn tooth, and two worn teeth.

Table 1

<table>
<thead>
<tr>
<th>Operating condition of the training samples</th>
<th>Corresponding neuron in the output layer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>1 0 0 0</td>
</tr>
<tr>
<td>Fault 1 (spalling)</td>
<td>0 1 0 0</td>
</tr>
<tr>
<td>Fault 2 (one worn tooth)</td>
<td>0 0 1 0</td>
</tr>
<tr>
<td>Fault 3 (two worn teeth)</td>
<td>0 0 0 1</td>
</tr>
</tbody>
</table>

Fig. 4. The third cumulants of the normal, spalling, one worn tooth and two worn teeth conditions.
tooth and two worn teeth. The third cumulant had been identified as the vital feature for diagnosing the health of the gear train. Therefore, the values of the third cumulant for each operating condition were acting as inputs to the RBF network. The results show that the RBF had correctly classified all conditions, but using much less time and computational resource as compared to the BP network. Because of the simple process in calculating the third cumulant, and the fast training speed and the high performance of the RBF network, the machine fault diagnosis could be designed as a real-time and on-line system that is urgently needed by the industry. Moreover, the study shows that the calculation of cumulants and the construction of the RBF network could be fully automated in the future.

<table>
<thead>
<tr>
<th>Test samples</th>
<th>Neuron 1 output</th>
<th>Neuron 2 output</th>
<th>Neuron 3 output</th>
<th>Neuron 4 output</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>1.0450</td>
<td>-0.0332</td>
<td>0.0087</td>
<td>-0.0205</td>
</tr>
<tr>
<td></td>
<td>1.0396</td>
<td>-0.0274</td>
<td>0.0078</td>
<td>-0.0201</td>
</tr>
<tr>
<td></td>
<td>1.0063</td>
<td>0.0090</td>
<td>0.0024</td>
<td>-0.0177</td>
</tr>
<tr>
<td></td>
<td>1.005</td>
<td>0.0153</td>
<td>0.0015</td>
<td>-0.0173</td>
</tr>
<tr>
<td></td>
<td>0.9609</td>
<td>0.0582</td>
<td>-0.0045</td>
<td>-0.0146</td>
</tr>
<tr>
<td>Fault 1 (spalling)</td>
<td>-0.1538</td>
<td>0.9612</td>
<td>0.1744</td>
<td>0.0182</td>
</tr>
<tr>
<td></td>
<td>-0.1644</td>
<td>0.9592</td>
<td>0.1881</td>
<td>0.0170</td>
</tr>
<tr>
<td></td>
<td>-0.1647</td>
<td>0.9592</td>
<td>0.1885</td>
<td>0.0170</td>
</tr>
<tr>
<td></td>
<td>-0.1684</td>
<td>0.9583</td>
<td>0.1936</td>
<td>0.0165</td>
</tr>
<tr>
<td></td>
<td>-0.1793</td>
<td>0.9546</td>
<td>0.2096</td>
<td>0.0151</td>
</tr>
<tr>
<td>Fault 2 (one worn tooth)</td>
<td>0.0253</td>
<td>-0.0710</td>
<td>1.0913</td>
<td>-0.0457</td>
</tr>
<tr>
<td></td>
<td>0.0485</td>
<td>-0.1192</td>
<td>1.1078</td>
<td>-0.0372</td>
</tr>
<tr>
<td></td>
<td>0.0685</td>
<td>-0.1589</td>
<td>1.1155</td>
<td>-0.0252</td>
</tr>
<tr>
<td></td>
<td>0.0917</td>
<td>-0.2004</td>
<td>1.1059</td>
<td>0.0026</td>
</tr>
<tr>
<td></td>
<td>0.0947</td>
<td>-0.2048</td>
<td>1.1002</td>
<td>0.0098</td>
</tr>
<tr>
<td>Fault 3 (two worn teeth)</td>
<td>0.1771</td>
<td>-0.3547</td>
<td>0.0368</td>
<td>1.1408</td>
</tr>
<tr>
<td></td>
<td>0.2239</td>
<td>-0.4340</td>
<td>0.1239</td>
<td>1.0862</td>
</tr>
<tr>
<td></td>
<td>-0.1111</td>
<td>0.2113</td>
<td>-0.0749</td>
<td>0.9747</td>
</tr>
<tr>
<td></td>
<td>0.0955</td>
<td>-0.1788</td>
<td>0.0708</td>
<td>1.0125</td>
</tr>
<tr>
<td></td>
<td>0.0748</td>
<td>-0.1411</td>
<td>0.0494</td>
<td>1.01698</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Training method</th>
<th>Network type</th>
<th>Training time (s)</th>
<th>Training steps</th>
</tr>
</thead>
<tbody>
<tr>
<td>K-means aggregation algorithm</td>
<td>RBF network</td>
<td>1.10</td>
<td>6</td>
</tr>
<tr>
<td>Improved BP algorithm</td>
<td>BP network</td>
<td>110.12</td>
<td>1916</td>
</tr>
</tbody>
</table>
Acknowledgements

This paper was supported by the National Ascend Plan Project of the People Republic of China (No. PD9521908).

References