An Active Learning approach for the condition monitoring of rotating machinery

Dimitrios Moshou, Xanthoula Eirini Pantazi, Dimitrios Kateris and Evangelos Tsikos, Aristotle University of Thessaloniki, School of Agriculture, Department of Hydraulics, Soil Science and Agriculture Engineering, Laboratory of Agricultural Engineering, P.0 275 54 124, Thessaloniki, Greece.
Nader Sawalhi, School of Mechanical Engineering, Prince Mohammad Bin Fahd University, Kingdom of Saudi Arabia

Abstract

Rotating machinery breakdowns are most commonly caused by failures in bearing subsystems. Consequently, condition monitoring of such subsystems could increase reliability of machines that are carrying out field operations. Recently, research has focused on the implementation of vibration signals analysis for health status diagnosis in bearings systems considering the use of acceleration measurements. Informative features sensitive to specific bearing faults and fault locations were constructed by using advanced signal processing techniques which enable the accurate discrimination of faults based on their location. In this paper, the architecture of a diagnostic system for detecting progressive faults in bearings is presented. Progressive appearance of faults in gear box bearings has been measured and different stages of faults corresponding to larger depths of cracks in inner and outer bearings have been identified. It is already known that novelty detection can be easily combined with machine learning techniques so as to detect abnormal events. For change detection, a healthy bearing state description was constructed. As a result, deviations in the vibrational behaviour were detected. Further, an active learning method in the form of one-class classifiers for the progressive detection of the advancement of faults based on vibrational features is proposed. This method learns to distinguish between different fault stages. The proposed active learning method detects different types of faults as outliers which are then augmented in final hybrid classifier which can learn by adding new types of faults. It was shown that the detection of the progressive fault stages was successful up to the level of 90% by utilizing One Class Classifiers (based on Self Organizing Maps and Support Vector Machines).

Keywords: Condition monitoring; Vibrations; Neural networks; Feature selection

Introduction

The deviation from a normal range of values of a measured or estimated sensor value originating from an on-going observed process can be spotted as a fault (Himmelblau, 1978). Following the above assumption a fault is considered as a warning or symptom of unusual behaviour. Faults usually emerge from different processes and they can be classified in three general types according to their cause, as described in (Craessaerts et al., 2010) "caused by malfunction of sensors and/or actuators, structural changes in a process, or a sudden change of model parameters". Manufacturing industry considers that ball bearings are of high importance and their task is to allow relative movement between two parts with least friction. However, demonstration of tolerance in carrying heavy loads and also efficiency and reliabil-
ity are expected. Considering the evaluation of industrial products and equipment, it is rather obvious that reliability belongs to one of the most important factors. A basic requirement for high product dependability is a reliable mechanical design. Even if the design of products is reliable, quality fades away as years are passing by, potentially due to operation under high load for extended periods of time. Therefore, maintenance is an effective way to ensure a satisfactory state of reliability. Extended operation under load is possible to cause faults in some ball bearing mechanisms. Rolling bearing faults result from many factors like incorrect design, acid corrosion, and plastic deformation, incorrect mounting and bad lubrication (Purushotham et al., 2005). Cracks in inner and outer race are the most frequent faults which constitute the 90% of all faults occurring in rolling bearings while cracks in balls or cage represent the 10% (Bently et al., 1989). Such problems are often responsible for economic damage and safety problems due to unforeseen and sudden production disruption. It is possible that minor faults in the ball bearing structure are capable of leading to a severe failure without warning (Choy, et al, 2005). The incidence of bearing faults results in severe vibrations of rotating machinery. When a rolling element passes a faulty point a transit impulse is generated. The impulse excites the structure and this response can be obtained with accelerometers. Vibration signals taken from bearings convey a great amount of information concerning machine health conditions.

