A Novel Feature Evaluation Methodology for Fault Diagnosis

Zacharias Voulgaris, Chris Sconyers

Abstract—**In this paper a novel approach for evaluating features is proposed, along with its applications in the diagnosis of faults. As most existing methods for this task make use of a number of assumptions about the features, their evaluations may not always be very reliable. The proposed methodology aims to address this issue. In this paper three distinct metrics based on an assumption-free framework have been developed and presented, along with a GUI to facilitate their use. The features assessed are based on a real world dataset from a project the authors are involved in. The results of the analysis conducted appear to be promising.**

Index Terms—**Feature Evaluation, Fault Diagnosis, Failure Prognosis, Classification.**

I. Introduction

THE issue of feature evaluation is as old as the field itself. Especially nowadays that the process of feature extraction has been more or less automated, an information overload of features is observed. To tackle this issue several feature selection methods have been developed. Yet, before an intelligent selection can be made, an equally intelligent evaluation of the features has to be conducted. The proposed methodology aims to contribute to this part of the field, by introducing a number of feature evaluation metrics and demonstrating how they can be used successfully, either independently or in combination.

II. Literature Review

A. Feature Extraction

The existence of a fault in a system or component produces uncharacteristic behavior that can be captured through certain sensing spectra, typically vibration [1]. 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 simplest and most direct being the time domain. From this domain, features such as root-mean-squared (RMS), an approximation of signal strength; Kurtosis, a measure of "peakiness"; and entropy can all be extracted [2-4]. 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 [5, 6]. Filtered orbit and sideband analysis on these harmonics can provide further fault information [7]. Including normalizing information such as total spectral energy can inhibit effects of operational variability.

Common frequency-based features may be robust, but require explicit knowledge or restrictive assumptions about the system or component under analysis. 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.

B. Feature Selection

All tables and figures you insert in your document are only to help you gauge the size of your paper, for the convenience of the referees, and to make it easy for you to distribute preprints.

Some evaluation techniques focus on the amount of information available in a set of data, such as Shannon's or differential entropy [8, 9] and Renyi's entropy [10, 11]. However, these measures, in simple terms, are measures of the "randomness" of a set of data, and may not provide specific information on the utility of the data as applied towards a specific goal.

To improve the performance metric, information or assumptions about the intent of the feature may be incorporated. Measures of correlation or dependence between the feature and a set of known data the feature is intended to represent (ground truth) can be used as a tracking metrics; these metrics analyze how the feature tracks the target data [12, 13]. Likewise, the same methods can compare two or more features to each other, giving insight into how much

Z. Voulgaris is a post-doc researcher with the Intelligent Control Systems Laboratory of the Georgia Institute of Technology, Atlanta, GA 30332 USA (phone: 404-894-9356; fax: 404-894-4130; e-mail: zvoulgaris@gatech.edu).

C. Sconyers is a PhD candidate with the Intelligent Control Systems Laboratory of the Georgia Institute of Technology. His main research topic is object tracking via particle filters and has been involved in a number of research projects (e-mail: csconyers@gatech.edu).

redundant information exists. Such measures on their own can be misleading, requiring accompanying analysis to determine the true effectiveness of the feature.

Other typical assumptions on the form of the data may be asserted through evaluators such as separability, Bhattacharyya distance, Fisher's Discriminant Ratio, and Discernibility [14-16]. These metrics analyze features best based on discrete classes of data, and are difficult to represent in a continuous sense. Also, like correlation, these tend to apply only broad restrictions to a feature. An evaluator may pass a feature that contains little to no useful information or Discernibility potential, or could fail a feature that may contain enough information to draw probable conclusions.

III. Methodology

The methodology proposed in this paper is quite simple and straight-forward. The idea is to use a number of assumption-free metrics for evaluating features, both independently and in combination. To attain this, these metrics must yield values on the same interval. By doing this, the evaluations of these metrics can be easily combined by taking their product, or their numeric average.

A. Proposed Metrics

Three metrics are proposed here, all of which make no assumptions about the distribution the feature values follow. These metrics were chosen for their high quality evaluation performance and for the fact that they are not closely correlated to each other. The proposed metrics are the Assumption Free Discriminant Ratio, the Absolute Density Correlation and the Density Monotonicity.

1) Assumption Free Discriminant Ratio (AFDR)

This metric is similar to the Fisher Discriminant Ratio, with the difference that it does not assume a normal distribution (or any other distribution for that matter) and it takes values in [0, 1]. In essence it performs a number of statistical tests, similar to the z-test that FDR performs. However, it makes use of the vectors' densities to form the pdf of each distribution, instead of the probabilistic densities that derive from a distribution model.

2) Absolute Density Correlation (ADC)

This metric is just like the Pearson Correlation one, with the difference that it takes into account the densities of the vectors under consideration, and uses them as weights. Also, the absolute value of their correlation is used, as this yields more meaningful information in the feature evaluation process. Just like the other metrics, it yields values in [0, 1].

