Applied Soft Computing 46 (2016) 459–468

Contents lists available at ScienceDirect

Applied Soft Computing

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

Comparison of four direct classification methods for intelligent fault diagnosis of rotating machinery

Dongyang Dou[a,b,c,*], Shishuai Zhou[a,b]

[a] Key Laboratory of Coal Processing and Efficient Utilization of Ministry of Education, China University of Mining and Technology, Xuzhou 221116, PR China
[b] School of Chemical Engineering and Technology, China University of Mining and Technology, Xuzhou 221116, PR China
[c] Jiangsu Key Laboratory of Large Engineering Equipment Detection and Control, Xuzhou Institute of Technology, Xuzhou 221111, PR China

ARTICLE INFO

Article history:
Received 7 December 2014
Received in revised form 1 February 2016
Accepted 8 May 2016
Available online 18 May 2016

Keywords:
Fault diagnosis
Rotating machinery
PNN
SVM
Rule

ABSTRACT

Condition monitoring of rotating machinery is important to promptly detect early faults, identify potential problems, and prevent complete failure. Four direct classification methods were introduced to diagnose the regular condition, inner race defect, outer race defect, and rolling element defect of rolling bearings. These include the K-Nearest Neighbor algorithm (KNN), Probabilistic Neural Network (PNN), Particle Swarm Optimization optimized Support Vector Machine (PSO-SVM) and a Rule-Based Method (RBM) based on the MLEM2 algorithm and a new Rule Reasoning Mechanism (RRM). All of them can be run on the Fault Decision Table (FDT) containing numerical variables and output fault categories directly. The diagnosis results were discussed in terms of accuracy, time consumption, intelligibility, and maintainability. Especially, the interactions of the systems and human experts were compared in detail. It was concluded that all the four methods can work satisfactorily on accuracy, in an order of the PSO-SVM ranking the first, followed by the RBM that functioned the friendliest. Moreover, the RBM had the ability of feature reduction by itself, and would be most suitable for real-time applications.

1. Introduction

Condition monitoring of rotating machinery contributes to detect early faults and potential problems promptly, in order to prevent complete failure. Bearing vibration can bring about noise and degrade the quality of a product line. Severe vibrations of bearings can even lead the entire system to function incorrectly and result in an unexpected downtime [1]. To keep the machine working at its optimal conditions, prevent personal casualties and reduce economic loss, various fault diagnosis methods have been developed and applied effectively to detect the machine faults at an early stage. By use of signal processing techniques, such as the Short Time Fourier Transform (STFT) [2], Wavelet Transform (WT) [3,4] and Empirical Mode Decomposition (EMD) [5–8], it is possible to obtain the vital diagnosis information from the vibration signals. However, many techniques currently available require a significant input of expertise to implement them successfully. This demands a domain-specific knowledge of maintenance and in-out of the system from an expert engineer. Often the expert is not immediately available [9]. Therefore, the easier approaches are required for relatively unskilled operators to make reliable decisions on the running health of machine without a diagnosis specialist.

When a rotating machine is subjected to some kinds of fault, several vibration characteristics measured from it will exhibit obvious changes from their reference levels. These changes will form a pattern called the fault signature of the machine [10]. Fault diagnosis of rotating machinery can be treated as a problem of pattern recognition. It consists of three steps: data acquisition, feature extraction and selection, and final condition identification. The purpose of the feature extraction is to expose fault patterns through a group of parameters from noisy signals. The parameters in time domain and frequency domain have been proved to be effective and practical due to their relative sensitivity to the early faults, and their robustness to various loads and speeds [6]. However, the presented frequency domain index is merely the statistical information of spectrums without clear physical meanings, which has a bad effect on the comprehension of results. In addition, the features extracted from the raw and/or preprocessed signals may sometimes have large dimensionality, which may increase the computational burden of a subsequent classifier, and degrade the generalization capability of the classifier. To overcome these shortcomings, a subset of features which obviously characterize the machine oper-

* Corresponding author.
E-mail addresses: ddy41@cumt.edu.cn (D. Dou), 1349437909@qq.com (S. Zhou).

http://dx.doi.org/10.1016/j.asoc.2016.05.015
1568-4946/

ational conditions is to be selected. However, if the feature set is designed elaborately with a moderate scale, the feature selection is not always necessary because it may take excessive time and introduce the arbitrary errors to the data. For example, only two features from autocorrelation over successive spectrograms are required to yield a near perfect classification accuracy [11].

To implement the condition monitoring operations for a non-expert engineer, research has been extended to make use of the clustering algorithms [3], Artificial Neural Network (ANN) [6–8,12], Support Vector Machine (SVM) [4,13], K-Nearest Neighbor algorithm (KNN) [14], fuzzy logic and genetic algorithms [10,15,16] and decision trees [9,12]. In addition, the Rough Sets Theory (RST) has also emerged as a tool for fault diagnosis, and proved to be successful in breaking the bottleneck of rule acquisition in the knowledge-based systems [5,15,17]. The Back Propagation (BP) and Radial Basis Function (RBF) neural network are usually the first choice of ANN. Although the multi-layer perception network such as the BP is widely used in engineering applications and several algorithms have been developed to reduce converge time or avoid trapping into local minimum, their effectiveness is often limited. Other problems in the structure of BP have been reported, such as the decision as to the number of hidden layers and the number of neurons in the hidden layer. The RBF network provides more effective approaches to train and organize its structure, but it still requires continuous iterations to achieve network convergence. Both methods were initially designed for predicting continuous values. Additional discretization is necessary for fault classification which may introduce errors.

