ARTICLE IN PRESS

Journal of Sound and Vibration ∎ (∎∎∎∎) ∎∎∎–∎∎∎

Contents lists available at ScienceDirect

Journal of Sound and Vibration

journal homepage: www.elsevier.com/locate/jsvi

# Dominant feature selection for the fault diagnosis of rotary machines using modified genetic algorithm and empirical mode decomposition

Lei Lu [a], Jihong Yan [a,*], Clarence W. de Silva [b]

[a] *School of Mechatronics Engineering, Harbin Institute of Technology, Harbin 150001, China*
[b] *Department of Mechanical Engineering, University of British Columbia, Vancouver, BC, Canada V6T 1Z4*

ARTICLE INFO

*Article history:*
Received 16 June 2014
Received in revised form
25 December 2014
Accepted 22 January 2015
Handling Editor: L.G. Tham

ABSTRACT

This paper develops a novel dominant feature selection method using a genetic algorithm with a dynamic searching strategy. It is applied in the search for the most representative features in rotary mechanical fault diagnosis, and is shown to improve the classification performance with fewer features. First, empirical mode decomposition (EMD) is employed to decompose a vibration signal into intrinsic mode functions (IMFs) which represent the signal characteristic with sample oscillatory modes. Then, a modified genetic algorithm with variable-range encoding and dynamic searching strategy is used to establish relationships between optimized feature subsets and the classification performance. Next, a statistical model that uses receiver operating characteristic (ROC) is developed to select dominant features. Finally, support vector machine (SVM) is used to classify different fault patterns. Two real-world problems, rotor-unbalance vibration and bearing corrosion, are employed to evaluate the proposed feature selection scheme and fault diagnosis system. Statistical results obtained by analyzing the two problems, and comparative studies with five well-known feature selection techniques, demonstrate that the method developed in this paper can achieve improvements in identification accuracy with lower feature dimensionality. In addition, the results indicate that the proposed method is a promising tool to select dominant features in rotary machinery fault diagnosis.



## 1. Introduction

Rotary mechanical equipment is widely used in engineering realms such as manufacturing, metallurgy, energy, and transportation. Malfunctions in rotors and the bearings are the major cause of unplanned downtime in rotary equipment [1,2]. The reasons for unexpected mechanical failures include mounting failure, low lubrication, fatigue, and wear and tear. These may result in formidable challenges in the accurate diagnosis of faults [3]. Digital signal processing methods provide a variety of efficient and effective ways to extract multiple features from machine signals such as mechanical vibration, sound, and temperature [4]. Selecting features of high information content is important for the analysis of mechanical failure, which is a crucial step in the design of a classification system. The use of artificial intelligence in advanced methods of dominant feature selection has much potential in computer-based mechanical fault diagnosis.

* Corresponding author. Tel./fax: +86 451 86402972.
*E-mail address:* jyan@hit.edu.cn (J. Yan).

http://dx.doi.org/10.1016/j.jsv.2015.01.037
0022-460X/

Please cite this article as: L. Lu, et al., Dominant feature selection for the fault diagnosis of rotary machines using modified genetic algorithm and empirical mode decomposition, *Journal of Sound and Vibration* (2015), http://dx.doi.org/10.1016/j.jsv.2015.01.037

Fault diagnosis systems can be roughly classified into two major types: model-based methods and data-driven approaches [5]. Model-based schemes utilize analytical models and techniques to simulate the normal condition, and identify faults by examining whether the monitored information is consistent with the predicted behavior. Data-driven approaches employ data mining methods such as signal processing, artificial intelligent, and expert systems [6,7] to extract sensitive information from collected response data and develop intelligent model to classify the normal condition and irregular behaviors. These two categories to some degree rely on the analysis of sensory information, and it has never been an easy task to identify dominant features from the investigated signals. This is particularly true with regard to key components that are decomposed by time–frequency methods such as wavelet analysis (WT) and EMD. Some literature exploited only a few decomposed sub-modes or empirically extracted fault features [8]. However, irrelevant, noisy, and high dimensional features may lead to a complicated classification model, low identification accuracy and inefficient learning speed.

