Machine learning and signal processing methods for anomaly detection in predictive maintenance (Project outline)

Yapi Donatien Achou

May 6, 2019

Contents

1 Introduction

Since the industrial revolution, engines and machines have been the driving force for economical growth across industries such as automotive, airline, oil and gaze, to name a few. However, machines are prone to failure and must be monitor and maintain regularly to avoid catastrophic failure leading to significant financial and human lost. To mitigate equipment and machines proclivity toward failure and the associated cost, a process called predictive maintenance has been developed within the industrial community. Predictive maintenance for machines and industrial equipments can be defined as a maintenance philosophy or more generally a framework with a set of methods used to predict and prevent machine failure in order to avoid unexpected downtime and reduce related human and financial cost. This maintenance philosophy, when correctly implemented, increases machine life time, and reduces maintenance cost [reference].

As a framework, predictive maintenance has 4 levels [reference]. In level 1, visual inspection of machines or equipments are performed in order to assess any damage. In addition, equipment must fail before they are replaced, which can incur a high production cost. In level 2 and 3, machines characteristics such as vibration, temperature, electrical current, voltage etc are monitored continually or periodically depending on their criticality. This is called condition monitoring. In condition monitoring, the goal is to detect any change in machine normal behaviour, in order to detect failure as early as possible and schedule maintenance accordantly. Maintenance actions are performed periodically regardless of machine health condition. In level 4, big data and machine learning are the main driving forces in detecting failure and planning maintenance. At this level, maintenance action are not performed periodically but are planned according to machine health condition derived from the application of anomaly detection techniques. This reduce unnecessary maintenance action and significantly cut down maintenance cost as well as increasing machine life time.

According to PriceWaterhouseCoopers (PWC), one of the four largest auditing and consulting company in the world, a survey from 280 companies in Belgium, Germany and the Nederland, revealed that only 11% of companies have reached level 4 [6]. The application of level 4 requires collecting, saving and analysing large amount of data, from which maintenance decision can be made. Anomaly detection methods and more generally supervised and unsupervised learning methods are used in the analysis phase.

In supervised learning, based on available failure data from defect machine, a learning algorithm is trained to recognise the failure pattern in the data. This is sometime achieved by fitting the algorithm parameters to the data, which result in a model called a classifier or a regressor. A classifier is a model derived from a classification algorithm while a regressor is a model derived from a regression algorithm.

In machine learning, the failure data is called a labeled data because we can assign a categorical label such as fail or a numerical label such as 1 to specify the condition of the machine through the data measuring its characteristics. In the absence of labeled data, unsupervised learning methods can be apply to detect pattern in the data without prior knowledge to classify data regions as anomalous or not. One such technique is clustering algorithm where the input data is separated into subregions.

The procedure of classifying a data region as anomalous or not is called anomaly detection. An anomaly is defined as a pattern in the data, that does not conform to expected normal behaviour [3]. A general anomaly detection strategy will first detect normal behaviour, secondly set a boundary around the normal behaviour and finally declare any data out of the boundary as anomaly.

Several factor make this general approach of anomaly detection challenging: The notion of anomaly is different for different application domain and not all application domain have enough labeled data to train a supervise learning algorithm [3]. An extensive survey from [3] revealed that the majority of research have been focussing on simple anomalies while most application domain are faced with complex anomalies. There are mainly three types of anomalies: Point anomalies which are simple anomalies, contextual and collective anomalies which are complex anomalies [3]. When one data point in a time series is anomalous with respect to the other data points, we have a point anomaly. In a contextual anomaly, a data instance is anomalous relative to a context. For example the vibration of a machine might be very high if the load increase suddenly, and decrease when the load goes back to normal. But if the vibration increases monotonously regardless of the load, then we have a contextual anomaly. If a collection of related data instance is anomalous with respect to the entire data set, it is termed a collective anomaly [3].

In rotating machines, more than 40 % of anomalies can be attributed to bearing defect [references]. Figure 1 shows the failure statistics for rotating machines.

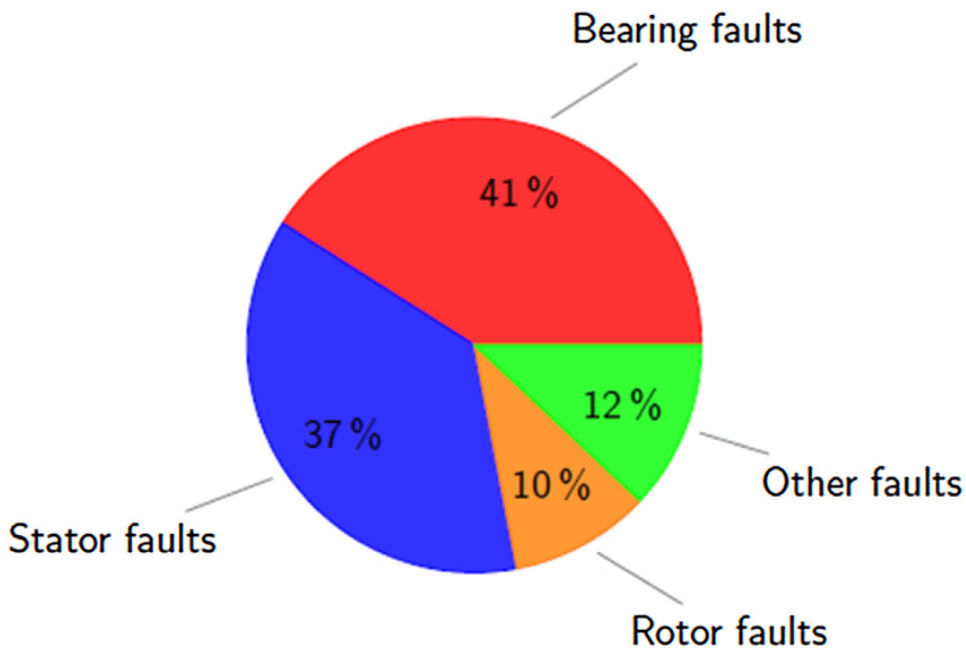

Figure 1: Defect statistics for rotating machines, taken from [reference]

In this project we present a mixed methodology to detect and predict bearing defects. The methodology consists of using signal processing for feature generation and data labelling, and machine learning for defects classification and failure prediction. For a given input dataset, the dataset is decomposed into it subcomponents or basis components. The basis components also called features or attributes are further used as input of a supervised learning algorithm for defect classification and failure prediction.

