FSM Online Internship Report

Project Title: Remaining Usable Life Estimation(Bearing Dataset)
Domain: Machine Learning
Phase2: Feature Engineering & Cleaning

Submitted by:
Akanksha Singh
Thapar Institute Of Engineering & Technology
Patiala, Punjab - 147004

Under Mentorship of:
Devesh Tarasia

IITD-AIA Foundation for Smart Manufacturing

[1-June-2022 to 31-July-2022]
Remaining Usable Life Estimation (Bearing Dataset)

DATA PREPROCESSING includes:
Data Filtering is the process of choosing a smaller part of your dataset and using that subset for viewing or analysis. Filtering is generally (but not always) temporary (= lasting only a short while, short-term, non permanent, provisional) – the complete dataset is kept, but only part of it is used for the calculation)
Data Ordering or Data Sorting is the process of arranging data into meaningful order so that we can analyze it more effectively

In data exploration part, we concluded that due to the loss of trend in the vibrational data, it was impossible to perform de-noising (smoothing) & consequently – RUL prediction directly for vibration parameters. So, instead of directly using measured vibration values, we proposed to use accumulated values. In current statistics, the vibration values are measured by means of an accelerometer as acceleration, in unit "g".
Physically the power at the moment t (both for platform and balls) is proportional to the acceleration, and current (instantaneous) degradation is proportional to the power.

There are two possibilities to handle measured Horizontal and Vertical Vibration parameters:
- Use an Integral value and perform accumulation for the integral value of the vibration – instead of the measured value of the Horizontal and Vertical Vibrations.
- Use one of the wide-known Machine Learning methods for RUL prediction (*for example, SVR – Support Vector Regression, RVM – Relevance Vector Machines, etc.)* for obtained multi-parameter data – after performing an accumulation separately for the Horizontal and Vertical Vibrations.

Generally speaking, the second approach can provide better accuracy, but it simultaneously leads to over-fitting due to the small number of training bearings in each of the operating conditions groups (two bearings per group)..

Features were extracted from vibration signals of the bearings in the time and frequency domains.
The time domain featured included root mean square (RMS), peak, crest factor, and kurtosis of the vibration signals. The frequency domain features included the magnitude at bearing defect frequencies.
In a preliminary analysis of the vibration signals of the six training bearings, we found that neither the time domain features nor the frequency domain feature exhibited a consistent trend of bearing degradation.
Time Domain analysis and frequency domain analysis; there are two such tools that can provide invaluable signal insight if properly used.A time-domain graph can show how a signal changes with time, whereas a frequency-domain graph will show how much of the signal lies within each given frequency band over a range of frequencies.

METHODOLOGY TRIED ADOPTING FOR FEATURE EXTRACTION

Finding Start Prediction Time (SPT)
In data-driven methods for bearing prognostics, there are several unresolved issues, such as deciding the time to start prediction (TSP), handling random anomalies in the measured data or features extracted from it, and determining the failure threshold. The proposed approach selects the best regression model to approximate the degradation behavior of a bearing based on the evolving trend in its health indicator. The approach has 2 distinct phases, TSP detection, and RUL estimation. The TSP indicates the onset of bearing degradation, and until it is detected the proposed algorithm does not estimate the RUL of the bearing. But due to its complicated algorithm and execution it is not adopted . [1]

METHODOLOGY ADOPTED FOR FEATURE EXTRACTION

Data-Driven approaches calculate the Health Indicator (HI) from sensor data like vibration signals & with machine learning methods and deep learning methods. HI differentiates a healthy operation & from a faulty operation of a machine. The features of the signals should be thoroughly investigated before extracting data from them. Most data analysis can be performed in two ways – time domain analysis & frequency-domain analysis. In time-zone analysis, vibration signals can be examined & determined based on time. Frequency domain analysis monitors vibration signals based on the frequency. With time-field analysis, the variation of vibration signals over time is visualized. Features derived from time-domain analysis perform well and are accurate for stationary signals. Although these features respond quickly to random variations within vibration signals. Features of frequency domain analysis can correctly interpret vibration signals because they determine more insight about signals than time-zone analysis features. Characteristics of frequency-domain analysis can assess and separate frequency parts, which helps to understand more insights than time-domain analysis. Mechanical bearing vibration signals are generally variable and at the same time these signals have undetermined faults in high noise environments. Therefore time-frequency domain analysis is an appropriate method to examine mechanical bearing vibration signals because the vibration signals of the mechanical bearing are interconnected. Features derived from time-frequency domain analysis have properties of both time components & frequency components.