Computational intelligence methods from which too main examples include artificial neural networks (ANN) (Hwang et al., 2009), and support vector machines (SVM) (Sugumaran et al., 2011,) have been effectively applied to automatic detection and identification of faults in machines. Some of them were applied to detect machine faults at an early stage, before they become severe. The learning ability of Artificial Neural Networks (ANNs) has led to their wide adoption as the most suitable for fault diagnosis in ball bearings. Miscellaneous types of neural networks and adaptive algorithms have been used in industrial fault diagnosis. Recently, several research problems that have arisen could be defined as classification problems. SVM is regarded as one of the most commonly used classification methods. According to Burgess (Burgess, 1998), SVM is applied in many cases where machine learning is needed, especially due to its high precision and good generalization based on statistical learning. Research that has been carried out recently examined the application of SVM for real time condition monitoring and its comparison with typical Neural learning algorithms (ANN) (Tyagi, 2008). It is proven that SVM is more efficient than typical ANN learning algorithms due to risk minimization. Several SVM works have been presented in fault diagnosis of rolling element bearings (Kankar et al., 2011). In order to train a common classifier, a neural network or Support vector machines, time and frequency domain features are expected to be extracted. A major burden concerns the computation related to the feature extraction procedure. Major defect frequencies are affected by the shaft speed, so determining actual features is too complicated because of side bands neighboring these frequencies. On the other hand time domain dimensionless features exhibit robustness in presence of load and speed variations (Li et al., 2000). Additionally, optimized feature selection results in increased fault detection performance (Nguyen et al., 2008). In order to facilitate the development of diagnostic and prognostic techniques for rolling element bearings different operating conditions rather than waiting for these to occur naturally, or alternatively having them seeded in the laboratory. Such fault simulation can be very valuable in machine diagnostics and prognostics in order to produce signals with well-defined characteristics. For example, the signals could be used to train neural networks to perform diagnostics and prognostics of a range of different fault types and locations in machines. These usually require so much data to train them that it would not be economical to actually experience the number of faults of each type required to accomplish the training. Some dynamic models are readily available for general use, an example of which is ADORE (Advanced Dynamics of Rolling Elements) (Nguyen et al., 2008). and BEAST (BEAring Simulation Tool). The program BEAST is developed by SKF and PELABs to perform simulations of bearing dynamics on any major bearing type. BEAST is a fully three-dimensional program used to perform simulations of bearing dynamics on any major bearing type which enables studies of internal motions and forces in a bearing under any given loading condition. As the dynamic modelling of the rolling element bearings has been extensively developed, different studies are emerging to combine the bearing model with
other rotating parts in the machine (rotors, gears, etc.) in order to study the effect of the vibration induced by bearings on other components. Sawalhi (2007) has introduced a comprehensive model for simulating bearing faults which has resulted in producing realistic results and have been validated in comparison with actual reproduced faults in laboratory conditions (Sawalhi and Randall, 2008). This work proposes detection of localized faults in rolling bearings by using Active Learning implemented on iterative application of novelty detection through One Class Classifiers utilizing two novelty detection methods based on one class classification, the one class Support Vector Machine (OCSVM) and one class Self-Organizing Map (OCSOM). The localized faults have been produced by the model presented in Sawalhi (2007) The suggested method introduces new signal processing techniques providing useful information to achieve fault recognition based on outlier detection. An original contribution includes the introduction of the new feature of line integral of the acceleration signal in combination with one class classification for novelty detection. The acceleration signals were analysed in order to obtain 12 time domain and 12 frequency domain features. The features were used in the Active Learning procedure to produce switching conditions for the different stages of the novelty detection algorithms which consisted of detection of outliers and subsequent augmentation of newly detected samples in an iterative scheme.

1 Materials and methods

1.1 Simulation model

The simulation model presented in Sawalhi (2007) was used in order to produce acceleration time series from different types of localized faults in the inner and outer ring of bearing. The simulation model displayed in Fig.1 has been programmed in Simulink (MathWorks Inc). A sampling frequency of 48 KHz has been used in order to produce an acceleration signal of duration of 1 sec with an angular speed of 2000 rpm, a load of 50 N×m. The simulated faults included localized faults in the outer ring of the bearing. The faults that were simulated had a depth of 300 μm and a width that took three different values assuming a temporal progression of fault severity from a smaller width to the larger width. In order to simulate the three different values were 0.8 mm, 1.6 mm and 3.2 mm. The fault conditions were also accompanied by the simulation of the healthy bearing where there are no faults. The assumed temporal progression of the localized fault development included the healthy condition as a starting point and it included the other three conditions of widening faults as the next stages of appearance of the fault occurring in an orderly manner.

![Simulation Model](image)

*Figure 1: The simulation model that was used in order to produce the localized faults in the inner and outer ring of the bearing.*
1.2 Signal processing and feature extraction

The first and maybe the most important step in any fault diagnosis problem, is the feature extraction from the raw signal. The aim of this is to reflect the general changes of the machine operation conditions. However, though some features are closely related to the fault, others are not. In this paper, twenty four (24) features parameters, twelve (12) time-domain (T1-T12) and twelve (12) frequency-domain (F1-F12) were selected (Lei, 2008; Moshou, et al., 2010).

