Robust performance degradation assessment methods for enhanced rolling element bearings prognostics

Hai Qiu\textsuperscript{a,\ast}, Jay Lee\textsuperscript{b,1}, Jing Lin\textsuperscript{c}, Gang Yu\textsuperscript{d}

\textsuperscript{a}NSF I/UCRC on Intelligent Maintenance Systems (IMS), Department of Industrial and Manufacturing Engineering, University of Wisconsin-Milwaukee, Milwaukee WI 53201, USA
\textsuperscript{b}NSF I/UCRC on Intelligent Maintenance Systems (IMS), University of Wisconsin-Milwaukee, Milwaukee WI 53201, USA
\textsuperscript{c}National Laboratory of Acoustics, Institute of Acoustics, Chinese Academy of Sciences, P.O. Box 2712, Beijing 100080, People’s Republic of China
\textsuperscript{d}Department of Mechanical and Industrial Engineering, Northeastern University, 360 Huntington Avenue 334 SN, Boston, MA 02115, USA

Abstract

Bearing failure is one of the foremost causes of breakdowns in rotating machinery and such failure can be catastrophic, resulting in costly downtime. One of the key issues in bearing prognostics is to detect the defect at its incipient stage and alert the operator before it develops into a catastrophic failure. Signal de-noising and extraction of the weak signature are crucial to bearing prognostics since the inherent deficiency of the measuring mechanism often introduces a great amount of noise to the signal. As a result, robust methods are needed to provide more evident information for bearing performance assessment and prognostics.

This paper introduces enhanced and robust prognostic methods for rolling element bearing including a wavelet filter based method for weak signature enhancement for fault identification and Self Organizing Map (SOM) based method for performance degradation assessment. The experimental results demonstrate that the bearing defects can be detected at an early stage of development when both optimal wavelet filter and SOM method are used.

\textcopyright 2004 Published by Elsevier Ltd.

Keywords: Prognostics; Predictive maintenance; Bearing prognostics

1. Introduction

Rolling element bearing condition monitoring and diagnostics has received considerable attention for many years because bearings are critical to almost all forms of rotating machinery and are among the most common machine elements. Bearing failure is one of the foremost causes of breakdowns in rotating machinery and such failure can be catastrophic, resulting in costly downtime. To prevent unexpected bearing failure, vibration analysis has been used extensively for various bearing condition monitoring techniques [1].

Many previously studies [2–4] have developed well-established theoretical foundation and tools to comprehensively describe bearing failure modes. Typically, when a defect in one surface of a bearing strikes another surface, an impulse is generated that excites bearing system resonance. The rotation of either inner race or outer race produce a series of successive impulses governed by the operating speed and geometry of the bearing, and decay in the vibration transmission path due to the system damping factor. Therefore, the signature of a damaged bearing consists of exponentially decaying ringing, which in the case of constant rotating speed occurs periodically at the characteristic frequency [3]. The vibration signal of a defective bearing usually considers being amplitude modulated at characteristic defect frequency. Matching the measured vibration spectrum with the defect characteristic frequency enables us to detect the presence of a defect and determine
where the defect is. A simplified diagnostic scheme can be described as in Fig. 1.

Driving by the desire of improved machine uptime and near-zero breakdown productivity, more and more attention has been put onto predictive maintenance, which necessitates advanced tools in prognostics. Prognostics is using predictive maintenance practices and tools to analyze the trends of machine performance against known engineering limits for the purpose of detecting, analyzing and correcting problem before failure occurs. More advanced prognostics methods are focused on performance degradation monitoring and assessment so failure can be predicted and prevented [4]. To fulfill the goal of prognostics, three crucial steps are needed. At first, the defect or abnormality should be able to detect at its early stage. Secondly, the machine or system performance needs to be assessed robustly and tracked continuously. Finally, a prediction with confidence interval needs to be generated estimating the remaining useful life and possible failure mode of the machine or system. A new prognostic scheme can be depicted as shown in Fig. 2.

As for the vibration signal of rolling element bearing, modulation [6] and noise are two major barriers in causing difficulties in incipient defect detection. To overcome the modulation issue, a large variety of signal demodulation methods have been developed over the years [8,9]. However, these methods did not successfully address how to enhance the weak signature from noisy signals and how to detect early-stage defect.

Normally, bearing vibration signals are collected with a vibration sensor installed on the bearing housing, where the sensors are often subject to collecting active vibration sources from other mechanical components. The inherent deficiency of the measuring mechanism introduces a great amount of noise to the signal. Therefore, the signature of a defective bearing is spread across a wide frequency band and hence can easily become masked by noise and low frequency effects [7]. One of the challenges is how to enhance the weak signature at the early stage of defect development. A signal enhancing method is needed to provide more evident information for bearing performance assessment and prognostics.

