Turbofan Engine Degradation Simulation

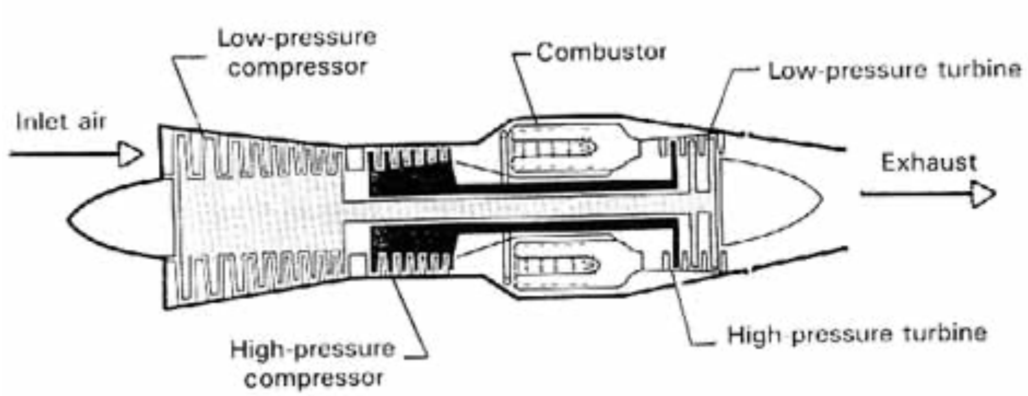

Figure 1. Turbofan in jet propulsion system. [Source: https://history.nasa.gov/SP-468/ch10-3.htm, last accessed: 15-Aug-23]

Contents

---

1. Dataset description

The engine degradation simulation was carried out with a simulator called C-MAPSS, which is NASA's aircraft engine simulator. The dataset contains four different scenarios, simulated under different combinations and fault modes. Different sensors acquire data to characterize the fault evolution.

The data is found near the dataset, in Moodle, in the project info page. The dataset has 26 columns, separated by spaces. Each column represents a different variable. Each row is a snapshot of data taken during a single operational cycle. The variables (columns) are:

1. Unit number.
2. Time, in cycles.
3. Operational Setting 1.
4. Operational Setting 2.
5. Operational Setting 3.
6. Sensor measurement 1.
7. Sensor measurement 2.

...

32. Sensor measurement 26

The dataset is found in Moodle, next to this document, in the General Project Work information page.

---

2. **Modelling goal**

   a. **Level B** (20p) Importance of sensors in identifying the fault

There are multiple sensors that are monitored, but there might be a difference in importance when describing the failure. Choose the appropriate number of latent variables, and the appropriate number of original variables to perform regression to predict the failure.
Note: Since this is a level B task, you can choose one of the four operational systems. You only need the training partition of one operational system.

   b. **Level B2 (20p)** Filling in missing data using multivariate approaces

One of the sensors has stopped emitting. Choose one sensor. How well can you estimate its values, using multivariate regression approaches? What is the appropriate number of latent variables needed, and what is the minimum number of sensors that are used to predict the faulty sensor?
Note: Since this is a level B task, you can choose one of the four operational systems. You only need the training partition of one operational system.

   c. **Level A** (30p) Forecast when the engine is due to maintenance

Predict the remaining useful life of test partition observations, based on the models calibrated with the training data. You know that in the training partition, the engine runs until failure. That means, the maximum operating cycle for a unit +1 was the failure operation. Make four prediction models and predict the remaining useful life for the test partition, using the model calibrated with the training data partition.
Note: You need all four operational settings datasets, both training and testing partitions for this task.

   d. **Level A2** (30p) Optimize the data acquisition

When creating a regression model to predict the RUL, you observe that certain sensors have more importance when it comes to the estimation of the RUL. Your task is to reduce the model to the sensors that are important to predicting the RUL, based on model diagnostics.
Note: You only need the four operational settings datasets, without the testing partition.

   e. **Level A3** (30p) Control charts for signaling incoming failure

Multivariate control charts can be used to predict when the engine is due to revision. Using multivariate control charts such as T2 and SPEx, signal when a certain number of operations prior to failure, for example 10 or 20 cycles prior to RUL are achieved. For individual observations, evaluate which are the sensors that drove the measurement to appear out of control limits when it is due to revision.
Note: You need both the training and test partitions for this task.

3. General hints on the dataset

- In the training partition, the RUL is (maximum operating cycles – current operating cycle);
- In the RUL file for the test partition, you have the remaining useful life = remaining cycles, after the last cycle of the unit. For example, the failure time is cycles elapsed + the remaining useful life at the end of these cycles = the cycle in which the engine stops (fails).
- You have different models for each operational setting = max 4 models, depending on what level of difficulty your group has.
- The sensor measurements are the only ones you need to predict the remaining useful life. If you take individual models for each setting, the columns 3-5 will nor vary (so we don't include constant variables in our model). Column 1 is used only for diagnostics and visualization, but not in modelling. Column 2 is only used to calculate the response variable.

References and aid sources

[1] - Feng, D., Xiao, M., Liu, Y., Song, H., Yang, Z. and Zhang, L., 2016. A kernel principal component analysis–based degradation model and remaining useful life estimation for the turbofan engine. Advances in Mechanical Engineering, 8(5), p.1687814016650l6.

[2] - Koen Peters, Time series analysis for predictive maintenance of turbofan engines [found online at: https://towardsdatascience.com/time-series-analysis-for-predictive-maintenance-of-turbofan-engines1b3864991da4, last accessed: 15.10.2023]