Optimal threshold selection in condition monitoring based on probability of false alarm

Nada Kocare[1], Đani Juričić[2], Pavle Boškoski[2]

[1]*Jožef Stefan International Postgraduate School, Jamova 39, Ljubljana*
[2]*Department of Systems and Control, Jožef Stefan Institute, Jamova 39, Ljubljana*
e-mail: nada.kocare@gmail.com
e-mail: dani.juricic@ijs.com
e-mail: pavle.boskoski@ijs.si

Abstract

Good quality of features and properly selected thresholds are essential for reliable performance of condition monitoring systems. Ideally, thresholds should be chosen so that false alarms are not triggered under normal conditions and faults are detected without delay. If the thresholds are set too high, missed alarms may occur while too low values implicate false alarms. Typical issue in the commissioning phase is that a skilled person is needed to tune the values of the thresholds, for each component of the machine. The task is nontrivial as the same feature may be related to several different faults with different sensitivities. Motivated by this issue, the aim of this paper is to derive algorithms for threshold selection on rigorous mathematical basis. The idea is to use the probability of false alarm (PFA) as a design parameter and then calculate the thresholds associated to the relative changes in condition indicators instead of tuning many thresholds separately. The performance of the algorithms is confirmed via simulations.

1 Introduction

Various maintenance strategies have been developed through the years, but one currently attracting attention is the condition based maintenance (CBM). CBM is a predictive maintenance strategy that recommends maintenance decisions based on condition monitoring (CM). In order to assess the system's health condition, several CM tasks have to be performed: (i) data acquisition, (ii) feature extraction, (iii) feature evaluation, (iv) fault isolation and (v) fault identification. The effectiveness of these tasks reflects in CM performance indicators of which sensitivity and accuracy are the most important. Ideally, one would prefer perfect accuracy, highest resolution and stable and reliable diagnosis; however in reality this cannot be fully achieved [1]. The performance significantly depends on thresholds used in diagnostic decision making. If the threshold value is set too high, missed alarms may occur, while too low values may implicate false alarms. Furthermore, typical issue in the commissioning phase is that a skilled person is needed to tune the values of the thresholds, for each component of the machine.

To illustrate this problem a drive train of a helicopter is analysed in [2]. This complex system has 25 shafts, 27 gears and 65 bearings, or total 117 monitored components. In most circumstances each component has a number of failure modes and a number of condition indicators (CIs) to detect those failure modes. For bearings, damages typically occur on their internal elements such as roller, cage, inner and outer race. Using envelope or cepstrum analysis normally 5 CIs are required for detection. The same applies in case of shafts where at least 3 types of failures can occur. However, for more complex components such as gears which have a number of failure modes we could suppose that only 6 CIs are needed. According to the example, considerable number of CIs is obtained and for each CI unique threshold has to be set, altogether $25 \cdot 3 + 27 \cdot 6 + 65 \cdot 5 = 562$ threshold values. Typically these threshold values are set statistically e.g. when single CI is 3 standard deviations above the mean value. Using the statistical rule mentioned before, the probability that any good CI (Gaussian distributed) will report false alarm is 0.0013. However, if we suppose that there are eight acquisitions per hour, or total $562 \cdot 8 = 4496$ trials, the probability of false alarm (PFA) just per acquisition would be 53%. Regrettably in reality this alarm rate is much higher. According to [2], due to Rayleigh distributed shafts and bearings CIs the false alarm rate per acquisition will raise to 83%, which is a huge percent of false alarms.

Motivated to address the above issues within a rigorous framework, this paper presents a novel thresholding approach, whose main idea is to use the PFA as only design parameter and then calculate the thresholds based on relative changes in CIs. This way a unique threshold is created, which leads to better performance and design of the condition monitoring system. At the beginning this paper overviews related references in this field. Then the design and implementation of our novel thresholding concept is explained and demonstrated via simulations in Matlab. At the end, ideas for further work are indicated.

2 Related work

The problem which is essential in diagnostic decision making is illustrated in Figure 1. Usually, the distinction between normal and abnormal values of features is performed by means of thresholds. If the numerical value of a feature does not exceed the threshold, this feature is considered to reflect normal condition of a system (or system component). If it exceeds the threshold, then the system (or component) is considered faulty. Although Boolean approach is used in many applications, still small changes around the threshold value could induce frequent change in the resulting diagnosis (Figure 1). This is called *diagnostic instability* [1].

An alternative solution is approximate reasoning, an approach in which the qualitative value of a residual is represented by a quantity between 0 and 1, indicating the degree of fault. In contrast to the previous approach, here incremental changes in residuals always result in incremental changes in the belief of fault candidates [1].

