Fault Detection of Rolling Bearings using Discrete Wavelet Transform and Neural Network of SVM

1O.R. Seryasat, 1M. Habibi, 1M. Ghane, 2H. Taherkhani

1 Department of Electrical and Computer Engineering, Takestan Branch, Islamic Azad University, Takestan, Iran
2 Department of Management, Qeshm International Branch, Islamic Azad University, Qeshm, Iran

INTRODUCTION

Rolling bearing is one of the most commonly used components in rotating machineries. Detecting Rolling bearing faults using methods of condition monitoring has been the subject of extensive research over the past two decades. There are a large number of methods for detecting faults in bearings, like [1-6]. Examining vibrations based on condition monitoring is the main method in this field. Faults in rolling bearings can be caused by many factors such as improper design or improper mounting, acid corrosion, improper lubrication, and plastic deformation.

The most common fault is due to material fatigue after a certain period of time. This phenomenon begins with observing small cracks below the surface of the bearing components during the work.

Cracks spread towards the surface as a result of the cyclic loads and cause perforation or crushing of the surface. Statistical data shows that 90% of all multiple faults in rolling bearing is related to cracking of the inner and outer ring, while cracking balls and racks are one other faults [7]. According to the importance this issue, detecting the first three types of faults was considered in this study.

Due to the importance of rolling bearings and their extended application, providing a convenient method for condition monitoring to prevent malfunction or failure during the work seems necessary.

Over the past three decades, various methods have been developed among which the vibrational analysis gives the best results. In this study, an advanced technique of signal analysis based on the evaluation of the discrete wavelet transform with neural network SVM was used. Wavelet analysis allows working in the scope of time - frequency and mapping with good accuracy at high frequencies the system being experimented As shown in Figure 1, experiment desk consisted of an engine with two rolling bearings, one in the output shaft and the other one is on other side of the fan. The results obtained from wavelet and neural network SVM analysis is consistent with the experimental observations.
Table 1: Characteristics of bearings being experimented.

Output side bearing - 6205-2RS JEM SKF, deep groove ball bearing
Size: inch

<table>
<thead>
<tr>
<th>Inner diameter</th>
<th>Outer diameter</th>
<th>Thickness</th>
<th>Diameter of spigots</th>
<th>Diameter of stride</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9843</td>
<td>2.0472</td>
<td>0.5906</td>
<td>0.3126</td>
<td>1.537</td>
</tr>
</tbody>
</table>

Destruction frequency (Hz)

<table>
<thead>
<tr>
<th>Inner ring</th>
<th>Outer ring</th>
<th>Rack</th>
<th>Rolling component</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.4152</td>
<td>3.5848</td>
<td>0.3982</td>
<td>4.7135</td>
</tr>
</tbody>
</table>

Bearing on the side of fan-6203-2RS JEM SKF, deep groove ball bearing
Size: inch

<table>
<thead>
<tr>
<th>Inner diameter</th>
<th>Outer diameter</th>
<th>Thickness</th>
<th>Diameter of spigots</th>
<th>Diameter of stride</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.6693</td>
<td>1.5748</td>
<td>0.4724</td>
<td>0.2656</td>
<td>1.122</td>
</tr>
</tbody>
</table>

Destruction frequency (Hz)

<table>
<thead>
<tr>
<th>Inner ring</th>
<th>Outer ring</th>
<th>Rack</th>
<th>Rolling component</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.9469</td>
<td>3.0530</td>
<td>0.3817</td>
<td>3.9874</td>
</tr>
</tbody>
</table>

Fig. 1: system under study.

Discrete Wavelet transform:

Wavelet Transform is a tool to cut data and functions or working with different frequency components and then studying each with a resolution corresponding its scale that gives about information signal in both frequency and time domain

\[
\text{CWT}(a,b) = \int_{-\infty}^{+\infty} f(t) \psi_{a,b}(t) dt
\]

Where

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

\[a, b \in \mathbb{R}; \ a \neq 0\]

In The above equation \(\psi_{a,b}(t)\) defines main wavelet function. Parameter \(a\) represents the characteristic of the scale, that is equal to the inverse of frequency. Parameter \(b\) specifies the displacement of time. DWT is obtained from discretization of CWT\((a,b)\) as follows:

\[
\text{DWT}(j,k) = \frac{1}{\sqrt{2^j}} \int_{-\infty}^{+\infty} f(t) \psi \left( \frac{t-2^j k}{2^j} \right) dt
\]

Where \(a, b\) are replaced by \(2^j\) and \(2^i\). An effective method for applying this formula is using filters stated by Mallat [8]. The main step in the wavelet algorithm is shown in Figure 2. In phase of separating, discrete signal is Cluttered with a low pass filter L and a high-pass filter H and as a result two Vectors A and D are created and components of vector D are called details’ coefficients the main characteristic of this stage is as follows [9-11]:

\[
s = D + A
\]

After storing signals, 12,000 primary data of each signal was used for training the proposed algorithm in this study and the other data for testing the accuracy and reliability of this algorithm. For analysis of signals after different experiments on various wavelets, wavelet of Biorthogonal 3.1 was selected and input signals were decomposed up to 5 steps (Figure 3) by this Wavelet and standard deviation data of each step were calculated (Figure 4) and they were used as \(x_i\) inputs for \(x_i\) input of the considered neural network. Here \(i\) represents the \(i_{th}\) component of the \(x_i\) input from the next N space; that in our N=5 based on stage 5 of signal decomposition.
**Fig. 2:** main signal $S$ that passes two complementary filter and exits as 2 signal of $D$ and $A$.

