Data preparation for training CNNs: Application to vibration-based condition monitoring

Vahid Yaghoubi[1,2], **Liangliang Cheng**[1,2], **Wim Van Paepegem** [1], **Mathias Kersemans** [1]
1-Mechanics of Materials and Structures, Ghent University, Technologiepark 46
B-9052 Zwijnaarde, Belgium
2-SIM M3 program, Technologiepark 48
B-9052 Zwijnaarde, Belgium
{vahid.yaghoubi, liangliang.cheng, wim.vanpaepegem, mathias.kersemans}@ugent.be

Abstract

Vibration data is one of the most informative data to be used for fault detection. It mostly employs in the form of frequency response function (FRF) for training deep learners. However, since normally the FRFs are measured at excessive numbers of frequencies, its usage not only enforces large computational resources for training the deep learners, but could also hinder proper feature extraction. In this paper, it is shown that given a predefined deep learning structure and its associated hyper-parameters, how proper data selection and/or augmentation could improve the performance of the trained model in classifying the samples. For this purpose, the least absolute shrinkage and selection operator (LASSO) and some generative functions are utilized respectively for data selection/reduction and augmentation prior to any training. The efficacy of this procedure is illustrated by applying it to an experimental dataset created by the broadband vibrational responses of polycrystalline Nickel alloy first-stage turbine blades with different types and severities of damages. It is shown that the data selection and augmentation approach could improve the performance of the model to some extent and at the same time, drastically reduce the computational time.

1 Introduction

Over years, the vibration-based fault detection become very popular, and thus, evolved from traditional methods to machine learning (ML)-based methods [1]. Deep learning (DL) is the most recent trend among researchers in the domain of condition monitoring of machinery due to its promising result and automated feature learning [2-4]. However, it mostly requires a complex model structures with a lot of epochs for proper convergence. This can readily reveals the need for huge computational resources. An extensive review of using deep learning for machine health monitoring is presented in [4].

Frequency response function (FRF) is one of the most informative response of a system that could be used for condition monitoring [5-7]. However, it normally measures or calculated in an extensive number of frequencies. Handling those frequencies to train the CNNs could cause several challenges such as requiring large computational resources to perform the training and introducing extensively large number of unknown parameters that in turn, could lead to another category of challenges related to the optimization process e.g. convergence, local minima, etc. Therefore, in this work we are proposing to employ a data selection/reduction step prior to conducting any optimization to select only the frequencies with vital information for classification. This could drastically reduce the size of dataset which in turn, could reduce the required computational time and memory. This is achieved here by employing the least absolute shrinkage and selection operator (LASSO) with

1st NeurIPS Data-Centric AI workshop (DCAI 2021).

the Least Angle Regression (LARS) method [8]. However, there is always a risk for losing some information. To minimize the risk and assure the presence of (almost) all important frequencies, some basic generative functions are utilized. Besides, these functions could also be very effective in exposing the information hidden in the data.

The efficacy of the suggested procedure is illustrated by its application to a real experimental dataset generated from broadband vibration response of first-stage turbine blades with complex geometry and different damage types and severities.

2 Data preparation

In this section, two basic but effective approaches for data selection and augmentation for the considered application is elaborated.

2.1 LASSO-based data selection

The LASSO method estimates the coefficient of a linear regression model under the constraint of having the sum of the absolute values of the coefficient less than a threshold. It is known for performing two tasks: regularization and variable selection. Recently, it attracts attention to be used as a feature selection method for classification [9]. Its procedure is as follows.

