Engineering Structures 99 (2015) 439–448

Contents lists available at ScienceDirect

Engineering Structures

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

Structural modification assessment using supervised learning methods applied to vibration data

Vinicius Alves [a,1], Alexandre Cury [b,*], Ney Roitman [c,2], Carlos Magluta [c,3], Christian Cremona [d,4]

[a] Graduate Program in Civil Engineering, Federal University of Ouro Preto, Brazil
[b] Department of Applied and Computational Mechanics, University of Juiz de Fora, Juiz de Fora, Brazil
[c] Federal University of Rio de Janeiro, Civil Engineering Department, COPPE, Rio de Janeiro, Brazil
[d] Technical Centre for Bridge Engineering, CEREMA/DTITM, Sourdun, France

ARTICLE INFO

Article history:
Received 4 June 2014
Revised 6 March 2015
Accepted 4 May 2015
Available online 28 May 2015

Keywords:
Pattern recognition
Damage assessment
Learning algorithms
Symbolic data
SHM

ABSTRACT

Structural systems are usually subjected to degradation processes due to a combination of causes, such as design or constructive problems, unexpected loadings or natural ageing. Machine learning algorithms have been extensively applied to classification and pattern recognition problems in the past years. Some papers have addressed special attention to applications regarding damage assessment, especially how these algorithms could be used to classify different structural conditions. Most of these works were based on the comparison of measured vibration data such as natural frequencies and vibration modes in undamaged and damaged states of the structure. This methodology has proven to be efficient in various studies presented in the literature. However, its application may not be the most adequate in cases where the engineer needs to know with certain imperativeness the condition of a given structure. This paper proposes a novel approach introducing the concept of Symbolic Data Analysis (SDA) to manipulate raw vibration data (signals, i.e. acceleration measurements). These quantities (transformed into symbolic data) are combined to three well-known classification techniques: Bayesian Decision Trees, Neural Networks and Support Vector Machines. The objective is to explore the efficiency of this combined methodology. For this purpose, only raw information are used for feature extraction. In order to attest the robustness of this approach, experimental tests are performed on a simply supported beam considering different damage scenarios. Moreover, this paper presents a study with tests conducted on a motorway bridge, in France where thermal variation effects also have to be considered. In summary, results obtained confirm the efficiency of the proposed methodology.

1. Introduction

Over the last years, numerous methods for structural damage assessment have been proposed in the literature. The basic (but not necessarily the easiest) objective is to determine whether a structure presents an abnormal behavior or not [1,2]. An indicator sensitive to damage is a quantity extracted from the measured system's response that is able to indicate the presence of a structural change. Identifying features that can accurately distinguish a damaged structure from an undamaged one is the focus of most Structural Health Monitoring (SHM) techniques [3,4]. One of the most common methods of feature extraction comes from correlating observations of measured quantities with posterior observations of the degrading system. Many techniques have been proposed to locate and quantify damage considering deviations from vibration data and more precisely from modal parameters [5–7]. The basic idea is that since they are functions of the physical properties of the structure (mass, damping, and stiffness), changes in these properties would lead to modifications in the modal parameters and the measured responses of the structure. Some other damage indicators can be built from the modal parameters (flexibilities, flexibility curvatures, modal curvatures, strain-energy ratios) that can enhance damage detection [8,9].

This paper attempts to address these different issues through the monitoring of a simply supported beam tested in laboratory

* Corresponding author.
E-mail addresses: vnichio@hotmail.com (V. Alves), alexandre.cury@ufjf.edu.br (A. Cury), ney@coc.ufrj.br (N. Roitman), magluta@coc.ufrj.br (C. Magluta), christian.cremona@developpement-durable.gouv.br (C. Cremona).
[1] Address: Campus Universitário, Morro do Cruzeiro, Ouro Preto, Minas Gerais, Brazil.
[2] Address: Rua José Lourenço Kelmer, Campus Universitário, Juiz de Fora, Minas Gerais, Brazil.
[3] Address: Campus Universsário, Ilha do Fundão, Rio de Janeiro, Brazil.
[4] Address: 110 rue de Paris, 77171 Sourdun, France.

http://dx.doi.org/10.1016/j.engstruct.2015.05.003
0141-0296/

and of a motorway box girder bridge located in France. As previously mentioned, most techniques for damage identification are essentially based on the determination of modal properties through an identification process. Nevertheless, the identification of modal parameters is a sort of filtering process, which leads, by nature, to a loss of information compared to raw data (signals). This data compression process can filter any small changes due to a structural modification. Furthermore, and this is certainly the major drawback when using modal parameters, modal components are essentially describing an equivalent linear behavior, a feature which may not be exact for the analysis of specific degraded systems. On the other hand, using raw dynamic measurements (especially if high sampling frequencies are used) leads to the storage of large data sets. Thus, dynamic measurements can easily contain over thousands of values, usually making an analysis process extensive and prohibitive. Despite the current processing power of computers, the necessary computational effort to manipulate large data sets remains a problem.