The signal processing methods used are Fourier transform, wavelet transform and Hilbert Huang transform. We focus on ensemble learning and feed forward neural network for classification. Furthermore we show that the back-propagation process in the feed forward neural network can be modelled by an ordinary differential equation, whose solution represents the path of the hidden and output layer weights.

This thesis is structured into six parts. To put the current work in prospective, chapter two outlines preview works, where machine learning and/or signal processing have been applied to anomaly detection for machines fault detection. Since the latter is the main goal, chapter three presents a general anomaly detection methodology, followed by a proposed methodology for predictive maintenance. The main methods used for anomaly detection, that is signal processing and machine learning are presented in chapter four and five respectively. The capstone of this work, to say the least, is revealed in chapter six in a form of three case studies, where we apply our proposed methodology to solve up to date anomaly detection problems encounter in a wide range of industries. As in any work, we conclude this thesis by summarising what have been done. The summary includes the strengths and weaknesses of our methodology, as well as what could be done to improve upon it.

2 Literature review (TO DO!!))

3 Anomaly detection methodology and methods (TO DO!!)

3.1 General anomaly detection methodology (TO DO!!)

3.2 A proposed methodology for predictive maintenance (TO DO!!)

4 Signal processing methods

4.1 Overview

In this section, we present three signal processing methods to generate new features also called attributes, for anomaly detection. In the context of this work, a feature or attribute is a numerical sequence that represents the state of a system. For example, voltage and vibration measurement taken from a motor over time are time series that we call features. These two features contain information about the state of the motor, the latter being the system in this case. By state we mean the current health of the motor. The health of the motor is represented by a health index which is a numerical value that quantifies the overall condition of the motor.

Generating new features from existing one, is a common procedure for anomaly detection. For example, by using Fourier transform, we can generate the frequency spectrum of a vibration time series. The frequency spectrum is the set of all frequencies from each vibration component of the original vibration time series. The importance of the frequency spectrum, lies in the fact that if a motor or machine is anomalous, the anomaly will generate an extra component in the vibration time signal, and the corresponding anomalous frequency will be visible in the frequency spectrum. Due to Fourier transform limitations, other signal processing techniques such as wavelet and Hilbert Huang transform are alternatives for generating new features to deal with more complex time series for anomaly detection.

In the rest of this section, we give an exposé of three signal processing methods, namely: Fourier analysis, wavelet transform and Hilbert Huang transform. For each transform, we cover the theoretical back bone in terms of mathematical constructs such as basis, vector space, orthogonality, existence, uniqueness, to name a few. Furthermore, these mathematical constructs will set the limitations as well as the strengths of each transform. Following the theoretical set up is a concrete application, where we generate new features and show how they can be used for anomaly detection.

4.2 Fourier Analysis

From solving differential equations to analysing sound wave, images and signal in general, Fourier analysis has a profound impact in science and engineering. It provides a convenient way to transform data from time domain to frequency domain, thus revealing unseen aspect of data. A time domain data can be viewed as a series of observations generated by a given process and recorded at discrete or continuous time interval. The underlying process might be the sum of subprocesses. In this case, the frequency domain will reveal all the subprocesses characteristics.

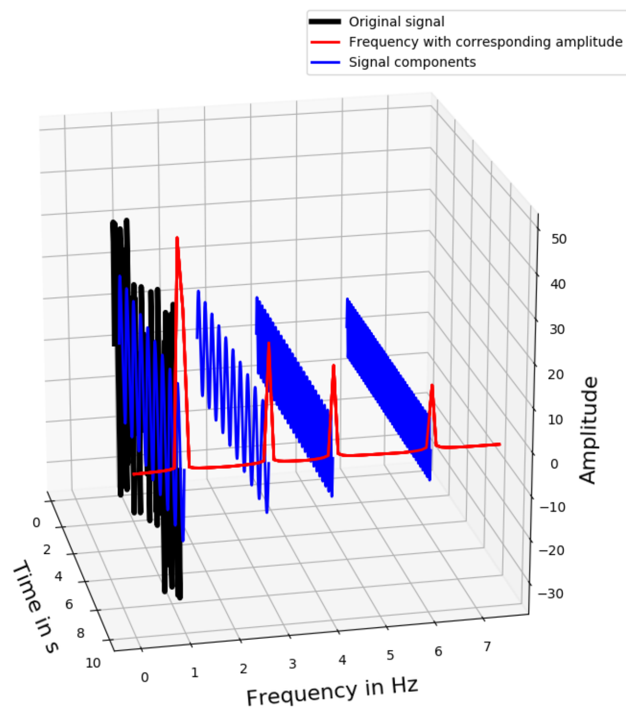

Figure 2:

Figure 2 Shows the time domain representation of the signal (in dark). The original signal is decomposed into four sub components (in blue) by using Fourier analysis. Each component frequencies and amplitude represent the frequency domain (in red).

Fourier analysis is concerned with the general problem of periodic and non periodic functions approximation. Fourier series on one hand addresses the former, while the latter is treated by Fourier transform. Given a periodic function, its Fourier series is given as a discrete superposition of exponential functions, and its Fourier transform is given by continuous superposition of exponential functions.

Fourier analysis is used in a wide range of application, including signal processing, data compression, image analysis. The main objective is to take a signal or more generally a function, and decompose its in trigonometric functions.

4.2.1 Fourier series

Looking at the bigger picture, Fourier series is concerned with the general problem of periodic functions approximation. The basic ingredients required to approximate a function in this scenario are: a vector space, a basis, which is a subspace of the vector space, and a mathematical operation, or more generally a function such as an inner product that maps two vectors to a real number. If a vector space has an inner product, we say that the vector space is an inner product space.

Before continuing, we see the need to clarify some abbreviations. We use the letters f, V, V_0 for an arbitrary function, a vector space, and a subspace of a vector space, respectively. Basis functions will be denoted by $\{\varphi_0, \cdots, \varphi_n\}$, where n can either be a finite integer or infinite. Having made this clarification, let explain the concept of function approximation.

