PHM08 Prognostics Data Challenge Dataset

Description

This dataset was used for the prognostics challenge competition at the International Conference on Prognostics and Health Management (PHM08). The challenge is still open for the researchers to develop and compare their efforts against the winners of the challenge in 2008. References to the three winner papers are provided below.

- Heimes, F.O., "*Recurrent neural networks for remaining useful life estimation*", in the Proceedings of the 1st International Conference on Prognostics and Health Management (PHM08), Denver CO, Oct 2008.
- Tianyi Wang, Jianbo Yu, Siegel, D., Lee, J., "*A similarity-based prognostics approach for Remaining Useful Life estimation of engineered systems*", in the Proceedings of the 1st International Conference on Prognostics and Health Management (PHM08), Denver CO, Oct 2008.
- Peel, L., "*Recurrent neural networks for remaining useful life estimation*", in the Proceedings of the 1st International Conference on Prognostics and Health Management (PHM08), Denver CO, Oct 2008.

Additionally, the following reference provides more in depth analysis of results from the competition and a comprehensive literature review on other efforts that have used these data for research.

- Ramasso, Emmanuel, and Saxena, Abhinav, "*Performance benchmarking and analysis of prognostic methods for CMAPSS datasets*", International Journal of Prognostics and Health Management Vol. 5(2), pp: 15, 2014 (http://www.phmsociety.org/node/1548)

Experimental Scenario

Data sets consist of multiple multivariate time series. Each data set is further divided into training and test subsets. Each time series is from a different engine – i.e., the data can be considered to 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 wear and 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 are contaminated with sensor noise.

The engine is operating normally at the start of each time series, and starts to degrade at some point during the series. In the training set, the degradation grows in magnitude until a predefined threshold is reached beyond which it is not preferable to operate the engine. In the test set, the time series ends some time prior to complete degradation. The objective of the competition is to predict the number of remaining operational cycles before in the test set, i.e., the number of operational cycles after the last cycle that the engine will continue to operate properly.

Usage

The data are provided as a zip-compressed text file with 26 columns of numbers, separated by spaces. Each row is a snapshot of data taken during a single operational cycle; 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

Users are expected to train their algorithms using data in the file named `train.txt`. They must then evaluate the RUL prediction performance on data provided in file `test.txt`. Associated true RUL values are not being revealed just like in the competition. Very soon the users will be able to use a web application to upload their results and get an aggregate score feedback. Please check back in June 2010 to get more details on the web application for automated scoring. Until then they can get a feedback by emailing their results through simple text files to any of the email addresses provided below. A set of representative top 20 scores obtained during the competition are included here to provide a reference to everyone.

No.	Score
1	436.841
2	512.426
3	737.769
4	809.757
5	908.588
6	975.586
7	1,049.57
8	1,051.88
9	1,075.16
10	1,083.91
11	1,127.95
12	1,139.83
13	1,219.61
14	1,263.02
15	1,557.61
16	1,808.75
17	1,966.38
18	2,065.47
19	2,399.88
20	2,430.42

Evaluation

The final score is a weighted sum of RUL errors. The scoring function is an asymmetric function that penalizes late predictions more than the early predictions. Following equations describe this function analytically.
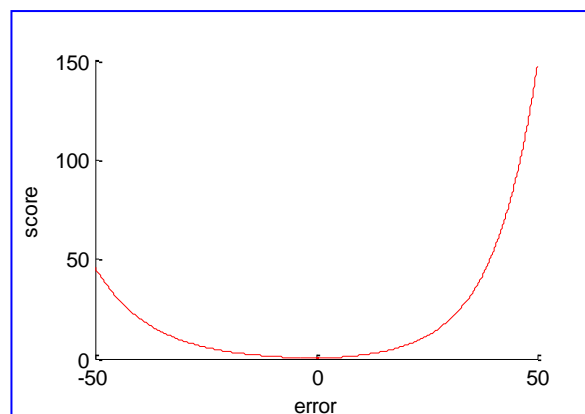
$$s = \begin{cases} \sum_{i=1}^{n} e^{-\left(\frac{d}{a_1}\right)} - 1 \text{ for } d < 0 \\ \sum_{i=1}^{n} e^{\left(\frac{d}{a_2}\right)} - 1 \text{ for } d \geq 0, \end{cases}$$

where,

    s is the computed score,
    n is the number of UUTs,
    $d =$ (Estimated RUL – True RUL),
    $a_1 = 13$, and $a_2 = 10$.

Once algorithms are trained to satisfaction, users can apply them to the final test dataset contained in the file named `final_test.txt`. Users should send the vector of RULs for the final test set to the PHM Society for evaluation. A score will be mailed back soon. Researchers are encouraged to publish their results regardless of the absolute performance if they believe there is novelty in their algorithm. The intent is to develop innovative approaches for prognostics.

Note: Any team or individual is allowed to submit their results on the final test set ONLY ONCE.

Data Set: train.txt, test.txt

Train trjectories: 218

Test trajectories: 218

final_test trajectories: 435

Contacts

Abhinav Saxena – asaxena@ge.com

Kai Goebel – kai.goebel@nasa.gov

References

A. Saxena, K. Goebel, D. Simon, and N. Eklund, "*Damage Propagation Modeling for Aircraft Engine Run-to-Failure Simulation*", in the Proceedings of the 1st International Conference on Prognostics and Health Management (PHM08), Denver CO, Oct 2008.