CILAMCE 2017
Florianópolis - Santa Catarina - Brazil

A COMPARATIVE STUDY OF CLUSTERING METHODS APPLIED TO STRUCTURAL DAMAGE DETECTION

Lucas S. V. Melo

lucas.senna@engenharia.ufjf.br

Undergraduate Student, Federal University of Juiz de Fora, Juiz de Fora, Brazil

Alexandre A. Cury

alexandre.cury@engenharia.ufjf.br

Adjunct Professor

Federal University of Juiz de Fora, Juiz de Fora, Minas Gerais, Brazil.

Abstract. *Structural Health Monitoring is one of the most promising and challenging areas of research in the field of Civil Engineering. Over the last decades, researchers have focused on the development of consistent and reliable indicators aiming to detect, locate, quantify or even predict damage. More recently, some researchers are focusing on the use of raw time histories extracted from structural dynamic monitoring to build damage indicators. In this sense, this paper has as its main interest the use of symbolic data analysis (SDA) coupled with clustering techniques i.e. the k-means and c-means algorithms to detect structural modification (damage). The approach is applied directly to dynamic measurements (accelerations) obtained on site. The efficiency of such methodology is attested by means of a numerical study performed on a model of a simply supported beam and a study based on a real case railway bridge, in France. Results show that HOS coupled with clustering techniques are able to differentiate damage scenarios with adequate classification rates.*

Keywords: *Damage detection, Clustering methods, Raw data, SHM.*

INTRODUCTION

Structural Health Monitoring (SHM) is of great importance to Civil Engineering, once it allows the detection of modifications in the physical properties of structures. SHM permits, if possible and/or necessary, the use of recovery procedures in suitable time. These procedures are often based on evidence collected from tests performed on the structure that would ideally allow 'detecting, locating, quantifying, and even predicting damages'. Recently, researches have focused on the dynamic evaluation as part of structural diagnoses [1][2], by the extraction of modal parameters or data built from these parameters that, so far, have provided promising results. However, some problems remain unsolved, such as the sensitivity of the damage identification methods, their need of a reference state and their reliability when it comes to the detection of false alarms [3].

Historically, damage detection techniques were developed under the premise that structural damage could be detected by observing the variation of the structure's vibration characteristics, i.e. natural frequencies, damping ratios, and mode shapes. This premise was and is correct, since these parameters are directly affected by modifications in the structure's mass and stiffness. However, the identification of modal parameters usually demand time and, ultimately, represent some sort of filtering process, yielding a loss of information compared to the raw acceleration data. This "filtering" process might erase any small tendencies, which could represent a structural modification. Moreover, modal components usually describe an equivalent linear structural behavior, which is not always the case after a damage occurs. On the other hand, raw data may contain thousands of accelerations values measured by numerous sensors. Thus, examining all of these data directly may take time or even be prohibitive, despite the current computers' processing power [4]. In this sense, it is imperative to transform these substantial amounts of data into a more compact yet rich descriptive type of data.

In this paper, raw data obtained from dynamic tests (acceleration measurements) are transformed into a more compact arrangement by computing symbolic data for each sensor e.g. accelerometer used. Then, these quantities are applied to two clustering algorithms in order to discriminate different structural states or, in other words, to detect damage. This paper is based on the conjecture that the symbolic data are sensitive to damage, meaning that they can provide information regarding variations on the physical properties of structures that could indicate the existence of damage. Thus, the clustering algorithms would eventually be able to identify different structural scenarios.

The key objective of this paper is to develop a methodology capable of detecting structural damage by clearly differentiating physical states that correspond to an "undamaged" configuration from a "damaged" configuration. However, it must also be reminded that this is a rather complex problem, since the proposed methodology is applied directly to the raw data i.e. the raw dynamic measurements.

METHODOLOGY

This section presents the main concepts within the framework of this paper. First, a brief description of SDA is presented. Then, a short explanation about k-means and c-means clustering algorithms is given (more details can be found in [5]).

Proceedings of the XXXVIII Iberian Latin-American Congress on Computational Methods in Engineering
P.O. Faria, R.H. Lopez, L.F.F. Miguel, W.J.S. Gomes, M. Noronha (Editores), ABMEC, Florianópolis, SC, Brazil, November 5-8, 2017.

Symbolic data analysis

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 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 n acceleration values measured by one single sensor. There are several ways to transform classical data into symbolic data. This signal can be represented by:

- a k-category histogram: $X = \{a_1(n_1), a_2(n_2), a_3(n_3), \ldots, a_k(n_k)\}$;
- an interquartile interval: $X = [a_{25\%}; a_{75\%}]$;
- a min/max interval: $X = [a_{min}; a_{max}]$.

Figure 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. 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. 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.

