VIBRATION-BASED STRUCTURAL DAMAGE IDENTIFICATION USING DATA MINING

Meisam Gordan, Zubaidah Ismail, Hashim Abdul Razak and Zainah Ibrahim

University of Malaya, Department of Civil Engineering, 50603 Kuala Lumpur, Malaysia
email: zu_ismail@um.edu.my, meisam.gordan@gmail.com

When a structure is damaged, the dynamic characteristics of the structure will change. Generally, main components of a structural health monitoring system are (1) data collection approach including a network of sensors for collecting the response data and (2) an extraction technique to obtain information on the structural health condition. Data mining (DM) is one of the emerging data extraction techniques. DM can play an important role to find out the hidden patterns in databases. Generally, this sophisticated tool is employed to find the relationship between data in datasets. Models and patterns, which are obtained from DM process, are used to make predictions. In this study, frequency response function (FRF) measurements obtained from experimental modal analysis of an intact and damaged composite girder deck are used as inputs for data mining to extract the principal components (PCs) of raw FRF data. Experimental modal analysis of the structure is carried out by exerting incrementally enhanced damage severity at specific location. Totally, 6 damage scenarios are considered with depth of 15 mm to 75 mm with the increment of 15 mm at the mid-span of the structure. In the modelling phase of DM process, principal component analysis (PCA) is employed to train a model. The performance of the model is illustrated by comparing the original FRFs and reconstructed FRFs with first 10 PCs.

Keywords: Frequency response function, data mining

1. Introduction

In civil, mechanical and aerospace in-service structures, damage can occur with changing the structural properties (mass, stiffness or damping) and lading to change in the dynamic responses, such as natural frequencies, mode shapes, damping ratios and frequency response functions (FFRs). The propagation of damage can lead to out-of-service conditions. Therefore, structural health monitoring (SHM) is a significant tool to ensure structural safety, integrity and minimal maintenance [1], [2]. Visual inspections are common damage evaluation techniques. However, they are time consuming, costly and inapplicable for continuous monitoring [3], [4]. Thus, a more accurate and faster method is needed to monitor structures for the occurrence, location and extent of damage.

The main components of a SHM approach are a data collection approach including a network of sensors for collecting the response data and an extraction technique to obtain information on the structural health condition [5]. Significant knowledge can be extracted from raw databases after their creation by monitoring SHM system. New and emerging computing and information technologies such as knowledge engineering, evolutionary computing and data mining can play a significant role in civil engineering. Therefore, civil engineers have slowly embraced computing and information technologies in the past decade [6]. Data mining is employed to analyse the raw datasets to find the most important relationship between data [7] and it presents a potential solution to SHM as a deeper data analysis approach [8]. Hence, data mining can be used as an extraction technique to obtain information on the structural health condition.

Directly-measured FRF is one of the easiest dynamic responses to obtain in real-time. It only requires a small number of sensors and straightforward in situ measurement. As modal parameters are indirectly-measured test data, they could be contaminated by measurement error and modal extraction error. In addition, modal parameters can provide smaller database in compare to FRF data [9]. For these reasons, it is more reliable to employ FRF as an input for data mining process.

This paper describes a data mining approach for the purpose of feature extraction and data reduction of FRF using principal component analysis (PCA) technique. The main focus of this paper is to illustrate the applicability of data mining for extraction the most important uncorrelated variables in FRF data obtained from experimental modal analysis of a healthy and damaged composite girder deck utilizing Cross-Industry Standard Process for Data Mining (CRISP-DM) model.

2. Frequency response function

Experimental modal analysis is used to identify damage by assessing structural responses before and after damage. FRF is the most important measurement of experimental modal analysis obtained from a data acquisition system. The time domain measured responses of structures are transformed into frequency domain utilizing Fast Fourier Transform (FFT). As shown in Fig. 1, FRF (H(ω)) is the ratio of the Fourier transform of output responses (X(ω)) to the Fourier transform of input excitation (F(ω)) [10]–[12].

