Mechanical Systems and Signal Processing 70-71 (2016) 87–103

Contents lists available at ScienceDirect

Mechanical Systems and Signal Processing

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

Fault diagnosis in spur gears based on genetic algorithm and random forest

Mariela Cerrada [a,b,*], Grover Zurita [a,c], Diego Cabrera [a], René-Vinicio Sánchez [a], Mariano Artés [d], Chuan Li [e,a]

[a] Mechanical Engineering Department, Universidad Politécnica Salesiana, Cuenca-Ecuador
[b] Control Systems Department, Universidad de Los Andes, Mérida, Venezuela
[c] Electromechanical Engineering Department, Universidad Privada Boliviana, Cochabamba, Bolivia
[d] Mechanical Engineering Department, Universidad Nacional de Educación a Distancia, Madrid, Spain
[e] Chongqing Key Laboratory of Manufacturing Equipment Mechanism Design and Control, Chongqing Technology and Business University, Chongqing, China

ARTICLE INFO

Article history:
Received 14 January 2015
Received in revised form
6 June 2015
Accepted 27 August 2015
Available online 26 September 2015

Keywords:
Gearbox
Fault diagnosis
Genetic algorithms
Random forest
Feature selection
Wavelet packets

ABSTRACT

There are growing demands for condition-based monitoring of gearboxes, and therefore new methods to improve the reliability, effectiveness, accuracy of the gear fault detection ought to be evaluated. Feature selection is still an important aspect in machine learning-based diagnosis in order to reach good performance of the diagnostic models. On the other hand, random forest classifiers are suitable models in industrial environments where large data-samples are not usually available for training such diagnostic models. The main aim of this research is to build up a robust system for the multi-class fault diagnosis in spur gears, by selecting the best set of condition parameters on time, frequency and time–frequency domains, which are extracted from vibration signals. The diagnostic system is performed by using genetic algorithms and a classifier based on random forest, in a supervised environment. The original set of condition parameters is reduced around 66% regarding the initial size by using genetic algorithms, and still get an acceptable classification precision over 97%. The approach is tested on real vibration signals by considering several fault classes, one of them being an incipient fault, under different running conditions of load and velocity.

1. Introduction

There are constant increasing requirements for the continuous working of the transmission machines. This is why new approaches to building up fault diagnostic systems with accuracy and reliability are highly valuable. There are invaluable studies to detect gear faults by using standard diagnostic techniques, such as Cepstrum Techniques and Envelope Data Analysis with Hilbert Transforms, among other techniques [1]. In recent years, several analysis techniques for gears faults diagnosis have used Wavelet Packet Transform (WPT), in order to enhance the information that is provided by the classical statistical parameters from the vibration signal, in time and frequency domains [2–4]. In case of machine learning-based

* Corresponding author at: Control Systems Department, Universidad de Los Andes, Mérida, Venezuela.
E-mail addresses: mcerrada@ups.edu.ec, cerradam@ula.ve (M. Cerrada), gzuritav@ups.edu.ec (G. Zurita), dcabrera@ups.edu.ec (D. Cabrera), rsanchezl@ups.edu.ec (R.-V. Sánchez), martes@ind.uned.es (M. Artés), chuanli@21cn.com (C. Li).

http://dx.doi.org/10.1016/j.ymssp.2015.08.030
0888-3270/

diagnosis, the most common approaches have been developed by using Neural Networks [5,6], Support Vectors Machines [7], Cluster Analysis [8] and Genetic Algorithms [9,10]. These approaches have been very useful for implementing condition-based maintenance (CBM), as presented in Jardine et al. [11].

Recently, Random Forest (RF), as a classification technique, has been used for fault diagnosis in several areas of engineering because it is a robust approach in case of having a large number of input attributes, a low number of available samples for learning and because the high interpretability of the tree-based models [12,13]. On the other hand, Genetic Algorithms (GA) have also been used for the optimization process in classification problems, one of them being the feature selection. In diagnosis of machinery condition, RF has been used with GA in order to improve the classification accuracy [14,15]. Yang et al. [14] have used GA to optimize two parameters of the RF algorithm: the number of trees and the split variable number at each node of the tree. In combustion engines, Karabadji et al. [15] have applied GA for selecting the best classifier based on trees for faults diagnosis in industrial fans, the outcome of the analysis was to find the classifier which appears most frequently in the population, and the classifier with the best accuracy in classification was also found.

