EFFECTS OF COMPRESSED SENSING ON CLASSIFICATION OF BEARING FAULTS WITH ENTROPIC FEATURES

M. L. D. Wong[1], *M. Zhang*[2], *A. K. Nandi*[2*]

[1]Swinburne University of Technology, Sarawak Campus, Jalan Simpang Tiga, Kuching, 93350, Sarawak, Malaysia
[2]Department of Electronic and Computer Engineering, Brunel University London, Uxbridge, UB8 3PH, United Kingdom

ABSTRACT

The ability of automatically determining the underlying fault type in-situ for a roller element bearing is highly desired in machine condition monitoring applications nowadays. In this paper, we classify roller element fault types under a compressed sensing framework. Firstly, vibration signals of roller element bearings are acquired in the time domain and resampled with a random Bernoulli matrix to emulate the compressed sensing mechanism. Sample entropy based features are then computed for both the normalized raw vibration signals and the reconstructed compressed sensed signals. Classification performance using Support Vector Machine (SVM) shows slight performance degradation with significant reduction of the bandwidth requirement.

Index Terms— Bearing Fault Classification, Compressed Sensing, Machine Condition Monitoring, Sample Entropy

1. INTRODUCTION

Machine condition monitoring research [1] has advanced in recent years and is moving from a predominantly labour intensive human monitoring based approach to a one that is highly automated where interventions are required only when a fault is detected [2-5]. This has driven the need for algorithms that automatically identify when a machine's operating characteristics digress from its normal operating conditions. Many solutions have been proposed in this respect, from basic decision making by observation of the various characteristics of the vibration time series, through to various machine learning based approaches.

The ability of identifying a fault condition in roller element bearings enables one to prescribe appropriate solution when a fault is detected, e.g. applying lubrication, dusting, replacing the faulty ball bearing, etc. However, in many industries, scheduling downtime for corrective maintenance is often a complex task. Furthermore, for large scale machineries, waiting for a fault to occur could be catastrophic to the entire operation. Therefore, in such cases, periodically scheduled preventive maintenance measures are often sought. Nevertheless, these preventive maintenance measures, in many cases, are not so cost effective, as quite often the optimal lifetimes of specific parts are not known. This drawback led the industries and the research communities to search for measures that are able to estimate and predict when a fault is likely to occur, i.e. predictive maintenance.

To enable predictive maintenance, one needs to be able to monitor the running condition of the machine online. A common automation approach is to utilise the classical pattern classification framework. Within this framework, we first extract discriminating features from the raw vibration signals and then employ a chosen classifier to classify the healthy and faulty patterns.

With the advancement of wireless sensors and communication technologies, it is now possible to monitor machines remotely. However, there is always a bandwidth limitation which limits the number of machines that can be monitored across the network. In this work, our primary aim is to investigate a mechanism to reduce the bandwidth requirement for such deployment. More precisely, compressed sensing [6] is utilised to enable us to sample at much less than the Nyquist rate and subsequently the effect on the classification performance is studied herein.

2. PROBLEM DESCRIPTIONS

Our aim is to distinguish faulty roller elements bearings from the healthy normal bearings and to ascertain which type of faults. The vibration data used in this paper have been taken from experiments on a small test rig, which simulates an environment for running roller bearings. Six

*Corresponding author: Asoke.Nandi@brunel.ac.uk

Professor Nandi is a Distinguished Visiting Professor at Tongji University, Shanghai, China.

978-0-9928626-3-3/15/$31.00 ©2015 IEEE

conditions have been recorded and tested. Two normal conditions -- a brand new condition (**NO**) and a worn but undamaged condition (**NW**); four fault conditions -- inner race (**IR**) fault, outer race (**OR**) fault, rolling element (**RE**) fault, and cage (**CA**) fault. Data was recorded over a range of 16 speeds. The variation of speeds adds some non-stationary characteristics to this problem. Fig. 1 depicts some typical time series plots for the six different aforementioned conditions.

Depending on the fault conditions, the defects modulate the vibration signals with their respective patterns. The inner and outer race fault conditions have a fairly periodic signal; the rolling element fault may or may not be periodic, dependent upon several factors including the degree of damage to the rolling element, the loading of the bearing, and also the track that the ball describes within the raceway itself. The cage fault generates a random distortion, which also depends on the degree of damage and the bearing loading.