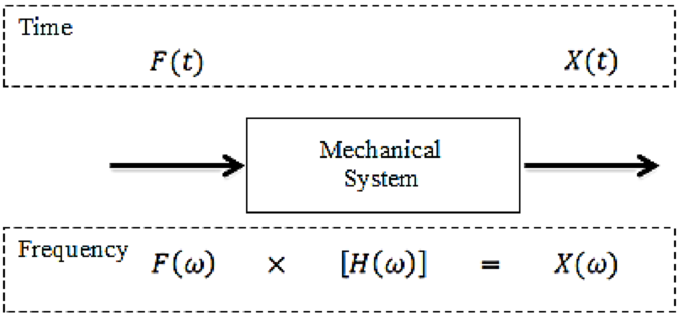

Figure 1: Black diagram of an FRF.

3. Feature extraction data mining process

Data mining is the analysis of datasets to discover the relationships, new correlations, trends and to extract the useful data in the form of patterns [13]–[15]. Data mining has different tools such as Knowledge Discovery in Databases (KDD), SEMMA, Cross-Industry Standard Process for Data Mining (CRISP-DM) ,etc. [16]. The most widespread application amongst the tools is CRISP-DM [17]. It was presented by a group which consist the firms of NCR System Engineering Copenhagen from USA and Denmark, Daimler Chrysler AG from Germany, SPSS Inc. from USA and OHRA Verzekeringen en Bank Groep B.V from Netherlands [18], [19]. As shown in Fig. 2, this model has six stages which are business understanding, data understanding, data preparation, modelling, evaluation and deployment.

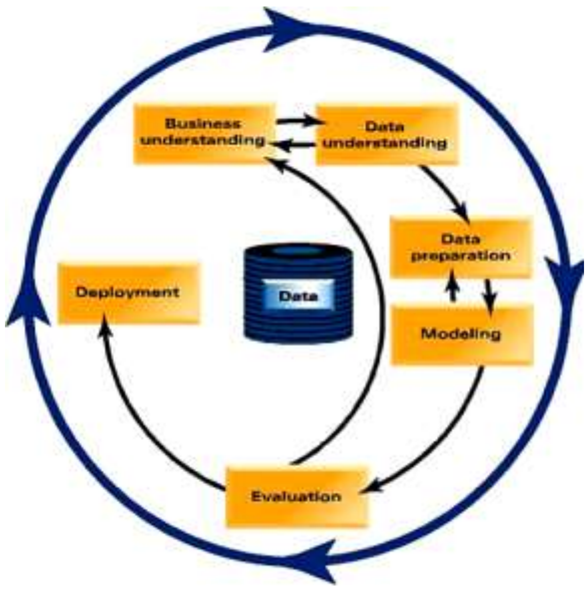

Figure 2: CRISP-DM data mining procedure [18], [20].

3.1 Business understanding

This phase emphases on understanding the project objectives and requirements from a business perspective, then converting this knowledge into a data mining problem definition and a preliminary plan designed to achieve the objectives [19], [21]. Overall business objective of this research is to detect damage in a composite girder deck structure. The model consists of three steel beams and a slab. Figure 3 illustrates the experimental setup of the structure in the laboratory. Also, a schematic diagram of the experimental test setup of the structure is demonstrated in Fig. 4. The length of the structure is 3200 mm. The dimensions of the beams include the flange width of 75 mm, section depth of 150 mm and thickness of 7 mm and 5 mm for the flange and web, respectively. The Young's Modulus of the steel is $2.1{*}10^{11}$ kg/m^2, with Poisson's ratio of 0.3 and density of 7,850 kg/m^3. The dimension of the slab includes the width of 1200 mm, the depth of 100 mm, and length of 3200 mm. The main purpose of this study is to extract the features of FRF data in the composite girder deck utilizing data mining.