1.2.1 Time-domain feature

The first eleven features were introduced by Lei et al., 2008. These were Mean value (T1), Standard deviation (T2), (T3), Root mean square (T4), Peak (T5), Skewness (T6), Kurtosis (T7), Crest factor (T8), Clearance factor (T9), Shape indicator (T10) and Impulse Indicator (T11). The twelfth one was introduced by Moshou et al., 2010 and regards the linear integral of the acceleration signal (Line integral, T12). All the used features provide statistical information about the nature of data, and were found to be reasonably good features for bearing fault detection. These features are shown in Table 1.

Table 1. Time-domain feature parameters.

\[

t_1 = \frac{\sum_{n=1}^{N} x(n)}{N} \quad (1)
\]

\[

t_2 = \sqrt{\frac{\sum_{n=1}^{N} (x(n) - t_3)^2}{N}} \quad (2)
\]

\[

t_3 = \left( \frac{\sum_{n=1}^{N} \sqrt{|x(n)|}}{N} \right)^2 \quad (3)
\]

\[

t_4 = \sqrt{\frac{\sum_{n=1}^{N} (x(n))^2}{N}} \quad (4)
\]

\[

t_5 = \max|x(n)| \quad (5)
\]

\[

t_6 = \sum_{n=1}^{N} \frac{(x(n) - t_1)^3}{(N-1)T_2^3} \quad (6)
\]

\[
t_7 = \int_{a}^{b} ds \approx \sum_{i=1}^{N} \frac{(t_i + T_s) - \bar{f}(t_i)}{T_s} = \sum_{i=1}^{N} \sqrt{(x(t_i + T_s) - x(t_i))^2 + T_s^2} \approx \sum_{i=1}^{N} |x(t_i + T_s) - x(t_i)| \quad (12)
\]

where x(n) for the time-domain feature is a signal series for n=1,2,.....,N, N is the number of data points. Especially for the line integral N is the number of sample points (equal to 500) in the non-overlapping windows used to calculate Kurtosis and the other features and the newly proposed line integral feature and Ts is the sampling period. Given the high sampling rate of 48 kHz and the domination of the signal from high frequencies (especially due to the presence of faults), the final approximation contains only acceleration values.

1.2.2 Frequency-domain feature

Frequency-domain is another description of a signal. This type of description includes some information that cannot be found in time-domain. In this study another twelve features (Lei et al., 2008) were used in order to feed the One Class Classifier with additional information with
respect to the time domain features. These twelve features were based on the Fourier transform of the vibration signals. Feature F1 may indicate the vibration energy in the frequency-domain. Features F2-F4, F6 and F10-F12 may describe the convergence of the spectrum power. Finally F5 and F7-F9 give information about the position change of main frequencies.

<table>
<thead>
<tr>
<th>Table 2. Frequency-domain feature parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>$F_1 = \frac{\sum_{k=1}^{K} s(k)}{K}$ (13)</td>
</tr>
<tr>
<td>$F_2 = \frac{\sum_{k=1}^{K} (s(k) - F_1)^2}{K-1}$ (14)</td>
</tr>
<tr>
<td>$F_3 = \frac{\sum_{k=1}^{K} (s(k) - F_1)^3}{K(\sqrt{F_2})^2}$ (15)</td>
</tr>
<tr>
<td>$F_4 = \frac{\sum_{k=1}^{K} (s(k) - F_1)^4}{KF_2^2}$ (16)</td>
</tr>
<tr>
<td>$F_5 = \frac{\sum_{k=1}^{K} f_k s(k)}{\sum_{k=1}^{K} s(k)}$ (17)</td>
</tr>
<tr>
<td>$F_6 = \frac{\sum_{k=1}^{K} (f_k - F_5)^2 s(k)}{K}$ (18)</td>
</tr>
<tr>
<td>$F_7 = \sqrt{\frac{\sum_{k=1}^{K} f_k^2 s(k)}{\sum_{k=1}^{K} s(k)}}$ (19)</td>
</tr>
<tr>
<td>$F_8 = \sqrt{\frac{\sum_{k=1}^{K} f_k^4 s(k)}{\sum_{k=1}^{K} f_k^2 s(k)}}$ (20)</td>
</tr>
<tr>
<td>$F_9 = \frac{\sum_{k=1}^{K} f_k^2 s(k)}{\sum_{k=1}^{K} \sum_{k=1}^{K} f_k^4 s(k)}$ (21)</td>
</tr>
<tr>
<td>$F_{10} = \frac{F_6}{F_5}$ (22)</td>
</tr>
<tr>
<td>$F_{11} = \frac{\sum_{k=1}^{K} (f_k - F_3)^3 s(k)}{KF_6^2}$ (23)</td>
</tr>
<tr>
<td>$F_{12} = \frac{\sum_{k=1}^{K} (f_k - F_5)^4 s(k)}{KF_6^4}$ (24)</td>
</tr>
</tbody>
</table>

