EARLY DETECTION OF BEARING DAMAGE BY MEANS OF DECISION TREES

Bovic Kilundu, Pierre Dehombreux, Christophe Letot, Xavier Chiementin

Abstract:

This paper presents a procedure for early detection of rolling bearing damages on the basis of vibration measurements. First, an envelope analysis is performed on band-pass filtered signals. For each frequency range, a feature indicator is defined as sum of spectral lines. These features are passed through a principal component model to generate a single variable, which allows tracking change in the bearing health. Thresholds and rules for early detection are learned thanks to decision trees.
Experimental results demonstrate that this procedure enables early detection of bearing defects.

Keywords: *damage detection, bearing damage, envelope detection, decision trees, preventive maintenance.*

1. Introduction

Rolling bearing degradation can be very detrimental in certain situations. Its progressive character raises the question of the determination of the right moment to perform their replacement, at the cost of stopping machines. If it is possible to detect incipient bearing damage and to identify all their evolution stages, one can estimate the reliability curve of the bearing and its remaining life and thus optimize maintenance schedule.

The use of vibrations for rolling bearing monitoring is explainable by the degradation process. Indeed, bearing degradation generally results in a subsurface or a surface fatigue of one of the races. Thus fatigue crack can occur and propagate until a large pit or spall occurs in the surface [1], [2]. This will generate repetitive impacts during the rotation of rolling elements (ball or roller) over the race. These shocks excite defective frequencies, which depend on the number of rolling elements, the rotational speed and the geometry of the bearing. These frequencies are given by the following expressions:

- Outer race defect frequency:

$$f(Hz) = \frac{n}{2} f_r \left(1 - \frac{BD}{PD} cos\beta\right) \quad (1)$$

- Inner race defect frequency:

$$f(Hz) = \frac{n}{2} f_r \left(1 + \frac{BD}{PD} cos\beta\right) \quad (2)$$

- Ball (roller) defect frequency:

$$f(Hz) = \frac{BD}{PD} f_r \left[1 - \left(\frac{BD}{PD} cos\beta\right)^2\right] \quad (3)$$

Where n is the number of balls or rollers, f_r is the relative rotational frequency between inner and outer races, BD is the ball (roller) diameter, PD is the pitch diameter and β the contact angle.

Consequently, vibration analysis is well indicated for monitoring and diagnosis tasks of bearing. Tandon and Choudhury [3] have presented an exhaustive review of methods for bearing monitoring and diagnostic of rolling bearings on the basis of vibration analysis.

For incipient damage, bearing defective frequencies are usually buried under noise and other frequency components in spectral representation. Denoising methods can be applied to improve damage detection.

According to industrial standards, the fatal size of spall is fixed at 6.25 mm² (0.01 in²) [4]. When a defect of the fatal size is detected, emergency stop, which likely involves expensive disturbances, must occur. Therefore, it is important to detect defects at their early phase.

In this work, we apply a procedure, which uses envelope analysis of band-pass filtered signals and decision trees to automate the detection of incipient defects. Features extracted from signals are processed by principal component analysis to define a residue, which accurately reveals the alteration of the bearing health. In section 2, envelope analysis is introduced. Section 3 presents the principles of principal component analysis, as well as its application in fault detection. In section 4, we present decision trees and their use in fault diagnostic. Experimental validation of these concepts and obtained results are discussed along the section 5.

2. Envelope analysis

Vibration signals raised on degraded bearings contain repetitive shocks, which excite high frequency resonances. A direct frequency analysis does not always give access to interesting information when the energy content of the signal, in consequence of these resonances, is located in these high frequencies. However, these repetition frequencies can be easily highlighted in the envelope signal.

Classically, the signal is first band pass-filtered around the frequency range where a significant broadband increase has been detected [5]. From the filtered signal, which must contains only the repetitive impulses; one performs envelope detection or amplitude demodulation, which gives the outline of the signal. Usual methods process by squaring and low pass filtering or by Hilbert transforms.

