Engineering Science and Technology, an International Journal 18 (2015) 59–69

HOSTED BY

ELSEVIER

Contents lists available at ScienceDirect

Engineering Science and Technology, an International Journal

journal homepage: http://www.elsevier.com/locate/jestch

Full length article

Determination of minimum sample size for fault diagnosis of automobile hydraulic brake system using power analysis

V. Indira [a], R. Vasanthakumari [b], R. Jegadeeshwaran [c,*], V. Sugumaran [c]

[a] Department of Mathematics, Sri Manakula Vinayagar Engineering College, Madagadipet, Puducherry, India
[b] Department of Mathematics, Villianur College for Women, Villianur, Puducherry, India
[c] School of Mechanical and Building Sciences, VIT University Chennai Campus, Chennai 600127, India

ARTICLE INFO

Article history:
Received 23 May 2014
Received in revised form
5 September 2014
Accepted 23 September 2014
Available online 4 November 2014

Keywords:
Fault diagnosis
Machine learning
Power analysis
Vibration signals
Minimum sample size
Statistical features

ABSTRACT

Hydraulic brake in automobile engineering is considered to be one of the important components. Condition monitoring and fault diagnosis of such a component is very essential for safety of passengers, vehicles and to minimize the unexpected maintenance time. Vibration based machine learning approach for condition monitoring of hydraulic brake system is gaining momentum. Training and testing the classifier are two important activities in the process of feature classification. This study proposes a systematic statistical method called power analysis to find the minimum number of samples required to train the classifier with statistical stability so as to get good classification accuracy. Descriptive statistical features have been used and the more contributing features have been selected by using C4.5 decision tree algorithm. The results of power analysis have also been verified using a decision tree algorithm namely, C4.5.

1. Introduction

The brake system is an essential component in an automobile to promote the highest degree of safety for the persons inside the vehicle and others moving on the road. Brake failure is crucial not only for the driver and passengers but also for automobile manufacturers. Fault diagnosis is an important process in preventive maintenance of hydraulic brakes. It avoids serious damage if defects occur to the component during operation. Prevention is better than cure. Early detection of the defects, therefore, is crucial to prevent the system from malfunction that could cause damage to entire system or accident. A fault diagnosis model can predict the condition of the system at any time and it avoids unexpected failures.

In this paper, only vibration signals of good and nine faulty conditions of a hydraulic brake system were considered for fault diagnosis. The characterization of the vibration signals was achieved by machine learning approach. The important two activities in machine learning approach are training and testing the classifier. To model the fault diagnosis problem as a machine learning problem, a large number of vibration signals are required for each condition considered for study. It may be possible to acquire any number of vibration signals for good condition; however, it is very difficult to acquire signals of different faulty conditions. Actually, the signals are to be taken from the system where the fault occurred naturally during operation. The difficulties involved in carrying out this process forces the fault diagnosis engineer to make a compromise. Taking many vibration signals from one specimen having a typical intended fault is one level of compromise in practice. Another level of compromise is that taking vibration signals from the system, where the required type of fault is simulated onto it. To overcome these problems, one should know the number of samples to be used for training to get good classification accuracy. Also, the signals can be acquired from the system where the fault has occurred naturally, only if one knows the minimum number of samples required to train the classifier to get good classification accuracy. Indeed, any results obtained out of these signals would be more practical and realistic in nature. Hence knowledge about the optimum number of samples required for building a model or training a classifier is very essential. In such situations, a study on determination of minimum sample size is highly desirable. Characteristics of the sample have a direct effect on power. Highly diverse samples will require adjustments in sample size. Adequate power is hard to achieve when results must

* Corresponding author. Tel.: +91 44 39931335; fax: +91 44 39932555.
E-mail addresses: indira_manik@yahoo.co.in (V. Indira), vasunthara1@gmail.com (R. Vasanthakumari), krjegadeeshwaran@gmail.com (R. Jegadeeshwaran), v_sugu@yahoo.com (V. Sugumaran).

Peer review under responsibility of Karabuk University.

http://dx.doi.org/10.1016/j.jestch.2014.09.007

be very accurate. Very high confidence levels require very large samples. A study with insufficient power may lead the researcher to abandon potentially useful samples. Power analysis is the best method to avoid these serious errors.

In machine learning, a model built with large sample size would be robust. During implementation, it becomes necessary to know the number of samples required to build a classifier with statistical stability. Many researchers have taken different approaches on minimum sample size determination reported in the field of bioinformatics and other clinical studies, to name a few, micro array data [1], cDNA arrays [2] transcription level [3] etc. Based on these works, data-driven hypotheses could be developed which in turn furthers vibration signal analysis research. Though it has been reported in [4,5] for fixing sample size to train the classifier for some particular applications of vibration analysis, the same sample size cannot be used for the present study. Hence, this paper focuses on determination of sample size to build a robust classifier for fault diagnosis. There are many ways available for determination of sample size viz. for tests of continuous variables [6], for tests of proportions [7], for time-to-event (survival) data [8], for receiver operating curve (ROC) analysis [9], for logistic and Poisson regression [10], repeated measurements [11], precision [12], paired samples [13], measurement of agreement [14], and power [15]. Studies were also carried out to discuss issues surrounding estimating variance, sample size re-estimation based on interim data [16], studies with planned interim analyses [17], and ethical issues [18]. However, there are certain issues to be addressed in implementation of such techniques to have better statistical stability.