The traditional approach for extracting signals from a noisy background is to design an appropriate filter, which removes the noise components and at the same time lets the desired signal go through unchanged. Based on noise type and application, different filters can be design to conduct the de-noising [10]. However, for a situation, where the noise type and frequency range are unknown, the traditional filter design could become a computationally intense process. The wavelet transform has been widely used in signal denoising due to its extraordinary time–frequency representation capability, which is discussed in detail later in this paper. While most of the signal de-noising approaches intend to detect smooth curves from the noisy raw signals, the vibration signal from mechanical failure, such as gears and bearings, are more impulse-like than smooth. Lin [11] developed a de-noising method based on Morlet wavelet analysis and applied the method to feature extraction of gear vibration signals. That method seeks the optimal wavelet filter that can give out the largest kurtosis value for the transformed signal. However, the defect signature of bearing is periodical impulses. The periodicity plays an important role in fault identification and should not be ignored in optimal wavelet filter construction.

Another challenge of bearing prognostics is how to effectively evaluate the system performance based on the extracted features. One the primary difficulties for effective implementation of bearing prognostics is the highly stochastic nature of defect growth [15,16]. Even though a large variety of features can be extracted to describe the characteristics of vibration signal from different aspects, such as RMS, Kurtosis, Crest Factor, Cepstrum, and Envelope Spectrum, etc. previous research work has shown that each feature is only effective for certain defect at certain stage. For example, spikiness of the vibration signals indicated by Crest Factor and Kurtosis implies incipient defects, whereas the high energy level given by the value of RMS indicates severe defects [17]. A good performance assessment method should take advantage of mutual information from multiple features for system degradation assessment. Feature map is one of widely used performance degradation assessment methods. Some successful applications can be found in [18]. It is based on the assumption that if an exclusive feature space could be established using physical models, expert experiences, and/or historical data, different fault modes should be able to find its corresponding mapping coordinates on it. With the exception of abrupt catastrophic failures, most of the faults have some kind of progression to failure. Consequently their coordinates in feature space should have traceable trajectories drifting from the normal operation region to various fault regions. By continuously tracking the trajectories, degradation detection and performance assessment in feature space can be accomplished. The challenge is how to construct a feature space that can consistently exemplify the degradation pattern.

This paper introduces the developed robust tools on signal de-noising and performance assessment for rolling element bearing prognostics. A wavelet filter based method
is used at first to address the signal de-noising and feature enhancement issue. The performance of wavelet decomposition-based de-noising and wavelet filter-based de-noising methods are compared based on signals from mechanical defects. The comparison result shows that wavelet filter is more suitable and reliable for de-noising of mechanical defect signals, whereas the wavelet decomposition de-noising method can achieve satisfactory results on smooth signals. To select the optimal parameters for the wavelet filter, a two-step optimization process is proposed to obtain an optimally designed wavelet filter.

Secondly, a Self Organizing Map (SOM) based method is proposed to address the problem of construction of feature space, degradation detection and performance assessment. The SOM neural network is a non-supervised learning neural network designed to organize itself according to the nature of the input data [25]. Two different methods of applying SOM in condition assessment and degradation detection, Trajectory method and Quantization Error method, are introduced.

Finally, a rolling bearing run-to-failure test is introduced, which is designed to generate multiple sets of full life cycle degradation data for validation of proposed methods. The result demonstrates that by designing an optimal wavelet filter, the bearing defects can be detected at an early stage of development and the proposed SOM based method can effectively and quantitatively describe the bearing degradation process. Combined with prediction technology, a robust bearing prognostic method can be developed.

2. Signal de-noising method

2.1. Wavelet transform

The wavelet is obtained from a single function \( \psi_{(a,b)}(t) \) by translation and dilation:

\[
\psi_{(a,b)}(t) = \frac{1}{\sqrt{a}} \psi \left( \frac{t - b}{a} \right) \tag{1}
\]

where \( a \) is the so-called scaling parameter, \( b \) is the time localization parameter and \( \psi(t) \) is called the ‘mother wavelet’. The parameters of translation \( b \in \mathbb{R} \) and dilation \( a > 0 \), may be continuous or discrete.

The wavelet transform of a finite energy signal \( x(t) \) with the analyzing wavelet \( \psi(t) \) is the inner product of \( x(t) \) with a scaled and conjugated wavelet:

\[
W(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t) \overline{\psi \left( \frac{t - b}{a} \right)} \, dt \tag{2}
\]

where \( \overline{\psi(t)} \) stands for the complex conjugation of \( \psi(t) \).