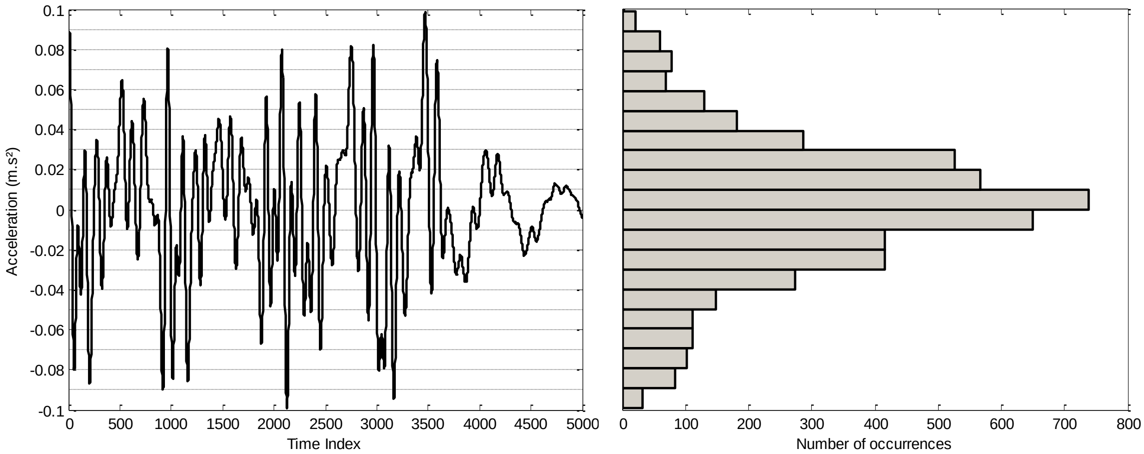

Figure 1 – Example of transforming classical signal to symbolic signal (20-category histogram).

K-means clustering algorithm

This clustering method was developed over a generalization of the classical dynamic clusters method. Clustering of a dataset is the partition of that data into groups, named clusters, so that the data inside a cluster has the highest degree of similarity among all possible combinations. What defines this 'similarity' is the clustering algorithm. In this paper, the chosen algorithm - k-means - uses the spatial distance between a data point (a dynamic test) to the centroid of the cluster. The idea is to identify the best combination of data that produces the clusters with the highest degree of similarity, that is, the smallest data-centroid sum of distances. The k-means algorithm has several metrics to calculate those spatial distances. In this paper, the metrics used are the square Euclidean and the cityblock, shown in Eq. (1) and (2), respectively:

Proceedings of the XXXVIII Iberian Latin-American Congress on Computational Methods in Engineering
P.O. Faria, R.H. Lopez, L.F.F. Miguel, W.J.S. Gomes, M. Noronha (Editores), ABMEC, Florianópolis, SC, Brazil, November 5-8, 2017.

$$d(E_i, E_j) = \sqrt{\sum_{k=1}^{n} (E_{ik} - E_{jk})^2} \tag{1}$$

$$d(E_i, E_j) = \sum_{k=1}^{n} |E_{ik} - E_{jk}| \tag{2}$$

Once the statistics were extracted, the resulting datasets were analyzed through clustering algorithm k-means.

C-means clustering algorithm (FCM)

As fully described in [8], "A fuzzy set is a class of objects that has a continuum of grades of membership. Such a set is characterized by a membership (characteristic) function that assigns a grade of membership that ranges between zero and one to each object. The FCM is an iterative algorithm clustering method that produces optimal c-partitions by minimizing the weighted dissimilarity function [7]

Let T_i $(i = 1,,2,...,n)$, which represents a m-dimensional vector (m is the number of categories of each dynamic test transformed into symbolic data). This notation is used to determine the cluster centers CC_{qr} for the q^{th} cluster and its r^{th} dimension by using the expression given below:

$$CC_{qr} = \frac{\sum_{i=1}^{n} U_{iqr}^{f} T_{ir}}{\sum_{i=1}^{n} U_{iqr}^{f}} \tag{3}$$

where C is the number of the cluster to be made $(2 \le C \le N)$, and f is an appropriate level of cluster fuzziness $f > 1$. U is the membership matrix, which has the size $n \times C \times m$ and is first initialized randomly such that $U_{iqr} \in [0,1]$ and $\sum_{q=1}^{C} U_{iqr} = 1$ for each i and a fixed value of r.

Then, the Euclidean distance is evaluated between the i^{th} data point and the q^{th} cluster with respect to the r^{th} dimension with the equation below:

$$D_{iqr} = \left\| T_{ir} - CC_{qr} \right\| \tag{4}$$

In the following step, CC_{qr} is updated in Eq. (4) by recalculating the fuzzy membership matrix U according to D_{iqr} (if $D_{iqr} > 0$), as follows:

$$U_{iqr} = \frac{1}{\sum_{c=1}^{C} \left(\frac{D_{ir}}{D_{icr}} \right)^{f^{2}-1}} \tag{5}$$

This updating iteration is repeated until the changes in U are sufficiently small such that $(U \le \varepsilon)$, where ε is a predefined termination criterion.