3) Density Monotonicity (DM)

This metric is also based on the density approach and provides insight to how monotonous a vector it compared to another one. This is particularly useful for cases of features that are used for fault diagnostics/prognostics purposes, as it is desirable that the features follow a monotonous trend compared to the wear level. In this approach, monotonicity is defined as the probability of inter-class distances being greater than zero. The inter-class distances are calculated by taking the difference between class i and the maximum of class $i-1$, for every class $i = 2 \dots m$, where m is the total number of classes. This metric can be used on its own with an additional parameter, namely a given confidence threshold, so that it can also yield a crisp output regarding whether the feature under consideration is monotonous or not. This can be particularly useful when examining a large number of features.

B. Overview of Integrated Feature Evaluation System

The aforementioned metrics are combined in a single feature evaluation program. This application takes as inputs the feature matrix and the class vector (which in the cases examined in this paper is the wear level vector). These vectors are given as two separate .csv files. Afterwards the evaluation metrics are selected. The results of the evaluation are given in the output of the program, which takes the form of another .csv file. If more than one metric are selected, the combination of all the selected metrics, in the form of their product, is also given in the results. It should be noted that all the features in the features matrix are evaluated independently, so for a set of m features and n evaluators, a $(n+1)$ x m matrix will be given. If a cumulative evaluation is desired, then a different approach must be taken (e.g. use of the SID measure on the whole feature set). A screenshot of the evaluation system developed can be viewed in the Appendix.

IV. Experiments and Results

A series of experiments were conducted, so as to investigate the relationship between the robustness of features and their evaluation. These were based on a real-world dataset using the sensor readings of a helicopter spline which is artificially worn with multiple levels of wear.

A. Experimental Setup

A set of four classifiers were used to perform fault diagnosis based on a set of three features based on phase data. These features exhibited a range of performance potential (mirrored in the accuracy rate of the classifiers) and measured by each one of the aforementioned metrics. The classifiers used were all parameter-free and included a distance-based one (k Nearest Neighbor – kNN), a statistical one (Linear Discriminant Analysis – LDA), a neuro-fuzzy classifier (ANFIS) and a Support Vector Machine (LS-SVM). The number of neighbors used for kNN was 5 while the number of epochs used for the training of ANFIS was 40. The features tested comprised of 5400 patterns and covered three distinct wear levels: baseline (0 mils of wear), light wear (2 mils) and moderate wear (3.25 mils), grouped in two classes, *healthy* and *worn*. This was done because the amount of wear is of secondary importance as regards fault diagnosis. Also, 50 rounds of 10-fold cross validations were carried out. Note that due to the nature of the LS-SVM classifier, the training and testing set were inverted, so only 10% of the data was used for training and the rest 90% for testing.

Parallel to the classification experiments, a number of evaluators are applied to the aforementioned features. These

ISBN: 978-988-17012-0-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

evaluators are the ones mentioned in Sec. 3A along with absolute correlation (AC) and a version of Fisher's Discriminant Ratio (FDR) yielding values in (0, 1]. The last two evaluators are quite wide-spread and constitute the core of the traditionally used feature evaluators, in the field of fault detection.

B. Results

The results of these experiments can be viewed on Figure 1 below.

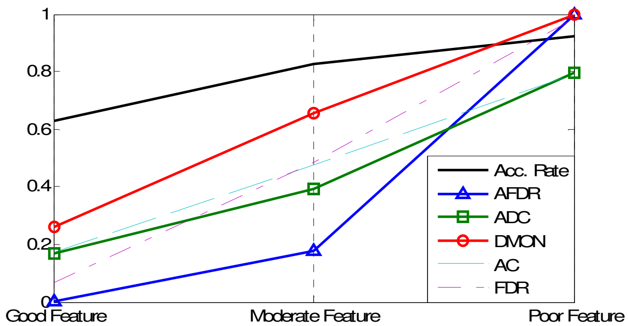


Fig. 1 Graphical comparison among the various feature evaluators tested and the average accuracy rate of the classifications.

It is noteworthy that the evaluator that stands out is Density Monotonicity, exhibiting an exceptionally promising trend in relation to the accuracy rate. In addition, this particular evaluator has a very low computational overhead, rendering it a very practical alternative to the traditional evaluators.

The other evaluators performed quite well too, as they managed to capture the quality of each feature, in terms of how accurate a fault diagnosis prediction it can yield. Interestingly all of them exhibited a good monotonicity in respect to the accuracy rate, so in practice they can all be used to evaluate a given feature. Also, as all of them yield values in [0, 1], they can be easily combined by taking their product. This way a more generic evaluator can be obtained, though this will come at additional computational cost.

V. Discussion

The proposed evaluators appear to map the accuracy rate of the classifications satisfactorily. This translates into a kind of insight to the quality of the examined features, a type of predictor of their classification potential. Of course one could employ a simple classifier to do that, using a k-fold cross validation scheme, but this would be more time-consuming and may not provide the most accurate insight (since some classifiers' performance heavily depends on the data). Also, the feature evaluators, particularly those yielding low computational cost, can be easily employed in an optimization scheme for selecting or generating new features. For this purpose, evaluators such as Density Monotonicity would be ideal.

