Bearing fault diagnosis using Hilbert-Huang transform (HHT) and support vector machine (SVM)

AIDA KABLA and KARIM MOKRANI

Laboratoire de Technologie Industrielle et de l’Information (LTII), Faculté de Technologie, Université de Bejaia, 06000 Bejaia, Algérie

Received 9 September 2013, Accepted 8 September 2015

Abstract – This work presents the application of the Hilbert-Huang transform and its marginal spectrum, for the analysis of the stator current signals for bearing faults diagnosis in asynchronous machines. Firstly, the current signals are decomposed into several intrinsic mode functions (IMFs) using the empirical mode decomposition (EMD). The Hilbert Huang spectrum for each IMF is an energy representation in the time-frequency domain using the instantaneous frequency. The marginal spectrum of each IMF can then be obtained. Secondly, the IMFs that includes dominant fault information are modeled using an autoregressive (AR) model. Finally, the AR model parameters serve as the input fault feature vectors to support vector machine (SVM) classifiers. Experimental studies show that the marginal spectrum of the second IMF can be used for the detection and classification of bearing faults. The proposed approach provides a viable signal processing tool for an online machine health status monitoring.

Key words: Signal processing / bearing faults / Hilbert-Huang transform / empirical mode decomposition / Support vector machine / AR model

1 Introduction

Asynchronous machines are widely used in the industry. They are used in many different applications: wind, military, electric drive for high speed train and pumping. To ensure continuity of operation, establishment of maintenance programs is required. Traditionally the maintenance procedure, known as repair procedure, was to repair or replace faulty equipment. A new approach, called predictive maintenance, is the detection and localization of faults and failures and act earlier to minimize their secondary effects [1]. There are many condition monitoring methods for detection of fault defects, such as vibration analysis, axial flux analysis, lubricating oil debris analysis and motor current signature analysis (MCSA). Studies on the analysis of the current signal offers the widest application range, which is ideal as a core detection technique for the condition monitoring strategy [2–6]. Moreover, in addition to the information contained in the vibrational signal, information specific to electrical phenomena appears in the stator current signal [7].

Many methods based on MCSA have been developed. These methods include the assessment of the power spectrum, the fast Fourier transform (FFT) and the spectrum analysis of the envelope. It turned out that they are effective in the detection of bearing faults. However, they are limited to stationary signals.

To treat non-stationary signals, several time-frequency analysis tools are commonly used such as the short-time Fourier transform (STFT) [8], the Wigner-Ville distribution (WVD) [9], the TFR (Time-frequency representation) of Cohen’s class and the wavelet transform (WT) [10]. The main drawback of these methods is that they depend on different parameters. For example, selection of a suitable window size is intended when applying the STFT to match with the specific frequency content of the signal, which is not known a priori. Wavelets require the specification of a core or a core function; and there is no universal core. In addition, one limitation of TFR, such as the WVD is the presence of interfering terms which affects the interpretation and the readability of the resulting representations [11]. The time-frequency smoothing can reduce the interferences but it introduces time and frequency localization errors.

In this work, we propose a combination of the Hilbert Huang Transform (HHT), autoregressive model (AR) and the support vector machine (SVM) for bearing fault diagnosis.

HHT is a method for analyzing non-stationary signals developed by Huang et al. [12]. HHT is a combination of Empirical Mode Decomposition (EMD) and the Hilbert
transform. EMD is a self-adaptive signal decomposition method, which is based upon the time scale local characteristic of the signal. It can decompose complex signals into a number of simple intrinsic mode functions (IMFs) [13]. The IMFs components can reveal the hidden information within the original signal. Moreover, the generated IMF components are stationary [14]. AR model is a time sequence analysis method whose parameters contain important information of the system condition. An AR model can accurately reflect the characteristics of a dynamic system [13]. Additionally, it is indicated that the autoregression parameters of AR model are very sensitive to the condition variation [15, 16]. However, AR model can only be applied to stationary signals, whereas the fault current signals of a rolling element bearing are non-stationary. Aiding at this problem, the EMD method is used as a pretreatment to decompose the non-stationary current signal of a roller bearing.