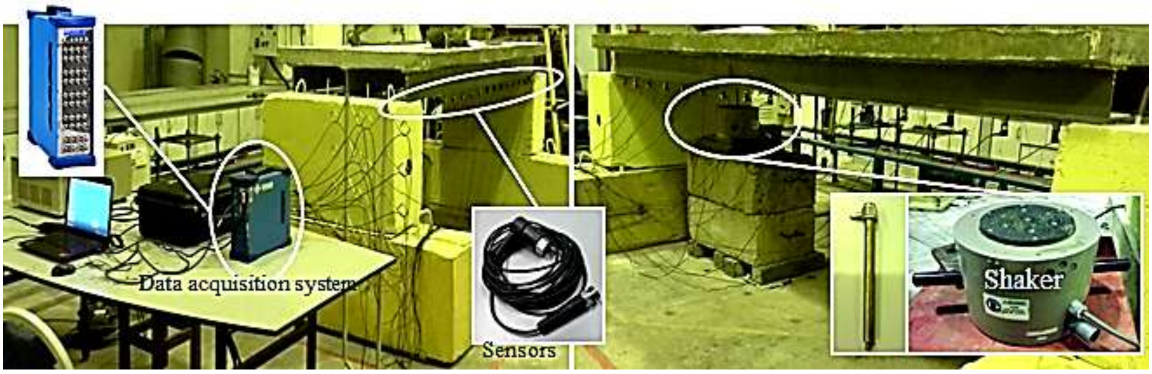

Figure 3: Experimental setup of the specimen.

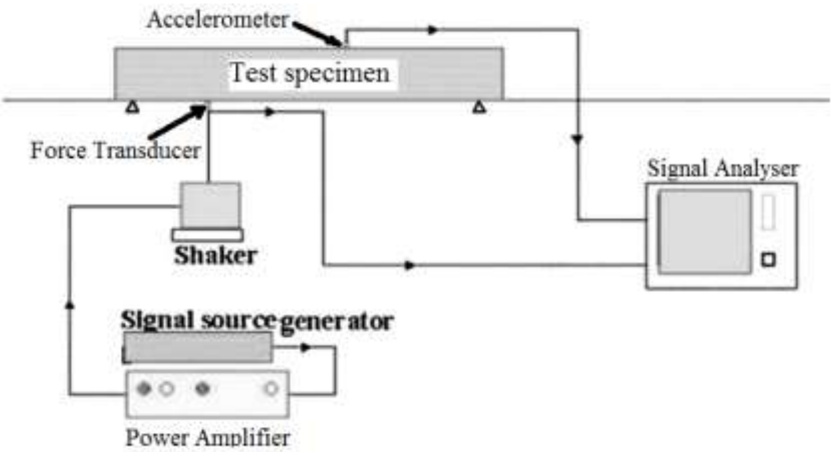

Figure 4: Schematic diagram of the experimental setup.

3.2 Data understanding

This stage starts with an initial data collection and proceeds with activities in order to get familiar with the data [22]. Focus is on feature extraction of measured FRF data using data mining. Therefore, experimental modal analysis was carried out to collect the FRFs of the intact and damaged structure. Measured FRFs generate a dataset for data mining process. The arrangement of the accelerometers was chosen to have 15 points in three sets between supports. Firstly, experimental modal test was implemented utilizing undamaged structure as a reference. Then, a number of damage scenarios were produced by introducing five damage severities at mid span of the structure, as shown in Fig. 5. The damage severities with 5 mm width and depth of 15-75 mm with the increment of 15 mm were gradually induced for each severity. Finally, the result of the modal testing is used as dataset for data mining. FRFs are measured 15 positions, as shown in Fig. 6.

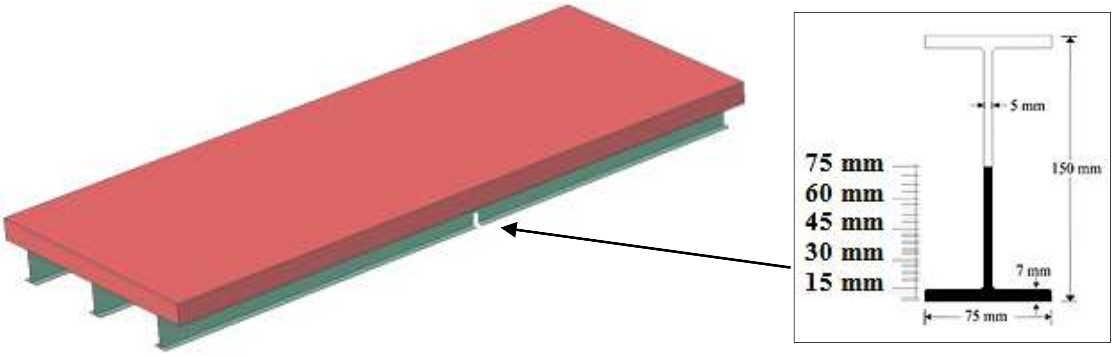

Figure 5: Location and severities of damage.

