Application of statistics filter method and clustering analysis in fault diagnosis of roller bearings

L Y Song, H Q Wang, J J Gao, J F Yang, W B Liu and P Chen
1 School of Mechanical & Electrical Engineering, Beijing University of Chemical Technology, 15 Beisanhuan East Road, ChaoYang district, Beijing, 100029, P. R. China
2 Graduate School of Bioresources, Mie University, 1577 Kurimamachiya-cho, Tsu-shi, Mie, 514-8507, Japan

E-mail: wanghq_buct@hotmail.com

Abstract. Condition diagnosis of roller bearings depends largely on the feature analysis of vibration signals. Spectrum statistics filter (SSF) method could adaptively reduce the noise. This method is based on hypothesis testing in the frequency domain to eliminate the identical component between the reference signal and the primary signal. This paper presents a statistical parameter namely similarity factor to evaluate the filtering performance. The performance of the method is compared with the classical method, band pass filter (BPF). Results show that statistics filter is preferable to BPF in vibration signal processing. Moreover, the significance level \( \alpha \) would be optimized by genetic algorithms. However, it is very difficult to identify fault states only from time domain waveform or frequency spectrum when the effect of the noise is so strong or fault feature is not obvious. Pattern recognition is then applied to fault diagnosis in this study through system clustering method. This paper processes experiment rig data that after statistics filter, and the accuracy of clustering analysis increases substantially.

1. Introduction

Rolling bearings find widespread industrial applications, and the detection of their defects is important. In the field of roller bearing fault diagnosis, filter analysis is often used to cancel the noise or extract the components of the signal due to the defect signal that is drowned in a noise. The classical method is that the unwanted noise is removed by passing the signal through a filter. However, the traditional filtering methods need to manually set the filter domains, e.g., band pass filter method that is widely used in bearing diagnosis. [1-2]

In this study, an estimated signal extraction method is investigated by genetic algorithm and statistical tests in the frequency domain for the detection of roller bearing defects. Each component of spectrum including reference signal and primary signal is compared by the hypothesis testing to identify the identical components. To some extent, the identical components are the noise that contaminates the measured signal. To investigate the performance of the technique, the similarity factor is defined. Moreover, the level of significance \( \alpha \) would be appropriately chosen using genetic algorithm to guarantee the excellent effects of statistics filtering.

To whom any correspondence should be addressed.
2. Principle of spectrum statistics filter (SSF) method

The proposed noise cancelling method is performed by significant difference testing based on statistics. The procedure of the filtering method is shown in figure 1.

Figure 1. Basic process of SSF.

The primary signal measured for diagnoses \( s(t) \) is corrupted by a noise. The reference input \( n(t) \) is measured in the normal state. First, the spectrum \( n(f) \) and \( s(f) \) are obtained by fast Fourier transform. Then, the average \( (\mu_1, \mu_2) \) and variance \( (\sigma_1, \sigma_2) \) is got through statistical analysis. The two alternate hypotheses are set as equations (1) and (2) [3]:

Hypothesis1: \( \mu_1(f) = \mu_2(f) \);  
Hypothesis2: \( \sigma_1^2(f) = \sigma_2^2(f) \);  

Hypothesis testing based on statistics is known as: if the assumption is logically impossible, the original hypothesis is proven [4]. Therefore, the null hypotheses are \( \mu_1(f) \neq \mu_2(f) \) and \( \sigma_1^2(f) \neq \sigma_2^2(f) \). The decision rule is to reject the null hypothesis and accept the alternate hypothesis if the observed value of test statistic is in the critical region.

If both hypotheses have been proven, this means that the spectrum component strength of failure signal is similar to that of the reference signal. The component did not contain useful information and is considered to be cancelled. Then, the spectrum component value of the estimated failure signal \( s^*(f) \) will be set to zero, otherwise the value of the estimated noise signal \( n^*(f) \) will be set to zero. The level of significance \( \alpha \) is optimized by genetic algorithm (GA). Through inverse fast Fourier transform, the fault signal is decomposed into extracted noise signal \( n^*(t) \) and filtered signal \( s^*(t) \) in the time domain.
Similarity factor is defined as the evaluation factor to represent the similar degree of extracted noise \( n^* (t) \) and reference signal \( n(t) \). The more similarity existed between them, the more accuracy of statistical test was likely to occur. This paper takes the reciprocal of similarity factor as the objective function and sets up the GA model on optimal the significance level. The similarity factor \( I_{pq} \) is defined as equation (3):