Research has been carried out into feature selection (FS) techniques over the past several decades. Fisher criterion and the sequential backward feature selection algorithm have been used to select features that are extracted from vibration signals of a rotor test-bed [9]. Wavelet transform has been used to analyze acoustic emission signals from a grinding process [10], and ant colony optimization and sequential forward floating selection methods have been utilized to optimize the extracted feature subsets. Sensitivity analysis and relational analysis-based statistical method have been used to select dominant features in different cases of gearbox condition monitoring [11]. Fault identification system based on wavelet packet transform and the artificial neural network has been developed for a glass transfer robot [12], and entropy based feature selection method has been used to remove uncorrelated nodes from the input vector of the classification model.

The process of utilizing FS approaches to search an optimal subset from the feature space of high dimensionality is essentially a NP-hard optimization problem. Obtaining the optimal solution cannot be guaranteed without performing an exhaustive search in the solution space. Furthermore, this technique is impractical for large scale features because of the unaffordable time cost [13]. The genetic algorithm (GA), a form of inductive learning strategy, has great advantages in feature selection, as it can generate near-optimal solutions in complex and nonlinear search spaces with time efficiency [14,15]. Considerable research has gone into employing GA-based techniques to select optimal feature subsets from a high dimensional feature space [16]. The Kernel-Genetic Algorithm (K-GA) technique has been developed to address the selection of nonlinear features [17]. An improved genetic algorithm (IGA) based on segmented management has been proposed to select an optimal feature subset from a multi-character feature set (MCFS) [18]. An advanced GA-based feature selection method has been used to select the optimal features and instance at one time [19]. However, there are still some shortcomings of GA-based feature selection methods. Most of them only employ a single binary number 0 or 1 to denote whether the feature has been selected [17,19], or use a fixed-length to encode the features [18], which are not efficient when dealing with a large number of feature candidates. Moreover, GA is relatively a heuristic search method, and the use of advanced method to obtain a more reliable and statistical results needs to be addressed. In the present paper, (1) in order to achieve the global optimal solution and select elite features at each dimensionality, one feature candidate in the chromosome is encoded with $n$ genes, which has a more flexible search range; (2) a length-variable encoding method and interactive relationships between features within the same chromosome are developed to design the chromosome, and the proposed method enables selection of the best feature combination for each feature dimensionality; (3) a mathematical model is developed to analyze the optimized feature subsets, which produces more reliable results.

In the context of feature selection, one crucial issue concerns generating a feature pool that covers positive candidates. Related studies have extract statistical features from initial raw signals [20]. State-of-the art FS techniques select features from sub-signals that are generated by time–frequency methods. These approaches transform sensory information into another space that could more clearly reveal properties of the original data. One problem with the wavelet transform (WT) method is the difficulty in choosing parameters such as the decomposition level and wavelet function. By contrast, empirical mode decomposition (EMD), which is a novel non-stationary signal processing method, uses a data-driven process, which avoids the tedious task of parameter selection and does not require any prior knowledge from the signal or the investigated object. Moreover, intrinsic mode functions (IMFs) derived from the original signal are characterized as zero-mean oscillations, and the negative influences of varying load on the IMFs are partially removed [8]. The importance of EMD in processing a mechanical vibration signal has been verified in [21,22]. In the present paper, EMD is employed to represent the investigated mechanical signal with intrinsic modes, and then feature candidates are derived from the decomposed IMFs.

In this paper, a new dominant feature selection scheme for rotary machine fault diagnosis is developed. It is based on an innovative variable-range genetic algorithm (VRGA), which employs a dynamic searching strategy by varying the length of chromosomes and changing the range of the feature candidates in real-time. Features extracted from sub-modes decomposed by EMD are optimized, and a performance-dimensionality relationship is developed by the VRGA model. Then dominant features are derived through ROC analysis and a statistical model. Comparative studies with other FS techniques are presented.

The remainder of this paper is organized as follows. Section 2 introduces the concept of empirical mode decomposition and methods for feature extraction. Section 3 reviews the theory of support vector machines. In Section 4, the VRGA model for feature selection is presented in detail. Section 5 describes the experimentation and discusses the results. Finally, Section 6 draws conclusions of the present work.

## 2. Empirical mode decomposition and feature extraction

The EMD method [23,24] is a relatively new approach of signal analysis, which seeks to identify the intrinsic oscillatory modes by the characteristic time scales of the data. It is derived from the idea that the interlaced local extrema and zero