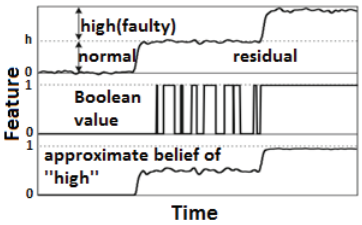

Figure 1: Boolean vs. approximate reasoning, [1]

Although many decision-making approaches have been developed, there are only few works dealing with the design of optimal threshold for maximal reliability. For example in [3], the fundamental change detection algorithms mainly based on hypothesis testing are proposed. However, in order to design a statistical decision rule that can detect changes in features, a priori formulation of the statistical properties describing the data in nominal condition must be known. Unfortunately, in reality this is not the case. Unlike the above mentioned approach, in [4] on the basis of CIs and without a priori known statistics an optimal threshold is designed. It is assumed that CIs for shaft magnitude and bearing envelope energy are distributed according to the Rayleigh distribution. These CIs for a given component are fused into one HI which is a function of distributions, where CIs are treated as random variables from Rayleigh distribution. Mathematical calculations are preformed and based of predefined PFA the critical value is calculated. However, to construct a function of distributions, the CIs must be independent and identically distributed from each other, and this cannot be performed just with decorrelation as in [4].

Motivated by [4] the aim of this paper is to revisit the problem and derive algorithms for thresholds selection on rigorous mathematical basis.

3 Optimal threshold selection

The acquired signal $y(t)$ is assumed to consist of a deterministic part corrupted with white Gaussian noise

$$y(t) = x(t) + w(t) \tag{1}$$

$$w(t) \sim N(0, \sigma^2) \tag{2}$$

where $x(t)$ is deterministic signal and $w(t)$ white Gaussian noise with zero mean and unknown variance. These types of signals are good model for explaining vibration measurements in rotating machinery. In many applications feature extraction in rotating machinery is carried out with linear signal processing techniques, like Fast Fourier Transform (FFT) and wavelet transform [5], [6]. Typically these features represent components of Fourier or wavelet spectrum.

3.1 Data acquisition setup

Data acquisition setup is explained in Figure 2. The logic behind the sampling protocol is that in the 99 % of the cases, the condition of the machines aggravates gradually. Hence, it is not needed to perform sampling all the time, unless the criticality conditions implicate continuous sampling. So we assume periodic occasional sampling within successive sampling sessions. Each session consists of N samples taken at a sampling rate $f_s = \frac{1}{T_s}$. The first N_{ref} sessions belong to the reference window and are taken while machine is in the healthy state. During the operation a sliding window with N_{cur} recent sampling sessions is used to decide whether there is a change in the features statistics. For the simplicity let us assume that after each measurement session labelled k, a feature being the m^{th} component of the Fourier spectrum is calculated as follows:

$$Z_{k,ref} = \sum_{t=0}^{N-1} y_k(t) e^{j2\pi\frac{mt}{N}} = X_{k,ref} + jY_{k,ref} \tag{3}$$

$$k = 1, ..., N_{ref}$$

Since Fourier transform (3) is linear transformation of normally distributed random signal (1), $Z_{k,ref}$ is also normally distributed but complex random variable. Furthermore, the same applies also for the current condition:

$$Z_{l,cur} = \sum_{t=0}^{N-1} y_l(t) e^{j2\pi\frac{mt}{N}} = X_{l,cur} + jY_{l,cur} \tag{4}$$

$$l = 1, ..., N_{cur}$$

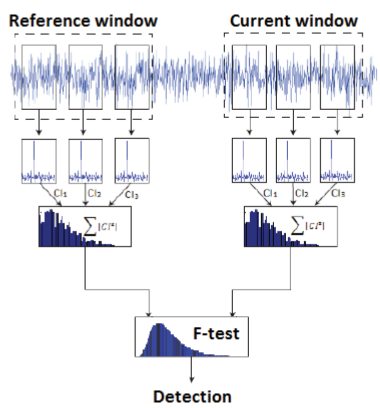

Figure 2: The computational scheme

After the set of features are formed, a batch of N_{ref} features obtained in nominal condition and a batch of

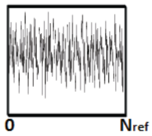

Figure 3: An example of vibrational signal from an acquisition session

N_{cur} features obtained at current condition are subjected to centering by subtracting the average from nominal condition from both batches

$$Z_{k,ref} \leftarrow (X_{k,ref} - \overline{X}_{k,ref} + j(Y_{k,ref} - \overline{Y}_{k,ref}) \quad (5)$$

where $\overline{X}$ denotes the expected value of X.