\[
I_{pq} = \sum_{i=1}^{K} \left[ \log \left( \frac{q_i}{q'_i} \right) \right] \cdot K^{-1} + \sum_{i=1}^{K} \left[ \log \left( \frac{p_i}{p'_i} \right) \right] \cdot K^{-1} \tag{3}
\]

where \( p_i = \frac{\sqrt{2\pi} \cdot q_i}{\sigma_v}, \quad q_i = \frac{\sigma_v}{2\pi \sigma_x} \exp \left( \frac{-n^2}{2\sigma_x^2} \right) \),

\[
\sigma_x^2 = \int_0^\infty F_x^2 (f_k) \, df_k, \quad \sigma_v^2 = \frac{(2\pi f_k)^3}{3} + \sigma_x^2.
\]

3. Experimental tests

The method has been tested on measured signal forming a roller bearing test rig. Two acceleration sensors are used to measure vibration signals, and the sensors are mounted on the bearing housing at the end of the shaft in vertical and horizontal directions, respectively. The signal of the vertical channel is used as an example. The sampling frequency is 100 kHz, and the sampling time is 10s. Defects have been artificially introduced on the roller bearings. Dozens of experiments had been done. The results show that SSF is effective. The three pass-frequencies of the typical faults of the roller bearings at a speed of 1300 rpm are listed in table 1.

| Fault types               | Frequency /Hz |
|--------------------------|---------------|
| Defect on inner race     | 145.84        |
| Defect on outer race     | 86.32         |
| Defect on the ball       | 102.26        |

3.1. Compared with BPF in spectrums

It is found that the noise vibration frequency of bearings is focused mostly on low frequency. Therefore, high-pass filter was used as a contrast to SSF. The vibration signals from roller bearings were processed by high-pass filter and SSF, respectively.

The coefficient optimization model is based on the basic theory of genetic algorithm with the population size 10, gene length 6, Gene iteration times 3, mutation rate 0.5, crossover rate 0.3, and random crossover point. High fitness is the indicator of the success of the system. The optimal value of significance level \( \alpha \) is found to be 0.012 for the inner race fault, 0.01 for the outer race fault, and 0.10 for the ball and inner race fault.

Signal-to-noise ratio (SNR) method of calculation is chosen to evaluate the performance of the methods on experiment data. SNR is defined as the ratio of the sum of the peak values of the defect frequency and its harmonics to the average value of the spectrum [5], as shown in equation (4).

\[
SNR = N \cdot \frac{\sum_{j=1}^{n} P_j}{\sum_{k=1}^{N} S_k} \tag{4}
\]

Signals characterized by 2048 data points of the power spectrum are used to calculate the ratio. The frequency spectrums after filter and envelop analysis are shown in figure 2, figure 3 and figure 4. The characteristic frequency of a defect on the outer race is 86.96 Hz, a defect on the inner race is 145.7 Hz, and defect on the ball is 103.3 Hz, which are extremely close to the calculated pass-frequencies.
SNR increases when both routines are applied to measured signals, as shown in table 2. The figures clearly show that the noise level is reduced in the filtered signal. In the case of the inner race defect, the SNR increases from 6.6 to 79.58, and the first harmonic component respectively increased from 2657 to 3322. In the case of the outer race defect, the SNR increases from 22.6 to 57.52, and the glitch pulse is significantly reduced. The rolling body fault is always difficult to detect. In the case of compound fault, the ball and inner race correspond to both ball and inner race pass-frequencies, and the inner race fault signal was taken as reference signal to extract the pass-frequencies of ball fault. The SNR increases from 7.9 to 41.2.

![Figure 2](image.png)

**Figure 2.** Spectrums of a bearing with a defect of 0.3 mm in width and 0.05 mm in depth on the inner race. (a) primary signal; (b) filtered signal after HPF; (c) estimated signal after SSF.