The function approximation process in light of Fourier series goes like this: Given an arbitrary function f that we seek to approximate, we pick an appropriate vector space which we call V, such that $f \in V$. We define a subspace V_0 of the vector space V and construct an inner product on V_0, if it does not exist. Furthermore, we fine an appropriate basis of V_0. A basis of V_0 is a set of linearly independent vectors $\{\varphi_0, \cdots, \varphi_n\}$ in V_0, that span V_0. This means that any vector in V_0 can be written as a linear combination of the basis vectors. Once we have all this in place, the best approximation of the function f is its orthogonal projection in the inner product space V_0. Figure 3 shows an illustration of a generic mechanism of function approximation by orthogonal projection, where f_0 is the orthogonal projection of f in the subspace V_0 of V.

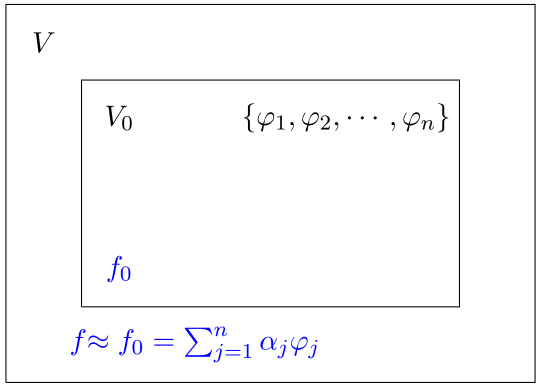

Figure 3: Illustration of a generic function approximation process of a function f into a subspace V_0 of a vector space V. φ_j are the basis functions and α_j are real numbers for $j = 1, \cdots, n$.

The Fourier space is a subspace of the space of all continuous functions of the interval $[0, T]$ and denoted by $C[0, T]$. Let $V_{N,T}$ be the subspace of $C[0, T]$ spanned by

$$\left\{1, \cos\left(\frac{2\pi t}{T}\right), \cdots, \cos\left(\frac{2\pi N t}{T}\right), \sin\left(\frac{2\pi t}{T}\right), \cdots, \sin\left(\frac{2\pi N t}{T}\right)\right\}. \tag{1}$$

The space $V_{N,T}$ is called the $N-th$ order Fourier series space. The Fourier series representation of an arbitrary periodic function f of period $T = 2L$ defined on an interval of length L is the best approximation of f in $V_{N,T}$. It is the the orthogonal projection of f into $V_{N,T}$ with respect to the inner product

$$<, f, g >= \frac{1}{T} \int_0^T f(t) g(t) \mathrm{d}t \tag{2}$$

Let f be an arbitrary periodic function of period $T = 2L$, defined on an interval of length L. Its Fourier series representation is

$$f(t) = \frac{a_0}{2} + \sum_{n=0}^{\infty} \left(a_n \cos\left(\frac{n\pi t}{L}\right) + b_n \sin\left(\frac{n\pi t}{L}\right)\right), \tag{3}$$

where the coefficients $a_0, a_1, \cdots, b_1, b_2, \cdots$ are given by

$$\begin{aligned} a_m &= \frac{1}{L} \int_{-L}^{L} f(t) \cos\left(\frac{m\pi t}{L}\right) \mathrm{d}t, \quad m = 0, 1, 2, \cdots \\ b_n &= \frac{1}{L} \int_{-L}^{L} f(t) \sin\left(\frac{n\pi t}{L}\right) \mathrm{d}t, \quad n = 1, 2, \cdots \end{aligned} \tag{4}$$

Theorem 1 *Convergence.*

Suppose that f and its derivative f' are piecewise continuous on the interval $-L \leq t \leq L$. Suppose also that f is defined elsewhere so that it is periodic with period $2L$. Then f has a Fourier series and associated coefficient given by (3) and (4). Furthermore, the Fourier series converges to $f(t)$ where f is continuous, and to

$$\frac{1}{2}\left(\lim_{t \to c^-} f(t) + \lim_{t \to c^+} f(t)\right) \tag{5}$$

at every point c where f is discontinuous.

In vu of Theorem 1, a periodic function and its Fourier series are equal if f is continuous on $-L \leq t \leq L$. If f is discontinuous on $-L \leq t \leq L$ then f and its Fourier series disagree at the discontinuities.

4.2.2 Fourier Transform

- General formulation
- Existence
- Uniqueness
- Convergence
- Stability

4.2.3 Fourier transform

4.3 Wavelet transform

4.3.1 Theory

4.3.2 Application

In the wavelet transform we generate two extra features from the vibration time signal, namely the discrete detailed coefficient cD and the approximate coefficient cA. The detail coefficient cD represents the hight frequency component of the vibration time signal and the approximate coefficient cA represents the low frequency component. For the mother wavelet we use Daubechies 20 or db20 shown in Figure 4

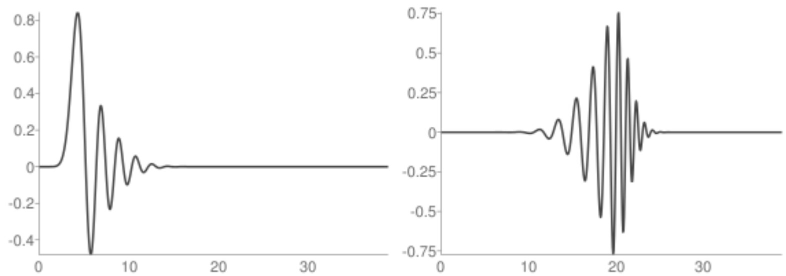

Figure 4: The scaling function φ (left) and the wavelet function Ψ (right)

Once we have the two extra features, we compute the dissimilarity between a reference sample and subsequent samples for each feature. This process generates a set of points (x, y) that represent the health index of each sample. From Figure 5 and 6, we can observe that bearing number four and bearing number three suffer from ball pass frequency outer race and ball pass frequency inner race defect respectively. We can also observe that a bearing can go through three main stages:

1. A healthy stage characterised by a low health index
2. A warning stage characterised by an increasing health index
3. An alarm stage characterised by a high health index

