Mechanical Systems and Signal Processing 40 (2013) 791–804

Contents lists available at ScienceDirect

Mechanical Systems and Signal Processing

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

A real-time approach to acoustic emission clustering

Eraldo Pomponi *, Alexei Vinogradov

Laboratory for the Physics of Strength and Intelligent Diagnostic Systems, Togliatti State University, 445667 Togliatti, Russia

ARTICLE INFO

Article history:
Received 26 August 2012
Received in revised form
11 February 2013
Accepted 15 March 2013
Available online 15 June 2013

Keywords:
Acoustic Emission (AE)
Real-time clustering
Sequential k-means
Fuzzy *c*-means
Machine learning
Bearings

ABSTRACT

A common target of clustering in acoustic emission (AE) non-destructive inspection technique (NDT) is to distinguish between the sources of different origin and to get a deeper insight into the interrelation between the underlying processes such as plastic deformation, crack initiation, corrosion cracking, etc. The major drawback of the most popular conventional schemes such as *k*-means and fuzzy *c*-means is that they are iterative in nature, which hinders their real-time applications. Inspired by the sequential *k*-means procedure, i.e. a non-iterative variant of the classic *k*-means, we present a novel classification technique designed for real-time applications. The proposed approach is "non-supervised", i.e., both the number of clusters and their elements are inferred from the data distribution in a multi-dimensional metric space. In its present form the approach is capable to adopt various dissimilarity measures to compare AE power spectral densities. A series of tests on different probing datasets has been performed to prove the efficiency of the proposed approach.

1. Introduction

The acoustic emission (AE) technique has emerged and developed primarily for the purpose of non-destructive testing (NDT) since about 1970s. The versatility of AE testing paired with rapid advances in computer science and signal processing are now bringing AE to the NDT forefront owing to its unique capacity for gaining crucial information regarding service conditions and damage accumulation in a broad variety of mechanical systems. In the wide landscape of NDT techniques, which are currently used in condition monitoring of mechanical systems, AE stands out against all others due to its unique capacity to monitor large scale facilities or mechanical systems in operation with location of structural damage in real-time. The identification—as early as possible—of the approaching critical stage is of primary importance to avoid premature failure. For efficient decision making, the incoming AE information must be processed quickly enough. The high data acquisition rate of the present apparatus (c.f. 10 Msamples/s per channel or faster) sets a new challenge for the entire processing chain including real time signal classification as a part. This industrially demanding and scientifically fascinating problem has motivated our research reported here. One of the most serious limitations in the routine NDT practice and academic research is that the AE signal is inevitably and considerably influenced by noise. The latter can be either a simple thermal noise generated steadily by electronic circuits or a transient noise which can be randomly induced by a variety of mechanical parts in the system under control. This transient noise and actual AE signals look pretty much alike so that much care must be taken to distinguish between them. Hence, the goal of the modern AE signal processing is twofold: (1) discrimination between AE signals arising on the background of noise and (2) identification of the active source among other possible multiple sources.

* Corresponding author.
E-mail address: eraldo.pomponi@gmail.com (E. Pomponi).

0888-3270/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.ymssp.2013.03.017

The AE signal can be represented in a broad variety of ways. The most common practice, which is used since the inception of the AE method, is to reduce the AE waveform to a set of few parameters calculated in time (peak amplitude, rms voltage, energy, number of counts, duration, rise time, etc.) or frequency domain (average frequency, partial energy, etc.). Depending on the generation of equipment and the feature extraction scheme, the set of variables can be wider or narrower. However, the point to be made is that many of these often loosely defined variables deliver very limited or no information about the emitting source itself. Moreover, some of them correlate strongly with each other (e.g. energy and rms voltage), making a large portion of variables redundant. Hence, the choice of the n-dimensional vector of descriptive variables is often poorly justified. Alternatively to this "parametric" description, the AE signal can be transformed into a frequency domain using one of the conventional transforms such as FFT, STFT, Wavelet, etc. The obtained frequency or time-frequency representation of the power spectral density (PSD) function can be used as a descriptive vector for further processing. In the present work we shall confine ourselves to the conventional Fourier transform albeit the proposed scheme can be extended, in principle, to other spectral techniques as well.