2.1. Squaring and low pass-filtering

This method proceeds by squaring the signal before low pass-filtering it. Squaring the signal effectively demodulates the input by using itself as the carrier wave. If necessary, one can correct the scale by using a gain of 2 on the signal. Since only the lower half of the signal energy is to be kept, this gain boosts the final energy to match its original energy. The square root of the signal is finally taken to reverse the scaling distortion from squaring the signal.

This method is easy to use, but one has to make judicious choice of cutoff frequency of the low pass-filter.

2.2. Hilbert transform

This approach creates the analytic signal from the input signal. The analytic signal is a complex signal. Its real part is the original real signal and its imaginary part is composed by the Hilbert Transform of the signal.

$$y(t) = x(t) + jH(x(t)) \tag{4}$$

The envelope of the original signal is obtained as the magnitude of the analytic signal.

2.3. The use of a root mean square detector

In a similar way to the squaring and low pass-filtering method, we have proposed the use of a root mean square detector to obtain envelope signal. This procedure is very simple to implement and to apply and proceeds as follows:

Let x_n be the vibration signal. One defines a sequence of vectors using a sliding window of judicious length and form. This window is applied with an overlap of 50%. Then, the root mean square value calculated on each of these vectors is assigned to the time position corresponding to its beginning. The time series made up of these root mean square values will represent on a certain scale the envelope (Figure 1).

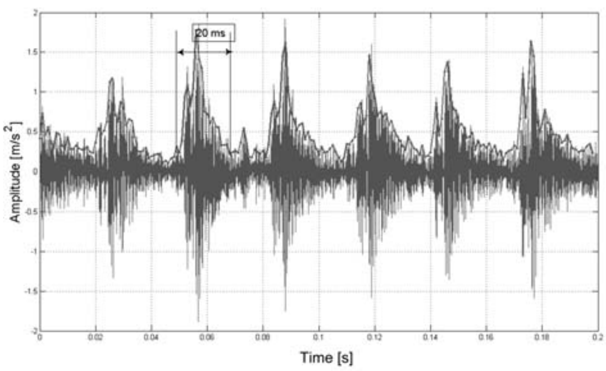

Fig. 1. Envelope detection by root mean square detector.

Envelope analysis is the FFT frequency spectrum of the modulating signal (the envelope of the original signal). In this work, this step is performed after filtering the envelope signal around the bearing frequency in order to emphasize the effect of bearing damage over other spectral lines. Thus low pass filtering is carried out in two steps: application of the sliding window then filtering around the bearing frequency. Figure 2 illustrates this process of envelope detection.

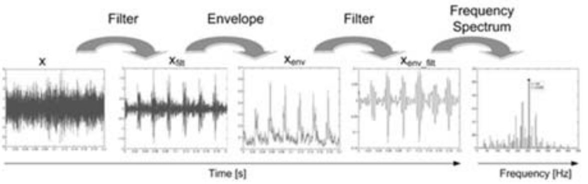

Fig. 2. Envelope analysis scheme.

3. Principal component analysis

Principal component analysis (PCA) is a way of transforming a set of data by finding, in the feature space, an orthogonal base, which dimension is determined by principal directions. PCA allows first to go from a set of vectors $x_1, x_2, \dots, x_M$ stored in matrix $X_{M \times N}$ to $Y_{M \times N}$, a set of vectors $y_1, y_2, \dots, y_M$. Components of vectors x_i are the original variables and those of y_i the factors or factor scores. The new variables will avoid any redundancy in the loaded information. Then, one will only retain in Y components that correspond to an informational criteria [6].