The main difference between hard clustering (k-means) and soft-clustering (FCM) relies on the fact that hard-clustering methods can only assign a given object to a unique cluster. Conversely, soft-clustering methods allow classifying a given object as a part of different the clusters. This allows defining the so-called pertinence values. The higher these values are, more likely a given object is to be classified into a cluster. In this paper, the aforementioned procedure is coupled with symbolic data transformation within Matlab® software framework. In fact, fuzzy c-means procedures are built-in in the Fuzzy Toolbox. The Euclidian distance was used and the level of fuzziness was set to 2."

RESULTS

It should be noted that the above-mentioned clustering methods were already applied to structural modal parameters (natural frequencies and mode shapes) obtaining very good results [1]. Thus, the main purpose of this paper is to explore the improvements/drawbacks of the proposed techniques when dealing with raw accelerations measurements directly.

Hence, the methodology follows two steps:

1. Evaluation of symbolic data for each accelerometer;
2. Use SD (step 1) as inputs to the clustering techniques (k-means and c-means).

It is worth mentioning that these steps strongly depend on the quality of the input data. In other words, if measured data are poorly sampled, or have missing or even incorrect values, the results obtained by the proposed approach will be surely compromised.

In this paper, two sets of data were used for the validation of the proposed methodology. In both cases, it was previously known which data corresponded to the 'undamaged' and 'damaged' states. Therefore, the goal for both clustering algorithms is to allocate the 'undamaged' tests into one cluster and the 'damaged' tests into another. Thus, it would show that this technique is able to distinguish between two groups related to two different physical structural conditions.

Numerical application – simply supported beam

The first dataset corresponds to a numerical simulation of the vibration response of a 6-meter simply supported steel beam. The beam is discretized into 200 finite elements using Matlab. A random force is applied at 0,69m from the right support as shown in Fig. 2. The beam has the following physical and geometric properties:

- Elastic modulus – 210 GPa;
- Volumetric mass – 7850 kg/m³;
- Cross-sectional area – 2,81 x 10^{-3} m²;
- Moment of inertia – 2,81 x 10^{-8} m^4.

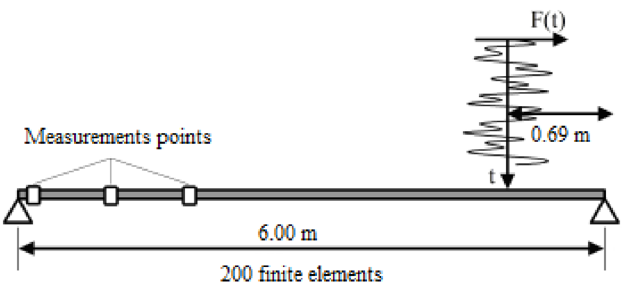

Figure 2. Discretized beam.

Dynamic measurements are taken at 10 equidistant points of the beam during 100s. The sampling frequency is 0.01 Hz, which corresponds to 10.000 acceleration measurements per sensor. Damage is simulated as the reduction of the elastic modulus of the beam in three progressive states, yielding three damage levels:

- Undamaged (D0): unaffected beam;
- Level 1 (D1): reduction of 20% in the elastic modulus of elements 96-105;
- Level 2 (D2): same as level 1 + reduction of 10% in the elastic modulus of elements 146-155.

Additionally, three levels of white noise (0, 5% and 10%) are added to the data.

The main objective here is to separate two different damage levels at a time (D0 from D1 and D0 from D2) for each noise level. Each dataset of the numerical simulations contains measurements from 10 accelerometers and, for each accelerometer, 10 SD are calculated (as shown in Table 1). Thus, each dataset yields a 20x100 matrix (20 tests by 100 SD). Tables 2 and 3 present the results obtained for the k-means method.

Table 1. Percentages of correctly classified data (levels D0 and D1, k-means).

	No noise		5% noise		10% noise	
	sqeuclidian	*cityblock*	*sqeuclidian*	*cityblock*	*sqeuclidian*	*cityblock*
D0	80%	80%	80%	80%	80%	80%
D1	80%	80%	80%	80%	80%	80%

Table 2. Percentages of correctly classified data (levels D0 and D2, k-means).

	No noise		5% noise		10% noise	
	sqeuclidian	*cityblock*	*sqeuclidian*	*cityblock*	*sqeuclidian*	*cityblock*
D0	70%	80%	70%	80%	70%	80%
D2	90%	70%	90%	70%	90%	70%

It is possible to notice that the proposed methodology is rather capable of distinguishing both damage levels, although the correct rates were lower for the second case. Tables 4 and 5 present the results obtained using the c-means technique.