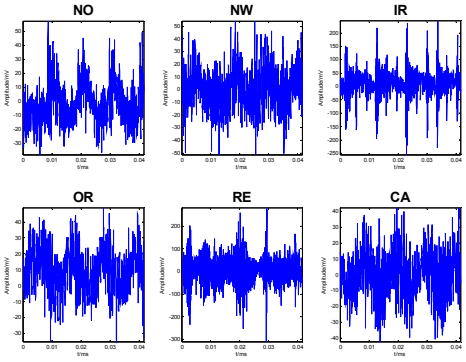


Fig. 1. : Typical vibration signals for the six different conditions.

The experimental data used in this work was acquired from a roller element bearing test rig. The test rig consists of a DC motor driving the shaft through a flexible coupling, with the shaft supported by two plummer bearing blocks. A series of damaged bearing were inserted in one of the plummer blocks, and the resultant vibrations in the horizontal and vertical planes were measured using two accelerometers. The output from the accelerometers was fed back through a charge amplifier to a Loughborough Sound Images DSP32 ADC card (using a lowpass filter with a cut-off 18 kHz), and sampled at 48 kHz, giving a slight oversampling. The machine was run at a series of different speeds ranging between 25 and 75 rev/s, and ten time series were taken at each speed. This gave a total of 160 examples of each condition, and a total of 960 raw data files to work with.

3. COMPRESSED SENSING

In the last decade, compressed sensing or compressive sensing [6], has received much attention for its ability to allow one to sample at much less than the Nyquist rate and recover the original signal afterwards. One preliminary requirement is that the signal needs to be compactly represented in some domain, e.g. time, frequency, Discrete Cosine Transform, etc. In our case, the vibration signals can be decomposed into several dominant frequency components and therefore one can represent the time domain signals using a finite sum of sinusoids, which satisfy the sparseness in frequency domain. When a fault occurs, it exhibits itself both in the time and the frequency domain.

Briefly, we present the simplified compressed sensing framework as follows:

Let $y = Mx$ be the compressed sensed signal vector where $x \in R^{n\times 1}$ denotes the original signal vector. The sensing matrix, $M \in R^{m\times n}$, is a chosen matrix (e.g. Fourier, Gaussian or Bernoulli Matrix) for which $m < n$.

If the signal is sparse in the frequency domain, then one can actually sample in the frequency domain which can be implemented in the time domain by carefully selecting the sensing matrix as the product of an inverse Fourier Transform Matrix, F^{-1}, and a sparse binary matrix, R, with at most a single non-zero entry in each of its column (see [7]). In other word, $M = F^{-1}R$.

Thus, reconstruction from the compressed sensed vector y can then be dealt with as a solution of underdetermined linear system of equations. A conventional convex optimization technique based on ℓ_1-norm was used to reconstruct a vector, $\hat{\mathbf{x}}$, as the estimation of the original x:

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{y}=\mathbf{M}\mathbf{x}} ||\mathbf{x}||_1 \qquad (1)$$

In our paper, we segmented a time series of original signals into non-overlapping segments of 512 data points. The ratio of down-sampling of x is signified by α, i.e. $m = \alpha n$. We set α to 0.5 and 0.25 respectively; thus the compressed sensed vectors are of lengths 256 and 128 respectively.

4. ENTROPICS FEATURES

For classifying the reconstructed signals, we lend ourselves to the entropic features proposed in [8] as introduced in this section.

The classical formulation of Shannon entropy, $H_n = -\sum_{x_i \in X} p(x_i) \log p(x_i)$, has long been interpreted as

a measure of system uncertainty. For a time indexed sequence of discrete random variables, such as the sampled and quantized vibration signals above, the joint entropy of each samples are defined as:

$$H_n = -\sum_{x_i \in X_0} \sum_{x_i \in X_{n-1}} p(x_0, \cdots, x_{n-1}) \log p(x_0, \cdots, x_{n-1}) \quad (2)$$

where p(x_1,⋯, x_n) is the joint probability of the n samples in the sequence. For characterizing the system dynamics, the Kolmogorov-Sinai (KS) entropy, which is defined as the average rate of new information generation, is usually used. However, it is difficult to estimate KS entropy within a satisfactory precision. However, for short, and noisy time series, Pincus [9] had proposed the approximate entropy (ApEn) to estimate the rate of generating new information. The notation ApEn(M, r, N) denotes the approximated negative natural logarithm of the conditional probability that a N-point sequence, having repeated itself for M points within a selected tolerance, r, will be repeating itself for (*M*+1) points. The tolerance parameter, r, by convention, is set to a fraction of the standard deviation of the sequence for convenience.