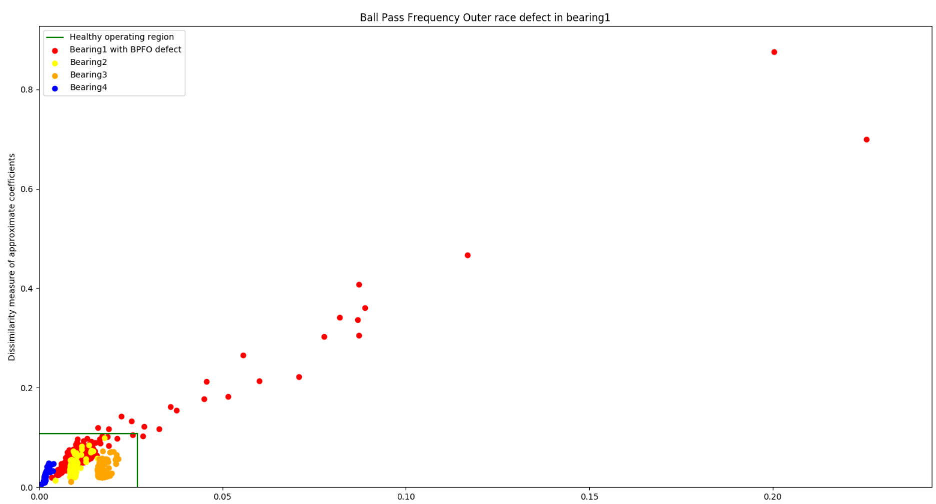

Figure 5: Ball past frequency outer race defect detection from wavelet transform

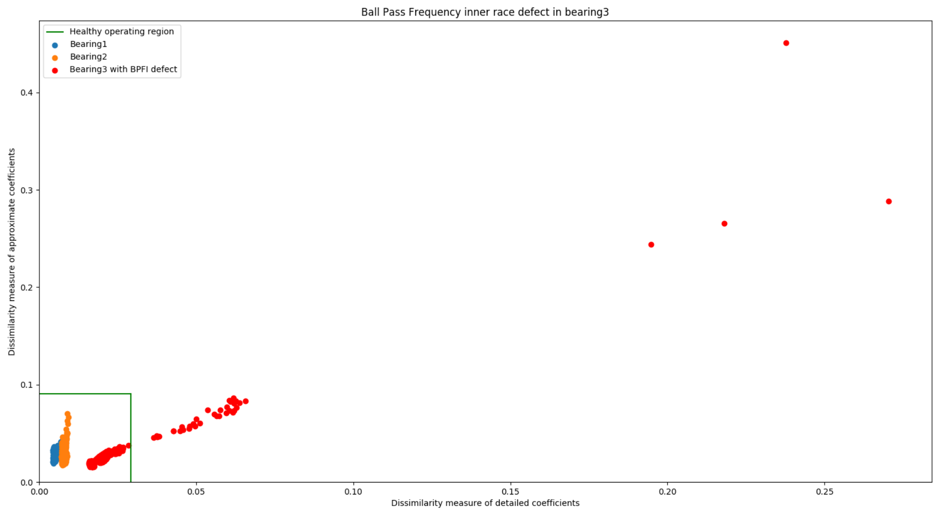

Figure 6: Ball past frequency outer race defect detection from wavelet transform

In the alarm stage, as the degradation becomes more severe, the distance between the points increases.

4.4 Hilbert Huang transform

4.4.1 Theory

The Hilbert-Huang transform is a data decomposition methods that consists of decomposing data in an adaptive fashion. Adaptivity means that rather than imposing an a priory basis such as trigonometric functions, a posteriori basis functions are derived from the data itself [1]. In doing so, the method deals better with nonlinearity and non stationarity which are inherently present in real world data.

This method gives an alternative approach of time-frequency-energy paradigm by using Hilbert spectral analysis and the so call empirical mode decomposition (EMD) to express the nonlinearity and the non stationary in data with instantaneous frequency and instantaneous amplitude [1].

The empirical mode decomposition (EMD) originated from the quest of functions that can be expressed by a time-frequency-amplitude expression, such that the frequency is physically meaningful. Consider a time series $x(t)$. Its Hilbert transform H(t) is given by

$$H(t) = \frac{1}{\pi} P \int_{\infty}^{\infty} \frac{x(\tau)}{t-\tau} \mathrm{d}\tau, \tag{6}$$

where P is the Cauchy principal value. The corresponding time-frequency-amplitude function of $x(t)$ is the analytical function

$$z(t) = x(t) + iy(t) = a(t)e^{i\theta(t)}, \tag{7}$$

where the instantaneous amplitude $a(t)$ and phase $\theta(t)$ can be computed by

$$a(t) = \sqrt{x(t)^2 + y(t)^2} \tag{8}$$

$$\theta(t) = \tan^{-1}\left(\frac{y(t)}{x(t)}\right). \tag{9}$$

Furthermore, the instantaneous frequency $w(t)$ can be derived from the phase $\theta(t)$ as

$$w(t) = \frac{d\theta}{dt}. \tag{10}$$

By setting

$$f(t) = \frac{y(t)}{x(t)},$$

the expression of the instantaneous amplitude $w(t)$ in (10) can be expanded as

$$w(t) = \frac{f'(t)}{1+f(t)^2} = \frac{y'(t)x(t) - y(t)x'(t)}{x(t)(x(t)+y(t)^2)}. \tag{11}$$

The instantaneous frequency $w(t)$ using the Hilbert transform is not always physically meaning. For example for an arbitrarily function, the instantaneous physical frequency values should be positive. However this is not always the case.

For example if

$$f(x) = \cos(ct) + d \tag{12}$$

where c and d are constants, the instantaneous frequency is given by

$$w(t) = \frac{-c\sin(ct)}{1+(\cos(ct)+d)^2} \tag{13}$$

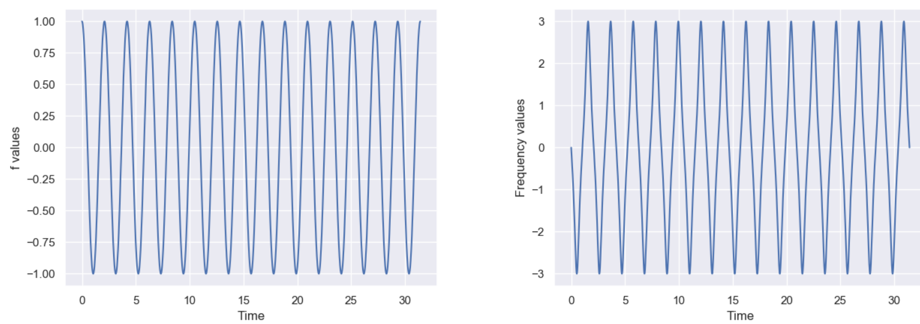

Figure 7: function f and its corresponding frequency

From Figure 7, we see that the instantaneous frequency takes negative values, which is not physically meaning full.