The wavelet transform \( W(a,b) \) can be considered as functions of translation \( b \) with each scale \( a \). Wavelet transform is also reversible, which provides the possibility to reconstruct the original signal. A classical inversion formula is:

\[
x(t) = C_\psi^{-1} \int \int W(a,b) \psi_{(a,b)}(t) \frac{da}{a} \, db \tag{3}
\]

where

\[
C_\psi = \int_{-\infty}^{\infty} \frac{\left| \hat{\psi}(\omega) \right|^2}{|\omega|^2} \, d\omega < \infty \tag{4}
\]

\[
\hat{\psi}(\omega) = \int \psi(t) \exp(-j\omega t) \, dt \tag{5}
\]

2.2. Wavelet decomposition-based de-noising method

Wavelet de-noising method is based on the principle of multi-resolution analysis [11]. By using multi-level wavelet decomposition the discrete detail coefficient and approximation coefficient can be easily obtained. Grossmann [13] proved that the variance and amplitude of the details of white noise at various levels decreases regularly as the level increases, whereas the amplitude and variance of the wavelet transform of the available signal are not related to the change of scale. According to this property, noise can be weakened or even removed by adjusting the wavelet coefficients properly.

In general, the de-noising procedure involves three steps as follows:

1. **Signal decomposition.** Choose a wavelet basis, choose a level \( N \). Compute the wavelet decomposition of the signal at level \( N \).
2. **Threshold detail coefficients.** For each level from 1 to \( N \), select a threshold and apply soft thresholding to the detail coefficients.
3. **Signal reconstruction.** Compute wavelet reconstruction using the original approximation coefficients of level \( N \) and the modified detail coefficients of levels from 1 to \( N \).

Generally speaking, this method performs very well on Gaussian noise and can almost achieve optimal noise reduction while preserving the signal. However, there are still unsolved issues. The first is how to select an optimum wavelet for a particular kind of signal. Currently there is no common guideline for neither how to select the optimum wavelet basis, nor how to select the corresponding shape parameter and scale level for a particular application. The second issue is how to perform thresholding. Despite a large variety of threshold selection strategies proposed in recent literatures [20], threshold selection for a specific application, where prior knowledge about the data is limited is still an open issue.

In summary, the wavelet decomposition-based de-noising method relies on the basic idea that the energy of a signal will often be concentrated in a few coefficients in the wavelet domain. Therefore, the non-linear thresholding function will tend to retain a few larger coefficients representing the signal and at the same time tends to reduce
the noise coefficients to zero. Most of the time the signals we intend to extract from the noisy background are smooth signals with none or only a few abrupt changes. By carefully selecting the wavelet basis, a sparse wavelet coefficient matrix can be achieved, which promises success for the de-noising process. However, if the signal consists of a lot of impulse components, which is often true in machinery diagnostic applications, where faulty gear and bearing signals are more ‘impulse-like,’ a sparse wavelet representation is very difficult to obtain [14]. This adds great difficulties to the wavelet de-noising process.

2.3. Optimal wavelet filter

2.3.1. The principle of wavelet filter

An important Fourier transform property is that convolution in one domain corresponds to multiplication in the other domain. So Eq. (2) can take the following alternative form:

\[
W(a,b) = \sqrt{a}F^{-1}\{X(f)\Psi^*(af)\}
\]  
(6)

where \(X(f)\) and \(\Psi(f)\) are the Fourier transforms of \(x(t)\) and \(\psi(t)\), respectively, and \(F^{-1}\) denotes the inverse Fourier transform. Eq. (6) shows that the wavelet transform can also be considered as a special filtering operation. The frequency segmentation is obtained by dilating the analysis wavelet. In other words, if we treat the daughter wavelet as a filter kernel, the convolution process in the wavelet transform is simply a filtering operation. The frequency response of the wavelet filter varies as the basis wavelet shape and scale changes, thus low-pass, high-pass, band-pass or even multiple-band pass filters can be built by reconstructing the wavelet coefficients at selected scales.

Another feature of Eq. (2) is that \(W(a,b)\) gives the information of \(x(t)\) at different levels of resolution and also measures the similarity between the signal \(x(t)\) and the wavelet function. This implies that a wavelet can be used for feature discovery if the wavelet used is similar to the feature components hidden in the signal. To some extent this convolution process of the daughter wavelet and the analyzed signal is similar to another classical concept of signal processing: matched filtering, which is originally derived from the correlation process.

2.3.2. Morlet wavelet

The Morlet wavelet was defined as [22]:

\[
\psi_{\alpha}(t) = \frac{1}{\sqrt{2\pi}}e^{-\beta^2 t^2/2}\cos(2\pi v_0 t)
\]  
(7)

where \(v_0\) is a constant, and \(\beta\) is the shape parameter which balances the time resolution and the frequency resolution of the Morlet wavelet.