To reduce the bias caused by pattern self-matching, Richman and Moorman [10] proposed the Sample Entropy (SampEn) for sampled time series data from a continuous process which give the precise negative logarithm intended for ApEn above. For the actual process of computation, the readers are referred to [10-11] for an in-depth analysis and discussion. A typical setting for the tolerance factor, i.e. r equals 0.2 times the standard deviation, is used in this work.

5. EMPIRICAL VALAIDATION

We validated our proposed method through computer experiments. To emulate the compressed sensing environment, each raw vibration acquired as described in Section 2 are first divided into ten non-overlapping segments of 512 samples, and each segment is then resampled according to the framework in Section 3 using values of α = 0.5 and 0.25. Afterwards the individual segments are reconstructed using L1 convex optimization implemented using the CVX MATLAB toolbox [12-13] and then concatenated back into a vibration signal of length 5120. With this, we have three sets of database for investigation, corresponding to the original data, data for α = 0.5, and data for α = 0.25.

Fig. 2 depicts an illustration of 512 points of the original data and their reconstructions for the six conditions. Original data is described by blue lines and reconstruction data by red ones. It is observed that, although there some reconstruction errors, the two signals are largely similar.

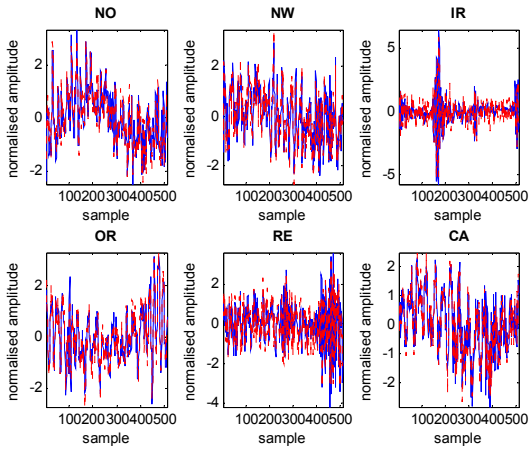


Fig. 2. : Comparisons of the original signals (in blue) versus the reconstructed signals (in red) for the six conditions.

5.1. EXPERIMENTAL SETUP

For the roller element bearing dataset described in Section 2, three SampEn (m = 0, 1, 2) features per vibration were extracted. The SampEn was extracted by utilizing the PhysioNet MATLAB scripts (available online at http://www.physionet.org/physiotools/sampen/matlab/).

The computing platform for the experiment reported in this paper is a desktop, with an Intel Core i7 quad core processor, 8GB RAM, professional graphic card with 1GB VRAM, running on 64-bit Windows 7. MATLAB 2013a is used as the main testing platform, with necessary classifier toolboxes.

The classification accuracy rates are obtained by averaging the results of ten experiments for each classifier and for each experiment, ten-fold validation was employed. The averaged accuracy rate is used to indicate the performance of the classifier.

5.2. EXPERIMENTAL RESULTS

For illustrative purposes, we chose the classical Support Vector Machine (SVM) [14] as the classifier for these experiments. Three sets of experiments are conducted by using features computed from i) the raw vibration signals, ii) 10 non-overlapping segments of 512 samples reconstructed using 256 compressed sensed data (α = 0.5), and iii) 10 non-overlapping segments of 512 samples reconstructed using 128 compressed sensed data samples (α = 0.25). For each experiment, the hyper-parameters of the

SVMs are chosen using a logarithmic grid selection method with a range of gamma and C values.

The overall classification results are presented in Table 1 and some sample confusion matrices are shown in Table 2. From the results, we achieved 92.4% accuracy with α set at 0.5 and 84.6% accuracy when α is reduced to 0.25.

	SVM's Hyper-parameters	Accuracy %
Raw Vibration	Gamma = 16 C = 32	98.9 (1.2)
Compressed Sensed (alpha = 0.5)	Gamma = 46 C = 35	92.4 (0.5)
Compressed Sensed (alpha = 0.25)	Gamma = 42 C = 36	84.6 (3.4)

Table 1. Overall classification accuracies and their associated standard deviations (shown in brackets)

6. DISCUSSION

We observed that it was possible to achieve good performance while reducing the bandwidth requirement by 50% and 75%. However performance degradation is observed, especially in the case of $\alpha = 0.25$, 10% loss was observed in terms of classification accuracy. From analysing the feature space for the three datasets as depicted in Fig. 3, we observed that the overlap among different conditions increases as the value of α reduces.