To circumvent this, the Hilbert-Huang transform offers a methodology to obtain from an arbitrarily function or time series $x(t)$ a set of finite subcomponents whose instantaneous frequency are physically meaningful. This methodology let to the empirical mode decomposition.

The necessary condition for obtaining a physical frequency is that $x(t)$ satisfies the approximate local envelope symmetry condition [2].
This condition is expressed in the empirical mode decomposition (EMD) such that an arbitrarily time series $x(t)$ can be decomposed by a sifting process into intrinsic mode function c_i

$$x(t) = \sum_{i=1}^{n} c_i + r_n \tag{14}$$

where the c_i satisfies the approximate local envelope symmetry condition

$$SD_k = \frac{\sum_{t=0}^{T}}{\sum_{t=0}^{T}} < \epsilon \tag{15}$$

where ϵ is a small predefined real number.

4.4.2 Application for bearings fault detection

we consider a vibration signal with sample frequency of 20000Hz rotating speed of 2000 RPM

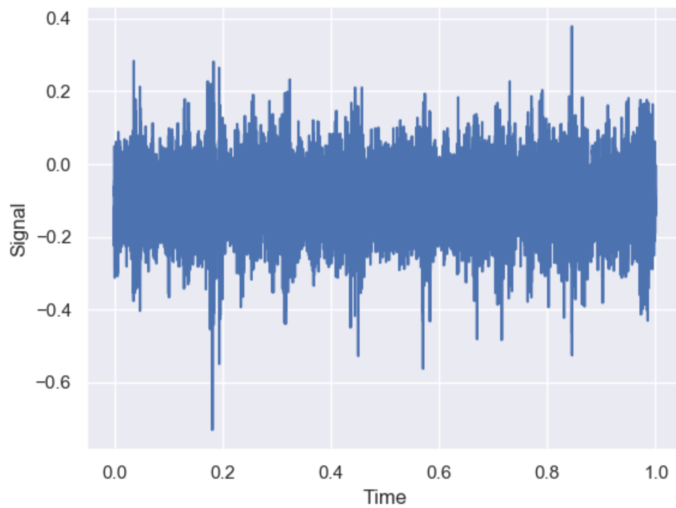

Figure 8: Vibration signal of 1 second snapshot

After applying the empirical mode decomposition on the vibration data from figure 8 we get sixteen intrinsic mode functions

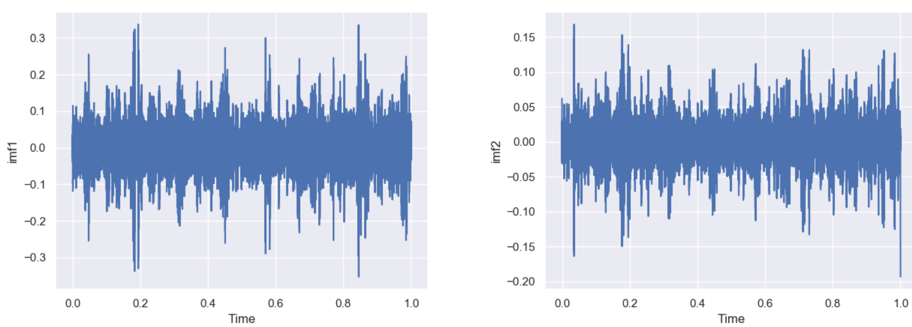

Figure 9: 1th and 2nd intrinsic mode function (imf)

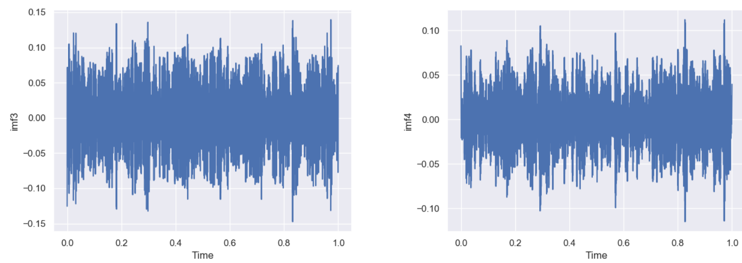

Figure 10: 3rd and 4th intrinsic mode function (imf)

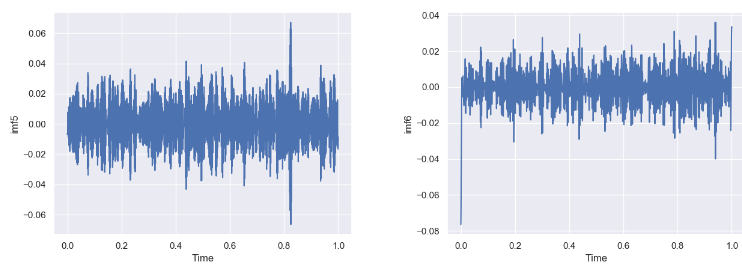

Figure 11: 5th and 6th intrinsic mode function (imf)

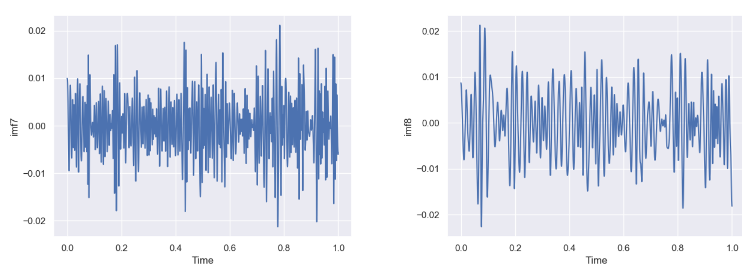

Figure 12: 7th and 8th intrinsic mode function (imf)

