Mechanical Systems and Signal Processing 50-51 (2015) 437–455

Contents lists available at ScienceDirect

Mechanical Systems and Signal Processing

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

A tri-fold hybrid classification approach for diagnostics with unexampled faulty states

Prasanna Tamilselvan, Pingfeng Wang*

Department of Industrial and Manufacturing Engineering, Wichita State University, Wichita, KS 67208, USA

A R T I C L E I N F O

Article history:
Received 10 September 2012
Received in revised form
20 August 2013
Accepted 12 May 2014
Available online 29 May 2014

Keywords:
Fault diagnostics
Health monitoring
Unexampled faulty

A B S T R A C T

System health diagnostics provides diversified benefits such as improved safety, improved reliability and reduced costs for the operation and maintenance of engineered systems. Successful health diagnostics requires the knowledge of system failures. However, with an increasing system complexity, it is extraordinarily difficult to have a well-tested system so that all potential faulty states can be realized and studied at product testing stage. Thus, real time health diagnostics requires automatic detection of unexampled system faulty states based upon sensory data to avoid sudden catastrophic system failures. This paper presents a trifold hybrid classification (THC) approach for structural health diagnosis with unexampled health states (UHS), which comprises of preliminary UHS identification using a new thresholded Mahalanobis distance (TMD) classifier, UHS diagnostics using a two-class support vector machine (SVM) classifier, and exampled health states diagnostics using a multi-class SVM classifier. The proposed THC approach, which takes the advantages of both TMD and SVM-based classification techniques, is able to identify and isolate the unexampled faulty states through interactively detecting the deviation of sensory data from the exampled health states and forming new ones autonomously. The proposed THC approach is further extended to a generic framework for health diagnostics problems with unexampled faulty states and demonstrated with health diagnostics case studies for power transformers and rolling bearings.

1. Introduction

Real-time health diagnosis of engineered systems interprets data acquired by smart sensors and distributed sensor networks and utilizes these data streams for critical lifecycle decision-makings [1–4]. System health diagnostics has been developed and promoted by various successfully engineered applications, including bearings [5,6], machine tools [7], transformers [8], engines and wind turbines [9,10] among many others. System health diagnostics provides diversified benefits such as improved safety, improved reliability and reduced costs for the operation and maintenance of complex engineered systems. Maintenance and life-cycle management is one area that is positioned to significantly benefit in this regard due to the pervasive nature of maintenance activities throughout both the manufacturing and service sectors. Maintenance and life-cycle management activities constitute a large portion of overhead costs in many industries [2]. These costs are likely to increase due to the rising competition in today's global economy. In the manufacturing and service sectors, unexpected breakdowns can be prohibitively expensive since they immediately result in lost production, failed shipping

* Corresponding author. Tel.: +1 316 978 5910.
E-mail addresses: pxtamilselvan@wichita.edu (P. Tamilselvan), pingfeng.wang@wichita.edu (P. Wang).

http://dx.doi.org/10.1016/j.ymssp.2014.05.016
0888-3270/

Nomenclature

AI	artificial intelligence
HS	health state
EHS	exampled health state
UHS	unexampled health state
MD	Mahalanobis distance
SVM	support vector machine
TMD	thresholded Mahalanobis distance
n	total number of initial EHS
r	total number of training data points
x_i	p-dimensional input vector, where $1 \le i \le r$
y_i	class label of ith data
MD_{ij}	Mahalanobis distance of the ith dataset with reference to jth HS, where $1 \le j \le n$
C	penalty parameter
μ_j	mean vector of the training data
s_j	variance matrix of the training data
w	normal vector perpendicular to hyper plane
b	offset of the hyper plane
α_i	lagrangian multiplier
ξ_i	slack variable
$k(\mathrm{x},\mathrm{x}_i)$	kernel function based on hyper plane
$\varphi(x)$	high dimensional feature space mapping