PCA goes about this transformation linearly. Factors are built as linear combination of variables. In this linear context, the non-redundancy condition of factors is simply the condition of non-correlation of factors. Spectral decomposition is thus applied to Σ the covariance matrix of X. The main idea behind PCA is that high information corresponds to high variance. To transform matrix into X into $Y=X.A$, $A_{N \times N}$ must be chosen in such manner that Y has largest variances. A will thus be the orthogonal matrix used in spectral decomposition of Σ. Columns in matrix A eigenvectors of Σ. Directions of largest variance are parallel to eigenvectors

As in practice Σ is not known, one uses the sample variance (covariance) matrix S, which is defined as:

$$S = \frac{1}{M-1} X^T X \tag{5}$$

Vectors y_i are such that their components are not correlated and they are characterized by the fact that high information is stored in a few components. Thus, only a reduced number of components can be considered to describe the data set. One will keep components with high variance, i.e. corresponding to largest eigenvalues. This means that principal components that contribute less than a given fraction (threshold) to the total variation in the data set are eliminated. This criterion can be written:

$$\frac{\lambda_\alpha}{\Sigma_{\alpha=1}^{N} \lambda_\alpha} \geq threshold \tag{6}$$

where λ_α represent eigenvalues of Σ.

After this, let assume that one has kept the P first factors; the transform matrix will be $E_{N \times P}$ instead of $A_{N \times N}$. These yields

$$Y'_{M \times P} = X_{M \times N} \bullet E_{N \times P} \tag{7}$$

3.1. Damage detection with PCA

PCA as a means of fault detection has already been intensively studied. But its application has mainly concerned the field of chemical process monitoring, where the number of sensors is generally significant [7], [8].

For detection of mechanical damage on the basis of vibration measurement, this technique has been rarely used probably because the number of variables to be supervised is not generally large. However, it can be very interesting to make use of this technique for detection of mechanical defects in combination with machine learning methods. The basic concept of the use of PCA for detection is summarized hereafter.

In multisensor context, or if several features are extracted from vibration signals, it is interesting to make use of PCA for damage detection. Let X be a data matrix representing normal (healthy) operation conditions. We can transform X by PCA to get Y. Retaining significant components in Y, we obtain Y'. Back transform of Y' to original variables gives

$$X^*_{M\times N} = Y'_{M\times P} \bullet E^T_{P\times N} \quad (8)$$

Since it was retained only significant components in the constitution of the transform matrix E, data in X^* are obtained with only significant variances, i.e. insignificant noise effects have been removed (9). The difference X-X^* between the two matrices will be insignificant.

Let now suppose that a set of new operation conditions is given in a data matrix X_1. One transforms X_1 by application of the transform matrix built on healthy data, then its back transform in the space of original variables will give a data matrix X_1*. The residual matrix is computed as

$$R = X_1 - X_1^* = X_1\,[I - EE^T] \quad (9)$$

This matrix indicates the deviation from the healthy state. For a vector x corresponding to a residual vector r the deviation is given by

$$R(x) = r_{1\times N} \bullet r^T_{N\times 1} \quad (10)$$

This number indicates how much an operation condition is far from the healthy one, and constitutes an ideal feature for detection. The detection process is illustrated in the Figure 3.

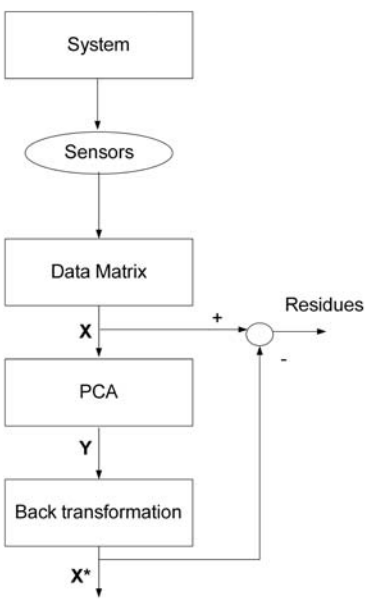

Fig. 3. Damage detection with PCA.

4. Decision trees

A decision tree is a hierarchical representation used to determine the classification of an object (observation) by testing the values of some of its attributes (variables). In a decision tree, final nodes are decision or classification nodes, they are called sheets. Intermediate nodes are nodes of test on the properties of the objects. The construction process of decision trees is recursive.