The SVM is a statistical learning method with good performances in many classification applications (Vapnik 1998). It can be claimed that the SVM classifier outperforms neural network classifiers in terms of generalization [17]. The last few years, SVMs have been found to be remarkably effective in many real-world applications. SVMs are introduced because of their high accuracy and good generalization for a smaller number of samples.

2 Hilbert-Huang transform (HHT)

The EMD is defined by a sifting process. It can decompose a multi-components signal into a series of IMFs. Huang et al. [12] have defined the IMFs as a function class that satisfies two conditions:

1. At any point, the mean value between the envelope defined by local maxima and the envelope defined by the local minima is zero.
2. The number of extrema and the number of zero-crossings are either equal to each other or differ by at most one.

To extract the IMFs, the sifting process used is defined in [12]. Having obtained the IMFs, we apply the Hilbert transform to each IMF.

To calculate the instantaneous characteristics (frequency and amplitude) of each IMF, the analytic signal \( z_i(t) \) associated to \( c_i(t) \) is used:

\[
 z_i(t) = c_i(t) + jH \left[ c_i(t) \right]
\]

where:

\[
 H \left[ c_i(t) \right] = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{c_i(\tau)}{t - \tau} d\tau
\]

and \( P \) is the Cauchy principal value.

The amplitude and instantaneous phase are defined by:

\[
a_i(t) = \sqrt{c_i^2(t) + H^2 \left[ c_i(t) \right]} \]

\[
\theta_i(t) = \arctan \left( \frac{H \left[ c_i(t) \right]}{c_i(t)} \right)
\]

The instantaneous frequency of \( z_i(t) \), is simply the derivative of the instantaneous phase:

\[
\omega_i = \frac{d\theta_i(t)}{dt}
\]

Thus, the original signal can be expressed as:

\[
x(t) = \Re \sum_{i=1}^{n} a_i(t) \exp \left( j \int w_i(t) \, dt \right)
\]

where the residue \( r_n(t) \) is omitted. \( \Re \{ \cdot \} \) denotes the real part of a complex quantity.

This time-frequency distribution is designated as the Hilbert-Huang spectrum \( H(w, t) \):

\[
H(w, t) = \Re \sum_{i=1}^{n} a_i(t) \exp \left( j \int w_i(t) \, dt \right)
\]

Equation (7) allows us to represent the instantaneous amplitude and frequency in three dimensions, in which the amplitude is the height in the time-frequency plane.

The time integral of Huang-Hilbert spectrum is the marginal Hilbert spectrum \( h(w) \) defined as:

\[
h(w) = \int_{0}^{T} H(w, t) \, dt
\]

where \( T \) is the signal duration.

The marginal spectrum offers a measure of the energy at each frequency. It represents the cumulated amplitude over the entire data span in a probabilistic sense [18].

Therefore, the marginal spectrum of each IMF can be defined, as:

\[
h_i(w) = \int_{0}^{T} H_i(w, t) \, dt
\]

3 Support vector machine (SVM)

SVM is a method of classification inspired by the statistical theory of supervised learning by Vapnik. It is considered today as one of the most powerful classification methods in many real applications [19, 20]. The SVM algorithm is based on the research of the optimal separating hyperplane which maximizes the training data margin.

For a given training sample set \( G = \{(x_i, y_i), \ i = 1, \ldots l\} \), each sample \( x_i \in \mathbb{R}^d \) belongs to a class by \( y \in \{+1, -1\} \). The boundary can be expressed as follows:

\[
xw + b = 0
\]
where $x$ is a weight vector and $b$ is a bias. The decision function can be done as:

$$f(x) = \text{sgn}(w \cdot x + b)$$  \hfill (11)

The optimal hyperplane separating the data can be obtained as the solution of an optimization problem.

Minimize

$$\frac{1}{2} ||w||^2$$  \hfill (12)

Subject to

$$y_i [(w \cdot x_i) + b] - 1 \geq 0, \ i = 1, \ldots, l$$  \hfill (13)

Introducing Lagrange multipliers $\alpha_i \geq 0$, the optimization problem can be rewritten as

Maximize:

$$L(w, b, \alpha) = \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{l} \alpha_i \alpha_j y_i y_j (x_i \cdot x_j)$$  \hfill (14)

Subject to:

$$\alpha_i \geq 0,$$  \hfill (15)

$$\sum_{i=1}^{l} \alpha_i y_i = 0$$  \hfill (16)

When perfect separation is not possible, for samples within the margin, slack variables are introduced. The optimization problem can be reformulated as

Minimizes

$$\frac{1}{2} ||w||^2 + c \sum_i \xi_i$$  \hfill (17)

Subject to constraints

$$y_i [(w \cdot x_i) + b] \geq 1 - \xi_i$$  \hfill (18)

Here $\xi_i$ are slack variables, which measure the degree of misclassification of the observation, and $c$ is a parameter which controls the trade-off between the slack variable $\xi_i$ and the margin $\omega$ [21].