The Morlet wavelet is a cosine signal that decays exponentially on both the left and right sides, and its function shape is very similar to an impulse. This similarity makes the Morlet wavelet very attractive and widely applied in mechanical fault diagnostic applications. A daughter Morlet wavelet is obtained by time translation and scale dilation from the mother wavelet,

\[
\psi_{a,b}(t) = \psi\left(\frac{t - b}{a}\right) = e^{-\beta^2(t-b)^2/2a^2}\cos\left[\frac{\pi(t - b)}{a}\right]
\]  
(8)

where \(a\) is the scale parameter for dilation and \(b\) is for time translation. By carefully choosing parameters \(a\) and \(\beta\), we can generate a daughter Morlet wavelet, as shown in Fig. 3(b), which closely matches the shape of mechanical impulse. It is not difficult to imagine that if we use this Morlet wavelet as the filter kernel and conduct the wavelet transform (or wavelet filtering), this wavelet filter should be able to detect the components that are similar to it within a noisy signal.

Again, the characteristics of impulse varies as the system parameters, stiffness, damping, and vibration transmissions path change from case to case. Therefore, an automatic process is needed that can generate an optimal wavelet filter for a specific application.

2.3.3. Optimal selection of shape factor

The sparsity of wavelet coefficients is often used as the rule for evaluating the efficiency of wavelet transforms. The wavelet corresponding to the fewest wavelet transformation coefficients of a signal is ideal. An optimal wavelet

---

**Fig. 3. Comparison of (a) mechanical impulse and (b) Morlet wavelet.**
transformation should be able to condense the signal into several large coefficients. The diversity of a series can be measured by Shannon Entropy, thus it can be used to evaluate the sparsity of wavelet coefficients [14,21]. Therefore, the shape factor $\beta$ can be determined according to the minimal Shannon entropy of wavelet coefficients.

### 2.3.4. Optimal selection of scale

After the shape factor $\beta$ is determined by the minimal Shannon entropy criterion, the next step is to decide the appropriate wavelet transformation scale $a$, in other words, the frequency range of the wavelet filter, so that the periodic pattern of the noisy signal can be detected.

Since the objective of de-noising is to identify the weak periodic components from the noisy signal, the periodicity of wavelet coefficients can be used as the criterion of selecting optimal scale $a$.

Singular Value Decomposition (SVD) can be applied to detect the periodicity of the time series [23]. The SVD of an $m \times n$ matrix $D$ is defined as the decomposition [24]

$$D = U E V^T$$

(9)

where $U$ is $m \times m$ square matrix and $V$ is $n \times n$ square matrix with orthogonal columns so that

$$U^T U = I, \quad V^T V = I$$

(10)

Additionally $E$ is an $m \times n$ diagonal matrix, $E = \text{diag} (\sigma_1, \sigma_2, \ldots, \sigma_p)$, with $p = \text{min}(m,n)$, and the diagonal elements $[\sigma_1, \sigma_2, \ldots, \sigma_p]$ of matrix $E$ are the singular values of matrix $D$ and $[\sigma_1, \sigma_2, \ldots, \sigma_p]$ are conventionally arranged as $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p$. The power of the SVD becomes apparent as its connections with other fundamental topics of linear algebra are explored. For example, if $D$ has rank $r$ and $r > 0$, then $D$ has exactly $r$ strictly non-zero singular values, so that $\sigma_r > 0$ and $\sigma_{r+1} = \cdots = \sigma_p = 0$. If $D$ has full rank, all its singular values are non-zero.

Consider a periodic signal $X = [x_1, \ldots, x_n]$ with a period of length $n$. A matrix $X$ can be formed by partitioning the series into periods and placing each period as a row of $X$, as:

$$X = \begin{pmatrix} x(1) & \cdots & x(n) \\ x(n+1) & \cdots & x(2n) \\ \vdots & \ddots & \vdots \\ x((m-1)n+1) & \cdots & x(mn) \end{pmatrix}$$

(11)

The matrix $X$ has $m$ repeated rows and is of rank 1. Therefore, it should have only 1 non-zero singular value $\sigma_1$ and $m-1$ zero singular values.

Now consider the case of a periodic waveform with time-varying amplitude plus noise. Assuming the period length of the time series $X = [x_1, \ldots, x_n]$ is still $n$, a different size matrix $X(\text{round}(l/2), i), 2 \leq i \leq l/2$ can be formed by dividing the time series into segments with different lengths $i$. The matrix $X$ may now be full rank due to the noise, but $\sigma_1$ would be very large compared to the rest of singular values when $i = n$. Hence, the ratio

$$\delta_i = \left( \frac{\sigma_1}{\sigma_2} \right)^2$$

will exhibit its maximal value at $i = n$. Then $\delta_i$ can be used to estimate the periodicity of the signal [23].