Condition parameters for fault diagnosis are mostly related to statistical measures from signals in time and frequency domains. Moreover, parameters associated to time–frequency domain have important information on machinery condition, and they are also used in order to extend the set of condition parameters that are processed in diagnostic algorithms. These condition parameters are called *the features* in classification problems. Taking into account the availability of a large number of feature candidates for fault diagnosis, the problem of feature selection, after the feature extraction process, is still an open research area in case of using machine learning-based diagnosis. The failure nature can lead to choosing certain condition parameters and, in case of incipient fault, the best condition parameters providing good diagnostic information are not easy to identify. An incipient fault is a fault that is just beginning to show symptoms. Sometimes, minor scratches and wear on the face of the tooth cannot be seen in scheduled maintenance inspections. This is a typical example of incipient fault. This type of fault commonly appears in rotating machinery and it is related to the first stage of severe situations that can lead to loss of the device function. Detection and diagnosis of incipient faults are still research problems that should be improved in real industrial applications. On the other hand, the availability of a large number of samples for classifier training is not usual in real industrial environments, in this case, RF-based models have better performance than other classical classifiers.

This work combines different methods for proposing best features and accurate models for spur gear fault diagnosis. The problem of feature selection is addressed by using GA, and the selected features are used as input attributes for the RF-based fault classifier. Feature selection is an optimization problem where the objective is to minimize certain classification performance metric, then, the major contribution is to have an efficient feature selection mechanism that can lead to a higher diagnostic performance, regarding the use of the whole set of features.

Fig. 1 shows the overview of our approach where the vibration is the analysed signal. After the data acquisition process, the feature extraction is carried out by calculating the most common statistical parameters on time and frequency domain of the vibration signal. Additionally, WPT are applied on the vibration signal to decompose the signal and later extract more condition parameters. Finally, a wide set of condition parameters is collected, called in the following *the attributes* of the

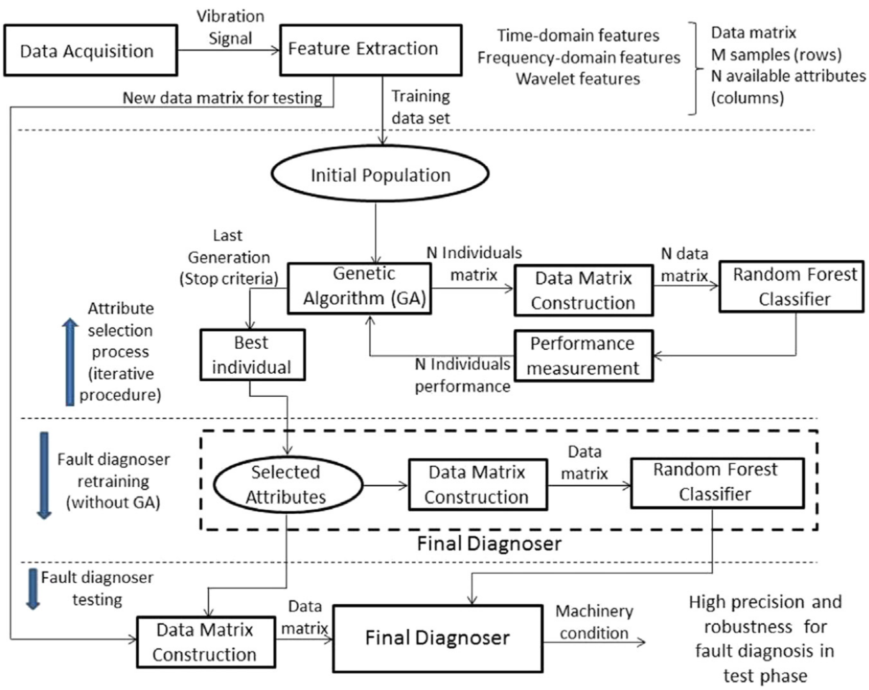

Fig. 1. An overview of the optimization approach for attribute selection in fault diagnosis.