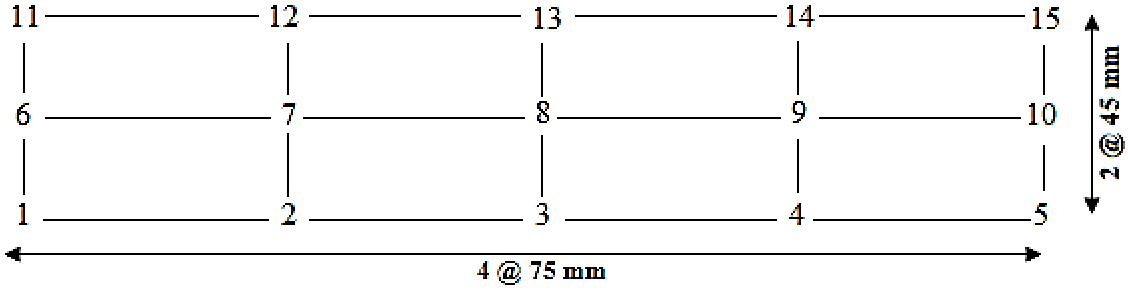

Figure 6: Accelerometers positions.

3.3 Data preparation

This stage covers all activities to construct the final dataset from the initial raw data. The final dataset is used for the modelling step. Different tasks of data preparation step include the selection of data including attribute selection and tabulate the data, the cleaning of data, the construction of data, the integration of data, the formatting of data, and transformation of data for modelling step [23]. The FRF data containing 3201 spectral lines from the test structure with undamaged case and different damaged cases are shown in Fig. 7 for both healthy and damaged cases.

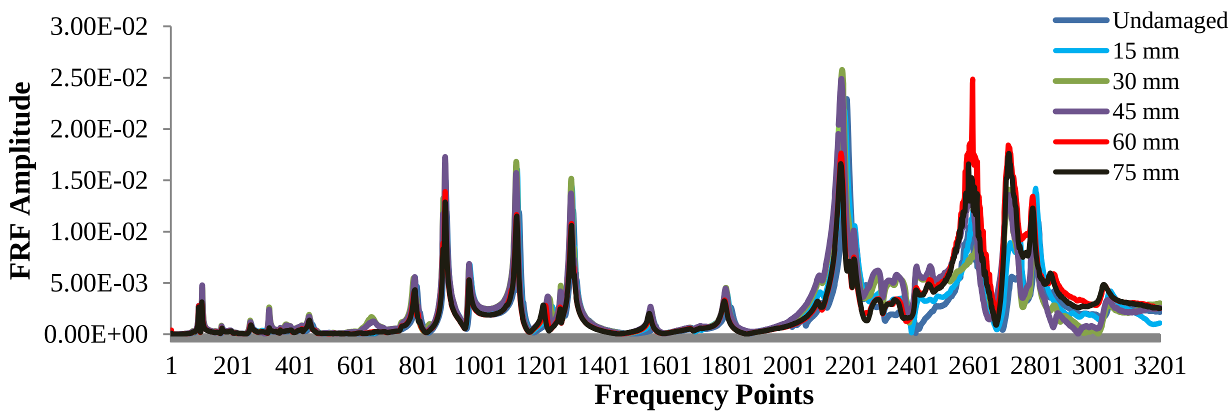

Figure 7: Undamaged and different damaged states of FRF at position number 8.