Let $X \in R^{(n_s \times n_f)}$ and $Y \in R^{(n_s \times 1)}$ be the matrices of input and output respectively. Here, n_s and n_f are the number of samples and features. The purpose is to find a linear regression model to fit the data, that is,

$$y_i = \beta_0 + \Sigma_{j=1}^{n_f}(x_{ij}\beta_j) \tag{1}$$

The LASSO estimate of the coefficients β is,

$$\beta^{lasso} = argmin(\Sigma_{j=1}^{n_s}(y_i - \beta_0 - \Sigma_{j=1}^{n_f}x_{ij}\beta_j)^2) + \lambda\Sigma_{j=1}^{n_f}|\beta_j| \tag{2}$$

here $\Sigma_{j=1}^{n_f}|\beta_j|$ is the penalty term, and λ is a tuning parameter to control the strength of the penalty term. The features could be ranked based on the magnitude of their associated coefficients β. One could relax λ to obtain all the important features. In this case, the number of selected features, i.e. the features x_{ij} with nonzero coefficient β_j, is thus limited to $min(n_s, n_f)$. In the current study, the amplitudes of the FRFs create the feature space therefore, the number of features n_f is equal to the number of frequencies available in the FRFs. Since normally $n_f \gg n_s$, the number of selected frequencies is upper limited to n_s. Therefore, there is always the risk of not including the important frequency into the selected frequencies. It can be treated by employing several generative functions.

2.2 Generative functions

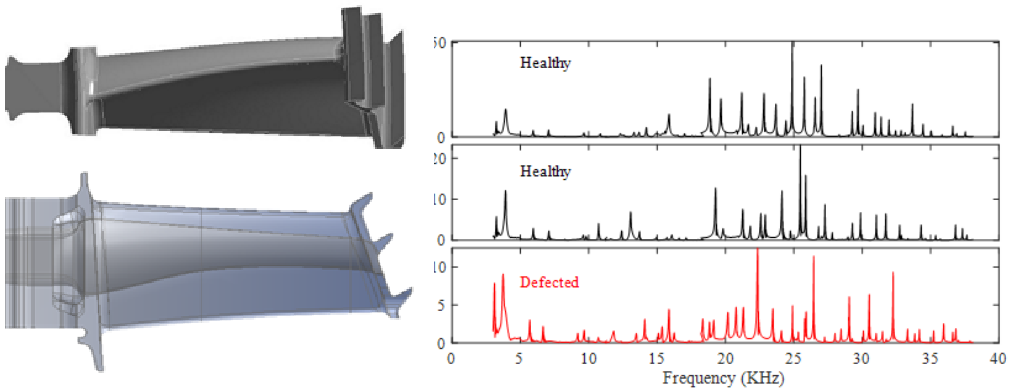
In this study, the generative functions are employed as a data augmentation technique. The augmentation is done by applying these function to the FRFs. This could enrich the data in two ways: (i) by generating new fictitious samples leading to selection of more frequencies. This could ensure the inclusion of the crucial frequencies in the selected set, and (ii) by unfolding the information hidden in the data resulting in less complex CNN structures and faster convergence during the optimization. Here, the following functions have been used:

Summation and *production* of the FRFs: to amplify the peaks and diminish the valleys. This, thus, increase the dynamic range of the FRFs and disclose the information at the high-amplitude frequencies.

Logarithm of the FRFs: this reduces the dynamic range of the FRFs. On the other hand, this reveals the information content at the low-amplitude frequencies.

3 Application

In this section, the procedure is applied to vibrational response data from Equiax Polycrystalline Nickel alloy first-stage turbine blades with complex geometry. Two views of its CAD model are shown in 1(a). The cooling channel in the middle can be seen in the transparent view. By using one actuator and two sensors, i.e. Single Input Multiple Output SIMO, the amplitude of the frequency response function (|FRF|) has been collected from each blade in the range of [3, 38] kHz at 11253 frequency lines, see Figure 1(b). To create a database 150 healthy and 79 defected blades have been measured. The damages in the blades range from microstructural changes due to over-temperature, airfoil cracking, inter-granular attack (corrosion), thin walls due to casting, to service wear.



(a) (b)

Figure 1: (a) Two views of the CAD model of the turbine blade. Bottom plot shows a transparent view to illustrate the cooling channel. (b) Three examples of the collected FRFs. Two top FRFs have been collected from healthy blades and the bottom FRF was from a defected blade.

