RUL Estimation of an Aircraft Turbine Engine

Abstract

RUL (Remaining useful life) estimation is the most common task in the research field of prognostics and health management. The data driven approach relies on the availability of run-to-failure data, based on which RUL can be estimated. The task of this project was to estimate the RUL following the similarity based approach[1] and to come up with a new novel method that can be used for RUL estimation.

Methodology

1. Calculating Health Index

The multi-dimensional sensor readings are first fused to produce a single Health Indicator (HI). We use linear multivariate regression model for estimating the HI, where x is input and i denotes the the ith sensor data .

$$y = \alpha + \sum_{i=1}^{N} \beta_i x_i + \epsilon \qquad ----(1)$$

A Linear Multivariate Regression(LMR) model is trained, one model for each of the six different operating region and is stored for transforming the input sensor values of the test data.

2. Experimental Data Description

The data set, provided by the 2008 PHM Data Challenge Competition, consists of multivariate time series that are collected from multiple units of an unspecified component. Each time series is from a different instance of the same complex engineered system, e.g., the data might be from a fleet of ships of the same type. There are three operational settings that have a substantial effect on unit performance. The data for each cycle of each unit include the unit ID, cycle index, 3 values for the operational settings and 21 values for 21 sensor measurements. The sensor data are contaminated with noise.

Each unit starts with different degrees of initial degradation and manufacturing variation which is unknown. This degradation and variation is considered normal. The unit is operating normally at the start of each time series, and develops a fault at some point during the series.

3. Operational Region Partition

Based on Operational settings the dataset can be divided into 6 modes, the experimental data consists of three settings, namely ‘Altitude’, ‘MachNo’ and ‘TRA’. We use ‘TRA’ settings to divide the dataset set into six different regions based on its distinct values of 0,20,40,60,80,100. This is done so that we preserve the behavior of the machine in different settings and train six different models for better HI calculation.

4. Sensor Selection

Most of sensors with continuous values exhibit a monotonic trend during the lifetime of the units. However, some of them show inconsistent end-life trends among the different units in the training data set which might indicate, for example, different failure modes of the system. It might be possible to first classify the units by failure modes based on these sensors and then process them using different prediction models; this strategy, however, will encounter two challenges. First, the end-life readings of these sensors spread out over a large range, which make it hard to quantize the failure modes without extra information. Second, the failure modes might not be unambiguously identifiable, if not completely indiscernible, at the early age of a unit, and thus might contribute little to RUL estimation when only early history of the unit is available. Therefore, only those continuous-value sensors with a consistent trend are selected for further processing. These sensors are indexed by 2, 3, 4, 7, 11, 12, 15