3.2 Detection

A statistical hypothesis test is derived from the previously obtained centred reference and current sets of features $Z_{k,ref}, k = 1, ..., N_{ref}$ and $Z_{l,cur}, l = 1, ..., N_{cur}$. Let us now note that the sum of squared modules of $Z_{k,ref}$ is compliant with the central chi-squared distribution with N_{ref} degrees of freedom.

$$CI_{ref} = \sum_{k=1}^{N_{ref}} Z_{k,ref} Z^*_{k,ref} = \sum_{k=1}^{N_{ref}} |Z_{k,ref}|^2 \quad (6)$$

$$CI_{ref} \sim \chi^2(N_{ref})$$

This result is obtained by assumption that in nominal condition small changes in the mean occur between different sampling sessions. Actually we suppose that the healthy features are zero mean and their chi-square statistics are obtained by mathematical calculation (6). In general, when the component becomes faulty, the mean of the features is no longer zero, meaning that the features in current condition are

$$CI_{cur} = \sum_{l=1}^{N_{cur}} Z_{l,cur} Z^*_{l,cur} = \sum_{l=1}^{N_{cur}} |Z_{l,cur}|^2 \quad (7)$$

$$CI_{cur} \sim \chi^2(N_{cur}, \lambda)$$

$$\lambda = \sum_{l=1}^{N_{cur}} (\frac{\mu_{l,cur}}{\sigma_{l,cur}})^2 \quad (8)$$

distributed according to the noncentral χ^2 distribution with N_{cur} degrees of freedom and noncentrality parameter λ. The noncentrality parameter is related to the means $\mu_{l,cur}$ and the variances $\sigma_{l,cur}^2$ of the random variables CI_{cur}. This is the case because of the non-zero mean features $Z_{l,cur}$ (5). As a result of the changes, for example in the amplitude of a signal, the statistics of the features are changing, which is an indication that a fault has occurred. From here it is easy to distinguish between different faults. Changes in the signal will reflect in large increase in the mean of the signal and by the ratio between these CIs the decision could be made. If there is no change in the system condition, then the statistical properties of $|Z_{l,cur}|^2, l = 1, ..., N_{cur}$ should be equal to the statistical properties of $|Z_{k,ref}|^2, k = 1, ..., N_{ref}$, i.e. should share the same χ^2 distribution. Let us assume $E|Z_{l,cur}|^2 = \sigma^2_{cur}$ and $E|Z_{k,ref}|^2 = \sigma^2_{ref}$ for all samples. If the system condition deteriorates, the statistical properties of the feature change, which normally affect the increase in feature variance. Since both sets of samples are independent, we can define the null hypothesis $H_0 : \sigma^2_{cur} = \sigma^2_{ref}$ versus the alternative hypothesis $H_1 : \sigma^2_{cur} > \sigma^2_{ref}$. We propose the test statistic

$$CI_F = \frac{CI_{cur}/N_{cur}}{CI_{ref}/N_{ref}} \quad (9)$$

which under H_0 complies with the central F-distribution with $N_{cur} - 1$ and $N_{ref} - 1$ degrees of freedom. So, given samples we reject H_0 if

$$CI_F \geq h = F_\alpha(N_{cur} - 1, N_{ref}) \quad (10)$$

where the term on the right side denotes the critical value of the distribution at the level of significance α. The level of significance $\alpha \cdot 100\%$ denotes the tolerated PFA. For example, in case of $N_{ref} = 400$ and $N_{cur} = 400$ and $PFA = 5\%$ from the table of low critical values for the F distribution it follows that the value of the threshold $h = 1,2290$.

3.3 Health index

In many applications more than one feature is usually extracted from available sensor systems. Hence a fingerprint of a fault can be characterized by several features with increased values. Moreover, in real applications the operators are interested to have one figure which will reflect the health states of the machine. This is called health index (HI) and takes values from 0 to 1. Since CIs reflect the level of fault in a particular component, the HI has to be a function of CIs. There are several options how to perform aggregation for example to use:

- order statistics (min/max values),
- weighted sum of features.

However, in order to make function of CIs, as stated in [7] the CIs must be independent and identically distributed. In our approach this has been achieved in the aggregation step, where every single feature is extracted from different successive acquisition sessions. In addition, we propose summation of n noncentral F distributed CIs. If the CIs are independent and identical then the function defines a unknown distribution. Due to the fact that the distribution of the obtained sum is unknown, we are proposing Monte Carlo simulation, where after several simulations the function defines a distribution with normal PDF.