In fault detection and identification [9], [10], [11], two questions arise for the tree structure building: which attribute to choose, and which value of that attribute will constitute the decision threshold for segmentation at a test node? The principle is to select, at each node, the variable, which presents the greatest information gain, called purity. The concept of purity just induces the fact that sets should contain only data of most similar type, ideally of only one single class membership [12]. Instead of purity, one calculates a measure of impurity given by the Shannon statistical entropy:

$$-\sum_i P_i \bullet log_2 P_i \quad (11)$$

where P_i s the proportion of data concerned by an attribute or a particular value of that attribute. The attribute, which presents the best gain, i.e. minimal entropy, will be selected as the root of the tree or as a test node. The process will be thus carried out in a hierarchical way until the final nodes are reached, i.e. nodes, which contains objects belonging to the same class. The most used algorithm to build decision tree is the C4.5 algorithm [13].

A decision tree can be used to learn the structure of monitoring data, and thus establish rules and thresholds to detect bearing damage at an early stage. The only requirement is the extraction of sensitive feature by means of adequate signal processing.

5. Experimental validation

5.1. Experimental setup

In order to apply this detection procedure, several sizes of faults were induced on the inner race of a FAG NU206 roller bearing. The test rig consists of a shaft supported by two roller bearings housed in a carter (Figure 4). During tests, the shaft is driven at three different speeds: 2000 rpm, 1500 rpm and 1000 rpm. Three different radial loads are applied to the shaft and bearing with the help of a hydraulic jack. The test bearing is not lubricated. Three accelerometers are used to measure vibration in horizontal, vertical and axial directions. Vibration data are collected with a sampling rate of 50 kHz. Table 1 gives the size of induced faults as well as the number of signals collected at each measurement point. In all, 353 signals are collected.

Table 1. Size of induced faults.

	Size (mm^2)	Level	Number of recorded signals
Healthy	0	0	8
Very slight fault	0.196	1	8
Slight fault	0.250	2	176
Advanced fault	1.50	3	71
Severe fault	3	4	90

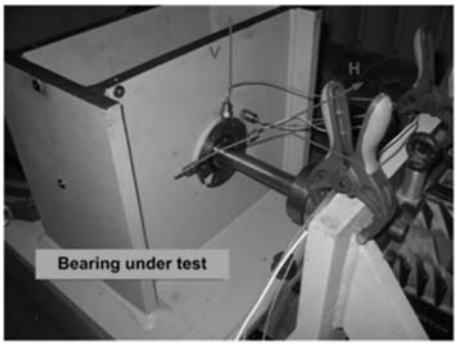

Fig. 4. The test rig.

5.2. Data preparation and feature extraction

Signals are band-pass filtered in 5 frequency ranges with a Butterworth filter: ≤1000, 1000-3000, 3000-5000, 5000-10000 and ≥10000 Hz. An envelope signal is extracted from each filtered signal by a root mean square detector with exponential sliding windows. Windows are used with an overlap of 50%. Window's length is chosen to be of 100 points, which allows an under-sampling of 100 and thus a frequency range, which goes up to 500 Hz.

The envelope signal is then filtered around the BPFI before a frequency spectrum is computed. To characterize a filtered signal, sums of spectral lines are considered. These sums were normalized by the load and the square of speed, and then fused by concatenation in a 15-dimension vector (3 directions x 5 frequency ranges), which represents the operation condition.

A PCA transform matrix was constructed from vectors representing healthy bearings, i.e. the 16 first one (level 0 and level 1). The transformation and back transformation of the other vectors gave a residue whose evolution is represented in the Figure 5. Significant components are retained on the basis of a scree test (Figure 6). This residue, denoted RFILT, will constitute the feature for the need of detection.

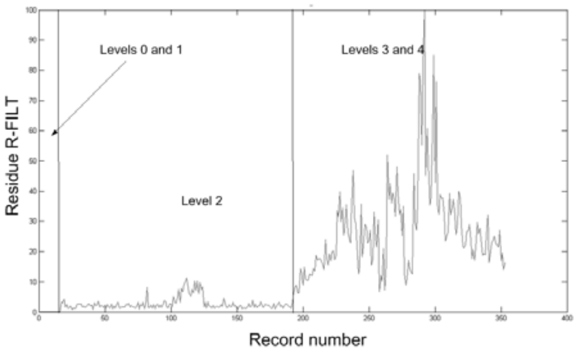

Fig. 5. Evolution of the residue with bearing damage.