Using Lagrange multipliers, we finally obtain

Maximize

$$L(w, b, \alpha) = \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{l} \alpha_i \alpha_j y_i y_j (x_i \cdot x_j)$$  \hfill (19)

Subject to constraints

$$\sum_{i=1}^{l} \alpha_i y_i = 0, \ c \geq \alpha_i \geq 0, \ i = 1, \ldots, l.$$  \hfill (20)

The decision function can be obtained as follow

$$f(x) = \text{sgn} \left( \sum_{i=1}^{l} \alpha_i y_i (x_i \cdot x) + b \right)$$  \hfill (21)

If the linear boundary in the input space is not sufficient to separate into two classes correctly, it is possible to create a hyperplane that allows linear separation in a higher dimension. The hyperplane is obtained by a transformation $\Phi(x)$ that maps the data from the input space to the feature space.

$$K(x, y) = \Phi(x) \cdot \Phi(y)$$  \hfill (22)

Introducing a kernel function, the basic form of SVM can be obtained:

$$f(x) = \text{sgn} \left( \sum_{i=1}^{l} \alpha_i y_i K(x, x_i) + b \right)$$  \hfill (23)

The most commonly used kernel functions are linear functions, radial basis functions, polynomials functions and sigmoid functions.

### 4 Description

#### 4.1 Main faults in asynchronous machine

A study conducted for IEEE [22] established statistical flaws that may occur on asynchronous machines: bearing, 41%, stator, 37%, rotor, 10%, and other, 12%.

This distribution shows that faults come mainly from bearings. Faulty bearings cause air gap eccentricities due to irregular motion of the rotor. These eccentricities affect the stator current due to the variations of the electromagnetic field.

#### 4.2 Characterization of ball bearings faults

A ball bearing fault is characterized by a continual repetition of faulty contacts with the bearing outer and inner cage. And, as the ball bearing supports the rotor, each fault will produce a radial motion of the rotor relative to the stator [23, 24].

According to Schoen [7], these variations generate stator currents at frequencies:

$$|f_a \mp k \cdot f_{\text{fault}}| \quad \text{were:} \ k = 1, 2, 3 \ldots$$  \hfill (24)

with:

- $f_a$: power source frequency
- $f_{\text{fault}}$: characteristic frequency induced by the fault.

The characteristic frequencies of the fault depend on the bearing dimensions and on the type of bearing defect [7, 25]. The bearing fault can be classified as inner ring, outer ring or ball. The characteristic fault frequencies are defined by the following equations:

Outer race bearing fault frequency:

$$f_a = (N/2) \cdot f_r \cdot (1 - (BD/PD) \cdot \cos \beta)$$  \hfill (25)
Table 1. Characteristics of the machine.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Power</td>
<td>1.1 kW</td>
</tr>
<tr>
<td>Power Frequency</td>
<td>51 Hz</td>
</tr>
<tr>
<td>Motor connection</td>
<td>Y</td>
</tr>
<tr>
<td>Phase voltage</td>
<td>400 V</td>
</tr>
<tr>
<td>Rated speed</td>
<td>1445 rpm</td>
</tr>
<tr>
<td>Number of pole pair</td>
<td>( P = 2 )</td>
</tr>
<tr>
<td>Number of rotor slots</td>
<td>28</td>
</tr>
<tr>
<td>Number of stator slots</td>
<td>48</td>
</tr>
</tbody>
</table>

Table 2. Bearing dimensions.