Therefore, scale $a$ that discloses the strongest periodicity from the wavelet coefficients will be selected as the optimal wavelet transform scale. Connecting the two optimization processes, a two-step optimal parameter selection algorithm can be designed as depicted in Fig. 4.

### 2.3.5. Comparison study on simulated signal

A simulated impulse signal consisting of 10 impulses with period of 100 data points is used to test the de-noising performance of the wavelet filter method proposed in this paper. White noise is added onto the simulated signal. Both Wavelet decomposition-based de-noising method and wavelet filter methods are applied to the simulated data. For wavelet decomposition-based de-noising method, different thresholds are applied as well.

Fig. 5 (a) and (b) show the de-noised signals by applying soft heuristic and soft universal threshold, respectively. Only the de-noised signal in Fig. 5(a) recovers the original signal partially, however, the periodic feature of the original signal is no longer noticeable anymore. In addition, there are other factors influencing the effectiveness of de-noising, such as wavelet decomposition level and threshold rescaling method selection, which make the de-noising problem even more intricate. Since there are no explicit guidelines for how

---

![Flowchart for selecting optimal shape factor and wavelet transform scale](image_url)
to tune the existing parameters, most of the time de-noising becomes a trial-and-error process.

For the comparison study, the wavelet filter method proposed in this paper is applied to the same set of simulated data. To find an optimal wavelet filter that can discover the periodic impulses from the noisy raw signal, the first step is to search for the optimal shape factor $\beta$. Increasing $\beta$ from 0.1 to 20 and calculating the entropy of the corresponding coefficients, the optimal shape factor $\beta$ leading to the minimal Shannon entropy relationship can be obtained. As depicted in Fig. 6, the entropy exhibits its minimal value at $\beta=0.54$. Therefore, $\beta=0.54$ is selected as the optimal shape factor.

After the shape factor $\beta=0.54$ is selected, a recursive route is carried out to find the optimal scale that can uncover the strongest periodicity from the wavelet transform results. The searching ranges for period and scale are set as $[2,300]$ and $[1,30]$, respectively. The measurement of periodicity $d_i$ is calculated and presented as a three-dimensional surface in Fig. 7. Strong periodicity is discovered when the period $i$ equals 100, 200, and 300, which matches the period of the simulated signal. In addition, the periodicity reaches its maximum value when the scale $\alpha=10$. Thus, we can conclude that conducting the wavelet transform at scale 10 can best reveal the periodicity of the signal.

Fig. 8 shows the de-noised signal by applying the Morlet wavelet filter with optimal shape parameter $\beta=0.54$ and scale parameter $\alpha=10$. The simulated impulses are represented clearly. Even though the noise level is still high, the periodic character of the simulated signal, which is the most important feature for fault diagnostics, is recovered.

3. SOM based performance degradation assessment method

3.1. Theoretical background of SOM

The SOM is a neural network concept developed by Kohonen in [25]. It forms a one or two-dimensional presentation from multi-dimensional data. The topology of the data is kept in the presentation such that data vectors, which closely resemble one another, are located next to each other.
Fig. 8. De-noised signal by wavelet filter with optimal parameter $\beta=0.54$ and $a=10$.

other on the map. In contrast to traditional methods, such as principal component analysis, the SOM can also be created from highly deviating, non-linear data [26].

The map units of SOM, or neurons, form usually a two dimensional regular lattice. Each neuron $i$ of the SOM is represented by a $n$-dimensional weight, or model vector, $m_i = [m_{i1}, m_{i2}, \ldots, m_{in}]^T$ ($n$ is the dimension of the input vector). The neurons of the map are connected to adjacent neurons by a neighborhood relation, which indicates the topology, or the structure, of the map. Usually rectangular or hexagonal topology is used.

After the input data is normalized, the SOM is trained iteratively. In each training step, one sample vector $X$ from the input data set is chosen randomly and the distance between it and all the weight vectors of the SOM, which is originally initialized randomly, is calculated using some distance measure, such as Euclidian distance. The Best Matching Unit (BMU) is the map unit whose weight vector is closest to $X$. After the BMU is identified, the weight vectors of the BMU as well as its topological neighbors are updated so that there are moved closer to the input vector in the input space. The vectors are updated following the learning rule:

$$m_i(t+1) = m_i(t) + \alpha(t) \cdot h(n_{BMU}, n_i, t) (X - m_i(t)) \quad (13)$$

where $h(n_{BMU}, n_i, t)$ is the neighborhood function, which is monotonically deceasing with respect to the distance between the BMU $n_{BMU}$ and $n_i$ in the grid, and the training time; $\alpha(t)$ is the learning rate, a decreasing function with $0 < \alpha(t) < 1$.