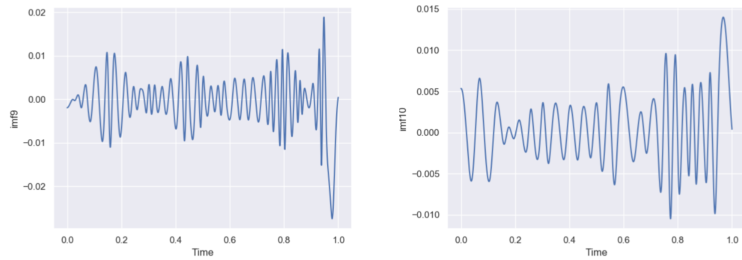

Figure 13: 9th and 10th intrinsic mode function (imf)

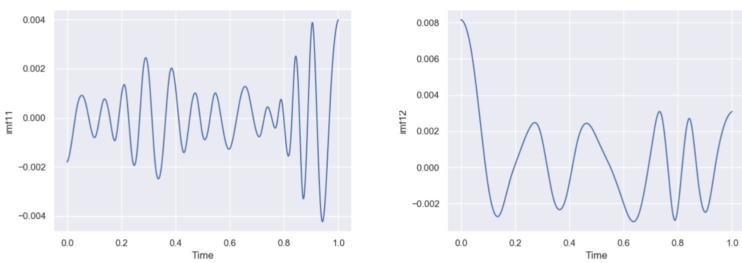

Figure 14: 11th and 12th intrinsic mode function (imf)

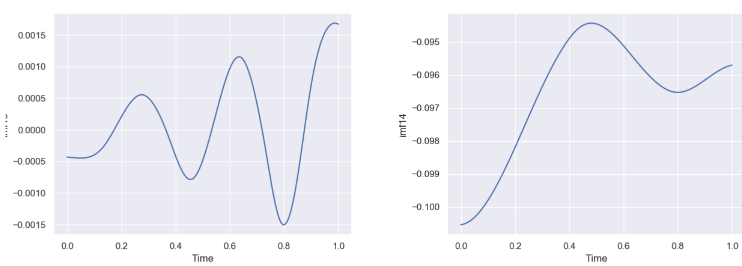

Figure 15: 13th and 14th intrinsic mode function (imf)

5 Machine learning methods (TO DO!!)

5.1 Overview (TO DO!!)

6 Machine learning and signal processing: Case study

6.1 Overview

In this section we show three case study where we use our propose methodology for anomaly detection in predictive maintenance.

6.2 Bearing fault detection

6.2.1 Problem formulation

Four bearings installed on the shaft of a motor exhibit different health conditions. The goal is to be able to use a machine learning model to detect the fault and prevent subsequent failures. Figure 16 shows the experimental setup with two accelerometers installed per bearing, measuring radial and axial vibration. An accelerometer is a sensor that measures the vibration of a given system. The radial vibration is the vibration experienced by the shaft perpendicular to it axis, while the axial vibration is the vibration experienced by the shaft along it axis.

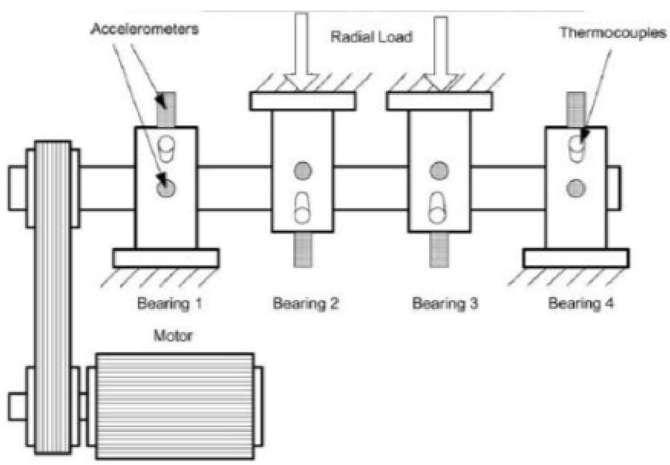

Figure 16: Experimental setup from [4]

The shaft is subjected to a radial load of 6000 Pound-force and the motor is rotating at 2000 RPM (Rotation Per Minute) or 33.33 Hz [5]. The bearings used are of type Rexnord ZA-2115.

6.2.2 Conceptual model

Figure 16 shows a very good description of the system and can serve as a conceptual model. Recall that the conceptual model is a diagram of the system explaining its behaviour. In this case the four bearings mounted on the shaft are subjected to a radial load of 6000 Pound-force or equivalently 26689.32 Newton. As the motor is rotating, this axial load will be will eventually cause the bearing to fail after a period of time.

We are interested in detecting failure in the four bearings if they occur or as early as possible. The common failure type are inner race defect, outer race defect, roller element defect, cage defect. Figure 17 shows a schematic description of a bearing. The inner ring is in contact with the shaft. The cage holds in place the roller elements which are spherical elements. As the shaft rotates, the roller elements, the cage, the inner race and the outer race are all under load.

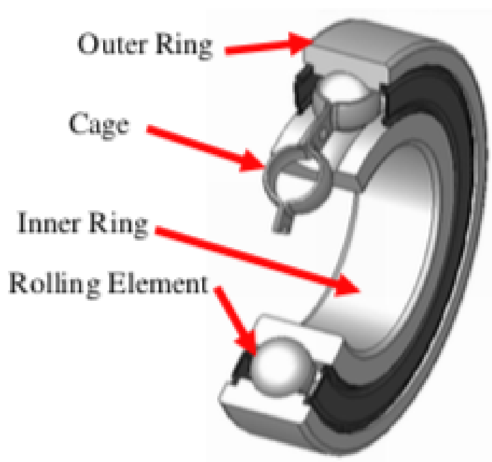

Figure 17: Schematic description of a bearing. Taken from [ref]

For inner race defect, a crack occurs in the inner ring, while for outer race a crack occurs at the outer ring. Similarly, cage and rolling element defect are defect happening at the cage and roller elements.

6.2.3 Data description

Three datasets resulting from three different experiment are available. In the first experiment run from October 22, 2003 at 12:06:24 to November 25, 2003 at 23:39:56, each of the four bearing axial and radial vibration were measured every 10 minutes. At the end of the experiment, bearing 3 exhibited inner race defect while bearing 4 exhibited roller element defect.