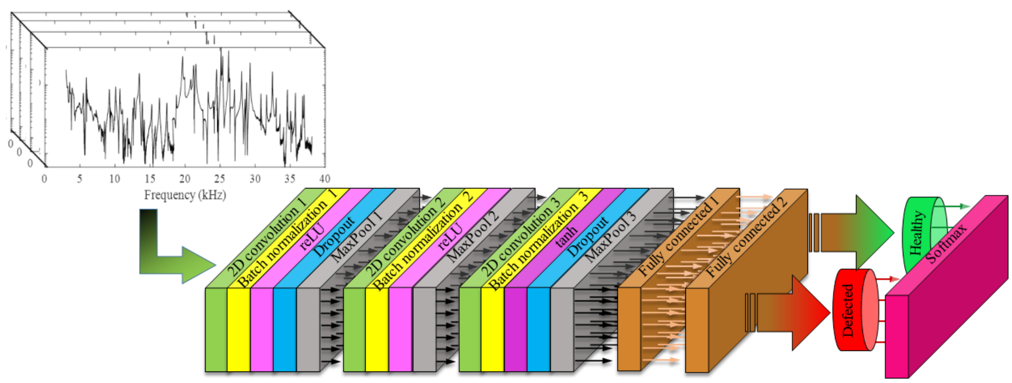


Figure 2: The chosen CNN model [2]

To commence the procedure, a CNN model as shown in Figure 2,with fixed hyper-parameter is chosen. Then the data was divided into 70% for training and 30% for Test. Since the number of defected blades is almost half of the healthy ones, some randomly selected defected samples were repeated such that an equal number of healthy and defected samples, i.e. 105 healthy and 105 defected blades, have been used for training the CNNs. Let X_1 and X_2 be respectively the two acquired |FRF|s. To illustrate the effect of generative functions, six functions are utilized as, summation $X_1 + X_2$, production $X_1 \cdot X_2$, square X_1^2,X_2^2, and logarithm $log(X_1)$, $log(X_2)$.

In the step of frequency selection, the LARS algorithm has been applied to each FRFs separately. Since 210 training samples were available, from each function 210 frequency lines could be selected. In the following two cases have been studied.
Case I) without generative functions: only the two main FRFs are used. In this case, in total $2 \times 210 = 420$ frequencies could be chosen.
Case II) with generative functions: eight functions of FRFs are used. In this case, in total $8 \times 210 = 1680$ frequencies could be chosen.

To provide statistical insight to the approach, the whole procedure has been repeated 10 times. For this purpose, the dataset was divided into 10 batches and at each iteration, 7 batches were used as the training dataset and the other 3 as the validation dataset. The result in the sense of simulation time and validation accuracy is presented in Tables 1 and 2 respectively, for Case I and II. For each case, two sets of CNNs have been compared: (i) the one trained by the |FRF| at all frequencies and (ii) the one trained by |FRF| at the selected frequencies by employing the LASSO. It can be observed that in both cases, the data selection could drastically reduce the computational cost of the training procedure, about 10 times for Case I and 7 time for Case II. It could also improve the classification performance to some extent.

To consolidate the analysis, the effect of noise on the whole procedure have been investigated. Therefore, the FRFs have been synthetically polluted with white Gaussian noise. Two noise levels have been investigated: 25% and 50% noise-to-signal ratio (NSR). The results indicate that, by adding some noise we will loose some accuracy but the efficiency of the suggested data preparation procedure can still be observed.

Moreover, comparing the results of the Case I with those of Case II reveals that employing the generative functions, although increases the overall computational time relatively, could improve the classification performances. Besides, it can be seen utilizing the LASSO in conjunction with the suggested generative functions could improve the average classification performances from 92.00% (without any data preparation) to 93.17% and at the same time, reduce the average computational time from 850.29 s (without any data preparation) to 306.56 s.

Table 1: Simulation time (in sec) and validation accuracies (in %) of the CNN models trained by the |FRF| polluted with different noise levels at the selected and whole frequencies for Case I: without the generative functions.

