Project I (EE47009) report on

Similarity based RUL Estimation

submitted by

Abhishek Patwari

Roll No: 16EE10002

Final Year Undergraduate Student

Department of Electrical Engineering

under the supervision of

Professor Debasis Samanta

Department of Computer Science & Engineering

Department of Electrical Engineering

Indian Institute of Technology, Kharagpur,

WB 721302, India

CONTENTS

1. 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. One method for this is to first convert the multiple sensor input data to a health indicator(HI), which will indicate the patterns in the dataset for the aircraft engine, here exponential degradation founds the best fit to the HI data and is used to estimate RUL based on similarity of the test data and the set of fitted exponential patterns from the training data, Linear Multivariate Regression is used for finding the HI from the input data.

2. INTRODUCTION

i) Run-to-failure historical data from multiple units of a system/component are recorded
ii) The historical data covers a representative set of units of the system/component;
iii) The history of each unit ends when it reaches a failure condition, or a preset threshold of undesirable conditions, after which no more runs will be possible or desirable.

Then, a library of degradation patterns can be created from these units with complete run-to-failure data (called training units). A unit whose remaining life will be predicted (called a test unit) also has its historical data recorded continuously. Instead of fitting a curve for a test unit and extrapolating it, the data will be matched to a certain life period of certain training units with the best matching scores. Finally, the RUL of the test unit can be estimated by using the real life of the matched training units minus the current life position of the test unit. The remaining life of a test unit is estimated based on the actual life of a training unit that has the most similar degradation pattern**.**

3. METHODOLOGY

3.1. 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 .

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

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.

3.2. 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.

3.3 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.

3.4. 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.

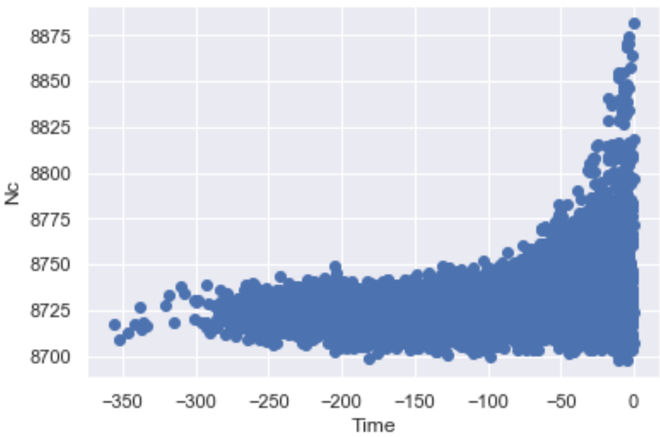

Fig1: Nc vs Time

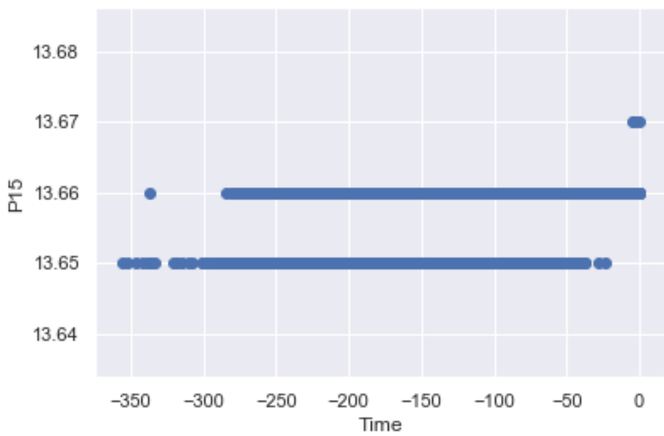

Fig2: P15 vs Time

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.

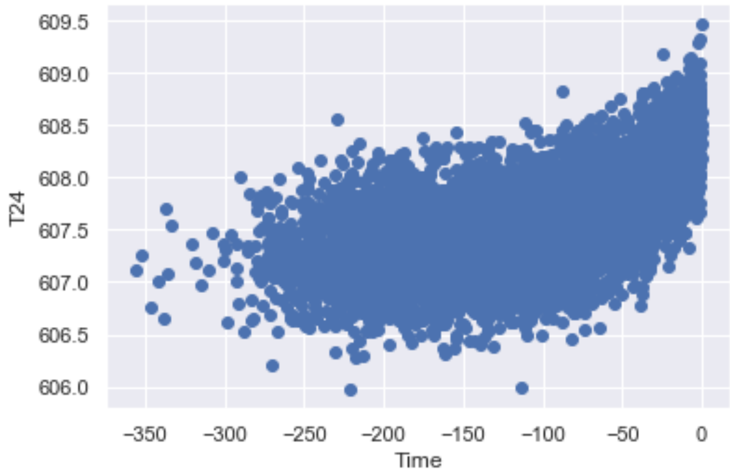

Fig3: T24 vs Time

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.

3.5 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 .

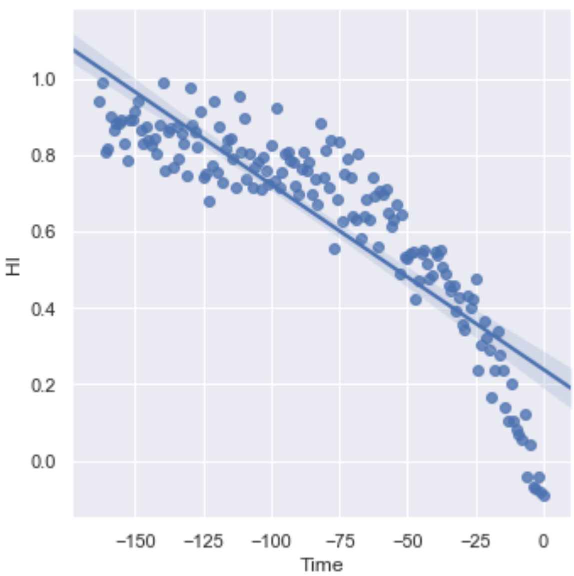

Fig4: Transformed HI time series data for unit 2 using the LMR Models

3.6. Model Identification

The exponential (nonlinear) regression models are used to describe the relationship between the adjusted cycle index C^{adj} *(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.

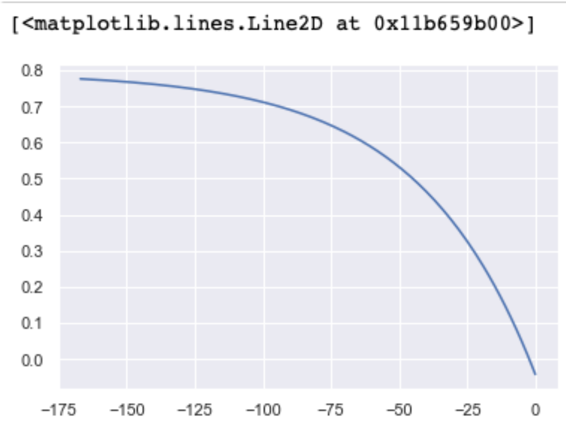

Fig5: Fitted exponential curve for the above pattern

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

3.7 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.

4. RESULTS

The final RUL Estimation and the error in estimation will be declared once the similarity measurement metric is defined and will be presented.

5. REFERENCES

1. System-level health assessment of complex engineered processes, https://www.researchgate.net/publication/50253040_System-level_health_assessment_of_complex_engineered_processes