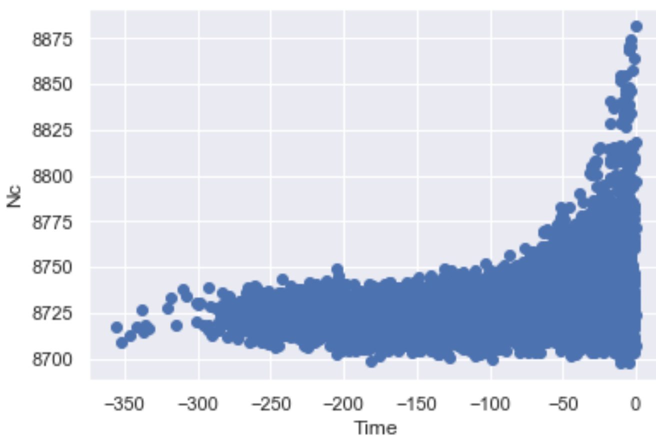

Fig1: Nc vs Time

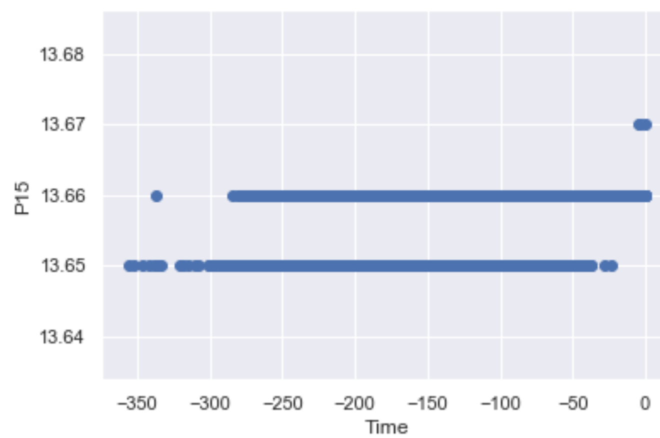

Fig2: P15 vs Time

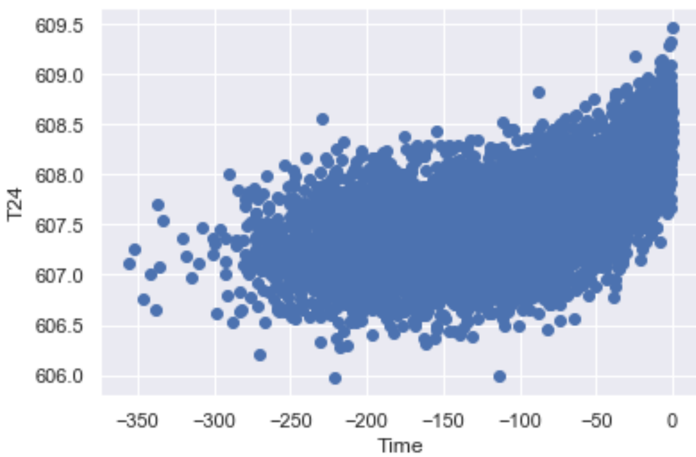

Fig3: T24 vs Time

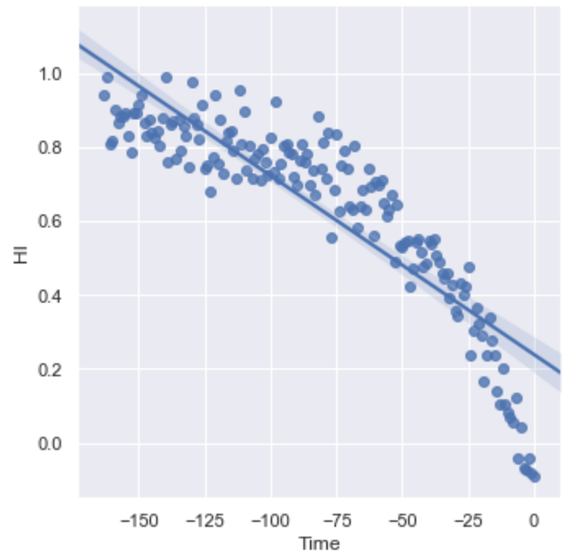

Fig4: Transformed HI time series data for unit 2 using the LMR Models

The above plot shows that the two sensors do not have a consistent trend during the lifetime and it becomes very difficult and complex to clear identify failure of these sensors at any given time, to simplify we consider the sensors that show consistent trend during failure and healthy time. See Fig3 for example. The seven sensors selected above show exponential trend makes it easy to identify the knee point for the start of the degradation in the engine.

5. Model Training:

We train six different LMR models on the selected sensors and obtain the unknown coefficients of equation 1, for this purpose we need to identify some healthy and unhealthy samples from the training dataset that can basically supervise the learning of the model, so we use the last five samples of each unit as having HI '0' meaning they have completely failed which is also known beforehand and we assume that those samples that correspond to time index before -300 are completely healthy having HI '1'. There is a scope for training non-linear model for the same which is yet to be explored .

6. Model Identification

The exponential (nonlinear) regression models are used to describe the relationship between the adjusted cycle index *Time* and the HI y, as it can be clearly seen the pattern that best fits the degradation of unit 2 in Fig4 is exponential and not linear.

```
[<matplotlib.lines.Line2D at 0x11b659b00>]
```

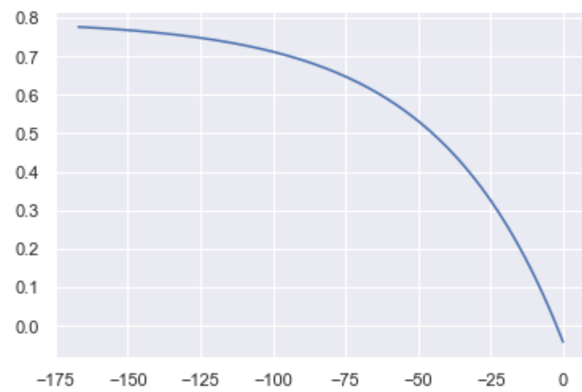

Fig5: Fitted exponential curve for the above pattern

A set of such degradation models is fitted and stored for each of the 218 aircraft in the training dataset.

7. RUL Estimation

For the test data, first it is classified based on operating regime and then transformed to HI values using the LMR models trained on the training dataset, then the similarity is measured between the set of models from the previous step and the test unit, it is used for estimating the RUL.

Proposed Method for calculating HI

One of the most important task in estimating RUL is to first transform the multiple sensor data to a corresponding HI value, which is an indicator of the degree/extent of the fault in the system. In this paper, Linear Multivariate Regression(LMR) is used to transform the multiple sensor data to HI, there is no clear indication of independency between sensors, so performance of LMR is not best. An alternative way to ensure that the features become independent of each other is by first applying PCA on the multiple sensor data to obtain principal components that are independent on each other and hence improve the performance of LMR.

The R2 score is obtained for the LMR, the score obtained for LMR on the dataset after applying PCA is higher compared to directly applying LMR on the dataset. Thus, this approach can give better end RUL estimation.

Proposed Alternative Solution

Sensor selection is an important task, some sensors are meaningful in RUL estimation and some are not. The sensors that are relevant to the RUL estimation are the ones that have some degradation introduced during the creation of the dataset, for this challenge the only sensors that have degradation introduced into them are the ones that correspond to HPC, so the sensors 3,7,11,12 should be selected for the RUL estimation instead of selecting the 7 sensors above.

Followed by the sensor selection, the sensor vs Time can be analyzed to see if there is some pattern/equation that fits the sensor degradation. As we know already that the last time cycle of each unit correspond to failure, so we can try to identify the failure pattern in each selected sensor and the corresponding range of values the sensor might give during healthy and failure modes respectively, so given the sensor value from the test set we can first comment if there is any failure or degradation in that sensor value or not. The difficulty here is that there might not be any fixed pattern of failure and also different failure/degradation might affect the sensor value differently which makes it difficult to identify a fixed failure pattern for the sensor, if the failure pattern of sensor can be modeled under some assumptions within certain error limits, then we can predict the remaining life within some error given the sensor value. The predicted RUL for each selected sensor can be weighed on a scale between 0 to 1, depending on how close that sensor value is to complete failure, thus if it is far away from failure it will have zero weightage and if it is very near to failure it will have 1 weightage, now all the weighted RUL from the selected sensors will be used to predict the final RUL.

References

1. A similarity-based prognostics approach for Remaining Useful Life estimation of engineered systems, Tianyi Wang , Jianbo Yu , David Siegel , Jay Lee