Here we extract the feature from Bearings Vibration Data by applying Continuous Wavelet Transform (CWT). Continuous Wavelet Transform is a transformation technique that transforms the signals into 2D Images (time-frequency features). Processing time is less with Wavelet transform for extracting Time-Frequency Features as compared to other methods such as Empirical Mode Decomposition. After converting the signal Data Normalization technique is applied to normalize the coefficients of 2D CWT data. To achieve Data Normalization we performed the Data Re-scaling method a.k.a min-max normalization to re-scale or adjust all 2D data to a particular range (say 0 to 1). Signal Processing visualizes, evaluates, & controls vibration signals. Here we extracted 2D features i.e. time-frequency domain features because time-frequency features contains more information about the vibration signals, in regression problem like Remaining Useful Life Prediction, more data means fast & easy analysis.

CWT (Continuous Wavelet Transform):

- The purpose of CWT is to generate a time-frequency depiction of a vibration signal which delivers an accurate positioning of both time & frequency of the signal.
- It is also used for computing the vibration signal varying features.
- The signals are represented as wavelets in Wavelet Transform.
- The CWT transforms the signals into wavelets
- Wavelets are wave-shaped vibrations of magnitude which initiates at zero, progresses & later reduces again to zero
- The wavelets produced from CWT provide well-defined and comprehensible interpretation of time & frequency components which helps in understanding the bearing deterioration process clearly.
- The CWT contains several in-built CWT wavelets like morlet wavelet, gaussian derivative wavelet, frequency b-spline Wavelet, Mexica hat wavelet, Shannon wavelet, complex morlet wavelet, etc. which can be accessed using python programming.
- And among them, we used morlet wavelet python implementation i.e. CWT Morlet PyWavelet as it shows better outcomes for regression problems & compared to other wavelet types & Fourier transform.
- Theoretically, the Morlet wavelet works well for time-frequency vibration signals.
- In the model, the CWT Morlet PyWavelet has been applied to reform the recorded 1D vibration signals of PROGNOSTIA Bearing Learning Dataset into 2D CWT image-like features or time-frequency domain image features.
- It is quite rough to interpret the signals in only one dimension (time), so converted the signals from 1D to 2D.

- 2D signals i.e. 2D CWT image features carry the information about the signals from both time & frequency domains which helps in visualizing the damage process of bearings perfectly.

IMPLEMENTATION

1. The entire feature engineering & cleaning is performed in a Collaboratory platform using python programming
2. After the data exploration and data merging the 6 pkz files are read where Continuous Wavelet Transform with Morlet Wavelet on 1D vibration signals was performed on 1D vibration signals which returned the 2D CWT feature images
3. Normalization was performed on the obtained feature images to a specific range, say 0 to 1
4. Then signal processing was applied to extract the time-frequency features which helps in differentiating the healthy operation and the faulty operation of mechanical bearing

5. Since or dataset is a run-to-failure dataset, at starting the functionality of bearing is in good condition & eventually fails while progressing.