The well established evaluator of (absolute) correlation with the wear exhibits a very good behavior in evaluating a feature. However, it may not always be as practical, particularly in cases where the wear growth follows a non-linear pattern. Therefore, it is advantageous to have alternatives to it, which as shown in the conducted experiments, exhibit an equally robust performance.

It is noteworthy that the proposed feature evaluators are accompanied by an easy to use graphical user interface. This program loads two files, one with the feature values and one with the wear levels (labels). Then, once the evaluators to be used are selected, it employs all of these evaluators and outputs another file containing a table with the results. All the aforementioned files are in .csv form, which renders them accessible by a variety of applications, including MATLAB, spreadsheet programs, and even basic text editors. The program is freeware and can be executed in any Windows OS.

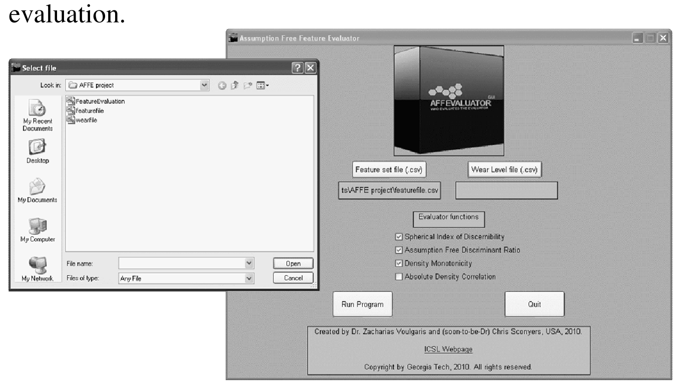
VI. Conclusions and Future Work

In this paper three alternative feature evaluation metrics were introduced and tested on three features of variable quality. Four classifiers were used for the experiments that were carried out, aiming at performing an anomaly detection task based on the aforementioned features. The proposed evaluators, which make use of no assumptions on the data they are applied on, appeared to perform satisfactorily in terms of how they related to the average accuracy rate of the classifiers. Two established evaluation metrics were also employed, exhibiting very similar results. Also, one of the proposed evaluators, Density Monotonicity, appeared to follow the accuracy rate trend very closely, while at the same time yielded a very low computational overhead. From all this, it can be concluded that the proposed metrics, especially the Density Monotonicity, are a viable alternative for the evaluation of features.

Future work in this field may include a sensitivity analysis of these evaluators in similar problems of anomaly detection as well as fine-tuning them so that their performance can be enhanced.

Appendix

This is a typical screenshot of the AFFE (stand-alone) application developed. Its user-friendly interface coupled with its simplicity of use makes it a viable alternative for feature evaluation.



ISBN: 978-988-17012-0-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

ACKNOWLEDGMENT

The authors would like to acknowledge *Impact Technologies* for carrying out all the experiments that yielded the project data, which they shared with us. Also, we would like to thank our sponsors of the AVDPIP project for which this research was conducted.

REFERENCES

[1] Vachtsevanos, G., Lewis, F. L., Roemer, M., Hess, A., Wu, B., *Intelligent Fault Diagnosis and Prognosis for Engineering Systems*. John Wiley & Sons, Inc, Hoboken, NJ, 2006.

[2] Mobley, K., *Vibration Fundamentals*, 1st ed., Butterworth-Heinemann, 1999.

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

[4] Decker, H., "Crack Detection for Aerospace Quality Spur Gears". *American Helicopter Society 58th Annual Forum*, Montreal. Canada, June 11-13, 2002.

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

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

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

[8] Lindley, D.V., "On a measure of the information provided by an experiment". *Ann. Math. Statist*, vol. 27, 1956, pp. 986-1005.

[9] Kovalevsky, V.A., "The problem of character recognition from the point of view of mathematical statistics". *Character Readers and Pattern Recognition*, V.A. Kovalevsky ed., Spartan, New York. 1968.

[10] Renyi, A. "On measures of entropy and information". *Proceedings of the Fourth Berkeley Symposium Math. Statist. and Probability*, vol. 1, 1960, pp. 547-561.

[11] Ben-Bassat, M. and Raviv, J., "Renyi's entropy and the probability of error". *IEEE Trans. Inform. Theory*, vol. 24, 1978, pp. 324-331.

[12] Dietrich, C.F., *Uncertainty, Calibration, and Probability,* 2nd ed., Boca Raton, FL, Taylor & Francis Group, 1991.

[13] Lewis, P.M., "The characteristic selection problem in recognition systems". *IEEE Transactions on Information Theory,* vol. 8, 1962, pp. 161-171.

[14] Ali, S.M. and Silvey, S.D., "A general class of coefficients of divergence of one distribution from another". *Journal of the Royal Statistical Society Series B.,* vol. 28, 1966, pp. 131-142.

[15] Lainiotis, D.G., "A class of upper bounds on probability of error for multihypothesis pattern recognition". *IEEE Transactions on Information Theory,* vol. 15, 1969, pp. 730-731.

[16] Voulgaris, Z. N., *Discernibility Concept in Classification Problems*. PhD thesis, University of London, 2009.