MDL AND SIGNAL CHANGE DETECTION IN MACHINE CONDITION MONITORING

Jenni Hulkkonen[1] *and Jukka Heikkonen*[2]

[1] Helsinki University of Technology
Department of Biomedical Engineering and Computational Science
P.O.Box 9203, FI-02015 TKK, Finland, Jenni.Hulkkonen@tkk.fi
[2] European Commission - Join Research Centre
Institute for the Protection and Security of the Citizen (IPSC)
Maritime Affairs Unit
TP 051, Via Fermi 1, I-21020 Ispra (VA), Italy, Jukka.Heikkonen@jrc.it

ABSTRACT

A minimum description length (MDL) based sequentially normalized maximum likelihood (SNML) approach combined with an autoregressive (AR) model is proposed for signal change detection in machine condition monitoring. The results showed the success of the method to detect signal changes and distinguish different ball bearing failures.

1. INTRODUCTION

Signal change detection in machine condition monitoring is a crucial task in the analysis of possible machine failures [1]. Traditionally detection of signal changes corresponding to incipient machine failures or other changes in conditions are based on rules and predefined failure types [1] resulting in rather rigid models. Statistical and information theoretical methods can alleviate some of the problems in traditional models.

In this paper a minimum description length (MDL) [2] and especially sequentially normalized maximum likelihood (SNML) [3] based method combined with an AR model is proposed for the signal change detection. The performance and behaviour of the approach are experimented by two real data sets.

2. MDL IN SIGNAL CHANGE DETECTION

Rissanen's MDL principle [2] is an approach trying to overcome typical model selection problems, such as overfitting. According to MDL the best model is the one that allows the shortest total code length for both the data and the model. In MDL, stochastic complexity (SC) is interpreted as the shortest achievable code length for encoding and hence it provides a measure for model comparison. SC is also suited for machine condition monitoring as the change of the signal (complexity) induced by a change in a machine's condition can be measured by it.

A recently proposed definition for SC is sequentially normalized maximum likelihood (SNML) [3]. SNML provides some advantages over traditional MDL formulations, and especially over the so-called normalized maximum likelihood (NML) [4] universal model, as with SNML there is no need for hyperparameters and the SC for time series data is computable. When SNML is combined with the AR process, we have a tool for signal change detection.

Consider a data vector $y^n = [y_1, \ldots, y_n]'$ modeled by AR(k) model as $y_t = b'\bar{x}_t + \epsilon_t = \sum_{i=1}^{k} b(i)x_{it} + \epsilon_t$, $t = 1, \ldots, n$, where $b(i), i = 1, \ldots, k$ are the model parameters, x_{it} are the components of columns $\bar{x}_t = [y_{t-1}, \ldots, y_{t-k}]'$, defining the regressor matrices X_t and ϵ_t is an iid Gaussian noise of zero mean and variance σ^2.

The idea of SNML lies on the sequentially maximized conditionals. Consider the maximization problem

$$\max_{\sigma^2} \prod_{t=1}^{n} f(y_t|y^{t-1}, X_t; \sigma^2, b_t) \;, \tag{1}$$

where b_t are maximum likelihood estimates, calculated from the data available up to t. With the solution $\hat{\sigma}_n^2 = \frac{1}{n}\hat{s}_n$, where the sequentially minimized sum of the squared deviations are calculated recursively as $\hat{s}_t = \sum_{j=m}^{t}(y_j - \bar{x}_j' b_j)^2 = \sum_{j=m}^{t-1}(y_j - \bar{x}_j' b_j)^2 + (y_t - \bar{x}_t' b_t)^2 = \hat{s}_{t-1} + \hat{e}_t^2$, where m is the smallest fixed number for which the ML estimate can be computed, the value of density function is $f(y^t|X_t) = (2\pi e \hat{\sigma}_n^2)^{-n/2}$, where parameter estimates $\hat{\sigma}^2$ and b_t have been dropped to keep the notation uncluttered. Now, we define

$$f(y_t|y^{t-1}, X_t) = \frac{f(y^t|X_t)}{f(y^{t-1}|X_{t-1})} \tag{2}$$

and so we can calculate the non-normalized conditional density function. The normalized conditional density distribution is

$$\begin{aligned} \hat{f}(y_t|y^{t-1}, X_t) &= \frac{f(y_t|y^{t-1}, X_t)}{K(y^{t-1})} \;, \\ K(y^{t-1}) &= \int f(y_t|y^{t-1}, X_t)\, \mathrm{d}y_t \;. \end{aligned} \tag{3}$$

By multiplying the normalized conditional density distributions we get the desired parameter free density function, called the SNML model

$$\hat{f}_{\mathrm{SNML}}(y^n|X_n) = q(y^m|X_m) \prod_{t=m+1}^{n} \hat{f}(y_t|y^{t-1}, X_t) \ , \tag{4}$$

where $q(y^m|X_m)$ is initial density function. The negative logarithm of the SNML model in Eq. 4 gives the stochastic complexity (SC) criterion for the model order selection to be minimized. The criterion is

$$-\ln \hat{f}_{\mathrm{SNML}}(y^n|X_n) = \frac{n}{2}\ln(2\pi e \hat{s}_n/n) - \sum_{t=m+1}^{n} \ln(1-d_t) + \frac{1}{2}\ln n + O(1) \ , \tag{5}$$

where $d_t = \bar{x}_t'(X_t X_t')^{-1}\bar{x}_t$.