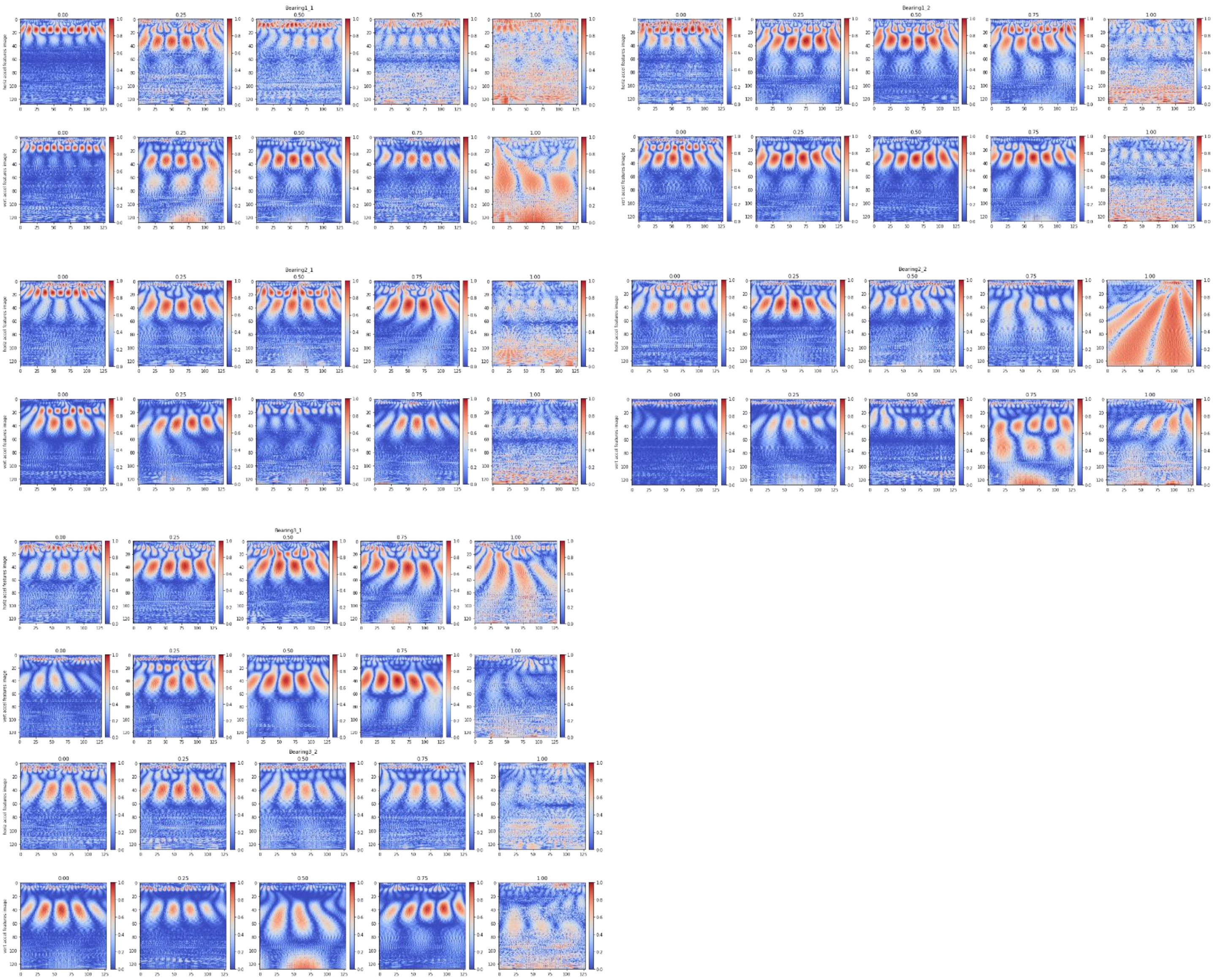

Fig1: 1D to 2D Continuous Wavelet Transform

Miscellaneous Observations in During Feature Extraction and research papers:

- The root mean square (RMS) is the most appropriate indicator of bearing health because it shows the trends of different health statuses of the bearing: in the normal stage, the RMS remains stable; in the initial degradation stage, the RMS starts linear growth; in the severe degradation stage, the RMS starts nonlinear growth, but there are some outliers caused by random noises in the RMS.

- The spurious fluctuation in RMS has a great impact on the RUL prediction performance.

- By reviewing the already done research work, we infer that RUL is a regression problem. And hence RUL prediction mechanical element (mechanical bearing) can be calculated using that mechanical bearing's previous performance data.

Code Implementation :
https://colab.research.google.com/drive/1-nIjz5iDKyPvWZVTThP2VLvF7O7Fh-Zw?usp=sharing

REFERENCE
[1] Wang, X.; Qiao, D.; Han, K.;Chen, X.; He, Z. Research on Predicting Remain Useful Life of Rolling Bearing Based on Parallel Deep Residual Network. *Appl. Sci.* **2022**, *12*, 4299. https://doi.org/10.3390/app12094299