3.4 Modelling

In this stage, selected data mining modelling technique is applied to database. Steps of this phase consist of the selection of data mining technique, the generation of test design, the creation of models and assessment of models [18], [23]. Principal component analysis (PCA), which is one of the statistical data mining techniques, is used for feature extraction and data reduction of FRF data.

3.4.1 Principal component analysis of FRF data

One of the popular multivariate statistical method is principal component analysis which generally applied to compress multidimensional datasets with correlated variables to smaller dimensions of uncorrelated variables to find features in the dataset [24], [25]. This is because of the fact that a small set of uncorrelated variables is much easier to understand and employ in future analysis than a large dataset of correlated variables.

PCA is applied here to extract the features of measured FRF data. Using an orthogonal projection, this can be achieved by transforming the original set of correlated variables (FRF measurement data) in an N-dimensional space to a new set of uncorrelated variables, the so-called PCs, in a P-dimensional space which are ordered so that the first few PCs remain most of variation presented in all of the original FRFs (P < N). Therefore, using all raw FRF data, form matrix $H=[h_{ij}(\omega)]_{M\times N}$, which has M rows of (spectral lines for each frame of FRF curve in frequency domain) FRFs and N frequency (measurement) points. The"autoscaling" of FRF matrix (H) is implemented to have zero mean variables and to have a unit variance by standard deviation. The mathematical theory behind the feature extraction of FRF using PCA is described as follows [26]–[28]:

The mean response of the *j-th* column is expressed as:

$$\bar{H}_j = \frac{1}{M}\sum_{i=1}^{M} h_{ij} \tag{1}$$

The corresponding standard deviation S_j is:

$$S_j = \sqrt{\frac{\sum_{i=1}^{M}\left(h_{ij}(\omega) - \bar{H}_j\right)^2}{M}} \tag{2}$$

To obtain a response variation matrix $\widetilde{H}$, an element $h_{ij}(\omega)$ of the FRF matrix $H=[h_{ij}(\omega)]_{M\times N}$ can be replaced by:

$$\tilde{h}_{ij}(\omega) = \frac{h_{ij} - \bar{H}_j}{S_j} \tag{3}$$

Using the response variation matrix, the covariance matrix $[C]_{M\times N}$, which is a square matrix, can be defined as follows:

$$[C]_{M\times N} = [\widetilde{H}]^T{}_{N\times M}[\widetilde{H}]_{M\times N} \tag{4}$$

According to the definition, the PCs are the eigenvalues λ_i and the correlated eigenvectors Ψ_i of the covariance matrix:

$$[C]\{\Psi_i\} = \lambda_i\{\Psi_i\} \tag{5}$$

The first PC, which includes the highest eigenvalue and its corresponding eigenvector, represents the direction and amount of maximum variability in the original raw FRF data. The next PC, which is orthogonal to the first principal component, illustrates the second most important contribution from the original raw FRF data, and so on.

The projection matrix $[A]_{M\times N}$ of the response variation matrix $\widetilde{H}$ on the N PCs is written as:

$$[A]_{M\times N} = [\widetilde{H}(\omega)]_{M\times N}[\Psi]_{N\times N} \tag{6}$$

The projection matrix *[A]* and the eigenvector matrix *[Ψ]* can be partitioned into two sub-matrices with K principal components (related to the important features) and *(N-K)* principal components (related to the uncertainty and noise), respectively.

$$\begin{aligned}[\widetilde{H}_R] &= [A][\Psi]^T \\ &= \left[[A_1]_{M\times N} \vdots [A_2]_{M\times(N-K)}\right]\left[[\Psi_1]_{M\times K} \vdots [\Psi_2]_{M\times(N-K)}\right]^T \\ &\cong [A_1]_{M\times K}[\Psi_1]^T{}_{K\times N}\end{aligned} \tag{7}$$

3.4.2 Feature extraction of FRF data

There are totally 90 FRFs, including 15 to each damage scenario. The number of lines are 3201 lines which from 0 Hz to 1000 Hz. Therefore, a 3201 × 90 matrix is formed. Equation (5) was used to extract the PCs of the FRF matrix. As shown in Fig. 8 and Table 1, it is observed that the percentage of variance of the first 10 principal components from the covariance matrix is 83.77% of total FRF values.

