A Novel Automated Feature Extraction Method for Fault Diagnosis of Rotating Mechanical Systems

Zacharias Voulgaris[1] and Chris Sconyers[1]

[1] *Georgia Institute of Technology, Atlanta, Georgia, 30332, USA*
{zvoulgaris, csconyers}@gatech.edu

ABSTRACT

A novel approach to feature extraction, capable of generating a number of robust features in an automated way, is introduced. Although the proposed method focuses on features on the frequency domain for vibration data related to rotating mechanical systems, it can be extended on different types of features. The method comprises of two simple models for the feature generation and a Particle Swarm Optimization system for establishing optimum or near optimum parameters for these models. The generated features are evaluated with a number of metrics, before they are used for diagnosis purposes. The features derive from real-world data related to a case of corroded bearings in a helicopter system. The extracted features of the proposed method are compared with some which were manually generated, and the former are found to be of superior quality. A series of diagnosis experiments based on the best extracted features was carried out. The results of these experiments appear to validate the performance of the automatically generated features.

1 INTRODUCTION

The feature extraction methodology is a major component of the fault diagnosis process and probably one of the most challenging ones. No matter how robust the feature selection and the feature fusion methods are, which follow the feature extraction one, if the extracted features are not potent enough, the diagnosis results may suffer. Also, it is important that a number of features are extracted, since even if they are not very powerful their combination may yield a quite potent feature. Hence, it is crucial that a sufficient number of relatively good features are generated for fault diagnosis purposes. And if this whole process can be done automatically, it can truly be an asset for the whole process.

2 LITERATURE REVIEW

2.1 Feature Extraction Methods

The existence of a fault in a system or component produces uncharacteristic behavior that can be captured through certain sensing spectra, typically vibration (Vachtsevanos et al., 2006). Where it may not be a simple task to detect and diagnose a fault, processing and analyzing such sensor data may provide information about anomalous system behavior, from which fault data may be gleaned. Such information is extracted in the form of features, scalar representations of signal information.

In rotating machinery, vibration signals can be analyzed and features extracted from numerous domains. The most simple and direct is the time domain, from which features such as root-mean-squared (RMS), an approximation of signal strength; Kurtosis, a measure of “peakiness”; and entropy can all be extracted (Mobley, 1999; Qu and Shen, 1993; Decker, 2002). These static features, however, are highly sensitive to operational variability, such as changes in loading conditions on the faulty component, and thus are less robust than features from more sophisticated domains.

Due to rotation, vibration effects caused by a fault may occur periodically, making the frequency domain a source of many useful features. FFT analysis of harmonics and subharmonics of main frequencies, such as bearing shaft speed or gear meshing frequency, is a typical source of features for both identifying an existing fault mode and diagnosing said fault. Typically, the first two harmonics, 1X and 2X, contain the most relevant information about the system behavior, in multiple frequency-based domains such as FFT, full-spectrum, auto-spectrum, or wavelet (Patel and Darpe, 2009; Downham, 1976). Filtered orbit and sideband analysis on these harmonics can provide further fault information (Shi et al., 2005). Including normalizing information such as total spectral energy can inhibit effects of operational variability. Other ways of pinpointing a fault using vibration data exist, yet they are not as effective as the aforementioned ones.

Common frequency-based features may be robust as they are common, but require explicit knowledge or restrictive assumptions about the system or component under analysis, at least in the case of rotating mechanical systems. Not only does this complicate the feature extraction process, but unforeseen changes in the system may not be accounted for, leading to degradation in detection or diagnostic accuracy. In these cases, an assumption-free, data-driven method is needed to improve the feature extraction process.

2.2 Particle Swarm Optimization

The PSO method was originally developed by J. Kennedy and R. Eberhart in 1995 (Kennedy and Eberhard, 1995) as an optimizer the simulated social behavior in a fast and easy to fine-tune way. Even at that time, different variations of the algorithm were developed and tested, yet the one that prevailed was the one known today as classical PSO. In short, the PSO method attempts to find a global optimum solution to a given problem by maximizing or minimizing a function (fitness function). This is done by populating the solution space with a number of possible solutions (particles) that interact with each other and eventually convert to a (usually good approximation of) the optimum solution.

3 METHODOLOGY

The Automatic Feature Extraction (AFE) method proposed in this paper is relatively simple in concept. Particularly it comprises of two basic parts: establishing a model for the features and finding the optimum (or at least near optimum) parameters for this model using a Particle Swarm Optimization (PSO) program. Naturally, more than one model can be used at the same time, yielding a greater diversity in the extracted features and a wider search over the feature space.

3.1 Feature Models

Two models for the extracted features were made and used in this method. The reason for that is that we wished to have a large variety of features in order to form a large enough subset through a relatively strict feature filtering method we have also developed.