**Fig. 3:** an example of decomposed Wavelet ($S=a_5+d_5+d_4+d_3+d_2+d_1$).

**Fig. 4:** Statistical data of Decomposed signal.

**Grouping:**

Using data obtained from the wavelet analysis derived in the previous step we could obtain statistical parameters necessary for fitting the probability density function related to data (for Gaussian distribution just the mean and variance are necessary to estimate). Statistical Decision optimal classifier selects Theories offer very general laws for the optimal classifying. Fisher showed that it maximizes conditional probability of $p(c_i|x)$ for a given data $x$ belonging to $c_i$. So $x$ belongs to $c_i$ category if:

$$P(c_i|x) > P(c_j|x) \text{ for all } i \neq j$$

The problem here is that this conditional probability cannot be measured directly but using Bayes law its value can be gained:
\[ P(c_i | x) = \frac{p(x | c_i) p(c_i)}{p(x)} \] (6)

Both \( p(c_i) \) and \( p(x | c_i) \) can be estimated from the calculated data, assuming probability density function. \( p(x) \) is normalising factor and can be removed in most instances of grouping (Figure 5).

\[ P(x) = \frac{1}{\sqrt{2\pi} \sigma} \exp\left(-\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2}\right) \] (7)
\[ \mu = \frac{1}{N} \sum_{i=1}^{N} x_i \] (8)
\[ \sigma^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i - \mu) \] (9)

Probability of Error is calculated by adding the sub-group area of \( i \) related to decision-making area of \( j \) group related to decision making area of group \( i \).

As can be seen in Figure 6 the error is dependent to the tail part of the Gaussian distribution. To estimate this error, we need to integrate areas listed in the input room; that in high dimension space is very difficult. The probability of error is equal to:

\[ P(error) = \int R_i P(x | c_i) P(c_i) dx + \int R_j P(x | c_j) P(c_j) dx \] (10)

That \( R_i \) and \( R_j \) are areas related to categories \( i \) and \( j \). Fukunaga showed [13] the least ratio of error is obtained by Bayes law, by choosing the boundary in a way that maximizes the conditional probability.

For a problem with given variance if we increase distance between mean of groups, Overlap will be reduced and grouping would be more accurate. However, we cannot look at group means of data to estimate the classification error since the error is related to the overlap between similarity of categories. Thus, we can have an example in which means are too far apart but the variances are so large that the overlap between the similarities are still large. On the contrary, it can mean of categories can be close to each other, but the variance of categories is small and grouping can still be done with a small error.

The \( p(x) \) depends on the distance of \( x \) from normalized mean by the variance. This distance is called Mahalanobis distance. Mahalanobis distance is representative of multiple values Gaussian distribution that is given by the following equation:

\[ P(x) = \frac{1}{(2\pi)^{D/2} |\Sigma|^{1/2}} \exp\left(-\frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu) \right) \] (11)

Where \( T \) that represents the transpose. \(|\Sigma|\), Determinant of \( \Sigma \) and \( \Sigma^{-1} \) is reverse of \( \Sigma \).

Note that in the equation \( \mu \) is a vector including mean of data of each dimension, so the vector has a dimension equal to \( D \).

\[ \mu = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_D \end{bmatrix} \] (13)
∑ Covariance is a matrix with dimension of $D \times D$ that dimension $D$ dimension is input space. $\Sigma$ Matrix is equal:

$$
\begin{bmatrix}
\sigma_{11} & \ldots & \sigma_{1D} \\
\vdots & \ddots & \vdots \\
\sigma_{D1} & \cdots & \sigma_{DD}
\end{bmatrix}
$$

(14)

And its elements are result of distribution among the data pair in coordinate $i_{th}$ and $j_{th}$ ($N$ is number of examples in groups).

$$
\sigma_{ij} = \frac{1}{N-1} \sum_{k=1}^{N} \sum_{m=1}^{N} (x_{i,k} - \mu_i)(x_{j,m} - \mu_j)
$$

(15)

Covariance measures the variance between the pairs of dimensions:

Structure of the covariance matrix for the location and shape of the selected function is critical. Since the distance for grouping is normalized by the covariance, if means of categories remain constant but covariance changes the location and shape of the selected function will change.

Fig. 6: view of error distribution due to data overlapping.

![Fig. 6: view of error distribution due to data overlapping.](image1)

Fig. 7: view of classifier to $c$ classes overlapping.

