Available online at www.sciencedirect.com
ScienceDirect
Mechanical Systems and Signal Processing 21 (2007) 1273–1284
Mechanical Systems and Signal Processing
www.elsevier.com/locate/jnlabr/ymssp

Fault classification using genetic programming

Liang Zhang, Asoke K. Nandi*

Signal Processing and Communications Group, Department of Electrical Engineering and Electronics, The University of Liverpool, Brownlow Hill, Liverpool L69 3GJ, UK

Received 20 December 2005; received in revised form 24 April 2006; accepted 24 April 2006
Available online 21 June 2006

Abstract

Genetic programming (GP) is a stochastic process for automatically generating computer programs. In this paper, three GP-based approaches for solving multi-class classification problems in roller bearing fault detection are proposed. *Single-GP* maps all the classes onto the one-dimensional GP output. *Independent-GPs* singles out each class separately by evolving a binary GP for each class independently. *Bundled-GPs* also has one binary GP for each class, but these GPs are evolved together with the aim of selecting as few features as possible. The classification results and the features each algorithm has selected are compared with genetic algorithm (GA) based approaches GA/ANN and GA/SVM. Experiments show that *bundled-GPs* is strong in feature selection while retaining high performance, which equals or outperforms the two previous GA-based approaches.

Keywords: Genetic programming; Condition monitoring; Multi-class classification; Fault classification; Roller bearing

1. Introduction

Condition monitoring is becoming increasingly important in manufacturing industry, through its ability to increase general reliability and uptime. However, many techniques available at present require much expertise to apply these successfully; new techniques are required which allow relatively unskilled operators to make reliable decisions without knowing the mechanism of the system and analysing the data. Such techniques must be reliable enough to use in real systems. Reliability of detection has come to be one of the most important criteria. Artificial neural networks (ANNs) and support vector machines (SVMs) are suitable for this kind of problem. These have been researched and applied to real systems [1–6]. This paper develops genetic programming (GP) to three such problems, and offers equal or better results compared with the results of ANN and SVM (in conjunction with genetic algorithms, GAs) in [7]. GP has the advantages of good performance, understandable solutions (Fig. 1) and feature selection, which makes it a very promising development.

GP, first introduced by John Koza [8], is derived from GA. In the short period of its development, GP has been successfully applied in many fields, including condition monitoring [9–11]. In this work, GP is developed

*Corresponding author. Tel.: +44 151 794 4525; fax: +44 151 794 4540.
E-mail addresses: a.nandi@liv.ac.uk, a.nandi@liverpool.ac.uk (A.K. Nandi).

0888-3270/$ - see front matter
doi:10.1016/j.ymssp.2006.04.004

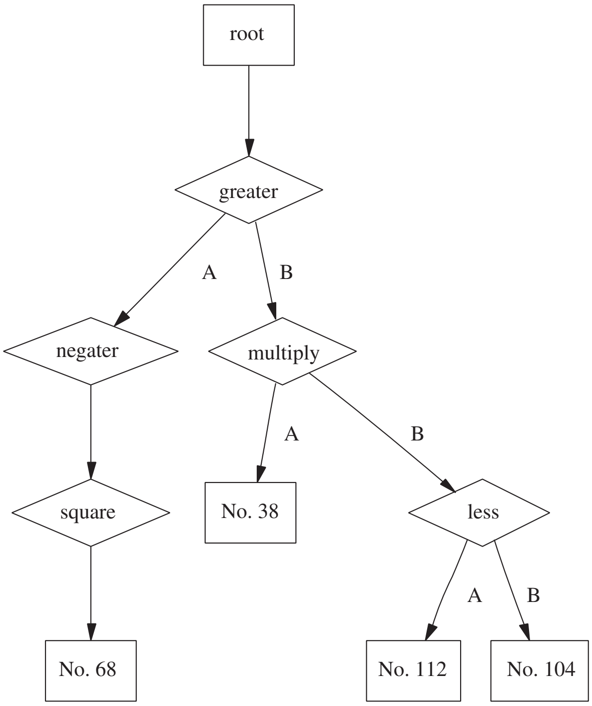

Fig. 1. A GP solution tree.

for multi-class classification of roller bearing faults. As a classifier, GP is often used for dual-class problems [12], but rarely for multi-class problems. Zhang et al. [13] mapped different intervals of GP output to different classes. However, the thresholds for the intervals have to be predetermined manually, which is hard to optimise. This approach, along with two other approaches that are proposed in this paper, to solving multi-class problems have been tested in this work. The results are compared with two other types of classification algorithms, GA/ANN and GA/SVM. It has been known for some time that ANN and SVM in conjunction with GA for feature selection are capable of improving the accuracy of classification systems by removing the features that may be confusing the classifier and hence improving the generalisation performance [14,15]. The GP approach has the ability of feature selection, which allows selecting those features with the most preferable characteristics, and the overall performance and robustness improve as a result.

2. The problems

Automatic monitoring of rolling element bearings (Fig. 2) is necessary in industry, as unexpected failure of machinery is a high cost problem; and manual checking of a number of rolling element bearings in a machinery may take an unacceptably long time, and consequently results in a great loss of money. Therefore, it is important in some areas to be able to detect the existence and severity of a fault of rolling element bearings in machinery automatically and reliably.

The vibration data used in the paper have been taken from three different sources. The first comes from the experiment on a small test rig, which simulates an environment for running roller bearings. Six conditions have been tested and recorded. There are two normal conditions—a brand new condition and a worn but undamaged condition; and four fault conditions—inner race fault, outer race fault, rolling element fault, and cage fault. Data have been recorded over a range of 10 speeds. Each condition has different characteristics, which make it possible to separate from others. Inner and outer race faults have a periodic signal; the rolling element fault may or may not be periodic, dependent upon several factors including the degree of damage to