ENRE 655 Project Description
Spring, 2024

1 Data for the Project

In this project, you will use the C-MAPSS data obtained from turbofan engine damage propagation simulation[1]. The C-MAPSS data are partitioned into four datasets. Each dataset consists of multiple multivariate time series. Each dataset is further divided into training and test subsets. Each time series is from a different engine, i.e., the data can be from a fleet of engines of the same type. Each engine starts with different degrees of initial wear and manufacturing variation which is unknown to the user. This variation is considered normal, i.e., it is not considered a fault condition. There are three operational settings that have a substantial effect on engine performance. These settings are also included in the data. The data is contaminated with sensor noise.

The engine is operating normally at the start of each time series and develops a fault at some point during the series. In the training set, the fault grows in magnitude until system failure. In the test set, the time series ends some time prior to system failure. The objective is to predict the number of remaining operational cycles before failure in the test set. A vector of true Remaining Useful Life (RUL) values for different engines in the test set, i.e., the number of operational cycles after the last cycle that each engine continues to operate, is provided.

Each training set is provided as a text file with 26 columns of numbers, separated by spaces. Each row is a snapshot of data taken during a single operational cycle, and each column is a different variable. The columns correspond to:

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

…

26) sensor measurement 26

[1] Saxena, A., Goebel, K., Simon, D. and Eklund, N., 2008, October. Damage propagation modeling for aircraft engine run-to-failure simulation. In *2008 international conference on prognostics and health management* (pp. 1-9). IEEE.

A summary of the datasets is provided in Table 1.

Table 1. C-MAPSS data summary.

	C-MAPSS data			
	FD001	FD002	FD003	FD004
Data provided	Training, Test, True RUL	Training, Test, True RUL	Training, Test, True RUL	Training, Test, True RUL
Engine units for training	100	260	100	249
Engine units for test	100	259	100	248
Number of operating conditions	1	6	1	6
Number of fault modes	1	1	2	2

2 Project Objectives

The overall objective is to develop machine learning models that can predict the RUL with high accuracy. In part 1 of the project, you are free to develop any model. In part 2, you are limited to convolutional neural networks (CNN) and long short-term memory (LSTM) networks, but still have large freedom of defining specifics of the model.

2.1 Part 1

Develop two models for RUL prediction and apply the models to the four datasets.

- You are free to choose all or a subset of the features in the datasets. Note that the values of certain variables may be constant and so you may want to remove these variables. Provide data preprocessing details.
- Use the root mean squared error (RMSE) as a performance evaluation metric.
- For each model, run 5 different instances to avoid the effect of randomness and report the mean RMSE value and its standard deviation. You should compare the performance of your models with the performance of models reported in the literature[2].
- Compare the results between the models and explain the difference if any.
- Compare the results between the four datasets and explain the difference if any.
- Any other meaningful analysis will be a plus.
- Make sure to compare with other candidate models and choose the top two models in your project.

[2] Li, X., Ding, Q. and Sun, J.Q., 2018. Remaining useful life estimation in prognostics using deep convolution neural networks. *Reliability Engineering & System Safety*, *172*, pp.1-11.

- Clarify your innovation if any, compared with existing methods in the literature. This may require a literature review. If you used someone else's work, provide a citation.

2.2 Part 2

Develop a model that integrates CNN and LSTM for RUL prediction and apply the model to dataset FD001.

- Report the model architecture and hyperparameters.
- Report the mean RMSE for 5 runs of the model.
- Compare the result of this model with the results of the two models developed earlier and explain the difference if any.

3 Expected Outcome

You will submit a report summarizing your approach, analysis, and results. This report should be treated as a technical report or a journal/conference paper. When you prepare your report, please follow the format below:

- The report itself should include the following sections: abstract; introduction (including literature and layout of the report), technical discussion and approach used to solve the questions, analysis and results, discussions, conclusions, references.
- The report is limited to 10 pages. You can add an optional appendix (not to exceed 5 pages) that includes examples of the data and scripts used.
- The font to use is Times Roman 12 point with 1.5-line spaces.
- Use IEEE Style for references.

Besides the report, you are also expected to submit slides summarizing the project in PowerPoint or PDF and give a 10-minute presentation on the last day of the class.

4 Project Evaluation

The following factors will be considered in evaluating your project.

- Presentation: The last day of the class, May 8 is the class presentations of your project. It is 10 minutes total that covers the approach you used to solve the problem, the results you obtained, discussions and any experience you gained in performing this project. The presentation is then followed by questions for about 5 minutes. You will be judged by the quality of the slides, clarity of the presentation and delivering it within the 10 min limit: Please upload your presentations and I will make it available to all class. 10%.

- Quality: Report’s clarity, meeting the report format given above, free of typos and grammatical errors, proper references, and citations: 15%.
- Approach: Innovation in the approach, appropriate uses of the approaches and solution employed, the relevance of methods and tools used or developed, and analysis performed in the project: 50%.
- Accuracy: Adequacy, reasonableness, and accuracy of the results, discussions, and conclusions: 20%.
- Timeliness: On-time submission of the presentation and the report: 5%