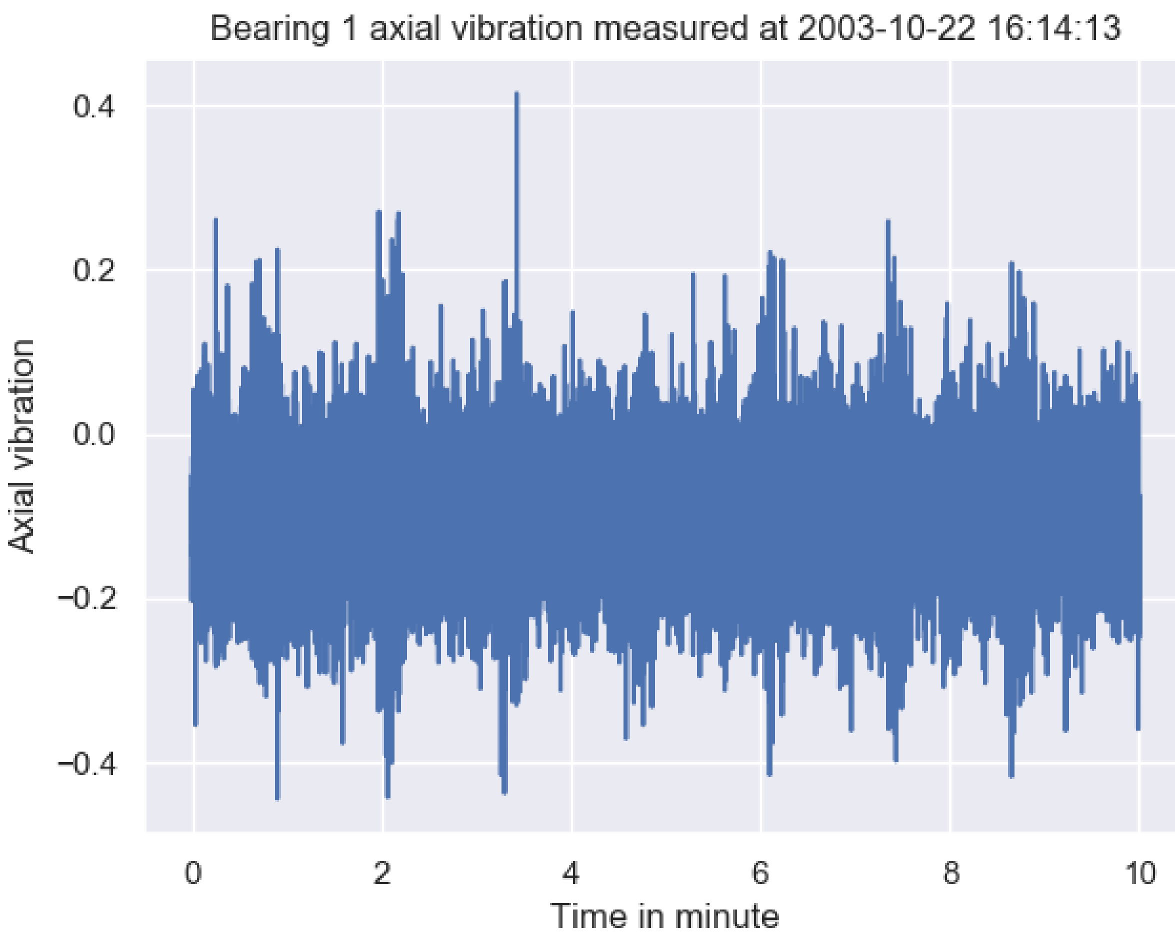

Figure 18: Axial vibration measurement for the first test

The second experiment run from February 12, 2003 10:32:39 to February 19, 2004 06:22:39. Radial acceleration were measured on bearing 1, 2,3 and four. The measurement were done every 10 minutes. At the end of the experiment, outer race defect occurred in bearing 1.