All these shortcomings are adverse to construct a high-performance system on traditional ANN. However, the Probabilistic Neural Network (PNN) only requires one epoch of training that can make the network quickly complete a study. Furthermore, the PNN can automatically organize its structure according to the information of input data without the help of users. Compared to the BP and the RBF networks, the PNN is more efficient in the training and generalization capabilities [18]. It has been used successfully in the earthquake magnitude prediction, medical science, and marketing research. However, it is surprising that – to our knowledge – there are limited publications in applying the PNN to the fault diagnosis of rotating machinery.

On the other hand, the distance-based methods are indeed a simple and intuitive way to make classification decisions. Especially, the KNN has been proved to be valid in the gear fault diagnosis [19], and it may become a promising way to the fault recognition of other types of rotating machinery. The SVM is very popular nowadays in the fault diagnosing field. To achieve good performance, optimization methods such as the Genetic Algorithm (GA) [20] and Particle Swarm Optimization (PSO) [13] are often adopted to tune the parameters of SVM.

In the recent years, knowledge-based diagnosis systems are developed quickly, among which the case-based methods and rule-based ones are the most outstanding techniques. In Ref. [15], a novel knowledge-based fuzzy neural network was proposed. However, the rule acquisition method requires data discretization first, because the basic rough set technique can only process the discrete variables. Discretization of numerous data in information systems may be the biggest obstacle to performing inductive learning from instances. It always reduces the expressiveness of data, and brings about a drop in final accuracy. The Case-based methods [20,21] reuse the past cases to find a solution to the new problem, so the prototypical previous cases should be artificially collected first. Rules could be regarded as the refined knowledge of the previous cases. They are more concise and much smaller in scale, so the reasoning would be more efficient.

The Rule-based system provides relatively easier knowledge acquisition from experts. The Rule-based methods imitate the same way of how a human expert does in a particular task. Normally, such systems include a knowledge base and an inference engine. The former usually stores fault features and corresponding results in the form of production rule, and the latter determines which rules will be used and how to resolve a conflict. Unfortunately, there exists a bottleneck of knowledge acquisition for building a well-acknowledged rule base in practical uses. The MLEM2 algorithm was adopted to excavate rules [5], but only dimensionless parameters were used which made the rules redundant. The reasoning process was not very robust as well. There are also fault detection approaches which are fully acting in unsupervised manner, i.e. no labeled data needs to be provided which is often time-intensive to collect and annotate with the help of some special models [22,23].

The aim of this paper is to compare the abovementioned classification methods in the fault diagnosis of rotating machinery. The KNN, PNN, PSO-SVM and a new modified Rule-Based Method (RBM) based on MLEM2 algorithm were employed respectively as typical representations of the distance-based, neural network-based, statistic-based and knowledge-based mechanisms. They have something in common: all of them can be carried out on numerous data without troubles of discretization, and the categories can be output directly. The paper is directed to providing help for mechanical engineers to make an easy choice in designing and implementing various diagnosis tasks. In view of the above analysis, the data acquisition and feature extraction process was designed first. EMD is a credible, self-adaptive signal processing method that can be used to preprocess the nonlinear and non-stationary signals perfectly. Subjective effects of parameter setting can be avoided and the preparation time can be reduced. Eight time-domain parameters were adopted, and two of them were dimensional. At the same time, five frequency-domain indexes were designed to transform the rich faulty information contained in the envelope spectrum of Intrinsic Mode Functions (IMFs) into operable measures. Thus, a fault decision table (FDT) was obtained. All the thirteen parameters were sensitive to specific faults theoretically. Their practical effects on identification were evaluated and a simple feature selection method was utilized for removing the useless ones. Then, basic KNN, PNN algorithms, PSO-SVM and RBM were performed on the FDT respectively for fault classification. The results were discussed in terms of accuracy, time consumption, intelligibility, maintainability, and the interactions between the system and human in particular.

2. Brief review of KNN, PNN, PSO-SVM and MLEM2

2.1. KNN algorithm

As P features are picked up from a sample, in the KNN classification method, each training record is described in a P-dimensional space based on the value of each of its P input characteristics. The testing sample is then expressed in the same form, and its K nearest neighbors are chosen. The category of each of these K neighbors is then counted, and the category with maximum vote is designated as the result of the unknown sample. The K nearest neighbors are generally decided by calculating the Euclidean distance between the testing record and each of the training ones. The Euclidean distance ED_q between the testing sample TE_p and the qth training sample TR_{qp} is defined as:

$$\mathrm{ED}_q = \sqrt{\sum_{p=1}^{P} \left(\mathrm{TE}_p - \mathrm{TR}_{qp}\right)^2} \tag{1}$$

$p = 1, 2, \ldots, P, q = 1, 2, \ldots, Q$