Predictive Maintenance

1 Task Description

In the following task, simulated run-to-failure data of commercial turbo fan engines, originally presented in (Saxena et al., 2008), is provided. The data was simulated with a software system called C-MAPSS, written in Matlab and Simulink, which was developed by the *National Aeronautics and Space Administration* (NASA) (Liu et al., 2012). The main two objectives are:

1. Assessing the current condition of a given turbo fan
2. Calculating remaining useful life (RUL) of a given turbo fan

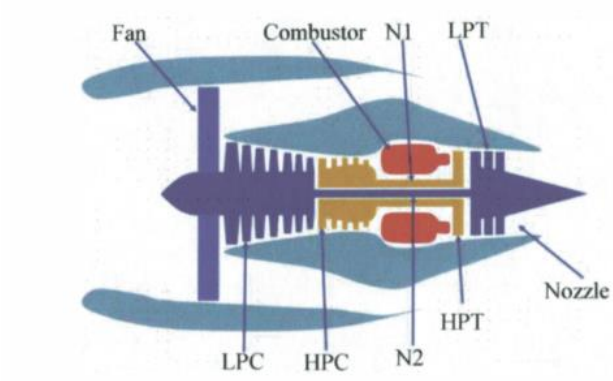

Figure 1: Simplified diagram of turbo fan engine simulated in C-MAPSS as in (Saxena et al., 2008).

1.1 C-MAPSS data

The original C-MAPSS dataset[1] consists of run-to-failure data of turbo fan engines. The data set is divided into four subsets: FD001, FD002, FD003 and FD004. Each of the subsets consists of a different number of observations, operation conditions and fault modes (see Table 1 for details). Furthermore, for each subset a training and test set is given in the original C-MAPSS dataset.

In each training data subset, 21 sensor signals were tracked over time until the considered turbo fan engine failed. The 21 sensor signals refer to physical units such as pressure and temperature at certain measurement positions. However, we are not provided with this information and have to work purely data-driven in this case.

[1] https://ti.arc.nasa.gov/c/6/

	FD001	FD002	FD003	FD004
Training observations	100	260	100	249
Test observations	100	259	100	248
Operation conditions	1	6	1	6
Fault modes	1	1	2	2

Table 1: C-MAPSS data overview

1.2 Objective 1: Prediction of current health condition of a given turbo fan engine

Based on the training dataset, the condition of a new engine, which was not observed before (test data), has to be assessed. Hence, it is necessary to construct a health indicator (HI). For an extensive literature review on health indicators see (Lei et al., 2018) and for a linear regression HI indicator in the context of machine tool RUL estimation the lecture slides and (Li et al., 2018).

Based on the learned HI, the test dataset should be used to assess the current health condition of the previously unseen turbo fan engines (i.e. test data).

1.3 Objective 2: Prediction of remaining useful life of a given turbo fan engine

Based on the training dataset, the remaining useful life of an unseen turbo-fan engine has to be estimated. As presented in the lecture, this can be done indirectly via a HI (that you learned in objective 1) or directly. Either one or both approaches can be chosen.

Based on a learned RUL-model, the test dataset should be used to calculate the future RUL of the turbo fan engines in the test dataset.

1.4 Research questions

As the provided data set allows a variety of analysis, some research questions are provided, which should guide the way to develop sensible machine learning models.

Step 1: Data exploration

What does the dataset look like? What kind of input features do you have (numeric, categorical, time series, etc…)? Do you have to conduct the train/test-split manually? How many sensor signals do not change at all over time? Are different sensor signals correlated with each other? What do the labels look like? Are the labels remaining useful life values or current degradation conditions? Will you have to extract extra features or labels in order to answer objectives 1 and 2?

Step 2: Feature and label engineering

In objective 1 you are asked to construct a HI, which represents a health status. What part of the training data can be considered “healthy” and what part of the training data can be considered “degraded”? How can you encode the two statuses?

In objective 2, you are asked to calculate RUL either directly or based on the HI you constructed. What are the advantages and disadvantages of the two approaches? When receiving test data, how does the process look like from receiving the raw test data up to the calculated RUL value? Do you use regression, classification or both in this process in order to predict RUL?

Step 3: Modeling

Which models can be used for the objectives? What are benchmark models for predictive maintenance tasks found in literature (have a look at Google Scholar, sciencedirect.com or ieeexplore.ieee.org)? After choosing a model, what are the critical parameters, hyper-parameters and how should you decide on them?

Step 4: Results and model insights

How good does the model perform on the test data? What are levers to increase the model performance? Is hyper-parameter tuning possible and sensible? Is your model sensitive to random seeds and train/test-splits? How robust is the model when reducing the number of training observations?

Step 5: Conclusion

If you had more time, what would your next steps be?

2 Literature

Lei, Y., Li, Naipeng, Guo, L., Li, Ningbo, Yan, T., Lin, J., 2018. Machinery health prognostics: A systematic review from data acquisition to RUL prediction. Mech. Syst. Signal Process. 104, 799–834. https://doi.org/10.1016/j.ymssp.2017.11.016.

Li, P., Jia, X., Feng, J., Davari, H., Qiao, G., Hwang, Y., Lee, J., 2018. Prognosability study of ball screw degradation using systematic methodology. Mech. Syst. Signal Process. 109, 45–57. https://doi.org/10.1016/j.ymssp.2018.02.046.

Liu, Y., Engineering, R., Heights, P., Frederick, D.K., DeCastro, J.A., Litt, J.S., Chan, W.W., 2012. User's Guide for the Commercial Modular Aero-Propulsion System Simulation (C-MAPSS). N. Y. 48.

Saxena, A., Goebel, K., Simon, D., Eklund, N., 2008. Damage propagation modeling for aircraft engine run-to-failure simulation, in: 2008 International Conference on Prognostics and Health Management. Presented at the 2008 International Conference on Prognostics and Health Management (PHM), IEEE, Denver, CO, USA, pp. 1–9. https://doi.org/10.1109/PHM.2008.4711414.

3 Presentation

- 10 minutes presentation
- Q&A session. Everyone should answer at least once

Introduction

- Introducing your topic
- The audience should understand the general topic and the motivation

Methods, Experiments

- Description of the entire data-handling and analysis process (for example application of KDD process)
- Description of applied methods and models
- Why did you choose the respective methods and models?

Results

- Clear presentation of the results.
- Precise and meaningful labeling of illustrations and diagrams

Discussion

- Interpretation and evaluation of results
- Direct references to your results

Summary

- Indicate the key findings you had
- Explain the future research you would conduct if you had more time