



Annual Conference of the Prognostics and Health Management Society 2014

29 September – 2 October, 2014

Fort Worth, Texas, USA

PHM Data Challenge 2014

The PHM Data Challenge is a competition open to all potential conference attendees. This year the challenge is focused on asset health calculation, a common problem in industrial remote monitoring and diagnostics. Participants will be scored on their ability to generate a health score that accurately segments a population of assets into high and low risks of failure.

This is a fully open competition in which collaboration is encouraged. The teams' may be composed of any combination of students, researchers, and industry professionals. The results will be evaluated by the Data Challenge Committee and all teams will be ranked. The top scoring teams will be invited to present at a special session of the conference and first and second place finishers will be recognized at the Conference Banquet event.

Teams

Collaboration is encouraged and teams may be comprised of one or more students and/or professionals. The team judged to have the first and second best scores will be awarded prizes of \$600 and \$400 respectively contingent upon:

- Having at least one member of the team attend the PHM 2014 Conference
- Presenting the analysis results and technique employed at a special session within the Conference program
- Submitting a peer-reviewed Conference paper. (Submission of the challenge special session papers is outside the regular paper submission process and follows its own modified schedule.)
- The top entries will also be encouraged to submit a journal-quality paper to the International Journal of Prognostics and Health Management (ijPHM).

The organizers of the competition reserve the right to both modify these rules and disqualify any team for any practices it deems inconsistent with fair and open practices.

Registration

Teams may register by contacting the Competition organizers (dustin.garvey@gmail.com, robert.wigny@pwc.ca) with their name(s) and a team alias under which the scores would be posted.

PLEASE NOTE: In the spirit of fair competition, we allow only one account per team. Please do not register multiple times under different user names, under fictitious names, or using anonymous accounts. Competition organizers reserve the right to delete

multiple entries from the same person (or team) and/or to disqualify those who are trying to "game" the system or using fictitious identities.

Data

There are 5 data sets. Due to proprietary concerns we cannot provide a detailed description of the data and the domain. If you have any questions, please contact Dustin Garvey (dustin.garvey@gmail.com).

Train – Part Consumption.csv

This file contains a record of what parts were replaced on an asset and the reason for their replacement. This file contains five columns. The first column is the asset ID. The second column is the time that the part(s) were replaced. The third column is a code that specifies the reason the asset was being worked on. The fourth column is a code that specifies the type of part that was replaced. The fifth column is the number of parts of the specified type that were replaced.

Train – Usage.csv

This file contains a record of a parameter that roughly measures asset usage. The units of this measure cannot be shared due to proprietary concerns, but think something along the lines of an odometer. This file contains three columns. The first column is the asset ID. The second column is the time the measurement was taken. The third column is the current usage of the asset.

Train – Failures.csv

This file contains a set of asset failures. The file contains two columns. The first is an asset ID and the second is the time of the failure. The units of time cannot be shared.

Test – Part Consumption.csv

This file contains the same type of information as the training file but over the test period.

Test – Usage.csv

This file contains the same type of information as the training file but over the test period.

Test Instances.csv

This file contains the asset ID and time combinations that should be used to generate the health scores in the submission. This file contains two columns. The first column is the asset ID. The second column is the time relative to which the health score should be calculated.

Example Submission.csv

This file contains an example submission with specified a segmentation of health scores of 0.9 (lower than 0.9 = low risk, greater than or equal to 0.9 = high risk) and randomly generated scores for the test instances.

Submission & Scoring

Each team is permitted one submission a week. A submission will be composed of a CSV file whose name is the team alias. If the alias is "example", then the filename will be "example.csv". The first line of the CSV file should be the value that segments the health scores into low and high risk. More specifically assets are low risk will have scores below the supplied limit and high risk will have scores above the limit. The remaining lines should have three columns for the prescribed set of assets and times. The first and

second column should be prescribed asset ID and calculation times. The third column should be a health score, with 0 meaning a healthy asset and increasingly large values meaning increasingly degraded.

The test assets and times contain a balanced sample of instances where the asset did and did not fail in the immediate future (within 3 time units). If N is the number of samples of each type (i.e. with and without a failure in the immediate future, total samples is $2N$), L is the number without an immediate failure and placed in the low risk category, and H is the number with an immediate failure and placed in the high risk category, the score of the submission is:

$$\text{Score} = (L / N) + (H / N)$$

Acceptable methods will use usage and part consumptions prior to the test times only.

Running Leaderboard

TBD

Key Dates

Competition Closed: 5 August 2014

Prelim Winners Announced: 6 August 2014

Winning Papers Due: 14 September 2014

Winners Announced: 21 September 2014