In machine learning approach, the vibration signals are typically subjected to analyses such as hypothesis testing, classification [19], regression and clustering that rely on statistical parameters to draw conclusions [20–23]. However, these parameters could not be reliably estimated with only a small number of vibration signals. Since the statistical stability of conclusions largely depends on the accuracy of parameters used, a certain minimum number of vibration signals are required to ensure confidence in the sample distribution and accuracy of parameter values. The objective of this paper is to determine the minimum number of samples required to separate the classes with statistical stability using *F*-test based statistical power analysis.

1.1. Methodology

The methodology has been illustrated with the help of a typical automobile hydraulic brake system. Fig. 1 shows the methodology of the proposed study. Referring Fig. 1, the vibration signal under different fault conditions were acquired from the brake setup. Number of descriptive statistical features were extracted from the vibration signal. Among them most important features were selected using decision tree. The minimum number of samples required for classification with statistical stability using *F*-test based statistical power analysis. The minimum sample size is also determined using an entropy based algorithm called ‘C4.5 decision tree’. The results of power analysis are compared with that of C4.5 decision tree algorithm and sample size guidelines are presented for the considered system at the conclusion section.

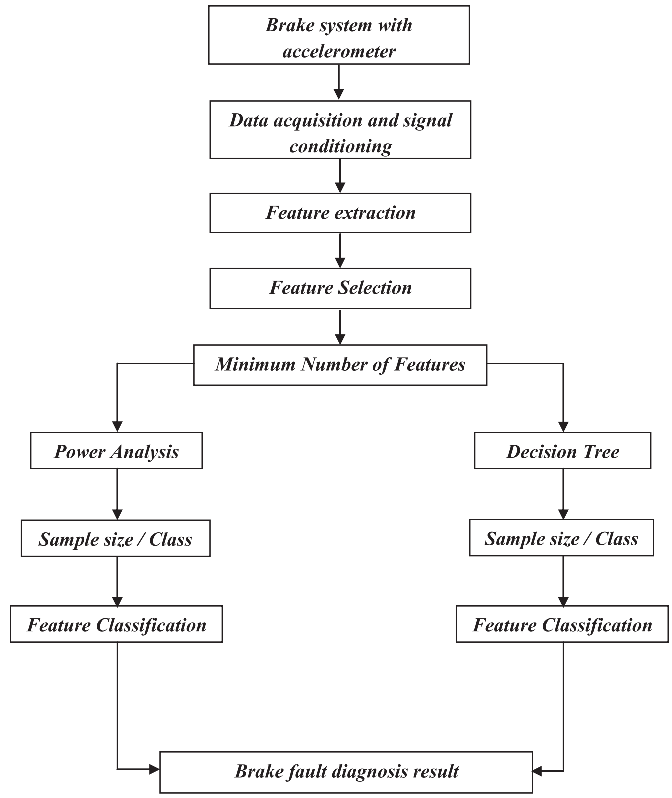

Fig. 1. Flow chart – methodology.

2. Experimental studies

Experimental study was carried out on a hydraulic brake system setup shown in Fig. 2 [24]. A commercial passenger car's (Maruti Swift) hydraulic brake system (shown in Fig. 2) was used to fabricate the brake test setup. The test rig consists of drum and disc brake coupled together by a shaft. The shaft is in turn run by a DC motor (1HP) coupled to a belt drive system. DC motor consists of an inbuilt drive. A lever is placed at the top of the motor which is connected to the accelerator pedal providing variable speeds up to 2500 rpm. If the accelerator pedal is pressed, the spring attached to the lever is compressed. Hence the pulley shaft is connected with the inbuilt drive. The power is transmitted from the drive to the pulley. When the pedal is released, the pulley shaft gets detached from the drive shaft through spring expansion. Brake pedal is provided to the left side of the accelerator pedal. It is attached to the piston in the master cylinder via a push rod. Master cylinder, the most important part of hydraulic brake is provided with pistons to move along the bore. Since hydraulic brakes are prominent brake system in medium motor vehicle like cars, in order to experiment with the components used in real world, branded vehicles (cars)

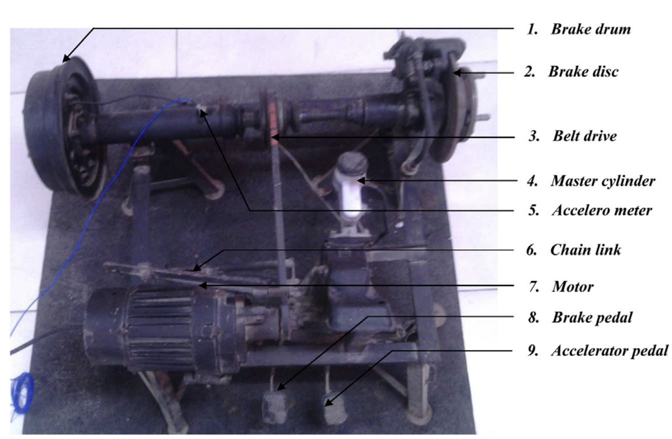

Fig. 2. Brake fault diagnosis – experimental setup.