Modified Fisher Ratio and Kernel Density Estimation Method for Efficient Feature Selection of Rotating Machine Diagnostics

Y. Ban[1], H. Oh[2*], M. Hwang[3], S. Kim[4] and M. Mo[5]

[1,2,3]School of Mechanical Engineering, Gwangju Institute of Science and Technology, Gwangju, 61005, Korea
[4,5]Technology Research and Development Institute, KEPCO Plant Service and Engineering Co., LTD, Naju, 58217, Korea

[*]Corresponding author: hsoh@gist.ac.kr

1. Introduction

Rotating machines are one of the most important systems in mission-critical engineering assets including power plants, manufacturing facilities, and high speed trains. To maintain the engineering assets, the fault diagnostics of the rotor systems was studied. As a result, a number of useful features are available for rotor diagnostics including RMS [1], kurtosis [2], and bearing defect frequency [3]. Nonetheless, it is challenging to determine which feature is most relevant for the diagnosis of a particular rotating machine of interest. A randomly selected feature set can have irrelevant and redundant information [4]. Therefore, feature selection is critical for rotor diagnostics.

Feature selection can be divided into filter and wrapper methods [5]. The filter methods rank features by evaluating the intrinsic characteristic of the data, whereas the wrapper methods use the interaction with classifiers to evaluate the feature subsets. For example, Lu et al. selected a subset of available vibration features for fault diagnostics of rotary machines using the Fisher ratio and genetic algorithms (GA) [6]. However, existing metrics such as Fisher ratio and mRMR often do not perform as intended due to challenges. First, available data from rotating machines in the field is sparse. Second, the feature values from a severely degraded rotating machines does not follow Gaussian distributions [7]. Therefore, a better metric is required to overcome the challenges. In this paper, a new wrapper method is proposed by combining modified Fisher ratio and kernel density estimation (KDE).

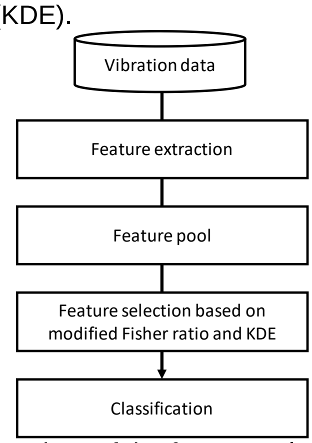

Fig. 1 Flow chart of the feature selection scheme

As shown in Fig. 1, the Fisher ratio were modified to detect outliers. Statistical parameters of the Fisher ratio can be estimated accurately using the KDE method. In addition, using simulation and testbed data, accuracy was evaluated.

2. Data acquisition and feature pool

A rotor testbed as shown in Fig. 2 was used to acquire the rotor vibration data. The testbed consists of two bearings and three disks to emulate low- and high-pressure steam turbines in power plants. For individual bearings, a pair of accelerometers were mounted in x- and y- axes.

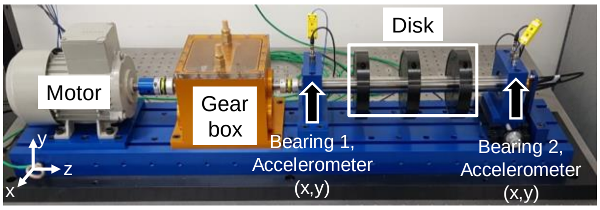

Fig. 2 Rotor testbed

As shown in Tables 1 and 2, a setup for data collection and various fault conditions are presented, respectively.

Table 1 Setup for data collection

RPM	Sampling rate	Duration	Interval
3,000	25.6 kHz	2 seconds	10 seconds

Table 2 Failure modes and severity

Condition	Unbalance (gram)			Misalignment (mm)		
level	3.0		4.5	0.4		0.6
Condition	Inner race fault (mm)			Outer race fault (mm)		
level	0.3	1	3	0.3	1	3

Twenty vibration features are extracted including time domain, frequency domain and bearing defect frequency. The time domain features are associated with statistical characteristics. The frequency domain features are associated energy of the vibration at the corresponding frequency. The bearing defect frequency features are the energy characteristics excited by defects in bearing elements.