To overcome these limitations, this paper proposes the use of a special set of data manipulation techniques. *Data mining* [10] is the process of extracting hidden patterns or features from data. As more data are gathered in monitoring, *data mining* is becoming an increasingly important tool to transform these data into information and is being used in a wide range of profiling practices, such as marketing, fraud detection and scientific discovery.

Different types of data can be employed and manipulated in *data mining*, such as a single quantitative or categorical values, interval-valued data, multi-valued categorical data, and modal multi-valued (histograms) [11,12]. These types of data are generally called "symbolic data" and they allow representing the variability and uncertainty present in each variable. The development of new methods for data analysis suitable for treating this type of data is the main issue of Symbolic Data Analysis (SDA). This paper proposes to combine SDA with classification methods in order to separate different structural states. The major advantage of such combined approach is that enhanced – yet raw – information is used, i.e. histograms, intervals, etc. and they can be applied to manipulate vibration data (acceleration measurements, for example). When these quantities are converted into symbolic data, this piece of information will be applied to three well-known supervised classification methods: Bayesian Decision Trees, Neural Networks and Support Vector Machines. The main objective here is to assess structural modifications due to damage or any other foreign event, such as reinforcement procedures, different types of traffic loads, among others. However, when applying vibration-based damage detection to SHM, changes in vibration signatures are not only based on changes in any physical property. Environmental changes, notably temperature variations, can have a significant effect and it is necessary to take this into account for the evaluation of structural integrity [13,14].

This combined methodology (SDA + supervised classification methods) has already proven its efficiency when modal parameters are used [11,15]. Now, this paper attempts to answer the following questions: (i) is it possible to classify different structural states using raw data (measured accelerations) only?; (ii) how important it is to consider the effects of temperature variation when it comes to classifying different structural conditions, i.e. detecting structural damage? Encouraging results are obtained and they show evidence that environmental effects plays an important role in the field of SHM.

2. Symbolic data analysis overview

In general, data acquisition campaigns in Civil Engineering structures gather thousands of accelerations values measured by several sensors. Consequently, analyzing all of these data (classical data) directly may usually be time-consuming or even prohibitive. In this sense, transforming this massive quantity of data into a compact but also rich and descriptive type of data (symbolic data) becomes an attractive approach. Let us consider, for instance, a signal X (which is part of a dynamic test) containing 5000 acceleration values measured by one single sensor (see Fig. 1 on the left). There are several ways to transform classical data into symbolic data. This signal can be represented by:

- a k-category histogram: $X = \{1(0.0025), 2(0.0721), 3(0.8546), 4(0.0626), \ldots, k(0.0082)\}$;
- an interquartile interval: $X = [-0.012; 0.015]$;
- a min/max interval: $X = [-0.025; 0.025]$.

Fig. 1 (on the right) shows how a classical signal (one sensor) is converted to a symbolic representation. In this case, all acceleration values are projected to the y-axis of coordinates and a 20-category histogram is constructed. In fact, it must be noted that the same representation could be applied to modal parameters, i.e. natural frequencies and mode shapes. In other words, both of these quantities can be represented by intervals or histograms. Transforming classical data to symbolic data is carried out almost instantaneously, which does not prohibit or make difficult the use of this methodology for a large ensemble of dynamic tests.

In fact, when this transformation procedure is carried out, two important aspects must be taken into account. The first one relies on the conservation of some statistical properties of the original data, i.e. the moments of first order (mean value) and second order (variance). Higher order moments (skewness and kurtosis) are not considered here. The second aspect refers to the number of non-zero categories. It is not of interest to keep categories with values equal to zero, since they will not contribute in the classification procedures. Thus, in this paper, 10-category histograms are used in the SDA process, since it has proven to be the most efficient transformation for this type of analysis [13].

3. Classification methods

This section presents an overview of three supervised classification methods. Firstly, some concepts regarding Bayesian Decision Trees (BDT) and its applicability are explained, followed by a brief discussion about how Neural Networks (NN) are used in this study. Finally, a general idea of Support Vector Machines (SVM) applied to classification problems is presented.

These methods were selected following the works of [12,13]. In those references, several supervised classification techniques coupled with SDA were tested using either artificial or real controlled (labeled) data. In general, BDT, NN and SVM achieved the best correct classification ratios. This is why they are used in this paper.

3.1. Bayesian Decision Trees

Bayesian Decision Trees (BDT) are a decision procedure that is able to solve classification problems [16]. The general idea of this method is to classify a particular object (dynamic test, in this case) into one of the classes (groups) previously defined i.e. in the training set. For instance, let $\Omega = \{T_1, T_2, \ldots, T_n\}$ be a set of n dynamic tests and C a class variable containing values varying within $\{1, \ldots, m\}$ where m is the number of classes. This discriminant analysis tries to predict the unknown value C for a given test $\widetilde{T}$ according to its p features[5] and a training set. The steps of this symbolic classification procedure are to represent a given partition in the

[5] Features are the symbolic representations of sensors (accelerometers).