schedules, and poor customer satisfaction. In order to reduce and possibly eliminate such problems, real-time health diagnostics plays an increasingly important role, which requires automatic detection of system anomalies and adverse events in an early stage so that potential catastrophic system failures can be avoided. Successful implementation of system health diagnostics relies on not only advanced sensing technology, but more importantly the understanding of relationships between the features of multidimensional sensory signals and the underlining system health conditions [1]. In health diagnostics, the condition of an operating system is usually described by different health states (HSs) (for example, healthy, faulty state due to different component failures, etc.), and in general there are two categories of HSs distinguished based on the prior knowledge of the system: namely the exampled HS (EHS) and the unexampled HS (UHS). The EHS represents a category of system health conditions that the features of the sensory signals from them are clearly understood, and normally shown as training data. On the contrary, the UHS represents a category of system health conditions that the features of the sensory signals from them are not understood and normally shown as no prior knowledge or training data available. With an increasing system complexity, it is extraordinarily difficult to have a well-tested system so that all potential faulty states can be realized and studied at product testing stage. Thus, real time health diagnostics requires not only accurate identification of EHS, but also automatic detection of UHS through multidimensional sensory signals, as the misidentification or ignorance of UHS could lead to catastrophic system failures.

Although it is important to handle UHS in system health diagnostics, however, a rich literature has been focused on the anomaly detection, also referred to as outlier detection. Different with the classification of different HSs for health diagnostics, the anomaly detection aims at detecting patterns in a given data set that do not conform to an established normal behavior [11–20]. In general, three broad categories of anomaly detection techniques exist. Unsupervised anomaly detection techniques detect anomalies in an unlabeled test data set under the assumption that the majority of the instances in the data set are normal by looking for instances that seem to fit least to the remainder of the data set. Supervised anomaly detection techniques require a data set that has been labeled as "normal" and "abnormal" and involves training a classifier. Semi-supervised anomaly detection techniques construct a model representing normal behavior from a given normal training data set, and then testing the likelihood of a test instance to be generated by the learnt model. Hodge et. al [21] and Varun et. al [22] provide a good survey of the state-of-the-art anomaly detection techniques and applications. Compared with anomaly detection, the UHS identification are more challenging in that (1) in particular in the context of UHS identification, the sensory signal features from UHS are often not rare occurrences, but they are from unexampled health states. This pattern does not adhere to the common statistical definition of an outlier as a rare object, and many outlier detection methods, in particular unsupervised methods, will fail on such data; and (2) the UHS identification generally involves more number of HSs than the "normal" and "abnormal" in the anomaly detection.

Despite the studies in the outlier detection, existing approaches have been mainly focused on correct identification of EHS from sensory signals, given the importance of handling both EHS and UHS in health diagnostics. Several advanced HS classification approaches have been developed to accomplish the task of EHS identification through comparing online sensory signal features with the offline training data. Based on the working principles, these approaches can be generally categorized into two categories: artificial intelligence (AI) based techniques [11–13] and statistics inference based techniques [14–20]. The AI based diagnostics techniques involve a process of learning the relationship between input parameters and the desired target values in the form of set of patterns. Examples of AI based diagnostics techniques out of many include back-propagation neural network [23–27], deep belief networks [28–30], self-organizing maps [27,31], and genetic algorithm [32,33]. There are also some kernel based machine learning techniques in which the support vector machine (SVM) is one of the most popular methods for health diagnostics applications [34–38]. The key advantage of employing the kernel based machine learning technique is achieving an optimized classification solution while introducing the desired sparseness of training samples. Besides the AI based diagnostics techniques, there are statistical inference based classification approaches, such as Mahalanobis distance (MD) classifiers [39], k-nearest neighbor method [40] and k-mean clustering [41]. These classification approaches partition a set of patterns into different disjoint clusters based on certain statistical distance measure. The MD classifier is one of the popular statistical inference approaches, which determines the