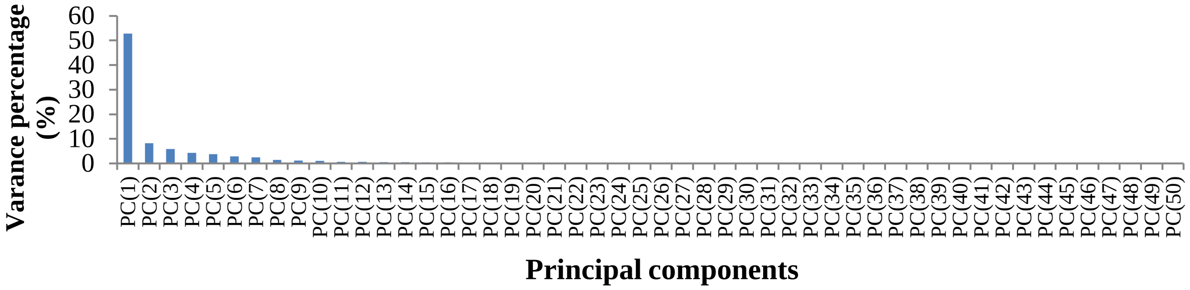

Figure 8: Percentage of variance of PCs of FRF data.

Table 1: Percentage of the first 10 principal components

Principal Components	PC (1)	PC (2)	PC (3)	PC (4)	PC (5)	PC (6)	PC (7)	PC (8)	PC (9)	PC (10)
Percentage	52.799	8.17	5.82	4.33	3.72	2.85	2.46	1.50	1.11	1.02

3.5 Evaluation

In this stage the extracted features of FRFs obtained from PCA are compared with original raw FRFs to be certain it properly achieves the main purpose of data mining modelling. Comparison of the raw and compressed FRFs with the first 10 PCs in healthy and 75 mm damage state are shown in Figs. 9 and 10, respectively.

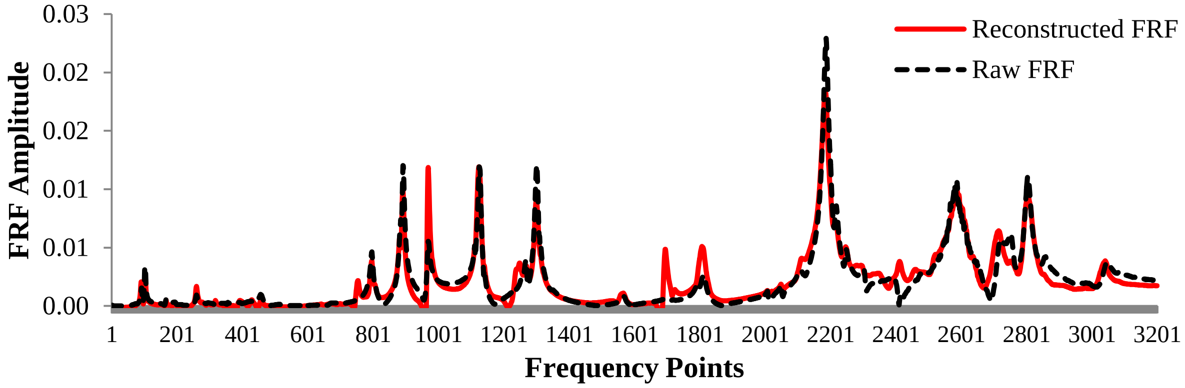

Figure 9:Raw and compressed FRFs of healthy state using 10 PCs.