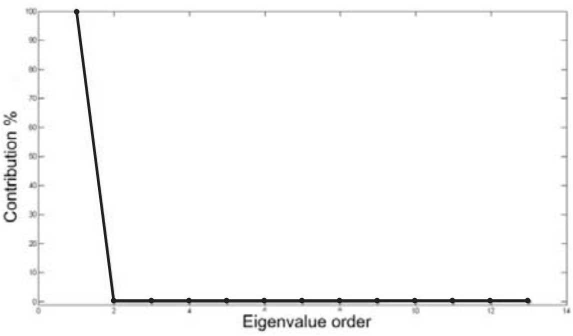

Fig. 6. Choice of significant components by scree test.

5.3. Results and discussion

Assuming that the evolution of bearing damage is continuous, early detection will be related to very low size of defect. As at level 1 the damage is very small, early detection will concern the transition from level 1 to level 2. With the aim of automating the task of detection, we propose the use of decision tree. The advantage in this choice is that the structure learned by a decision tree can easily be traduced in rules, and from rules one can define decision thresholds.

Figure 7 shows the decision tree learned from data. One can observe that the thresholds for the transition from level 1 to level 2 is located at RFILT=0.055102. This decision tree has allowed early detection with an error rate of 0.2%. The learning evaluation is made by cross validation. The confusion matrix (Figure 8) shows that levels 0 and 1 are well separated from the other levels excepted one object of level 1 which is recognized as non defect-free one.

This demonstrates that the methodology proposed in this work allows early detection of bearing degradation. Another fact observed from the confusion matrix is that level 0 and level 1 are not very different since the decision tree failed to separate them accurately.

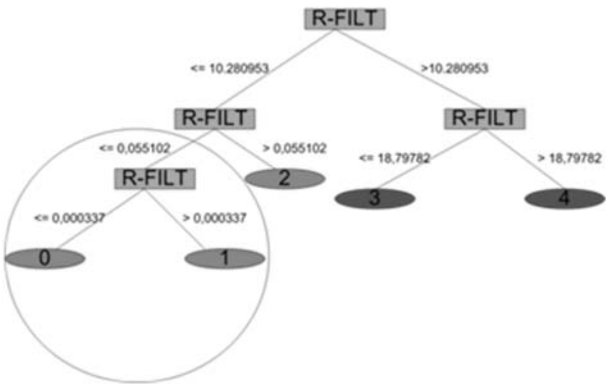

Fig. 7. Decision tree.

↱	Level 0	Level 1	Levels 2,3,4
Level 0	7	1	0
Level 1	2	5	1
Levels 2,3,4	0	0	337

Fig. 8. Confusion matrix for early fault detection with decision tree.

For each new operation condition, data will be processed as represented in Figure 9.