Various dissimilarity measures can be introduced to discriminate quantitatively between the AE PSD functions; we shall exemplify our approach with only one measure—the correlation distance—without loss of generality. For two vectors x_s and x_t, the correlation distance $d(x_s,\ x_t)$ is defined as a measure of statistical dependence between them as

$$d(x_s,\ x_t) = 1 - \frac{(x_s-\overline{x}_s)(x_t-\overline{x}_t)^T}{\sqrt{(x_s-\overline{x}_s)(x_s-\overline{x}_s)^T}\sqrt{(x_t-\overline{x}_t)(x_t-\overline{x}_t)^T}} \tag{1}$$

here the upper bar denotes the average of the variable.

Several data clustering techniques have been developed and proved effective for a huge variety of applications. k-means and fuzzy c-means schemes are among the best known and most widely used for this purpose. However, both are *iterative* processes. This hinders their application in real-time workflow. Hence, there is an apparent tradeoff between the great potential of the AE technique capable of working on line and clustering techniques which can be used "*post-mortem*" only. To meet the requirements for AE real time monitoring any clustering algorithm should create and assign clusters adaptively as AE signals appear one after another. The proposed algorithm inspired by the *sequential* k-means, i.e. a non-iterative variant of the classic k-means scheme, has this capacity.

Besides, clustering is computationally expensive, particularly for large data sets, and is time consuming mostly due to the iterative nature of schemes such as k-means, c-means or Self Organizing Maps (SOM). Another sensible issue to be considered in the choice of the categorization procedure is the amount of *a-priori* information which is required by supervised approaches. For instance, such popular classifiers as Neural Networks (NNs) or Support Vector Machine (SVM) need to be exposed to large enough training sets representing all anticipating source mechanisms, which might be impossible or expensive to obtain. The proposed non-supervised adaptive algorithm whereby the number of clusters and their elements are data driven, i.e. they are inferred from the data distribution in a multi-dimensional metric space, is flexible and efficient enough to overcome the above limitations.

2. Motivation

In application to AE, "batch" clustering of events has gained greatest popularity (e.g. [1–6], etc.) in attempts to group the signals originating from the same source and to separate that group from others. Two important points are to be borne in mind: (1) batch clustering requires all data to be available from the beginning until the end of the procedure; (2) the clustering problem is inherently ill-posed, and the quality of clustering is an ill-defined concept, i.e. multiple solutions can be expected [5,7,8] without a clear recipe for success and preference. Developing a robust and noise-resistant clustering procedure is a challenging task [9] which has not been elaborated enough as yet particularly for complex safety-critical real-time applications. Hence, the present work aims at reducing this obvious deficit. Being general in nature, the proposed scheme has been designed specifically for AE purposes by taking into account actual testing conditions and acquisition features. k-means is one of the most appealing clustering techniques for the real-time use due to its high speed paired with simplicity [10]. However, the conventional k-means has two major shortcomings:

1. k-means is an iterative procedure
2. The number of clusters k must be specified *a-priori*

The first issue is crucial for real-time application because it strongly affects the speed of the algorithm. The second issue is particularly troublesome, since, in general, there is no way to envision how many clusters and underlying processes can be faced in the course of the actual NDT test. Notwithstanding, MacQueen in his seminal paper [11] has outlined a possible way to address this issue by introducing two parameters: C ("coarsening") and R ("refinement"). The C parameter controls when two clusters have to be merged by "averaging" them so that the partition defined by the means is coarsened. The R parameter states that a certain data point should be assigned to a seed of a new cluster if the distance from this point to the closest mean is greater than R. As a result, k is increased and the partition is refined. Unfortunately, the issue is not resolved with the aid of these additional parameters, but it is rather transferred to a proper choice of C and R. Hence, in reply to a strong demand for efficient clustering procedures, the strategic idea is to develop a progressive clustering procedure