where $s(k)$ is a spectrum for $k=1,2,\ldots,K$, $K$ is the number of spectrum lines and $f_k$ is the frequency value of the $k$th spectrum line.

### 1.3 One Class Classifiers

One-class classification can be characterized as follows:

- Only data belonging to the target class (not outlier class) is available for calibration;
- The limit demarcating the border between the two classes has to be calculated from data stemming from the target class only; and
- The main task concerns the definition of a boundary surrounding the target class (to classify as many as possible of the target examples correctly, while simultaneously minimizing the prospect of accepting outlier examples).

The one-class SVM (OCSVM) constructs a model from performing calibration using normal data. At the second stage, classifies test data based on the deviation from normal training data as either normal or outlier (Scholkopf et al., 2001). The influence of the RBF’s spreading parameter in $K(x,z) = \exp\left(\frac{||x-z||^2}{\sigma^2}\right)$ can be estimated considering that a large spread indicates a linear class of target data while on the other hand, numerous support vectors combined with a small spread indicate a highly nonlinear case.

Similarly to the OCSVM, a one class SOM (OCSOM) is calibrated using normal operation data. Subsequently, the feature vector that corresponds to a new measurement is examined in order to assess its similarity to the weight vectors of every other map unit. If the smallest distance exceeds a predetermined threshold, it is assumed that the process belongs to a fault situation. This result emanates from the assumption that quantization errors exceeding a certain value are associated with the operation points that are external to the region that has been covered by the training data. Hence the situation is novel and raising the possibility of abnormality detection. Depending on the magnitude of deviation from the normal operation state, a degradation index can be calculated. The one-class SOM (OCSOM) constructs a
model from healthy bearing data and subsequently classifies new data according to its deviation from the healthy training data. During novelty recognition, novel examples from bearings of not definable health state are used to formulate the input to the network while the SOM algorithm selects the best matching unit. The assumption used here is that if the quantisation error that results from the comparison between the new exemplar data \((x_j^{\text{NEW}})\) and best matching unit (bmu) is larger than a pre-specified threshold \((d)\) then the example is considered as novel. Eq. 25 represents the minimum distance for the bmu and examines it against the threshold.

\[
\min_{j \in M} \sum_{j=1}^{M} (x_j^{\text{NEW}} - m_j)^2, \quad i \in M \tag{25}
\]

Where \(M\) denotes the SOM grid of neurons similar to equation (25).

There are various heuristics to determine a threshold based on the usefulness of the threshold and the specific structure of the data set. A simple way to define a threshold \((d)\) depends on the similarity between the SOM centroid vectors and target training vectors that have selected them as best matching units which determines the quantization error. These distances have to be estimated according to Eq. 26:

\[
\text{distances} = \min_{i \in M} \sum_{i=1}^{M} (x_k^{\text{TARGET}} - m_i)^2, \quad i \in M \tag{26}
\]

The threshold is estimated by utilizing the Matlab code which is presented below:

1. `Data_distances_sorted=sort(distances);`
2. `Fraction =round(fraction_targets*length(target_set));`
3. `Threshold=(Data_distances_sorted(fraction)+ Data_distances_sorted(fraction+1))/2;`

### 1.4 Active Learning Procedure based on One Class Classifiers

An active learning method comprising of novelty detection and incremental class augmentation based on OCSOM and OCSVM, was implemented. A baseline set of vibration signatures consisting of 24 features (time and frequency) from healthy bearings were used as initial set. Then at each step, vibration signatures from faulty bearings were presented to the one-class classifier and after outlier detection, the fault type class was incrementally added in a multi-class classifier based on one-class binary classifiers. This step was repeated until all fault type classes were augmented. In this way, the active learning scheme can achieve continuous incremental learning upon encountering new fault bearing types.

