Applied Acoustics 77 (2014) 169–177

Contents lists available at ScienceDirect

Applied Acoustics

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

Dimensionality reduction via variables selection – Linear and nonlinear approaches with application to vibration-based condition monitoring of planetary gearbox

A. Bartkowiak [a,b], R. Zimroz [c,*]

[a] Institute of Computer Science, University of Wroclaw, Joliot Curie 15, Wroclaw 50-383, Poland
[b] Wroclaw School of Applied Informatics, Wejherowska 28, Wroclaw 54-239, Poland
[c] Diagnostics and Vibro-Acoustics Science Laboratory, Wroclaw University of Technology, Plac Teatralny 2, Wroclaw 50-051, Poland

ARTICLE INFO

Keywords:
Dimensionality reduction
Feature selection
Linear and nonlinear approach
Least square regression
Lasso
Diagnostics
Planetary gearbox

ABSTRACT

Feature extraction and variable selection are two important issues in monitoring and diagnosing a planetary gearbox. The preparation of data sets for final classification and decision making is usually a multistage process. We consider data from two gearboxes, one in a healthy and the other in a faulty state. First, the gathered raw vibration data in time domain have been segmented and transformed to frequency domain using power spectral density. Next, 15 variables denoting amplitudes of calculated power spectra were extracted; these variables were further examined with respect to their diagnostic ability. We have applied here a novel hybrid approach: all subset search by using multivariate linear regression (MLR) and variables shrinkage by the least absolute selection and shrinkage operator (Lasso) performing a non-linear approach. Both methods gave consistent results and yielded subsets with healthy or faulty diagnostic properties.

1. Introduction, statement of the problem

Multidimensional feature space is frequently used in many fields of sciences, including advanced condition monitoring of the observed rotating machinery. Diagnostics of the objects condition usually uses a model for the investigated phenomenon; in the simplest one dimensional (1D) case this may be a probability density function for healthy and faulty conditions, in more complex problems after initial preprocessing of the gathered data complex mathematical modeling of different kind is applied [11,25,14,15,1,2,8].

Generally, when analyzing vibration signals from a rotating machinery working with installed gearboxes, the methods of analysis fall into three broad categories (i) time domain analysis; (ii) frequency domain analysis; (iii) simultaneous time–frequency domain analysis. Each domain may use many specific multivariate methods originating from statistics, pattern recognition, artificial neural networks and artificial intelligence; for examples see the invited paper by Jardine et al. [16] with its 271 references, also two other invited papers, a little bit more specific [18,20], with 122 and 119 references to the mentioned topics.

In the case of multidimensional data space it is very important to decide how many measured variables should be used to build the model. It is not reasonable to use all of available variables. Reducing dimensionality of data set can be carried out in many ways. Optimal (in sense of classification ability, contained information, etc.) features set allows to classify data with computational efforts and maximal stability/efficiency of classification results.

In particular, when monitoring the rotating machinery, one may want – on the base of the recorded data – to come to know about the state of the machine, in particular to find out if it is in a healthy or a faulty state.

The assumed model may be related to number of sensors, different physical parameters used for diagnostics, for example temperature, vibration, acoustics signals (including noise or acoustics emissions, etc.) or multidimensional representation of a single signal (statistical descriptors of the process, 1D spectral representation, 2D time–frequency representations and other) [3,24,10,11,25,14,15,1,2,8]. Having too many variables included in the model may not be convenient for the following reasons: Some variables may be not relevant to the problem at hand and contain a large amount of redundant information; taking them for the analysis may introduce noise and unexplained fluctuations of the output. Also, using some more complicated nonlinear equations, the necessary parameters may be extremely difficult to estimate. In other words, redundant and irrelevant variables may cause considerable impediment in the performed analysis. Therefore, before

* Corresponding author. Tel.: +48 71 320 68 49.
E-mail addresses: aba@ii.uni.wroc.pl (A. Bartkowiak), Radoslaw.Zimroz@pwr.wroc.pl (R. Zimroz).

0003-682X/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.apacoust.2013.06.017

starting the proper analysis, a responsible researcher should find out what kind of data will be analyzed. The first question should be about the intrinsic dimensionality of the data. The second question should be if the number of variables at hand might be somehow reduced without losing the overall information hidden in the data. These are difficult questions and need expert guide.

Such a guide may be obtained from a special issue of Journal of Machine Learning Research, in particular from the first paper of that issue authored by Guyon and Elisseeff [12]. The contributions in the mentioned issue consider such topics as: providing a better definition of the objective function, feature construction, feature selection, features ranking, efficient search methods, and features validity assessment methods.

When considering dependencies among variables, we may consider either linear or non-linear dependencies. The same concerns prediction models. A comprehensive introduction to non-linear methods serving for dimension reduction may be found in the paper by Yin [26] (contains 62 bibliographical references). The author discusses various non-linear projections, such as nonlinear principal component analysis (PCA) obtained via self-associative neural networks), kernel PCA, principal manifolds, isomaps, local linear embedding, Laplacian eigenmaps, spectral clustering, principal curves and surfaces. He emphasizes the importance of visualization of the data by topology preserving maps, like the Visualization-induced Self Organizing Map (ViSOM).