The proposed signal change detection algorithm is following. First, the signal is windowed, i.e. it is processed in smaller segments. This helps in recognizing short and long-term signal complexity changes. For each window an optimal AR model order and SC value are computed by minimizing the SC criterion (Eq. 5). As a result we have a description (features) for each signal sample (i.e. time step) for further analysis. Here we used the Self-Organizing Map (SOM) [5] for the post-processing task due to its well known clustering and visualization abilities. Thanks to SOM the operator can easily evaluate the state of the machine visually on a computer screen.

3. EXPERIMENTAL RESULTS

The experiments were performed using two data sets. The first data set is based on a laboratory test where the measured signal was corrupted by an external source. The second data set consists of real industrial measurements, allowing more realistic validation of the method.

First, a movable drawer unit on castors was used to generate a signal with a 3-axis accelerometer (SCA3000-E04, VTI Technologies Oy, [6]). On the drawer we attached a mobile phone that vibrates as a signal for an incoming call. Our goal was to recognize seven mobile phone vibrations hindered by the oscillations due to the drawer movement. Only the vertical acceleration component was used and the signal consisted of 6390 samples (sampling rate 200Hz). The following window sizes were employed: 50, 100, 250 and 750 samples. Figure 1 shows the original signal and SCs as a function of window size. These results can be compared to AR model order results (by SNLS) in Figure 2.

From Figure 1 it can be observed that the windows of 100 and 250 samples reveal the vibrations well. With the window size of 50, the resulting SC feature is relatively ”noisy”, i.e. the size of the window is too short. When the window size is too long (750) the resulting SCs are only able to detect some general changes in the signal. Although one would expect that a change in signal complexity affects the optimal AR model order accordingly, this does not apply generally, as seen in this case.

The ball bearing fault data set (provided by Neurovision Oy [7] company) consists of three different typical

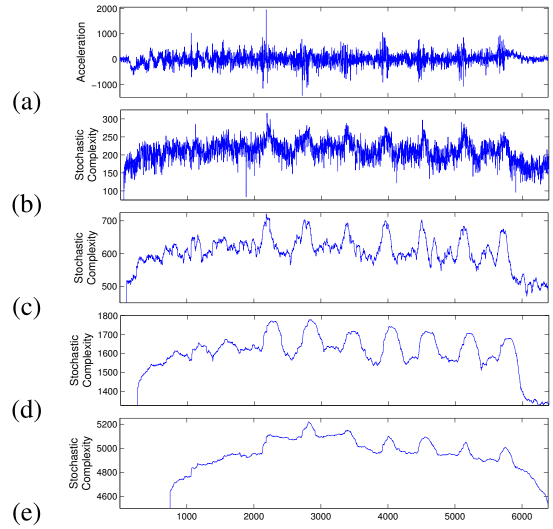

Figure 1. Results of the drawer unit case. (a) The original signal, (b)-(e) stochastic complexity with different window sizes: (b) 50, (c) 100, (d) 250 and (e) 750 samples.

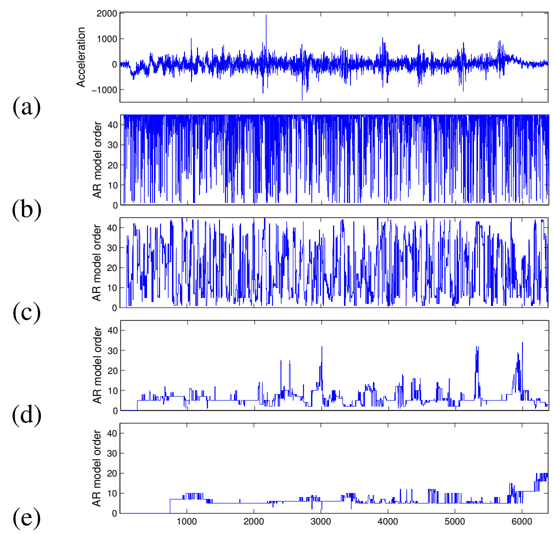

Figure 2. Results of the drawer unit case. (a) The original signal, (b)-(e) AR model order chosen by SNML with different window sizes: (b) 50, (c) 100, (d) 250 and (e) 750 samples.

ball bearing faults: inner ring, outer ring and ball failures. Each failure was measured by piezoelectric accelerometer from three different axes, vertical (V), axial (A) and horizontal (H). As a result, we got 4097 data samples for each failure signal.