3. Modified Fisher ratio and KDE

Fisher ratio is defined as the ratio of the sum of the distances between class means to the sum of the class variability [8]. Fisher ratio is represented.

$$\text{Fisher ratio} = \frac{\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}(\mu_i - \mu_j)^2}{\sum_{k=1}^{n}\sigma_k^2} \tag{1}$$

where μ_i and μ_j are the mean vector of class i and j, respectively; σ_k is the variance of class k; n is the total number of conditions; i = 1, 2, …, n-1, j = 2, 3, …, n and k = 1, 2, …, n.

Feature values calculated with signals from degraded rotors sometimes have outliers. The outliers can decrease the diagnostic accuracy. Thus, it is necessary to consider the effect of the outliers on the Fisher ratio. Kurtosis, 4th statistical moment, was known to be an effective measure to address the effect of the outliers [9]. With this theoretical background, a modified Fisher ratio is proposed.

$$\text{Modified Fisher ratio} = \frac{\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}(\mu_i - \mu_j)^2}{\sum_{k=1}^{n}\sigma_k^2 + \sum_{k=1}^{n}\mathrm{Kurt}_k} \tag{2}$$

where Kurt_k is the kurtosis of the kth class. The proposed metric can select a set of relevant features compared to the original Fisher ratio.

KDE is a non-parametric method to estimate the probability density function of a random variable. The density function using KDE is represented [10].

$$f_h(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right) \tag{3}$$

where f_h is the estimated density function; n is the number of samples; h is the bandwidth; K is the kernel; and x_i is the samples. The KDE can be used to estimate the variation accurately.

4. Conclusion

This paper proposed a modified Fisher ratio metric and KDE method for feature selection in rotor diagnosis based on the wrapper method. The modified Fisher ratio was effective to increase the diagnostic accuracy in the existence of outliers. The KDE method provided to calculate the statistical moments accurately. The proposed method can help to select more relevant features in rotor diagnostics. Therefore, it can increase the diagnostic accuracy and reduce the computational cost for diagnostics.

Acknowledgment

This work was supported by the Korea Institute of Energy Technology Evaluation and Planning (KETEP) Grants No. 20172010105590 funded by the Korean Ministry of Trade, Industry, & Energy (MOTIE).

This research is part of the projects that are financially supported by the KEPCO KPS (Project No.: C101800708).

References

[1] N. Tandon and B. Nakara, Detection of defects in rolling element bearings by vibration monitoring, *Journal of the Institution of Engineers (India): Mechanical Engineering Division*, 73 (1992) 271-282.

[2] D. Dyer and R. M. Stewart, Detection of rolling element bearing damage by statistical vibration analysis. *Journal of Mechanical Design*, 100 (2) (1978) 229–235.

[3] T. Igarashi and H. Hamada, Studies on the vibration and sound of defective rolling bearings (first report: vibration of ball bearings with one defect). *Bulletin of JSME*, 25 (204) (1982) 994–1001.

[4] B. Li, P. Zhang, H. Tian, S. Mi, D. Liu and G. Ren, A new feature extraction and selection scheme for hybrid fault diagnosis of gearbox, *Expert Systems with Applications*, 38 (2011) 10000-10009.

[5] B. Kumari and T. Swarnkar, Filter versus wrapper feature subset selection in large dimensionality micro array: a review, *International Journal of Computer Science and Information Technologies*, 2 (3) (2011) 1048-1053.

[6] L. Lu, J. Yan and C and de Silva, Dominant feature selection for the fault diagnosis of rotary machines using modified genetic algorithm and empirical mode decomposition, *Journal of Sound and Vibration*, 344 (2015) 464-483.

[7] J. Mathew and R. J. Alfredson, The condition monitoring of rolling element bearings using vibration analysis. *Journal of Vibration Acoustics Stress and Reliability in Design*, 106 (3) (1984) 447–453.

[8] Y. Huang, C. Liu, X.F. Zha, and Y. Li, A lean model for performance assessment of machinery using second generation wavelet packet transform and Fisher criterion, *Expert Systems with Applications* 37 (5) (2010) 3815–3822.

[9] F. Velasco and S. P. Verma, Importance of skewness and kurtosis statistical tests for outlier detection and elimination in evaluation of geochemical reference materials, Mathematical Geology, 30 (1) (1998), 109-128.

[10] M. Rosenblatt, Remarks on some nonparametric estimates of a density function, *The Annals of Mathematical Statistics*. 27 (3) (1956), 832–837.