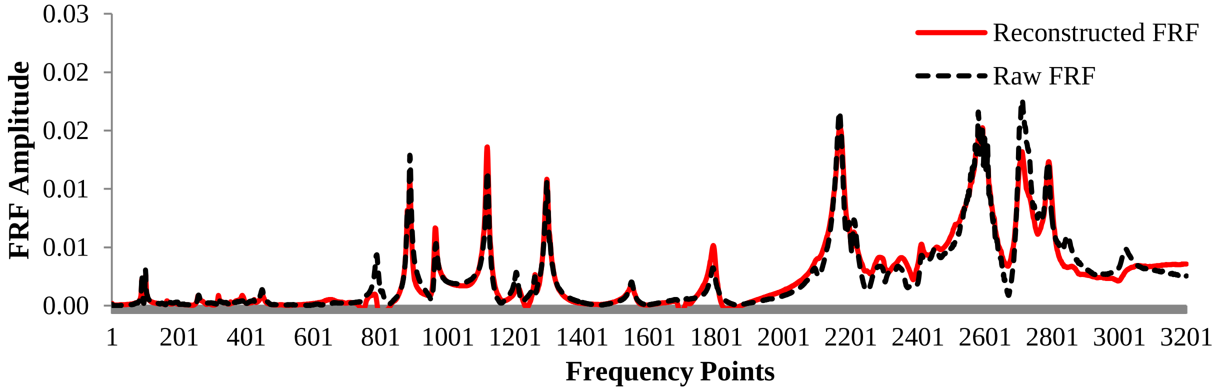

Figure 10: Raw and compressed FRFs of damaged state using 10 PCs.

3.6 Deployment

Feature extraction of FRFs is not the end of the project. The reconstructed FRF data can be used as an input data instead of raw FRF data in the feature plan to develop a robust damage detection system. Therefore, de-noising methodology based on PCA can be employed for dimensionality reduction and noise elimination of FRF data.

4. Conclusion

This paper presents a data mining-based damage identification technique using frequency response functions measurements obtained from experimental modal analysis of a composite girder deck structure in healthy and damaged states. To the best knowledge of the authors, Cross-Industry Standard Process for Data Mining (CRISP-DM) model is derived for the first time in structural damage identification domain. Consequently, the following conclusions can be written:

- The results indicate the applicability of data mining for the purpose of data reduction of FRF.
- PCA technique can be successfully applied for extraction of the most important uncorrelated variables in FRF data obtained from experimental modal analysis.

Acknowledgment

The authors would like to express their sincere thanks to University of Malaya (UM) for the support given through research grants FP028/2013A and PG144-2016A.

REFERENCES

[1] S. W. Doebling, C. R. Farrar, and M. B. Prime, "A Summary Review of Vibration-Based Damage Identification Methods," *Shock Vib. Dig.*, vol. 30, no. 2, pp. 91–105, 1998.

[2] Wei Fan and Pizhong Qiao, "Vibration-based Damage Identification Methods: A Review and Comparative Study," *Struct. Heal. Monit.*, vol. 10, no. 1, pp. 83–111, 2011.

[3] E. Aktan, S. Chase, D. Inman, and D. Pines, "Monitoring and Managing the Health of Infrastructure Systems And an NSF-Supported Workshop on Health Monitoring of Long-Span Bridges," in *Proceedings of the 2001 SPIE Conference on Health Monitoring of Highway Transportation Infrastructure*, 2001.

[4] X. He, "Vibration-Based Damage Identification and Health Monitoring of Civil Structures," Doctoral dissertation, University of California, 2008.

[5] Z. Hou, A. Hera, and M. Noori, "Wavelet-Based Techniques for Structural Health Monitoring," in *Health Assessment of Engineered Structures: Bridges, Buildings and Other Infrastructures*, World Scientific, 2013, pp. 179–202.

[6] H. Adeli, "Sustainable Infrastructure Systems and Environmentally-Conscious Design — A View for the Next Decade," *J. Comput. Civ. Eng.*, vol. 16, no. 4, pp. 231–233, 2002.

[7] C. A. Edeki, "A Comparative Study of Data Mining and Statistical Learning Techniques for Prediction of Canser Survovability," Capella University, 2012.

[8] Z. Duan and K. Zhang, "Data Mining Technology for Structural Health Monitoring," *Pacific Sci. Rev.*, vol. 8, pp. 27–36, 2006.

[9] J. A. Pereira, W. Heylen, S. Lammensa, and P. SAS, "Influence of the number of frequency points and resonance frequencies on model updating techniques for health condition monitoring and damage detection of flexible structure," in *Proceedings of the 13th International Modal Analysis Conference*, 1995, pp. 1273–1281.