The first model has to do with taking a number of windows of a certain bandwidth of frequencies around a non-integer multiple of the shaft speed. A certain overlap between two consecutive windows is allowed and is part of the feature as a parameter. This model has a total of 4 parameters: center frequency, bandwidth, time-series window size, and window overlap.

The second model used is a more sophisticated one as it takes two bands of frequencies at the same time and calculates their ratio. Again some overlap is allowed between two windows and the frequency bands can be defined around different shaft speed multipliers and be of different width. This model has 6 parameters in total: two center frequencies and bandwidths, and one time-series window size and window overlap (to preserve ratio normality).

3.2 Parameter Finding Using PSO

The aforementioned features are quite generic and unless a good selection of values for the involved parameters is given, their potency is questionable. Yet, they were made generic on purpose so that they can cover a large range of solutions. To find the right values for the parameters involved a PSO method was implemented. This is a viable alternative for pinpointing the optimum values of the parameters since all of these parameters are continuous and belong to a well-defined interval. Also, PSO's ability to avoid "getting trapped" in local minima allows it to function well for this kind of problem.

This method functions as follows: the members of the group of solutions (particles) are accelerated towards the *pbest* and *gbest* locations (which correspond to the best solution of the individual particle and the whole group). The accelerations involved are weighted by random numbers, which are different for the *pbest* and *gbest* locations. The velocity of each particle is limited to a given number, *Vmax*, so that the particles converge more easily to an optimum. This is an important parameter of the PSO method since it determines the resolution of the search space (high values of *Vmax* may result to particles passing by a good solution, while very small values may limit the search to locally good regions). According to (Kennedy and Eberhard, 2001) a good value for this parameter is ±4.0. Other parameters of this method are the acceleration constants (c1 and c2), which correspond to the *pbest* and the *gbest* positions and are related to the tension of the system (which controls how much the swarm explores the search space before converging to an optimum solution). In most applications these parameters are set to the value of 2.0.

A number of fitness functions can be used for the evaluation of the various solutions for the particles of PSO. Some of these are the Spherical Index of Discernibility (Voulgaris, 2009), the Fischer Discriminant Ratio, the Exponential Correlation, a Monotonicity metric or the well-established and widely used Pearson Correlation. In this research the absolute value of Pearson Correlation was used as a fitness function. This is defined in Eq. 1 as follows.

$$\rho = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{(n-1)\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2 \sum_{i=1}^{n}(Y_i - \bar{Y})^2}} \qquad (1)$$

where X is the feature value
Y is the wear level, and
n is the number of test points

4 EXPERIMENTS AND RESULTS

4.1 Data Description

The data used for this research come from a series of experiments conducted for a project coded AVDPIP, involving corroded bearings for a helicopter module. These experiments involve a combination of video files and a set of sensor readings. In this research the focus is on the latter, a variety of vibration data files, for various operating conditions and wear levels. Particularly, the data are split into two types, *axial* and *radial*, corresponding to the two directions of the sensors that were placed to pick up the vibrations from the bearings. Also, the data involve two load levels, one of *500 lbs* and one of *975 lbs*. There were 10 data files for each case and 4 wear levels, including a ground truth baseline (no corrosion in the bearings). Furthermore, 3 rounds of experiments were carried out, the data from which were used in our analysis.

4.2 Experiment Setup

A number of features were extracted based on the given dataset, for both the axial and radial data. Afterwards, this feature set was filtered so that only the most potent and uncorrelated features remained. From the original 32 extracted features (16 for each data type), four were in the final feature set. All of these features were based on multiples of the shaft frequency, as this type was the easiest to model. The selected features were then fused (combined) into a single feature (FF_1) using a robust method introduced and described in (Voulgaris et al., 2010). The fused feature was then used for diagnosis using a Particle Filter (PF) based method.

Parallel to all this, a set of five manually extracted features based on the same data and frequency domain was used. To obtain this feature set, all possible known features that could be manually derived from the vibration data were extracted. From this original feature set which comprised of about 10 features, the best performing ones were selected using a recently developed automated feature selection method (Voulgaris and Sconyers, 2010). Note that the methodology for extracting these features included, but was not limited to, investigating various multiples of the shaft frequency. However, it must be stressed that due to the vast amount of features that could be derived from the harmonics and sidebands, only the most common ones were manually extracted.

The performance of the extracted features was based on their correlation with the wear level and the correlations with each other, yielded another fused feature (FF_2) which was also used for diagnosis.