At the end of the learning process, the weight vectors are grouped in clusters depending on their distance in the input space. The SOM can be interpreted by labeling the units according to input vectors, whose type or operation state is known. The Unified distance Matrix (U-matrix) method [31], showing the distances between neighborhood units, is widely used to visualize the cluster structure of the SOM. In graphic presentation, the darker the color between two map units is, the larger is the relative distance between them, and thus indicate cluster boarders.

Unlike networks based on supervised learning which require that target values corresponding to input vectors are known, the SOM can be used for clustering data without knowing the class membership of the input data [32]. It can be used to detect feature inherent to the problem. Together with proper visualization methods the SOM is a powerful tool for discovering the visualizing general structures of the state space, and therefore it is an efficient tool for visualizing the system behavior and an effective tool for condition monitoring and system degradation detection.

3.1.1. Use trajectory for operation state monitoring

If data from multiple operation regions as well as faulty operation states are available, the trained SOM can be treated as a state space, where different clusters represent different operation states. The condition of the machine can be described by its matching region in SOM. And the operation state changes can be described by the trajectory of its BMUs in SOM. In normal operation, the BMUs should follow well-defined paths or trajectories in normal regions. When an incipient fault appears, its BMU would deviate from the normal region. The deviation will depend on the type and severity of the abnormality. By plotting the trajectory of current data on a labeled map, the machine condition can be followed over time. Furthermore, if a probability of the next BMUs is available, a prediction of the next possible machine state can be assessed.

3.1.2. Quantization error for fault and degradation detection

Given the fact that most of the time, it is hard to acquire a dataset representative of the whole failure space, whereas the normal operation space can be characterized very accurately, fault detection can be based on the quantization error away from the normal feature space. At first, the SOM is trained with normal operation data. Then the feature vector corresponding to the unidentified measurement is compared with the weight vectors of all map units, and if the smallest difference exceeds a predetermined threshold, the process is probably in a fault situation. This conclusion is based on the assumption that a large quantization error corresponding to the operation point belonging to the space not covered by the training data. Therefore, the situation is new and something is possibly going abnormal. Depending on how far away the current process is deviating from the normal operation state, a quantitative degradation index can be calculated.

Practically, the quantitative degradation assessment can be fulfilled by calculation of minimum quantization error (MQE) of the new measurement data to a SOM trained using normal operation data sets. From the degradation monitoring point of view, the distance between the BMU...
and input data actually indicates how far away the input data is deriving from the normal operation region. Thus, the MQE can be defined as:

$$MQE = \| D - m_{BMU} \|$$  \hspace{1cm} (14)

where $D$ is the input data vector and $m_{BMU}$ stands for the weight vector of BMU. Extremely high MQE value may occur for two reasons: either the testing feature vector is an outlier or it belongs to a fault class [27]. Therefore, the condition degradation can be quantized and visualized by following the trends of MQE.

4. Experimental verification using roller element bearing

4.1. Experimental setup

Most bearing diagnostics research involves studying the defective bearings recovered from the field, where the bearings exhibit mature faults, or from simulated or ‘seeded’ damage. Simulated damage is typically induced by scratching or drilling the surface, introducing debris into the lubricant, or machining with an electrical discharge. Experiments using defective bearings have less capability to discover natural defect propagation in the early stages. In order to validate the wavelet filter methodology and truly reflect the real defect propagation processes, bearing run-to-failure tests were performed under constant load conditions on a specially designed test rig.

The bearing test rig hosts four test bearings on one shaft. The shaft is driven by an AC motor and coupled by rub belts. The rotation speed was kept constant at 2000 rpm.

Fig. 9. Bearing test rig.

Fig. 10. Photo of bearing components after test (a) inner race defect in bearing 3, test 1 (b) roller element defect in bearing 4, test 1 (c) outer race defect in bearing 1, test 2.

Fig. 11. Time feature (a) RMS of bearing 3 (b) RMS of bearing 4 for the whole life cycle.
A radial load of 6000 lbs is added to the shaft and bearing by a spring mechanism. All the bearings are forcibly lubricated. An oil circulation system regulates the flow and the temperature of the lubricant. A magnetic plug installed in the oil feedback pipe collects debris from the oil as evidence of bearing degradation. The test will stop when the accumulated debris adhered to the magnetic plug exceeds a certain level and causes an electrical switch to close.

Four Rexnord ZA-2115 double row bearings were installed on one shaft as shown in Fig. 9. The bearings have 16 rollers in each row, a pitch diameter of 2.815 in., roller diameter of 0.331 in., and a tapered contact angle of 15.17 degrees. A PCB 353B33 High Sensitivity Quartz ICP® Accelerometer was installed on each bearing housing. Four thermocouples were attached to the outer race of each bearing to record bearing temperature for monitoring the lubrication purposes. Vibration data was collected every 20 min by a National Instruments DAQCard™-6062E data acquisition card. The data sampling rate was 20 kHz and the data length was 20,480 points. Data collection was conducted using National Instruments LabVIEW software.