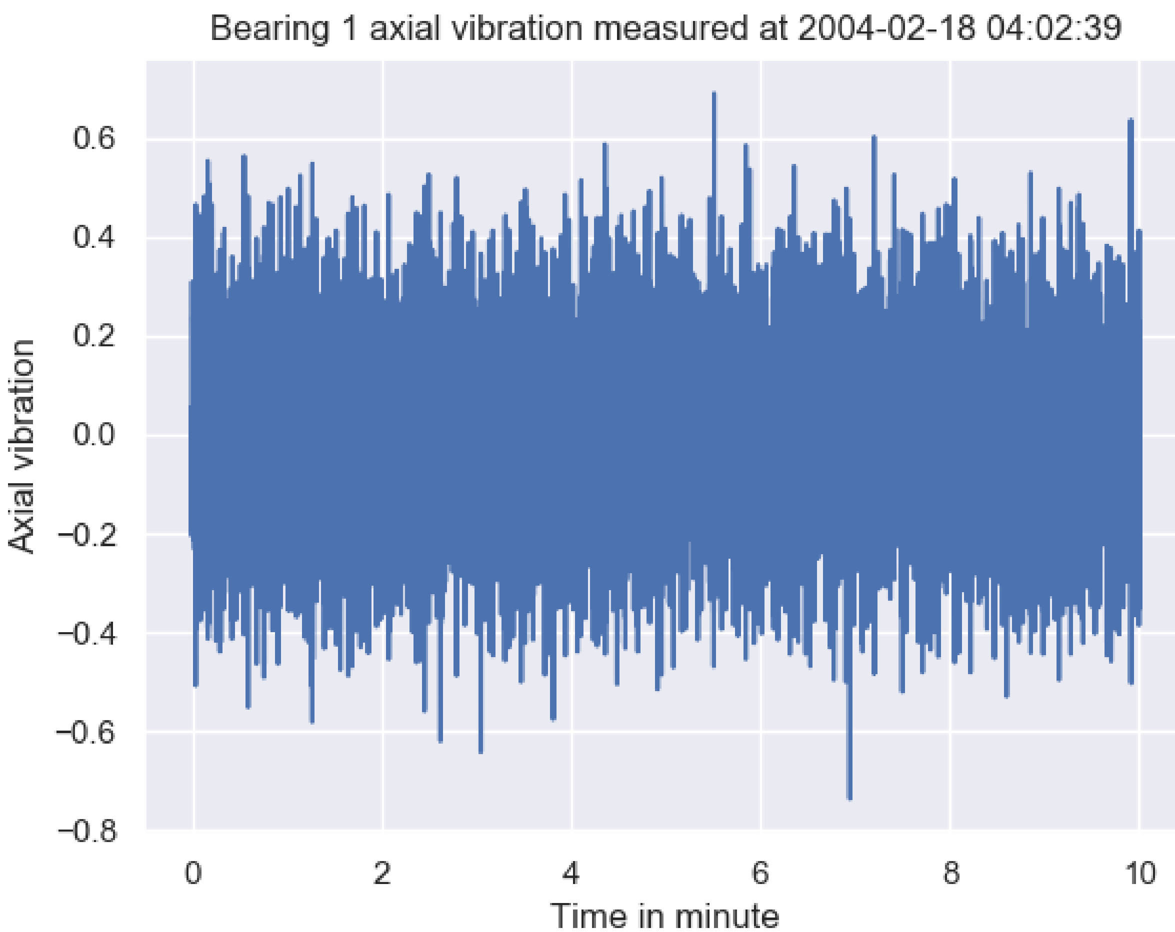

Figure 19: Axial vibration measurement for the second test

The third experiment run from March 4, 2004 09:27:46 to April 4, 2004 19:01:57 and the measurement were done every 10 minutes. At the end of the experiment, outer race defect occurred in bearing 3.

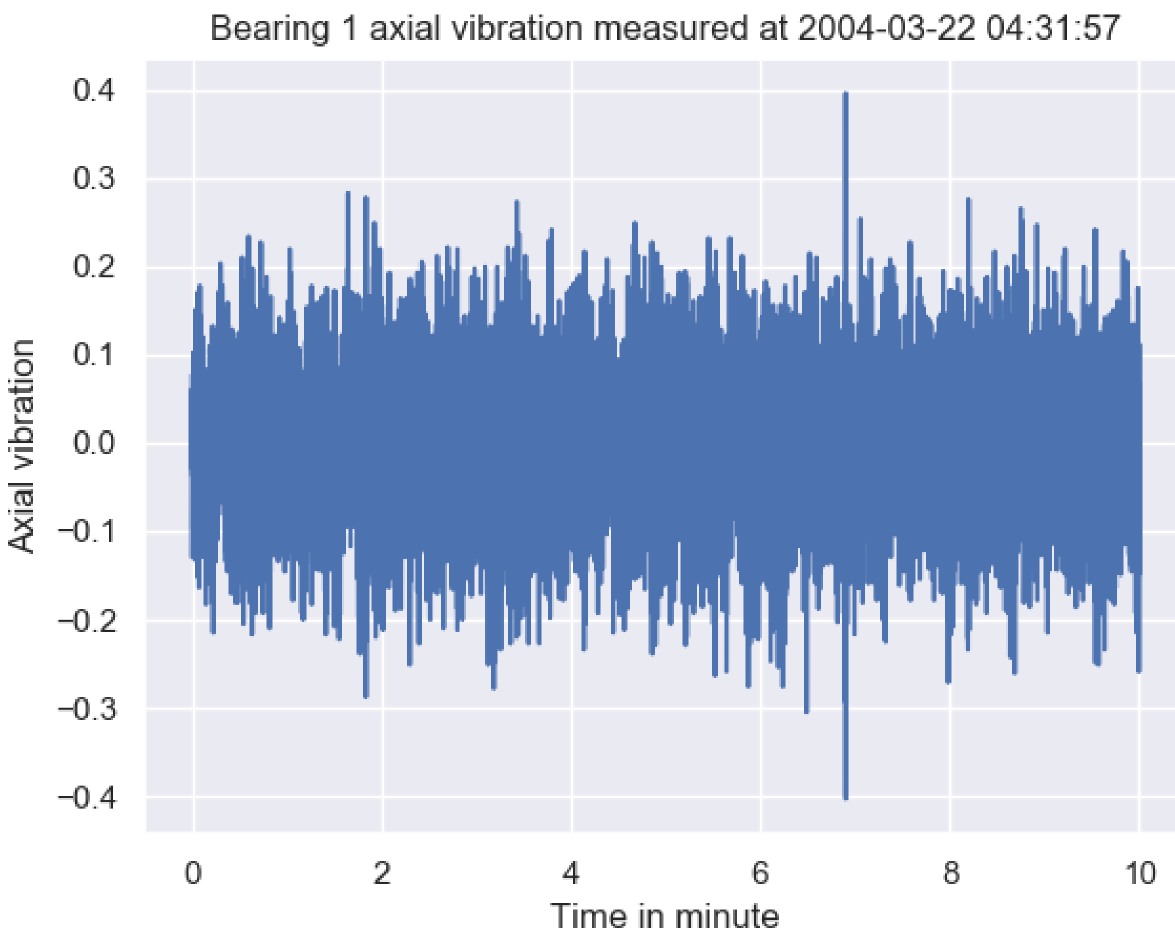

Figure 20: Axial vibration measurement for the third test

6.2.4 Data driven model

The objective now is to apply Fourie

7 Conclusion (TO DO!!)

References

[1] Norden E. Huang and Zhaohua Wu *A review on Hilbert-Huang Transform: Methods and its Application to Geophysical Studies*. Review of Geophysics, 2008

[2] Norden E. Huang, Zheng Shen, Steven R. Long, Manlic C. Wu, Hsing H.Shih, Quanan Zheng, Nai-Chyuan Yen, Chi Chao Tung and Henry H. Liu *The Empirical mode decomposition and the Hilbert Spectrum for nonlinear and non-stationary time series analysis* Proc. R. Soc. Lond. A (1998) 455, 903:995.

[3] Varun Chandola, Arindam Banerjee, and Vipin Kumar. *Anomaly detection: A survey* Proc. R. Soc. Lond. A (1998) 455, 903:995.

[4] Hai Qiu, Jay Lee, Jing Lin *Wavelet Filter-based Weak Signature Detection Method and its Application on Roller Bearing Prognostics* Journal of Sound and Vibration 289 (2006) 1066-1090

[5] *IMS Bearing Data*

[6] Michel Mulders and Mark Haarman *Predictive maintenance 4.0. Predict the unpredictable, June 2017*

[7] Hui Li, Yuping Zhang and Haiqi Zheng *Hilbert-Huang transform and marginal spectrum for detecting and diagnosis of localized defects in roller bearings* Journal of Mechanical Technology, 2007