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ball diameter (BD)</td>
<td>7.9 mm</td>
</tr>
<tr>
<td>Inside diameter (D)</td>
<td>20 mm</td>
</tr>
<tr>
<td>Outside diameter (d)</td>
<td>47 mm</td>
</tr>
<tr>
<td>pitch diameter (PD)</td>
<td>((D + d)/2) 33.5 mm</td>
</tr>
</tbody>
</table>

Inner race bearing fault frequency:

\[ f_i = \left(\frac{N}{2}\right) \cdot f_r \cdot \left(1 + \left(\frac{BD}{PD}\right) \cos \beta \right) \]  

Ball defect frequency:

\[ f_b = \left(\frac{PD}{BD}\right) \cdot f_r \cdot \left(1 - \left(\frac{BD}{PD}\right) \cos \beta \right)^2 \]  

with:
- PD: pitch diameter,
- BD: ball diameter
- \( \beta \): Contact angle of the ball on the race
- N: number of balls

4.3 Description of the monitoring system

The tests were carried on the test bed at the LEG Laboratory of Grenoble. The defects were created artificially by an electrical erosion of a 1 mm diameter hole at the outer ring, inner ring or ball [26].

The monitoring system of the stator current is composed of:
- A three-phase power source with a frequency of 51 Hz.
- A sampler which aims to acquire the three phase voltages and currents of the stator. After a low pass filtering, signals are sampled at 10 kHz.
- An asynchronous squirrel cage with the characteristics given in Table 1.
- Bearing type SNR 6204, with the following characteristics:
  - 8 Balls.
  - Contact angle \( \beta = 0 \).

5 Characterization of bearing fault using the HHT

The purpose of this section is to describe the signature bearing fault as it has been observed on the steady state signals. Analyses were performed on blocks of 60,000 points; the 10 kHz sampling frequency allows a spectral resolution of 1/6 Hz.

5.1 Motor with healthy bearing

To the data of Figure 1, the EMD algorithm is applied. Figure 2 displays the empirical mode decomposition in seven IMFs of the current signal. The decomposition highlights seven modes: IMF1 \( \sim \) IMF7 and the residue. IMF1 contains the highest signal frequencies band, IMF2 the next higher frequency band and so on.

The marginal spectrum of IMF 2 and IMF 3 are presented in Figures 3 and 4, respectively. We can note that the marginal spectrum for a healthy motor reveals spectrum lines at frequencies \( f_a \) and \( 2f_a \).

5.2 Motor with outer race fault

Motor with outer race fault generates, in the stator current spectrum, lines at frequencies shown in Table 4.

Figure 5 represents the decomposition of the current signal with outer race fault. The decomposition identifies nine modes: IMF1 \( \sim \) IMF9 and the residue. The marginal spectra of the IMF 2 are shown in Figures 6 and 7. We show that the marginal spectrum of the IMF2 produces an amplitude variation of the spectrum lines at \( f_a \) and \( 2f_a \), corresponding to the theoretically calculated frequencies.
The marginal spectrum of the IMF 3 (Fig. 8) illustrates spectrum lines at frequencies \((f_a - f_{rot})\) and \((f_a + f_{rot})\), characteristics of an air gap eccentricity. We note that the marginal spectrum of the IMF 3 can’t highlight the characteristic frequencies of an outer race fault. From now on, we rely only on the analysis of second IMF.

5.3 Motor with inner race fault

Theoretically, an inner race fault generates spectrum lines at frequencies shown in Table 5.

Figure 9 shows the decomposition results for a bearing with inner race fault. Nine modes: IMF1 \(\sim\) IMF9 and the residue are identified.
Fig. 6. Marginal spectrum of IMF 2.

Fig. 7. Marginal spectrum of IMF 2 for different frequency range.

Table 5. Theoretical frequencies of inner race fault.

| k | $f_{\text{fault}} = |f_a + k f_i|$ | $f_{\text{fault}} = |f_a - k f_i|$ |
|---|---|---|
| 1 | 175.3 | 73.3 |
| 2 | 299.6 | 197.6 |
| 3 | 423.9 | 312.9 |

Figures 10 and 11 show the marginal spectrum of the IMF 2. Peaks of the spectrum lines at $(2f_a)$, $(f_i)$ and $(f_a + 2f_i)$, corresponding to the theoretically calculated frequencies, are revealed.

5.4 Motor with ball fault

Theoretically it generates, in the stator current spectrum, lines at frequencies as shown below.