Table 4. Percentages of correctly classified data (levels D0 and D1, c-means).

	No noise	5% noise	10% noise
D0	80%	80%	80%
D1	80%	80%	80%

Table 5. Percentages of correctly classified data (levels D0 and D2, c-means).

	No noise	5% noise	10% noise
D0	70%	70%	70%
D2	90%	90%	90%

From Tables 4 and 5 one notices that the c-means algorithm yields better results compared to the k-means method. Furthermore, it is important to remark that both methods are completely insensitive to noise, since all rates remained unchanged throughout the simulations.

Experimental application – TGV viaduct

The second application comprehends the measurements taken in the PK 075+317 viaduct in Southeast France, between Paris and Lyon, over which passes high-speed rails for TGV (*Train à Grande Vitesse*) trains. The viaduct is 17.5 meters wide (Fig. 3).

Figure 3. Side view of viaduct.

This bridge was built in the early eighties; the increase of the operating speed of TGVs has moved the excitation frequency of the trains close to the first natural frequency of the bridge. This risk of resonance was furthermore increased by the uncertainties in the mass of the ballast disposed on the bridge. The first natural frequency was 5.86 Hz and the excitation frequency was around 4.0 Hz. The French railways SNCF considered that this difference was not enough and ballast recharging in connection with new operating speeds could reduce it even more. Therefore, SNCF set up a system of rods near the bearings tightened by torque wrench (Fig. 4); this strengthening brought stiffness and increased the natural frequencies. In 2003, a strengthening intervention was scheduled and led to a change in natural frequencies.

Figure 4. Tightening procedure.

In order to assess the efficiency of the tightening procedure, eight vertical accelerometers were installed under the bridge's deck, having the sampling frequency fixed at 4096 Hz. 28 dynamic tests were carried out i.e. 15 tests before and 13 after the procedure. Thereby, this dataset can be represented by a 28x80 matrix containing 10 SDA for each accelerometer.

Table 6 presents the results obtained by both clustering techniques. It is possible to observe that such approaches were not able to differentiate the two structural scenarios adequately.

Table 6. Percentages of correctly classified data within cluster.

	k-means		*c-means*
	sqeuclidian	*cityblock*	
Before	100%	86,67%	86,67%
After	7,69%	69,23%	53,85%

CONCLUSIONS

This paper introduced a novel approach based on the coupling of Symbolic Data Analysis with k-means and c-means clustering algorithms. The main goal was to discriminate different structural behaviors using only raw information for feature extraction.

To assess the efficiency of the proposed approach, two applications were studied: the cases of a FEM beam model and of a real-case railway viaduct. For the first case, different damage and noise levels were simulated. It was noticed that although the procedure was insensitive to noise, it was not properly capable of distinguishing and separating the structural states. For the experimental application, the same methodology was applied, yielding poor results. Moreover, the c-means presented a better performance compared to the k-means.

It is important to highlight the complexity of such an analysis, since it deals directly with raw data measurements. However, the authors are aware that results must be improved before using this methodology to untested structures.

ACKNOWLEDGMENTS

The authors would like to thank UFJF (Universidade Federal de Juiz de Fora - Federal University of Juiz de Fora), CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior), CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico - "National Council of Technological and Scientific Development") and FAPEMIG (Fundação de Amparo à Pesquisa do Estado de Minas Gerais) for the financial support.

REFERENCES

[1] Alves, V., Cury, A., Roitman, N., Magluta, C., Cremona, C., "Structural modification assessment using supervised learning methods applied to vibration data", Engineering Structures, 99, 439-448, 2015.

[2] Cury, A., Cremona, C., "Assignment of structural behaviors in long-term monitoring: Application to a strengthened railway bridge". Structural Health Monitoring, 11, 422-441, 2012

[3] Cury, A., Cremona, C., Diday E., "Application of Symbolic Data Analysis for structural modification assessment", Engineering Structures, 32(3), 762-775, 2010.

[4] Santos, J.P., Cremona, C., Orcesi, A.D., Silveira, P., "Multivariate statistical analysis for early damage detection", Engineering Structures, 56, 273-285, 2013.

[5] Madhulatha, T.S., "An overview on clustering methods", IOSR Journal of Engineering, 2(4), 719-725, 2012.

[6] Farrar, C., Worden, K, "Structural Health Monitoring: a machine learning perspective." Chichester. Wiley. 2013.

[7] Bezdek, J.C. Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum, New York. 1981.

[8] Alves, V., Cury A., Roitman N., Magluta C., Cremona C., "Novelty detection for SHM using raw acceleration measurements", Structural Control and Health Monitoring, 22, 1193-1207, 2015.