The feature fusion method used was based on the relevant research presented in (Voulgaris et al., 2010). Briefly it can be summarized as follows: a model for the combined feature is given, comprising of all the available features and a corresponding weight parameter for each one of them (this takes the form of an exponent in the model). Then the PSO algorithm is applied on the parameters space to obtain a set of parameters such that it maximizes one of the feature evaluation metrics defined by the user. The one most commonly used is the absolute correlation with the wear level, so that the fused feature bears a monotonic relationship with the wear and closely follows its trend.

4.3 Diagnosis Process

The fault diagnosis performed involved a PF based technique, which is widely used for this purpose as well as for failure prognosis. This method involves the approximation of the probability distribution of the condition states using a swarm of points (particles) and a set of weights denoting discrete probability masses. These particles are easy to generate and update in real time, as they are based on a nonlinear dynamic growth model and a measurement model relating the system states with the observed fault indicators. Two process modes, healthy and faulty, are represented by a binary state (0 for healthy, 1 for faulty), that enables the PF to detect the moment a fault is instantiated, and subsequently track the fault dimension (Orchard, 2007).

This PF method was implemented in a .NET framework, which operated in real-time as newer values from FF_1 and FF_2 were fed into the diagnosis system.

4.4 Results

The absolute correlations of the various extracted features are exhibited in Table 1. It is clear that the fused feature based on the selected automatically extracted feature was relatively better in terms of absolute correlation with the wear, than the other fused feature or any other feature for that matter. However, as this is based on the result of a single experiment (additional experiments will yield the exact same result due to the nature of the Fault Diagnosis method), it is not possible to obtain a statistical analysis of the results. Yet, even if someone is not convinced that the result of proposed methodology is better than that of the manually selected feature set, and believes that both methods yield results of the same quality (i.e. the null hypothesis is not rejected), the proposed method is still better. This is because one does not have to undertake the time-consuming process of extracting the features from the vibration data one by one, with the risks of error that this process may have.

It must be stressed that although a lot of features were extracted in these experiments, this must not be confused with the amount of data available. The latter comprised of a small number of samples that rendered the possibility of a proper statistical analysis infeasible.

The fused features were validated based on a two-fold approach: first the absolute correlation with the wear was calculated and then a PF-based Fault Diagnosis was conducted.

Table 1: Absolute Correlation of Automatically and Manually Extracted Features with Level of Wear. The feature showing the highest absolute correlation with the wear level is depicted in bold.

	Feature	|Correlation with Wear|
AFE	f_3	0.8542
	f_4	0.6242
	f_{29}	0.6312
	f_{30}	0.7229
	Fused (FF_1)	**0.9661**
Manual	3x Axial	0.3421
	2x Radial	0.0256
	Total Energy Radial	0.7259
	Tot. En. Ax. / 1x Axial	0.1006
	Tot. En. Rad. / 1x Radial	0.5228
	Fused (FF_2)	0.9302

The two fused features, FF_1 and FF_2, were used for the diagnosis of the corrosion of the bearings, using a Particle Filtering based technique. A screenshot of the fault diagnosis program for the manual, fused feature, FF_2 can be seen in Figure 1 below.

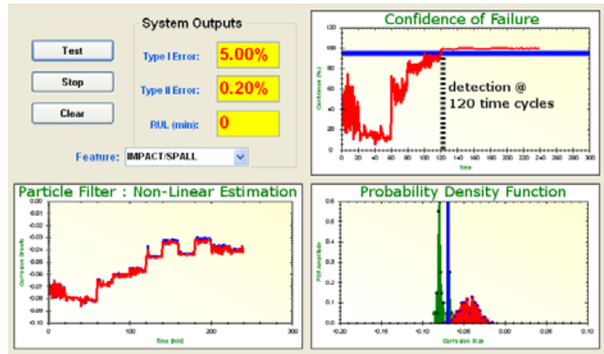

Figure 1: Fault Diagnosis screenshot

The plot of the confidence of failure (upper right graph) shows the increase of the detection confidence as more feature data is input to the diagnosis program, with fault detection occurring when the confidence exceeds the detection threshold (blue line at 95%).

Interestingly, the FF_1 feature yielded a much earlier detection of the fault at the 64th time cycle using the same diagnosis program, almost twice as quick as the FF_2 feature detection (at the 120th time cycle).

4.5 Discussion

It is clear from the aforementioned results that the features extracted using the proposed method yield better correlation with the wear level, rendering them a better basis for fault diagnosis. Especially the fused feature that derives from the selected automatically extracted features yields an outstanding performance, something reflected in the performance of our particle filtering based fault diagnosis system. The early fault detection is of crucial significance, as it allows the commencement of the failure prognosis at an earlier stage. This may prove vital for the proper preparation for the failure, which translates to a more efficient resource management and the limiting of the risk of accident.

5 CONCLUSIONS AND FUTURE WORK

From the research conducted it can be concluded that the proposed method for feature extraction is quite promising and able to provide substantial aid in finding worthwhile features which can perform satisfactorily in a fault diagnosis scheme.

