Data Science Assignment

Predicting Compressor and Turbine Decay of Naval Vessel

About Dataset:

Link to dataset: https://kemsys-my.sharepoint.com/:x:/p/anand_borad/EZ9jKMmx94hInA1YELbZJmABDuPmfHcnNi69J4s1hJ9FOQ?e=NpblOc

This particular dataset includes information about the propulsion plants of naval vessels that undergo condition-based maintenance. It consists of 16-feature vectors of GT measures in a steady state.

This dataset allows for the examination and identification of the decline in performance of the plant's components, including the GT compressor and turbines, over time. These insights are extremely valuable in understanding a crucial aspect of naval vessel operations.

The dataset includes the following information:

- Ship speed, which is a linear function of lever position LP
- Compressor degradation coefficient kMc
- Turbine degradation coefficient kMt
- Ship speed from 3 knots to 27 knots with a granularity resolution equal to three knots
- Thrusts torque (GTT) [kN m]
- Revolutions per minute in gas generator (GGn) [rpm]
- Turbine injection control (TIC) [%]
- Various other essential engine metrics

The dataset gives valuable information about a vessel's health, helping to improve its safety under challenging environmental or traffic conditions. Extensive testing has been done using this numerical simulator, proving it to be effective in aiding maintenance activities.

Real-time monitoring of these engine parameters can prevent mechanical failures, which can result in expensive repairs or even casualties in international waters.

Goal:

Predict compressor and turbine decay coefficient based on 16 features.

Assignment:

Notes:

1. Language to be used: Python (Jupyter notebook or Google Colab is preferred)
2. Machine learning library : Any library of choice (sklearn, tensorflow, pytorch, or anything)
3. Try to create modular code (ref: https://www.geeksforgeeks.org/understanding-code-reuse-modularity-python-3/)
4. Documentation is must (ref: https://realpython.com/documenting-python-code/)

Phase 0: Familiarize yourself with the features in the data

Before beginning your analysis, it is important that you familiarize yourself with the various parameters included in this dataset.

It includes 16 features related to the GT (gas turbine) measures of a naval vessel's propulsion plant along with decay state coefficients for GT compressor and turbine. This includes Lever position, Ship speed, Shaft torque, Rate of revolutions for GT and GG (Gas Generator), Starboard and Port Propeller Torques, Turbine exit temperature/ Pressure measurements as well as Fuel flow rate etc..

Make sure you have an in-depth understanding of each parameter before moving forward with your analysis.

Phase 1: EDA - Visualization & correlation between features

Once you have familiarized yourself with each feature included in the data set, it is useful to explore any potential correlations between them by plotting visualizations such as scatter matrix, correlation coefficient or block histograms etc..

This can help identify any areas where Machine Learning algorithms could beneficially work together or where further investigation will be required.

Phase 2: Algorithm Development & Model Training (Magical part)

A range of appropriate Machine learning philosophy and algorithms to be applied to achieve the goal.

Phase 2.1: Data cleaning, feature removal is required

Phase 2.2: Train, test, and validation split

Phase 2.3: Select right optimization metric (e.g. loss and error)

Phase 2.4: Predict compressor decay coefficient using Linear regression (1st model) (ref: https://towardsdatascience.com/linear-regression-detailed-view-ea73175f6e86)

Phase 2.5: Predict turbine decay coefficient using any of supervised learning algorithm of your choice. (2nd model)

Reach out to us:

Please reach out to us in case of any queries.

Mokshad: mokshad.navan@kemsys.com

Govind: govind.patidar@kemsys.com

Anand: anand.borad@kemsys.com