Methodology for Optimal Feature Selection for Bearing Fault Diagnosis

S. Ham[1], S. Kim[2], H. J. Park[3], S. H. Cho[4] and J. H. Choi[2*]

[1]Dept. of Aerospace and Mechanical Engineering, Korea Aerospace University, Goyang-City, Gyeonggi-do, 10540, Republic of Korea
[2]School of Aerospace and Mechanical Engineering, Korea Aerospace University, Goyang-City, Gyeonggi-do, 10540, Republic of Korea

[*]Corresponding author: jhchoi@kau.ac.kr

1. Introduction

Recently, the economic loss and catastrophe caused by unexpected failures have been recognized as a important problem. To overcome those problems, numerous studies are conducted in the filed of prognostics and health management (PHM). Most of researches in which features are calculated by using the expert or domain knowledge have followed 4 steps, which are sensing, feature engineering, diagnosis and prognosis. In the second step, feature selection is important to select the optimal feature and scatter matrix is used as the separability criterion. In most researches, the distribution of calculated features is supposed as the normal distribution and scatter matrix has to be used in the normal distribution. However, the real distribution is often non-normal which is the skewed distribution or two normal distribution for one class. In this study, the separability criterion is proposed for the non-normal distribution.

2. Separability criterion using CDF

In case of the normal distribution, scatter matrixes are used as the separability criterion [1]. Scatter matrixes can be used in assumption that features are in normal distribution. Especially, for two class problem, scatter matrixes can be named in fisher discriminant ratio (FDR) which is shown in Eq. (1)

$$FDR = \frac{(\mu_1 - \mu_2)^2}{\sigma_1^2 + \sigma_2^2} \quad (1)$$

In Eq. (1), mean and variance of features are used. However, above two values are not defined in non-normal distribution. So, the separability criterion is proposed by using cumulative density function (CDF). CDF can be defined for any distribution. So, each feature which has non-normal distribution can be used as the separability criterion. In this case, the separability criterion is calculated as follows. First, to obtain probability density function (PDF), density function from each feature is obtained and scaled that area is 1. And PDF is cumulative to obtain CDF. After obtaining CDF, the area difference between each CDF from classes is calculated and used for the separability criterion.

In this study, to verify the usefulness of the proposed separability criterion, two criterions are used for feature selection by using bearing vibration signal. To make the non-normal distribution data sets, two conditions are applied in bearing testbed which is shown in Fig. 1.

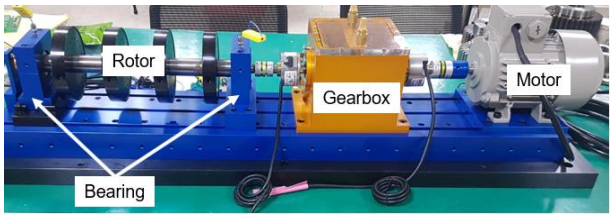

Fig.1 Bearing testbed

features are calculated and selected by using the FDR value and proposed criterion. By using selected features from each criterion, a classification is conducted, and the classification performance is compared.

3. Conclusion

Unexpected failures which cause the enormous management costs and human hazard are key issue in the industrial filed. So, PHM technique is developed in many researches. However, most researches are supposed that features from data have the normal distribution. There can be lots of case that have the non-normal distribution. In this study, the separability criterion used in feature selection is proposed and compared with the traditional criterion. This study is only about fault diagnosis. prognosis which has the non-normal distribution is studied in future work.

Acknowledgment

This research was supported by a grant from business of snetsystems.

References

[1] C. M. Bishop, *Pattern Recognition and Machine Learning Springer Mathematical notation Ni*. 2006.