A more detailed elaboration on variable selection may by found in the book by Clarke et al. [7]. Chapter 10 of this book is devoted to 'Variable Selection'. The authors describe there in more than 100 pages such topics as linear regression, subset selection, some classical and recently developed information criteria (Akaike's AIC, Bayesian BIC, deviance information DIC). They discuss how to choose the proper criterion and assess the appropriate model. Apart from the model selection methods, they consider also some shrinkage methods penalizing the risk: ridge regression, nonnegative garrotte, least absolute selection and shrinkage operator (Lasso), elastic net, least angle regression, shrinkage methods for support vector machines and Bayesian variable selection. Computational issues of the methods are also discussed and compared.

Some comparative reviews on dimensionality reduction methods may be found in van der Maaten et al. [23] (Technical Report of Tilburg University with 149 bibliographic references). and Parviainen [17] (Ph.D. dissertation with 218 bibliographic references, where a taxonomy of the existing methods is built). Recently, Pietila and Lim have published a review of intelligent systems approaches when investigating sound quality [19]. They state that the most common model used today is the Multiple Linear Regression and non-linear Artificial Neural Network. They go into shortcomings that are associated with both the current regression and neural network approaches. They mention robust approach as a new thought for improving the current state-of-art technology.

Generally there are many methods and their success depends on the data and the problem to be solved. The first crucial step is data acquisition and extracting from them variables (traits) for further analysis. A number of approaches can be used to obtain variables for further elaboration. The most popular are: plain statistical variables of vibration time series (treated as record of unknown process), spectral representation of the vibration time series or other advanced multidimensional representations. In condition monitoring, the preparation of data sets for final classification and decision making is usually a multilevel (multi-step) process.

After recording the experimental data, the first step of the analysis makes a kind of preprocessing of the raw vibration data by averaging them, de-noising and segmenting. Often the feature extraction process is carried out by transformation of the time series to other domain (frequency, time–frequency, wavelets coefficient matrix, etc.). After that, selection of particular components (for example mesh frequency) or aggregation of group of components (energy in band) is done. Final features set preparation is aimed at minimizing redundancy of data, which results in reducing computational efforts at classification phase (both in the training and the testing phase). There is no univocal and clear answer which representation of raw vibration signal is better for condition monitoring. This may be the reason that some authors use not one, but more of them. Giving a few examples of dealing the problem: For reducing generally the dimensionality of the data, many authors use: principal component analysis (PCA), independent component analysis (ICA), isomap, local linear embedding, kernel PCA, curvilinear component analysis, simple genetic algorithm, adaptive genetic algorithm with combinatorial optimization and others. These methods serve generally for reduction of dimensionality of the recorded data with the intention to obtain a subspace representation, in which the fault classes are more easily discriminable.

In the second stage, having the variables for the analysis fixed, a prediction algorithm, usually neural networks like radial basis functions (RBF) or support vector machines (SVM) are applied to perform the monitoring. Some authors use also hybrid models combining multiple feature selection to obtain the input variables for the second stage. It is also possible to use a combined approach by analyzing frequency spectra obtained by Fourier analysis in time (to monitor changes in spectra along lifetime and fault development, it can be seen as kind of time–frequency analysis). Often wavelets decomposition is used as a preprocessor, wavelets are used before calculation and comparing frequency spectra to decompose raw time series into set of sub-signals with simpler frequency structure. Such an analysis was performed by Eftekharnejad et.al. [10] when considering shaft cracks in a helical gearbox.

In this paper we will focus on the most informative variables selection from 15D data vectors using linear and nonlinear techniques. The basic data origin from spectral representation of vibration time series measured on two planetary gearboxes, mounted in bucket wheel excavators used in a mining company (for more details, see [3,27,29]). A planetary gearbox is really a complex device and it is difficult to deal with. In our machine, planetary gearbox was part of complex (multi-stage) gearbox. The purpose of the experiment was to asses the planetary stage as a key element of the system.

After gathering the vibration signals they were segmented (divided into short sub-signals) and analyzed in frequency domain using power spectral density to obtain an array of real values – amplitudes of isolated components indicating for high energy at some frequencies (planetary gear mesh frequency and its harmonics).

Based on some a priori knowledge related to machine design, 15 parameters from vibration spectrum have been extracted. These 15 parameters are expected to describe the technical condition of a planetary gearbox. It should be added that the same method might be used for a distributed form of change of condition, not for the localized one. A feature extraction procedure used here might be interpreted as time frequency approach, because for each short segment of signal, frequency domain features were extracted. It could be seen as calculation of spectrogram (without overlapping) – the simplest time–frequency representation. For each slice of spectrogram 15 features are extracted.

It should be emphasized that such method of feature extraction was proposed already by Bartelmus and Zimroz [3]. From the obtained power spectra, they have retained 15 components and have called them pp1, pp2, ..., pp15 appropriately. The distribution of these variables, characterizing the observed two gearboxes – one healthy and one faulty– working with and without load, was also considered in [4,28].

In a previous work the authors [3] used sum of amplitudes of selected components as an aggregated measure of gearbox condi-