4.2. Experimental result analysis

Three sets of tests are carried out. All tests are stopped until a significant amount of metal debris was found on the magnetic plug of the test bearing.

Test one ended up with an inner race defect in bearing 3 and a roller element defect in bearing 4. Test two and three ended up with an outer race defect in bearing 1 and 3, respectively. Three inspection photos from test one and two are shown in Fig. 10. Fig. 11 depicts the time domain features, root mean square (RMS) for the entire life cycle from bearing 3 and bearing 4 of test 1, respectively.

Fig. 11(a) reveals that for the inner race defect, the change of RMS can be divided into two stages. In the first stage, during the first 30 days of operation, no underlying trend can be observed. After the test had been carried out for 30 days (approximately 86.4 million cycles), the RMS started to increase and the rate of change also increased significantly.

The time domain feature also shows that most of the bearing fatigue time is consumed during the period of material accumulative damage, while the period of crack propagation and development is relatively short. The experiment in Ref. [28] also verified this discovery. This phenomenon is obvious if all the RMS curves are plotted in one chart as shown in Fig. 12. This means that if the traditional threshold-based condition monitoring approach is used, the time available for the maintenance crew to respond prior to catastrophic failure after a defect is confirmed is very short. An early warning approach that can detect the defect at the early stage is demanded so that enough buffer time is available for maintenance and logistical scheduling. This requirement is extremely important for some mission-critical situations, such as power plants and continuous production lines.

Another important information Fig. 12 revealed is the inconsistent degradation pattern. Even all the test bearings are the same type and are tested under the same operational
condition, their RMS trends still exhibit strong inconsistency. It would be very difficult to establish a deterministic model that can accurately describe the variable process of defect propagation. Therefore, a practical approach to describe the bearing degradation should base on the combination of trend analysis and robust performance assessment that derives from the understanding of the historical behaviors and in-process condition symptoms.

4.2.1. Signal de-noising

Fig. 13 presents the vibration waveform collected from bearing 3 of test 2 at the last stage of the bearing test. The signal exhibits strong impulse periodicity because of the impacts generated by a mature outer race defect. The bandpass frequency of the outer race (BPFO) is 236.4 Hz, so we expect to see a duration between the two conjoined impulses of 1/236.4 = 0.0042 s. The vibration waveform clearly verifies the calculation.

However, when examining historical data and observing the vibration signal three days before the bearing failed, the periodic impulse feature is heavily masked by the noise as shown in Fig. 14.

The de-noising method proposed in this paper is used to enhance the signal shown above. At first, the optimal wavelet shape factor $\beta = 1.3$ is found by minimal entropy method. Then scale $a$ from 1 to 5 was scanned to find the optimal scale that can reveal the signal periodicity most clearly. Fig. 15 demonstrates the periodicity measurement when the scales for the wavelet filter are chosen as $[1, 5]$. Notably, the selection of search range for optimal scale $a$ is very important. That is the same meaning as the selection of variable boundary in an optimization problem. It should be determined by checking the reasonable band pass width of corresponding wavelet filter, otherwise a misleading result might be generated.

The periodicity of the de-noised signal reaches its maximum value using scale $a = 2.6$. Fig. 16 further illustrates that when scale $a = 2.6$ the periodicity measurement exhibits its maximum value at period $n = 89$ data points. Given the data sampling rate of 20 kHz and signal length of 20,480 data points, period $n = 89$ actually means...
frequency $20,480/89 = 230$ Hz, which is very close to the BPFO 236.4 Hz.

Applying the wavelet filter with carefully selected shape factor $\beta = 1.3$ and scale $a = 2.6$, the de-noised signal can be obtained as shown in Fig. 17. The periodic impulse feature is clearly discovered. The period of the impulse is 230 Hz, which is strong evidence of bearing outer race degradation. Comparing Figs. 17 and 14, the wavelet filter-based de-noising method successfully enhanced the signal feature and provided strong proof for a more proactive and prognostic decision-making.

4.2.2. SOM application

The SOM toolbox [29] developed by Helsinki University of Technology is used in this research. The authors adopt the parameter selection guidance from Ref. [30]. The number of map units is approximately $m = 5 \sqrt{n}$, where $n$ is the number of data samples. The map shape is rectangular with hexagonal lattice and the ratio of side lengths corresponds to the ratio between two greatest eigenvalues of the covariance matrix of the input data.