The detailed active learning scheme comprised the following steps:

1. The initial training set comprised of feature vectors which were consisting of 24 features extracted from the vibration signals.
2. Each of the different types of one class classifiers were trained with the training vectors from the target set (healthy bearing).
3. The trained one-class classifiers were tested with feature vectors from one type of fault that normally occur an initial stage of fault development which corresponds to a width of 0.8 mm. The criterion of success is the ability to classify the unknown fault as outlier in comparison with the healthy bearing condition which is considered as the baseline set.
4. The initial target set is augmented with outlier values from the just detected outlier fault condition and the resulting set is considered as the new baseline set. The procedure of outlier detection is repeated but the criterion of success is the ability to classify new types of faults as outlier a comparison with the newly augmented baseline set that includes the healthy condition and the incorporated faults.
5. In the case that the new sample belongs to a class that is already included the target set the detection of outliers is executed per class internally in the baseline set and the sample is classified in one of the existent subclasses of the baseline set.
6. Steps 4 and 5 are repeated and more specifically, the detection of the outliers and the augmentation are executed for unknown data that could belong to the already existent fault condition categories or could concern new unclassified categories of fault types. In the current work the repetition of the steps terminated after the creation of 4 different classes that correspond to the healthy condition and the three fault types that corresponded to widening faults. It must be noted that the overall procedure from steps 1 to 6 does not require external
intervention but relies solely on outlier detection and augmentation steps that are performed automatically.

2 Results and Discussion

The Active Learning procedure described in previous section was implemented by using OCSOM and OCSVM. The number of samples was 92 per class which meant that a total of 368 samples were available for implementing the Active Learning procedure. The OCSOM was tested with different sizes between 5×5 and 25×25 units with a step of 5 per dimension. The arrangement of units was rectangular. The training was implanted in a batch mode. The threshold was set at 10% which meant that the tolerance to outliers was set at 0.1 in the novelty detection algorithm. The results for OCSOM are shown in the Table 3. The SVM had a tolerance to outliers equal to 0.1 and a spread of 1. Other spreads were tested but the results were inferior compared to a spread of 1.

Table 3 The results of Active Learning using OCSOM augmentation

<table>
<thead>
<tr>
<th>Real</th>
<th>Estimated by augmented OCSOM (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Healthy</td>
</tr>
<tr>
<td>Healthy</td>
<td>90.22</td>
</tr>
<tr>
<td>Outer race fault 0.8mm width</td>
<td>0</td>
</tr>
<tr>
<td>Outer race fault 1.6mm width</td>
<td>0</td>
</tr>
<tr>
<td>Outer race fault 3.2mm width</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 4 The results of Active Learning using OCSVM augmentation

<table>
<thead>
<tr>
<th>Real</th>
<th>Estimated by augmented OCSVM (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Healthy</td>
</tr>
<tr>
<td>Healthy</td>
<td>83.70</td>
</tr>
<tr>
<td>Outer race fault 0.8mm width</td>
<td>3.26</td>
</tr>
<tr>
<td>Outer race fault 1.6mm width</td>
<td>0</td>
</tr>
<tr>
<td>Outer race fault 3.2mm width</td>
<td>1.09</td>
</tr>
</tbody>
</table>

It is observed that the OCSOM has given better results than OCSVM method, reaching 90.22% of successful identification for all tested bearing conditions (healthy and 3 types of widening faults), while OCSVM was able of reaching only up to 75% in the worst case. This can be explained due to the large number of OCSOM units and the local nonlinear behavior of data which made difficult for OCSVM to find the proper spread that can result in a correct class classification.

3 Conclusions

In this paper, the architecture of a diagnostic system for detecting progressive faults in bearings is presented. Progressive appearance of faults in gear box bearings has been measured and different stages of faults corresponding to larger depths of cracks in inner and outer bearings have been identified. The OCSOM has given better results than OCSVM method, reaching 90.22% of successful identification for all tested bearing conditions (healthy and 3 types of widening faults), while OCSVM was able of reaching only up to 75% in the worst case. Other types like mixture of Gaussians will be tested in future implementations because SOM problems tend to be very promising.
4 References