From empirical studies, we also noticed that the SampEn features require large sample size for stable computation. Owing to page constraint, we did not present the results for the feature values computed using different segment size. It also appears that the higher order features are less stable for the reconstructed compressed sensed vibration signal, especially for lower ratio.

There are several outstanding interesting issues remain to be investigated: 1) What is a good value of segment length for adequate reconstruction? In this work, we have only presented the results for n = 512. 2) What is the optimum number of segments required for stable features computation? We have used 10 segments in this work. 3) Do we need to compute features in the reconstructed domain? Would it be feasible to compute directly in the compressed domain? In our future works, we plan to take a closer look at these open issues.

7. CONCLUSION

We have demonstrated that it is possible to sample the vibration data of roller element bearings at less than the Nyquist rate and reconstruct the signal for fault classification. For simplicity, we have opted for the entropic features and empirical data shows that classification results suffer from a slight degradation but requires much less bandwidth.

Predicted Conditions	True Conditions						class Precision (%)
	NO	NW	IR	OR	RE	CA	
NO	160	0	0	0	0	0	100.0
NW	0	156	0	0	0	7	95.7
IR	0	0	160	0	0	0	100.0
OR	0	0	0	160	0	0	100.0
RE	0	0	0	0	160	0	100.0
CA	0	4	0	0	0	153	97.5
class recall (%)	73.1	95.6	99.4	91.3	100.0	92.5	

(a) Raw Vibration Signal

Predicted Conditions	True Conditions						class Precision (%)
	NO	NW	IR	OR	RE	CA	
NO	117	0	0	14	0	0	89.3
NW	0	153	0	0	0	9	94.4
IR	0	0	159	0	0	0	100.0
OR	42	0	0	146	0	3	76.4
RE	0	0	1	0	160	0	99.4
CA	1	7	0	0	0	148	94.9
class recall (%)	73.1	95.6	99.4	91.3	100.0	92.5	

(b) Compressed Sensed Data $\alpha = 0.5$

Predicted Conditions	True Conditions						class Precision (%)
	NO	NW	IR	OR	RE	CA	
NO	110	0	0	13	0	8	84.0
NW	0	149	0	2	0	6	95.0
IR	0	0	158	0	7	0	95.8
OR	24	1	0	116	0	20	72.0
RE	0	0	2	0	153	0	98.7
CA	26	10	0	29	0	126	66.0
class recall (%)	68.8	93.1	98.8	72.5	95.6	78.8	

(c) Compressed Sensed Data $\alpha = 0.25$

Table 2. Sample confusion matrix for the three scenarios

REFERENCES

[1] P. J. Tavner, *Review of condition monitoring of rotating electrical machines*. IET Electrical Power Applications, Vol. 2, IET (2008), pp. 215-247.

[2] M. L. D. Wong, L. B. Jack, A. K. Nandi, *Modified self-organising map for automated novelty detection applied to vibration signal monitoring*, Mechanical Systems and Signal Processing, 20(3), 593–610 (2006).

[3] R. Yan, R. X. Gao, *Machine health diagnosis based on Approximate Entropy*. in: Proceedings of the 21st IEEE Conference of Instrumentation and Measurement Technology Conference 2004, Vol. 3, Como, Italy (2004), pp. 2054—2059.

[4] J. Tian, J.-J. Gu, X.-Z Peng, Z.-M Qing, *A fault diagnosis method based on wavelet ap-proximate entropy for fan*, in: Proceedings of 6th International Conference on Machine Learning and Cybernetics, Hong Kong, (2007) pp. 519—523.

[5] M. L. D. Wong, S.-H. Lee, A. K. Nandi, *Anomaly Detection of Rolling Elements using Fuzzy Entropy and Similarity Measures*, in: Proceedings 10th International Conference on Vibrations and Rotating Machine (ViRM 2012), London (2012), pp. 693-702.

[6] D. L. Dohono, *Compressed sensing*, IEEE Transactions on Information Theory, Vol. 52, No. 4, IEEE (2006), pp. 1289-1306.

[7] M. Rudelson, V. Roman, *On sparse reconstruction from Fourier and Gaussian measurements*, Communications on Pure and Applied Mathematics Vol 61, No. 8, Wiley Inc. (2008), pp. 1025-1045.