Selection Functions:

Suppose that we $N$ measures of $x_1, x_2, ..., x_n$ that each measure of $x_k$ has a vector with component $D$.

$$
x_k = [x_{k1}, x_{k2}, ..., x_{kD}]
$$

(16)

And it can be shown in a space of $D$ dimension pattern. Determining category is performed by Bayes law based on similarities comparison scaled by conditional probability.

Thus, measuring $x_k$ will belong to the category $i$ if:

$$
g_i(x_k) > g_j(x_k) \text{ for all } i \neq j
$$

(17)

Each scaled similarity can be assumed as a selection function of $g(x)$ that is function that determines the target for each point in the input space. Each category has its own special target function that reaches its maximum value for a point belonging to the category.
Optimal grouping compares just Selection functions and selects the category according to selection $g_i(x)$ that produces the largest value for the measured $x_i$. (figure 4)

$$g_i(x) = -\frac{1}{2}(x - \mu_i)^T \Sigma_i^{-1}(x - \mu_i) - \frac{D}{2} \log(2\pi) - \frac{1}{2} \log |\Sigma_i| + \log P(c_i)$$

(18)

SVM Neural Network:

SVM is one of the new algorithms used teachable machines and neural network community. Development of this algorithm originates from work on the of learning theory by Vapnik [14,15] in the last decade of the twentieth century; that is used both for grouping and regression.

In grouping state, the general idea is finding and super plane that maximizes the separation between categories.

An attractive feature of SVM is its Strong mathematics for analyzing and performance of its categorizing, clearly comparable to a Feedback Neural Network. In addition to this, controlling capability of minimizing training error and overall performance of SVM grouping is easily possible. Finding the optimal super-plane in top dimension space is done by that

Separation of two categories as maximizing problem can in form of the following formula:

$$Q(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j d_i d_j \phi^T(x_i) \phi(x_j)$$

(19)

Under condition of:

$$\sum_{i=1}^{N} \alpha_i d_i = 0, \quad 0 \leq \alpha_i \leq C$$

(20)

Here $x_i \in \mathbb{R}^n$ is teaching model and $d_i \in \{-1,1\}$ are teaching patterns, N is the number of teaching patterns, C is parameter of controlling complexity of machine - efficiency of category of training data and $\alpha$ is variable used to create the optimal super plane. In the above equations Bayes or transferring optimal super plane is not stated explicitly, but optimum value of $\alpha$ is calculated for balance state as the sum of $\alpha$ and $d$ for Bayes.

SVM uses LEM foundation, which means we do calculations in high dimension space without explicitly stating Variables in that space. The function $\phi(x)$ takes the dimension of low space pattern of $x$ and then puts it into the top dimension space. In maximizing problem mentioned above the only calculation performed in dimension space in mentioned problem is calculating internal multiplication of $\phi^T(x_i) \phi(x_j)$.

This calculation of internal multiplication can be done by a foundation $K(x,x_i)$, which can be done in low dimension space of $X$. Throughout this paper we will use foundations that have an internal Bayes sentence, So the first sentence of the inner multiplication is added to 1. Two types of these fundamentals are polynomial foundations.

$$(x^T x_i + 1)^p$$

(21)

Where p is multinomial selection power, another foundation is radial basis function (RBF):

$$\exp\left(-\frac{1}{2\sigma^2} ||x - x_i||^2\right)$$

(22)

When the foundation is used with an internal Bayes sentence, as already noted, the following maximization problem appears:

$$Q(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j d_i d_j K(x_i,x_j)$$

Under condition that:

$$0 \leq \alpha_i \leq C$$

(23)

In this study, RBF foundation is used for SVM network:

The network that was implemented had $x_i$ input with 5 components and 8 output categories. In Figures 8 and 9, algorithm of work and inputs and outputs of the network shown.
**Results:**

SVM Neural networks with 5 dimensional inputs that each dimension corresponded to the standard deviation of each stage of the wavelet decomposition, and 8 outputs, each of which represented a fault in the system under test were taught; diagram of network learning is shown in Figure 10.

In Figure 11 also sensitivity bigrams of each output in relation to input data is shown, then. To evaluate the accuracy, from every 5 faulted data samples, a total of 40 states, were tested and as it is seen in figure 11, the Network separated this data into output groups of each fault quite accurately; that has demonstrated 100% accuracy.

---

**Fig. 8:** the algorithm used for fault detection.

**Fig. 9:** the output and input of the used neural network.
Fig. 10: diagram of learning neural network of SVM.

Fig. 11: diagram of sensitivity of neural network of SVM.

Fig. 12: diagram of networks’ output against the data of experiment.

Summary:
In this study, a new method for condition monitoring of bearing using the SVM classifier was proposed. Bearing vibration data obtained was analyzed using wavelet Biorthogonal 13.1 and using the data of this process, parameters for grouping faults of rolling bearings were derived. These data were used for training by SVM network. For verification of this algorithm, a number of experimental data was tested and results obtained in this way represented 100% accuracy for detecting faults in this system. Hence, this method can be used for troubleshooting complex and sensitive systems that need accurate condition monitoring with confidence.

REFERENCES