Signal changes corresponding to failures are not observable in time domain. Thus, spectrograms based on short-time Fourier transform were computed to visualize the frequency content of the signal over time [8]. The spectrogram of V-axis signal is shown in Figure 3. The faults are more visible and some frequency rules for faults could be tried to derive. In our method we applied five different windows: 64, 128, 256, 512 and 1024 samples, resulting 3072 data points for each failure with all SC val-

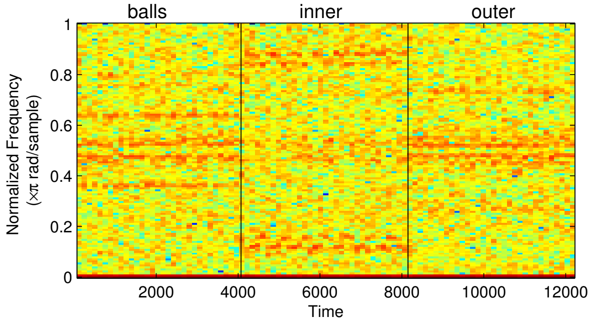

Figure 3. Spectrogram of V-axis data. The first third on the left is from the ball failure, the second from the inner ring failure and the last from the outer ring failure of the ball bearing.

ues.

In Figure 4 SOM results from V-axis data are visualized by a U-matrix and best matching units plots. The U-matrix represents the distances as colours between the map units and their neighbouring units and helps to see the emerging higher-level clusters of the data. Here we can see that the SC feature vectors form two rather clear clusters of units. The best matching unit plots are presented for three different ball bearing failures. The number of samples with known failure labels in each SOM unit is correlating with the size of the coloured circle of the unit; the more mapped samples, the bigger the coloured circle is. It can be observed that the cluster border in the U-matrix separates inner failures from the ball and outer ring ones whereas the ball and outer ring failures are slightly overlapping. This is also supported by the confusion matrix in Table 1 based on classification of the samples according to the labelled SOM units.

Measurements from two other axes, axial (A) and horizontal (H), were also analysed, see results in Table 1. A-axis measurements can separate the outer ring failures from the other failures. Thus based on V- and A-axis measurements three different ball bearing failure types can be recognized. The results of the third axis (H) were rather similar, although the failure types were slightly overlapping. The order of the failures was similar in all of the V, A and H-axes SOM results. The inner and outer failures are the most distant on the SOM lattice and the ball failure stays in the middle overlapping the inner or outer failure, or both.

4. CONCLUSION

We have proposed an MDL, and especially SNML, based method for machine condition monitoring. SNML based stochastic complexity measure combined with an AR model was shown to possess potential in signal change detection tasks as observed via our examples. The results should be of interest to all persons working with machine condition monitoring and signal change detection applications.

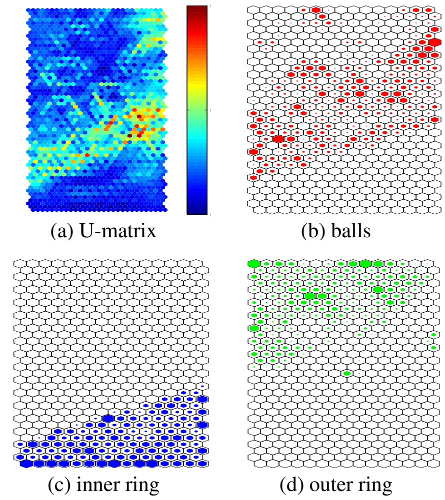

Figure 4. U-matrix and the best matching units for different failure types from V-axis results.

Table 1. Confusion matrixes in percents of V, A and H-axes results.

		Predicted		
	Actual	balls	inner	outer
V axis	balls	96.0	0	4.0
	inner	0	100	0
	outer	3.6	0	96.4
A axis	balls	87.3	12.7	0
	inner	9.3	90.7	0
	outer	0	0	100
H axis	balls	87.6	8.4	4.0
	inner	3.8	96.2	0
	outer	4.1	0	95.9

5. ACKNOWLEDGEMENTS

We thank Veli-Erkki Lumme from Neurovision Oy for proving the ball bearing data set and Dr. Jorma Rissanen for valuable comments.

6. REFERENCES

[1] B. K. N. Rao, Ed., *Handbook of condition monitoring*, Elsevier Advanced Technology, 1996.

[2] J. Rissanen, “Modeling by shortest data description,” *Automatica*, 1978.

[3] J. Rissanen and T. Roos, “Conditional NML universal models,” *Information Theory and Applications Workshop*, pp. 337–341, 2007.

[4] J. Rissanen, “Fisher information and stochastic complexity,” *IEEE Transactions on Information Theory*, vol. 42, pp. 40–47, 1 1996.

[5] T. Kohonen, *Self-organizing maps*, Springer, 2001.

[6] “VTI Technologies Oy,” referenced 18th February, 2008, URL: `http://www.vti.fi/en`.

[7] "Neurovision Oy," referenced 9th January, 2008, URL: http://www.neurovision.fi.

[8] L. Cohen, "Time-frequency distributions - a review," *Proceedings of the IEEE*, vol. 77, no. 7, pp. 941–981, 7 1989.