[8] M. L .D. Wong, L. Chao, A. K. Nandi, *Classification of Ball Bearing Faults using Entropic Measures*, Proc. Surveillance 7, (2014). Available on-line: http://surveillance7.sciencesconf.org/resource/page/id/20

[9] S. M. Pincus, *Approximate entropy as a measure of system complexity*, Proceedings of the Natioal Academy of Sciences USA, Vol. 88, National Academic of Sciences (1992), pp. 2297—2301.

[10] J. S. Richman, J. R. Moorman, *Physiological time-series analysis using approximate entropy and sample entropy*, American Journal of Physiology Heart and Circulatory Physiology, Vol. 278, No. 12, The American Physiological Society (2000), pp. H2039-H2049.

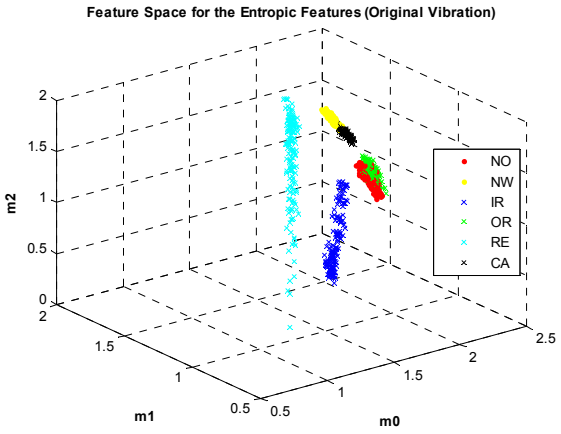
[11] D. E. Lake, J. S. Richman, M. P. Griffin, J. R. Moorman, *Sample entropy analysis of neonatal heart rate variability*. American Journal of Physiology Regulatory Integrative Comparative Physiology, Vol. 283, The American Physiological Society (2002), pp. R789—R797.

[12] CVX Research, Inc. *CVX: Matlab software for disciplined convex programming*, version 2.0, (2011). http://cvxr.com/cvx

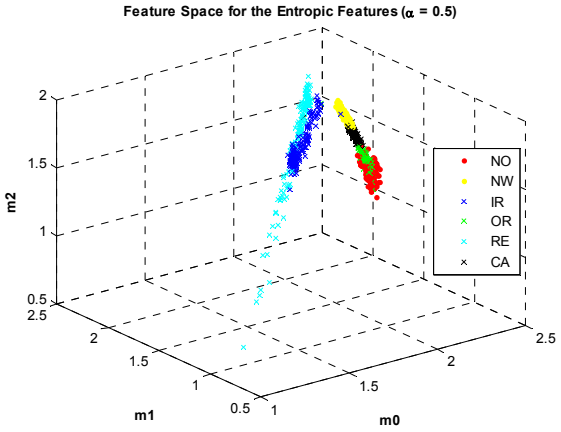
[13] M. Grant and S. Boyd. *Graph implementations for non-smooth convex programs*, Recent Advances in Learning and Control (a tribute to M. Vidyasagar), V. Blondel, S. Boyd, and H. Kimura, editors, Lecture Notes in Control and Information Sciences, Springer, (2008), pp. 95-110.

[14] Chang, Chih-Chung, Lin, Chih-Jen, *LIBSVM: a library for support vector machines*. ACM Transactions on Intelligent Systems and Technology, Vol. 2, No. 2, ACM (2011), pp. 27:1--27:27, 2011. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm

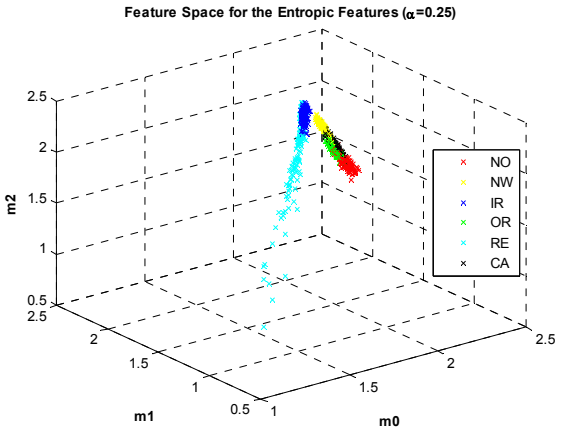
[15] T. Mu, A. K. Nandi. *Automatic tuning of L2 - SVM parameters employing the extended Kalman filter*" Expert Systems Vol. 26, No. 2, Elsevier B.V. (2009), pp. 160-175.



a) Original Vibration Data



b) Compressed Data α = 0.5



c) Compressed Data α = 0.25

Fig. 3. A comparison of the three data sets in feature space.