The data collected from bearing 4 of test one, which exhibits roller element defect, is used for SOM training at first. The whole data set includes 2342 data points. The feature vector is constructed by the following time domain features, RMS, Kurtosis, Crest Factor, and corresponding RMS, Kurtosis and Crest Factor of envelope signals. Fig. 18 presents the U-matrix map of the trained SOM. By investigating the U-matrix map, three clusters can be identified from the gray color difference, as pointed out in Fig. 18. Those identified clusters in U-matrix can be hypothetically associated with three stages in run-to-failure test, which are normal, degradation and failure. Cluster one, the biggest cluster in upper left area, could be the normal operating condition.

To validate the hypothesis, trajectory from normal operating condition to final failure is plotted on the U-matrix map. In order to show the trajectory clearly, one of every 100 data points is picked up from whole data sets. Fig. 19
shows the trajectory and arrow presents the direction of trajectory. It clearly confirmed that operating condition has slowly moved from cluster 1 to cluster 2 and eventually stopped at cluster 3. Therefore, the three identified clusters can be labeled as normal operation, degradation and failure, respectively. By means of labeling, normal and abnormal regions can be identified visibly. Moreover, degradation monitoring can be realized based on observation of the trajectory.

The data of bearing 3 of test 1 and bearing 1 of test 2, which exhibit inner race failure and outer race failure, respectively, are also used to train the corresponding SOM. The U-matrix maps and trajectories are shown in Figs. 19 and 20. The final U-matrix maps are different because those three data sets are corresponding to three different failure modes. However, the normal condition is always identical no matter what kind of fault the bearing eventually has. Furthermore, the migration tendency of trajectory also provides potential information about failure modes. Thus, if data from different degradation processes are used jointly to train a SOM, no only the incipient degradation could be detected, but also a prediction of possible final fault mode could be generated based on the trend of trajectory.

To validate the MQE’s capability of degradation detection, SOM is firstly trained by the selected normal data sets from bearing 1 of test 2. Then the full life cycle data from bearing 3 of test 1, bearing 4 of test 1 and bearing 1 of test 2 are input to the trained the SOM and the corresponding MQE are calculated. The results are shown in Fig. 21.

The bearing degradation processes are clearly presented in Fig. 21 by plotting MQE versus time. Comparing Fig. 21(a) to Fig. 11(a) which are both from bearing 3 of test 1, the MQE index has much more sensitivity than the RMS feature in terms of identifying incipient defect. Therefore, an early warning is possible if MQE is calculated and monitored in real time. In addition, although the bearing lives and failure modes are different from each test and each bearing, the MQE chart still consistently depicts the bearing degradation behavior in the whole run-to-failure test.
A consistent bearing performance curve is essential to an accurate prediction of the bearing remaining useful life. The unique robust feature of MQE chart facilitates a reliable rolling bearing performance prediction.

5. Conclusions

This paper addressed challenging issues on de-noising and extraction of the weak signature from the noisy signal for bearing prognostics. The enhanced robust methods include a wavelet filter based method for weak signature enhancement and SOM based method for performance degradation assessment.

The performance of traditional wavelet decomposition-based de-noising methods are greatly impacted by relative energy levels of signal and white noise coefficients. When dealing with smooth signals, satisfactory results can generally be achieved by manipulating the threshold. However, it is much more challenging to de-noise impulse series signals, where wavelet coefficients are not so concentrated. Morlet wavelet filter based de-noising method is based on the idea of detecting the ‘similar’ impulse components from the noisy signal by designing a daughter Morlet wavelet with specific shape factor \( \beta \) at certain scale \( a \). This method is well suitable for detecting the weak signature from the defective bearing signal, where defect features are impulse-like. By applying the minimal Shannon entropy criterion, an optimal wavelet shape factor \( \beta \) can be obtained. The optimal scale \( a \) can be determined by SVD-based periodicity evaluation of wavelet transform results based on the assumption that the undetected signature is periodic.

The experimental results verified that the weak periodic impulse signature is successfully revealed and enhanced, therefore the degradation at its early stage can be detected. It also showed that SOM based performance assessment and degradation detection approach provides a means of enhancing the condition monitoring of roller bearing. It provides a comprehensible indication of current operation state. Performance degradation can be measured by monitoring the trajectory and quantitative performance assessment is possible by MQE calculation. Combined with time series prediction methods such as Auto Regressive and Moving Average (ARMA) [33] and recurrent neural network [34], the prediction of remaining useful life of rolling bearing can be realized.

Acknowledgements

This research was funded by the National Science Foundation Industry/University Cooperative Research Center (NSF I/UCRC) on Intelligent Maintenance Systems (IMS) under Grant No. 0117518. Rexnord Technical Services, an IMS sponsor, provided the testing facility for this research.

References