![Figure 3](image.png)

**Figure 3.** Spectrums of a bearing with a defect of width 0.3 mm and depth 0.25mm on the outer race. (a) primary signal; (b) filtered signal after HPF; (c) estimated signal after SSF.
3.2. Compare with BPF in clustering analysis

The application of clustering has been widely used in the study of fault diagnosis method. The common methods such as SVM, ACO and PSO, have been researched extensively [6]. It is difficult to detect the types of weak fault signals or low rotation speed signals only from time domain waveform or frequency spectrum, such as at the speed of 500 rpm in the following study. Then, the state identification is converted to the clustering problem [7].

The simplest clustering analysis based on the Euclidean distance between samples, system clustering method, is combined with SSF to prove the effectiveness of SSF on useless information cancelling. Four frequency domain parameters commonly used for fault diagnosis of roller bearings were chosen according to Ref. [8], as in equations (5)-(8). P1 to P4 are calculated from the spectrum of the filtered signals in each state as the object of clustering, respectively. This paper provides clustering results comparison between signals after SSF, HPF and signals without being processed. It was found that the accuracy of clustering analysis is increasing after SSF, as shown in figure 5.

\[
P_1 = \left( \frac{\sum_{i=1}^{N} f_i^2 S(f_i)}{\sum_{i=1}^{N} S(f_i)} \right)^{1/2}
\]

\(5\)
Then, three parameters with good performance in training are used to typical faults diagnose of roller bearings. Under each type, fifteen signal conditions unknown samples are taken for diagnosis. The accuracy of clustering is shown in table 3. It is demonstrated that using extracted signals through SSF to cluster analysis has a significant effect for roller bearings typical faults diagnosis.

**Table 3.** Results of fault diagnosis using system-clustering.

| Fault types           | Accuracy (%) |
|-----------------------|--------------|
| Defect on inner race  | 100          |
| Defect on outer race  | 100          |
| Defect on the ball    | 100          |

4. Conclusions

This paper presents a SSF method to cancel the noise of roller bearings’ vibration signals. The similarity factor is defined to evaluate the effect of the filter and taken as the object function of GA to optimize significance level. A number of experiments were carried out to check the effectiveness of the method. Defective bearings that have typical simulated defects were used. The measured signals were processed by SSF and HPF separately. The frequency-domain modified SNR was calculated to assess the performances, and it was proven that performance of SSF is better than that of the HPF. However, to better understand which kind of signals is more suitable for the method, further studies are required because the adaptability was different under particular conditions. In addition, this study demonstrates that the method combining SSF and system-clustering can further increase the accuracy of clustering. It would be conclusion that combining SSF and intelligent clustering analysis methodology that can ease clustering and enhance the accuracy.
Acknowledgements
This study was supported by National Program on Key Basic Research Project (Grant No. 2012CB026000), National Natural Science Foundation of China (Grant No.51075023, 51135001), and Fundamental Research Funds for the Central Universities (Grant No. ZZ1124) and SRF for ROCS, SEM.

References
[1] Tandon N and Choudhury A 1999 A review of vibration and acoustic measurement methods for the detection of defects in rolling element bearings Tribol. Int. 32 469-80
[2] Bozchalooi I S and Liang M 2007 Identification of the high snr frequency band for bearing fault signature enhancement Mechatronics and machine vision in practice Int. Conf. (Xiamen: IEEE) pp 39-43
[3] Ryota I, Wang H Q, Tetsurou M and Chen P 2010 Advances in maintenance and condition diagnosis technologies towards Sustainable Society ed O Susumu and K Tadao (Nara: Sunrise Publishing Limited) pp 227-30
[4] David W S 1996 Introductory statistics: concepts, models and applications (Missouri: Missouri State University)
[5] Imed K and Mnaouar C 2005 Detection of rolling element bearing defects by adaptive filtering Eur. J. Mech. A/Solids 4 293-303
[6] Li K and Chen P 2011 Intelligent Method for Diagnosing Structural Faults of Rotating Machinery Using Ant Colony Optimization Sens. 11 4009-29
[7] Song L Y, Wang H Q, Pan J, Gao J J and Li K 2011 Application of ant colony algorithm in fault diagnosis of roller bearing Adv. Mater. Res. 291-4 1957-60
[8] Wang H Q and Chen P 2008 Condition diagnosis of blower system using rough sets and a fuzzy neural network. WSEAS Trans. Bus. Econ. 5 66-71