[10] P. Avitabile, "Experimental Modal Analysis," *Sound Vib.*, vol. 35, no. 1, pp. 1–15, 2001.

[11] B. J. Schwarz and M. H. Richardson, "Experimental Modal Analysis," *CSI Reliab. Week*, vol. 35, no. 1, pp. 1–12, 1999.

[12] Z.-F. Fu and J. He, *Modal Analysis*. 2001.

[13] D. J. Hand, H. Mannila, and P. Smyth, *Principles of Data Mining*. MIP press, 2001.

[14] GartnerGroup, “Gartner Group Advanced Technologies and Applications Research,” 1995. [Online]. Available: http://www.gartner.com.

[15] J. Han, J. Pei, and M. Kamber, *Data mining - Concepts and Techniques*. Morgan Kauffman, 2001.

[16] A. Azevedo and M. F. Santos, “KDD, SEMMA AND CRISP-DM:A Parallel Overview,” in *IADIS European Conference Data Mining*, 2008, pp. 182–185.

[17] M. Saltan, S. Terzi, and E. Ug, “Backcalculation of pavement layer moduli and Poisson’s ratio using data mining,” *Expert Syst. Appl.*, vol. 38, no. 3, pp. 2600–2608, 2011.

[18] P. Chapman, J. Clinton, R. Kerber, T. Khabaza, T. Reinartz, C. Shearer, and R. Wirth, “CRISP-DM 1.0 Step-by-step data mininng guide,” 2000.

[19] I. B. Fernandez, S. H. Zanakis, and S. Walczak, “Knowledge discovery techniques for predicting country investment risk,” *Comput. Ind. Eng.*, vol. 43, no. 4, pp. 787–800, 2002.

[20] S. C. Chen and M. Y. Huang, “Constructing credit auditing and control & management model with data mining technique,” *Expert Syst. Appl.*, vol. 38, no. 5, pp. 5359–5365, May 2011.

[21] R. Wirth and J. Hipp, “CRISP-DM : Towards a Standard Process Model for Data Mining,” in *Proceedings of the Fourth International Conference on the Practical Application of Knowledge Discovery and Data Mining*, 2000, pp. 29–39.

[22] Z. Tang and J. Maclennan, *Data Mining With SQL Server 2005*. Wiley, 2005.

[23] C. Shearer, H. J. Watson, D. G. Grecich, L. Moss, S. Adelman, K. Hammer, and S. a Herdlein, “The CRIS-DM model: The New Blueprint for Data Mining,” *J. Data Warehousing14*, vol. 5, no. 4, pp. 13–22, 2000.

[24] O. R. de Lautour and P. Omenzetter, “Nearest Neighbor and Learning Vector Quantization Classification for Damage Detection using Time Series Analysis,” *Struct. Control Heal. Monit.*, vol. 17, no. 6, pp. 614–631, 2010.

[25] N. Kwak, “Principal Component Analysis by L p -Norm Maximization,” *IEEE Trans. Cybern.*, vol. 44, no. 5, pp. 594–609, 2014.

[26] L. Hao, P. L. Lewin, J. a Hunter, D. J. Swaffield, a Contin, C. Walton, and M. Michel, “Discrimination of multiple PD sources using wavelet decomposition and principal component analysis,” *Dielectr. Electr. Insul. IEEE Trans.*, vol. 18, no. 5, pp. 1702–1711, 2011.

[27] X. G. Hua, Y. Q. Ni, J. M. Ko, and K. Y. Wong, “Modeling of Temperature – Frequency Correlation Using Combined Principal Component Analysis and Support Vector Regression Technique,” *J. Comput. Civ. Eng.*, vol. 21, no. 2, pp. 122–135, 2007.

[28] K. Ku and Z. Waszczyszyn, “Neural Networks and Principal Component Analysis for Identification of Building Natural Periods,” *J. Comput. Civ. Eng.*, vol. 20, no. 6, pp. 431–436, 2006.