Future work on this topic will include more extensive testing of the method on other frequency data, combination of this method with other feature-related modules developed, and fine-tuning of the method to improve its efficiency and applicability. Also, it will be applied on a larger dataset, one that can be partitioned appropriately so that a more thorough k-fold cross validation analysis can be conducted.

ACKNOWLEDGMENT

The authors would like to take this opportunity to extend their gratitude to the sponsors of the AVDPIP project and our teammates at Impact Technologies who conducted the related experiments and provided the data upon which this research was based.

REFERENCES

Decker. H. (2002) Crack Detection for Aerospace Quality Spur Gears, in *American Helicopter Society 58th Annual Forum*, Montreal. Canada, June 11-13, 2002.

Downham, E. (1976) Vibration in rotating machinery: malfunction diagnosis—art and science, in *Proceeding of the Institution of Mechanical Engineering—Vibration in Rotating Machinery*, London, UK, pp. 1–6.

Kennedy, J. and Eberhard, R. (1995) “Particle Swarm Optimization,” in *Proceedings of IEEE International Conference on Neural Networks*, IEEE Press, Piscataway, NJ, 1995, pp. 1942-1948.

Kennedy, J. and Eberhard, R., Shi. Y. (2001) *Swarm Intelligence*, Morgan Kaufmann, San Francisco, CA, 2001.

Mobley, K. (1999) *Vibration Fundamentals*, 1st ed., Butterworth-Heinemann.

Orchard, M. E. (2007) *A Particle Filtering-Based Framework for On-line Fault Diagnosis and Failure Prognosis*. Ph. D. thesis, Georgia Institute of Technology.

Patel, T. and Darpe, A. (2009) Experimental investigations on vibration response of misaligned rotors, *Mechanical Systems and Signal Processing*, vol. 23, pp. 2236-2252.

Qu, L. S. and Shen, Y. D. (1993) Orbit complexity: a new criterion for evaluating the dynamic quality of rotor system, in *Proceedings of the Institution of Mechanical Engineers Part C 207*, pp. 325–334.

Shi, D. F., Wang ,W. J., Unsworth, P. J., Qu L. S. (2005) Purification and feature extraction of shaft orbits for diagnosing large rotating machinery, *Journal of Sound and Vibration*, vol. 279, pp. 581-600.

Voulgaris, Z. N. (2009) *Discernibility Concept for Classification Problems.* Ph. D. thesis, the University of London.

Voulgaris, Z., Sconyers, C., Vachtsevanos, G. (2010). A Particle Swarm Optimization Approach to Feature Fusion for Failure Prognosis of Engineering Systems, *in Proceedings of 18th Mediterranean Conference on Control and Automation*, Marrakech, Morocco, 2010, pp. 773-777.

Voulgaris, Z. and Sconyers, C. (2010) Z. Voulgaris, C. Sconyers. A Novel Feature Selection Method for Fault Diagnosis, in *Proceedings of 6th IFIP Conference on Artificial Intelligence Applications & Innovations*, October 2010 [pending publication].

Vachtsevanos, G., Lewis, F. L., Roemer, M., Hess, A., Wu B., (2006) *Intelligent Fault Diagnosis and Prognosis for Engineering Systems*, 1st ed. Hoboken, New Jersey: John Wiley & Sons, Inc.

Zacharias Voulgaris was born in Greece where he did his first degree, in Production and Management Engineering, at the Technical University of Crete. He graduated in early 2003 and later that year took a Masters course at City University of London, in Information Systems and Technology. Upon his graduation in 2004 he remained in London where he started his PhD in AI based Classification the following year (2005) at the University of London. Right after he finished, in 2009, he went to the USA for work, accepting an offer for a Research Fellowship at Georgia Institute of Technology, where he has been working until the present day. His work experience also includes TA and teaching work at the University of London, and financial analyses for a company in Greece. His research interests include classification, AI techniques and applications, clustering, urban logistics, optimization, data analysis/synthesis, evolutionary computing, geometry (particularly computational), fault diagnosis and failure prognosis, reasoning, and statistics.

Chris Sconyers received a B.S. in Electrical Engineering and a B.S. in Computer Science from Texas A&M University in 2004. After a brief time spent working as a contractor for Verizon Directories to fund his graduate studies, he entered the Georgia Institute of Technology, where he received his M.S. in Electrical and Computer Engineering in 2006. He is continuing his research at Georgia Tech, pursuing his Ph.D. Currently, Chris is working in the field of UAV-based surveillance and reconnaissance. His interests are in data and image processing for the purposes of target tracking and subject localization. His proposed research focuses on geo-location of ground targets utilizing video sensors and multiple UAVs.