4 Simulation study

Preliminary results are shown on the simulated two-component signal with additive white Gaussian noise.

$$y(t) = A \cdot (sin(2\pi \cdot 20 \cdot t) + sin(2\pi \cdot 50 \cdot t)) + w(t) \quad (11)$$

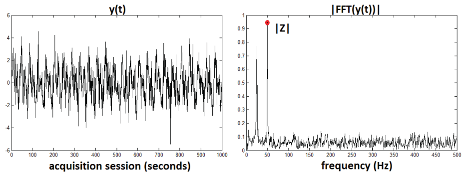

Figure 4: Signal from an acquisition session (left) and calculated feature as a component from the Fourier spectrum

Changes in amplitude A simulate increase in a component of a vibrational signal due to a particular fault (e.g. unbalance). This change influences the current feature statistics. Let us take the referent set with $N_{ref} = 400$ sampling sessions and let be sliding window in on-line operation $N_{cur} = 100$. Each session contains $N = 1000$ samples. Prior to calculating the FFT of sampled record in a measurement session the record is multiplied by Hamming window. From a batch of features collected in nominal condition and a batch corresponding to the current condition, a statistical hypothesis test is derived. Figure 5 shows the distribution of summed squares of features in nominal condition obeying central chi-squared distribution and a batch of features in current condition obeying noncentral chi-squared distribution. Furthermore, noncentral F-test is preformed where

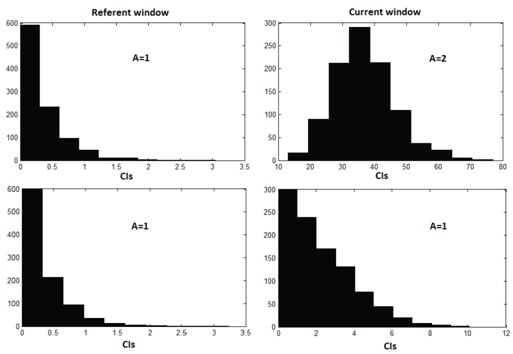

Figure 5: Histogram of the CIs in the nominal and current condition

the resulting CIs are used for obtaining the threshold values by inverse cumulative density function. In this sim-

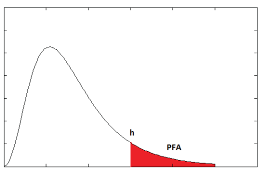

Figure 6: PDF of Noncentral F distribution and decision threshold

ulation the frequency component of interest is the one at the frequency $f = 50Hz$. For prescribed $PFA = 5\%$ the optimal threshold value is $h = 1.283$. We see that the test reliably detects change in the feature. However,

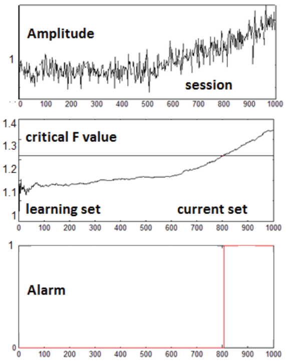

Figure 7: Performace of the detection algorithm

the alarm is triggered with a delay defined by the window length.

5 Conclusion

The paper presents a novel approach, setting the thresholds to n of condition monitoring algorithms by using only one ”tuning knob” i.e. the allowed probability of false alarm. Feature extraction based on linear signal processing results in a F-test, which triggers alarm. By using this approach, many disabled machines and economic losses due to false alarms and total shut downs due to missed alarms will be avoided. However, further improvements are needed for implementing this concept on real applications. Our idea is to use other types of features (envelope, entropy indices and etc.) and also to improve the HI procedure.

Acknowledgment: The first author is indebted to Slovene Human Resources Development and Scholarship Fund (Ad Futura) for providing the financial means for this research project.

References

[1] Rakar, A., Juričić, Đ., and Balle P., Transferable belief model in fault diagnosis, Engineering Applications of Artificial Intelegence, Vol.2(5), 555-567, October 1999.

[2] Bechhoefer, E and Bernhard , A.P.F., A generalized process for optimal threshold setting in HUMS, Aerospace Conference, 2007 IEEE, 1-9, March 2007.

[3] Basseville, M., and Nikiforov, I., Detection of Abrupt Changes – Theory and Applications. Prentice Hall. 1998.

[4] Bechhoefer, E., He, D., and Dempsey, P., Gear Health Threshold Setting Based On a Probability of False Alarm, Proceedings of the 2011 Annual Conference of the Prognostics and Health Management Society, Montreal, Canada, September 23 – 28, 2011.

[5] Boškoski, P. Condition Monitoring of Mechanical Drives: Feature Extraction and Fault Diagnosis Methods : Doctoral Dissertation, Ljubljana, December 2011, 117 pages.

[6] R. B. Randall. Vibration-based Condition Monitoring. John Wiley, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, United Kingdom, 2011.

[7] Wackerly, D., Mendenhall, W., Scheaffer, R.,(1996), Mathematical Statistics with Applications, Buxbury Press, Belmont, 1996.