NSR (%)	All frequencies		Selected frequencies		
	Sim. time	Acc	LARS time	Sim. time	Acc
0	850.29 ± 7.42	92.00 ± 2.37	45.71 ± 3.31	30.44 ± 1.19	92.73 ± 2.25
25	743.47 ± 5.87	90.27 ± 3.56	24.55 ± 4.64	29.16 ± 2.24	90.97 ± 3.34
50	744.98 ± 4.07	89.64 ± 4.31	18.62 ± 1.21	29.31 ± 1.47	90.23 ± 3.42

Table 2: Simulation time (in sec) and validation accuracies (in %) of the CNN models trained by the |FRF| polluted with different noise levels at the selected and whole frequencies for Case II: with the generative functions.

NSR (%)	All frequencies		Selected frequencies		
	Sim. time	Acc	LARS time	Sim. time	Acc
0	2050.62 ± 45.50	92.86 ± 1.97	75.75 ± 15.75	230.81 ± 6.38	93.17 ± 1.15
25	1995.10 ± 23.51	91.24 ± 1.56	55.37 ± 12.51	226.24 ± 4.27	92.85 ± 3.28
50	2001.48 ± 15.37	90.77 ± 2.15	56.88 ± 8.71	220.45 ± 2.12	93.00 ± 3.46

4 Conclusion

In this paper, given a CNN model, the importance of data preparation prior to training CNN models for vibration-based condition monitoring has been investigated. For this purpose, a LASSO-based method was employed for data selection/reduction and some generative functions were used for data enrichment. The procedure was applied to an experimental dataset when it was synthetically polluted with different noise levels. It was shown that the data selection step drastically reduce the computational cost, and improve the classification performance to some extent. Besides, the effectiveness of the generative functions in the has also been illustrated. The overall results suggested that jointly using the data selection and data enrichment could improve the classification performance and reduce the simulation time at the same time.

Acknowledgments and Disclosure of Funding

The authors gratefully acknowledge the ICON project DETECT-ION (HBC.2017.0603) which fits in the SIM research program MacroModelMat (M3) coordinated by Siemens (Siemens Digital Industries Software, Belgium) and funded by SIM (Strategic Initiative Materials in Flanders) and

VLAIO (Flemish government agency Flanders Innovation & Entrepreneurship). Vibrant Corporation is also gratefully acknowledged for providing anonymous datasets of the blades.

References

[1] O. Avci, O. Abdeljaber, S. Kiranyaz, M. Hussein, M. Gabbouj, D.J. Inman, A review of vibration-based damage detection in civil structures: From traditional methods to Machine Learning and Deep Learning applications, Mechanical Systems and Signal Processing. 147 (2021) 107077.

[2] V. Yaghoubi, L. Cheng, W. van Paepegem, M. Kersemans, CNN-DST: ensemble deep learning based on Dempster-Shafer theory for vibration-based fault recognition, Structural Health Monitoring. (2021).

[3] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, R.X. Gao, Deep learning and its applications to machine health monitoring, Mechanical Systems and Signal Processing. 115 (2019) 213–237.

[4] P. Seventekidis, D. Giagopoulos, A. Arailopoulos, O. Markogiannaki, Structural Health Monitoring using deep learning with optimal finite element model generated data, Mechanical Systems and Signal Processing. 145 (2020) 106972.

[5] V. Yaghoubi, L. Cheng, W. van Paepegem, M. Kersemans, A novel multi-classifier information fusion based on Dempster-Shafer theory: application to vibration-based fault detection, Structural Health Monitoring. (2021). https://doi.org/10.1177/14759217211007130.

[6] V. Yaghoubi, L. Cheng, W. van Paepegem, M. Kersemans, An ensemble classifier for vibration-based quality monitoring, Mechanical Systems and Signal Processing. 165 (2022) 108341. https://doi.org/10.1016/j.ymssp.2021.108341.

[7] D.J. Ewins, Modal Analysis: Theory, Practice and Application, 2nd ed., John Wiley & Sons, Ltd, 2009.

[8] B. Efron, T. Hastie, I. Johnstone, R. Tibshirani, Least angle regression, The Annals of Statistics. 32 (2004) 407–499.

[9] C. Kang, Y. Huo, L. Xin, B. Tian, B. Yu, Feature selection and tumor classification for microarray data using relaxed Lasso and generalized multi-class support vector machine, Journal of Theoretical Biology. 463 (2019) 77–91.