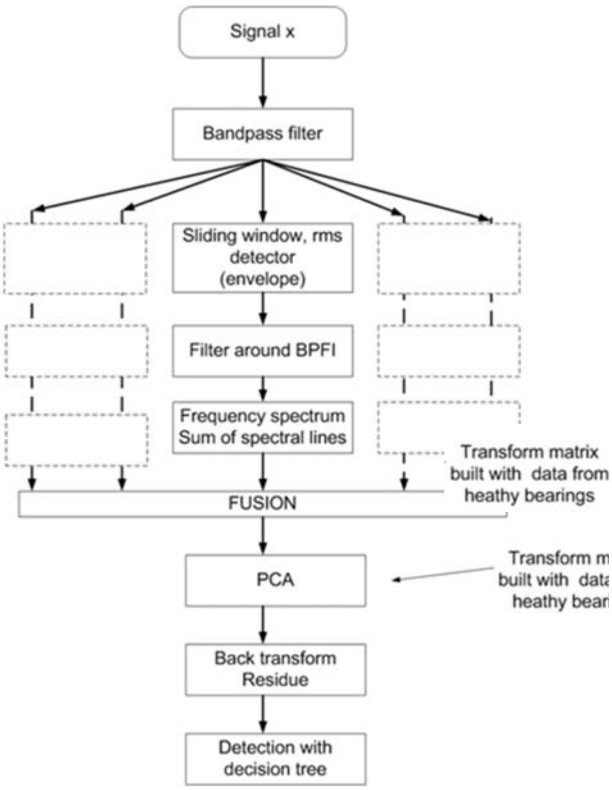

Fig. 9. Data processing.

6. Conclusions

This paper addressed the important issue of early detection of bearing fault, which can allow an optimal organization of maintenance interventions. Principal component analysis was used to construct a single detection feature from which a decision tree learned rules and thresholds for an early detection. Envelope analysis was used to emphasize the effect of bearing damage over other spectral lines in the frequency domain. The results obtained in this study show that it can be possible to detect incipient defects of bearing by the using the decision trees with the proviso that a suitable signal processing have been is carried out.

AUTHORS

Bovic Kilundu*, Pierre Dehombreux, Christophe Letot - Pôle Risques, Faculté Polytechnique de Mons, rue du Joncquois, 53 B-7000 Mons, Belgium.
E-mail: bovic.kilundu@fpms.ac.be.
Xavier Chiementin - Université de Reims Champagne-Ardenne, BP 1039, 51687 Reims, Cedex 2, France.
* Corresponding author

References

[1] Gebraeel N., Lawley M., Liu C.R., Parmeshwaran V., "Residual life predictions from vibration-based degradation signals: A neural network approach", *IEEE Transactions on Industrial Electronics*, vol. 51, issue 3, 2004, pp. 694-700.
[2] Harris T., Kotzalas M., *Advanced Concepts of Bearing technology*. s.l.:CRC Press, Taylor & Francis Group, 2007.
[3] Tandon N., Choudhury A., "A Review of Vibration and Acoustic Measurement Methods for the Detection of Defects in Rolling Element Bearing", *Tribology International*, vol. 32, no. 8, 1999, pp. 469-480.
[4] Li Y., Zhang C., Kurfess T. Danyluk S., Liang S., "Adaptative Prognostics for Rolling Element Bearing Condition", *Mechanical Systems and Signal Processing,* vol. 13, no. 1, 1999, pp. 103-113.
[5] Courrech J., Eshleman R., "Condition monitoring of machinery". In: *Harri's Shock and Vibration Handbook*. s.l.: McGrawHill, 2002, pp. 16.1-16.25.
[6] Hand D., Manilla H.,Smyth P., *Principles of Data Mining*. s.l., MIT Press, 2001.
[7] McGregor J., Kourti T., Nomikos P., "Analysis, monitoring and fault diagnosis of industrial processes usind multivariate statistical projection methods". In: *Proceedings of IFAC Congress. San Francisco 1996*. vol. M, 1996, pp. 145-150.
[8] Mohamed-Faouzi H., *Détection et localisation de défaut par analyse en composantes principales.* Institut National Polytechnique de Lorraine. 2003. Ph.D thesis (in French).
[9] Iserman R., *Fault-diagnosis System. An introduction from fault detection to fault tolerance.* s.l.: Springer, 2006.
[10] Sugumaran V., Ramachandran K., "Automatic rule learning using decision tree for fuzzy classifier in fault diagnosis of roller bearing", *Mechanical Systems and Signal Processing,* vol. 21, no. 5, 2007, pp. 2237-2247.
[11] Chen Y., "Impending failure detection for a discrete process", *Mechanical Systems and Signal Processin*,. vol. 7, no. 2, 1993, pp. 121-132.
[12] Chen Z., *Computational Intelligence for Decision Support,* New York: CRC Press LLC, 2000.
[13] Quilan R., *C4.5: Programs for Machine Learning*, San Mateo: Morgan Kaufmann Publishers, 1993.