Fig. 8. Marginal spectrum of IMF 3.

Fig. 9. Decomposition by EMD.

Table 6. Theoretical frequencies of ball fault.

| k | $f_{\text{fault}} = |f_a + k f_b|$ | $f_{\text{fault}} = |f_a - k f_b|$ |
|---|---|---|
| 1 | 150.6 | 48.6 |
| 2 | 250.2 | 148.2 |
| 3 | 349.8 | 247.8 |
The current signal is decomposed into eight IMF (Fig. 12). The marginal spectrum of the IMF 2 is presented in Figures 13–16 for different frequency ranges. We see clearly peaks at frequencies \((f_a + f_b)\), \((f_a + 2f_b)\), \((f_a - 2f_b)\) and \((f_a - 4f_b)\), which are related to the fault characteristic frequencies.

From the tests results, we can say that:

- bearing faults affect the marginal spectrum of the IMF 2 of the stator current;
- the ball fault is easier to detect;
- ball faults are characterized by frequencies corresponding to the theoretical outer and inner ring faults. The ball faults can be seen as outer and inner ring faults [7];
- the amplitude of the ray corresponding to faults is very small compared to the harmonics of the stator current.

This could be due to the fact that the faults are not severe enough to generate large amplitude streaks.

### 6 Implementation of the classification by SVM

#### 6.1 Feature extraction

According to expression (25), a fault creates a frequency shift of the stator current spectrum; this shift is proportional to the characteristic frequency of the fault.
We have shown in Section 5 that the marginal spectrum of IMF 2 contains the key fault information of roller bearing: the fault characteristics could be extracted. The AR model of IMF 2 is used as input vector of the SVM classifier. The fault diagnosis method is given as:

1. Roller bearings in four conditions (normal bearing, bearing with inner-race fault, bearing with out-race fault and bearing with ball fault) are tested respectively and 10 current signals of roller bearings in each condition are obtained.

2. Each signal is decomposed into IMFs using EMD, the second IMF that includes the most dominant fault information is chosen to extract the classification features.

3. The IMFs 2 of each signal are represented by their AR models; we estimate the order \( p \) and the coefficients \( (a_i) \) of each signal based on the FPE (Final Prediction Error) criterion. The parameters are obtained by averaging the estimated parameter over 300 periods of 200 points.

4. The model order estimation curves of the four conditions are shown in Figure 17. We can see that when the model order is 4, each model’s residual tends to be stable. Therefore, the model order is selected as 4 for the database.

A total of 480 feature vectors were collected (120 normal bearing, 120 bearing with inner-race fault, 120 bearing with out-race fault and 120 bearing with ball fault). 400 of the feature vectors (100 feature vectors for each condition) were used for training the classifier and 80 (20 feature vectors for each condition) as the test feature vectors.

Tables 7 list classification results for a linear Kernel functions. The average classification rate is 100%. For radial basis and polynomial function Kernel, the same average classification rate achieved.

The classification using SVMs allowed us to obtain very good result. The classification rate obtained is higher than the one in [27] where the AR model of the envelope of the stator current and SVMs were used.

### 6.2 Linear-SVMs

The results are obtained using the optimal value for the parameter \( c = 10 \).
6.3 RBF (Radial Basis Function)-SVMs

The results are obtained using the optimal value for the parameter $c = 1$.

6.4 Polynomial-SVMs

The results are obtained using the optimal value for the parameters $c = 1, d = 3$ (polynomial degree).

7 Conclusion

In this work, to overcome the limitations of traditional time/frequency analysis methods, we applied a new method for the detection of bearing faults namely the Hilbert Huang Transform and its marginal spectrum. Using EMD method, the current signal of the bearing fault can be decomposed into intrinsic modes. Therefore, we have a better understanding of the nature of the fault information within the current signal. According to the marginal spectrum of IMF 2, the characteristic frequencies of the bearing fault can be easily recognized. The experimental result have shown that HHT and its marginal spectrum can be used as an effective diagnostic method for bearing faults detection.

Classification of bearing faults using AR model of the IMF 2 of the stator current and SVM was presented. It has been shown that the proposed approach can be applied to classify the bearing fault pattern. This approach offers a new method for the diagnosis of bearing faults.

References


