Presentation Outline

• A Similarity Based Prediction Approach
  ◦ Analysis Approach
  ◦ Pre-Processing and Feature Selection
  ◦ Health Index Model
  ◦ Prediction Method and Settings

• Review of Top Winner Methods

• Perspectives on the Data Challenge
A Similarity Based Prediction Approach
Overview About Approach and Processing Steps

Objective: Remaining Useful Life (RUL) Prediction of Aircraft Engines

Aircraft Engine

Data-Preprocessing

Feature Selection

Objective: Remaining Useful Life (RUL) Prediction of Aircraft Engines

Model Training

Health Index

Life Prediction

Objective: Remaining Useful Life (RUL) Prediction of Aircraft Engines

Model Training

Health Index

Life Prediction

Objective: Remaining Useful Life (RUL) Prediction of Aircraft Engines

Model Training

Health Index

Life Prediction

Objective: Remaining Useful Life (RUL) Prediction of Aircraft Engines

Model Training

Health Index

Life Prediction

Objective: Remaining Useful Life (RUL) Prediction of Aircraft Engines

Model Training

Health Index

Life Prediction

Objective: Remaining Useful Life (RUL) Prediction of Aircraft Engines

Model Training

Health Index

Life Prediction

Objective: Remaining Useful Life (RUL) Prediction of Aircraft Engines

Model Training

Health Index

Life Prediction

Objective: Remaining Useful Life (RUL) Prediction of Aircraft Engines

Model Training

Health Index

Life Prediction

Objective: Remaining Useful Life (RUL) Prediction of Aircraft Engines

Model Training

Health Index

Life Prediction

Objective: Remaining Useful Life (RUL) Prediction of Aircraft Engines

Model Training

Health Index

Life Prediction
## Data Characteristics and Analysis Selection Table

<table>
<thead>
<tr>
<th>Data Characteristics</th>
<th>Time Statistics</th>
<th>Stats per Time Slice</th>
<th>High Frequency Methods</th>
<th>Residual Based</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Low frequency signals</td>
<td><strong>H</strong></td>
<td><strong>L</strong></td>
<td><strong>L</strong></td>
<td><strong>M</strong></td>
</tr>
<tr>
<td>2 Data from healthy and degraded state</td>
<td>Expert</td>
<td>Filter Methods</td>
<td>Trendability Metrics</td>
<td>Wrapper methods</td>
</tr>
<tr>
<td>3 Large amount of historical degradation data</td>
<td><strong>M</strong></td>
<td><strong>H</strong></td>
<td><strong>H</strong></td>
<td><strong>M</strong></td>
</tr>
<tr>
<td>4 Six different discrete operating regimes</td>
<td>Feature Summation</td>
<td>Distance from Normal</td>
<td>Hypothesis Testing</td>
<td>Regression</td>
</tr>
<tr>
<td>5 One failure mode</td>
<td><strong>M</strong></td>
<td><strong>M</strong></td>
<td><strong>L</strong></td>
<td><strong>H</strong></td>
</tr>
<tr>
<td>6 Desired output is estimate of remaining life</td>
<td>Curve Fitting</td>
<td>Neural Networks</td>
<td>Stochastic Filtering</td>
<td>Similarity Based</td>
</tr>
<tr>
<td></td>
<td>RUL Prediction</td>
<td><strong>M</strong></td>
<td><strong>H</strong></td>
<td><strong>M</strong></td>
</tr>
</tbody>
</table>
Data Analysis Flow Chart for Aircraft Engine

- **Training Data**
  - Pre-Processing / Variable Selection
  - Health Assessment
  - Model Fitting
  - Save Degradation Patterns and Health Models
- **Testing Data**
  - Pre-Process and Select Key Variables
  - Health Assessment Using Training Models
  - Similarity Evaluation Using Pattern library
  - Weighted RUL Estimation
Pre-Processing / Operating Regime Consideration

• For each operating regime, we calculated the mean value for each sensor using all the training data.

• For each sensor data point, the raw value was subtracted by the mean value in that regime; this provided a variable that could be trended with time.
Feature Selection

<table>
<thead>
<tr>
<th>Number</th>
<th>Variable Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>T24 Total temperature at LPC outlet °R</td>
</tr>
<tr>
<td>2</td>
<td>T30 Total temperature at HPC outlet °R</td>
</tr>
<tr>
<td>3</td>
<td>T50 Total temperature at LPT outlet °R</td>
</tr>
<tr>
<td>4</td>
<td>NRc Corrected core speed rpm</td>
</tr>
<tr>
<td>5</td>
<td>Ps30 Static pressure at HPC outlet psia</td>
</tr>
<tr>
<td>6</td>
<td>BPR Bypass Ratio</td>
</tr>
<tr>
<td>7</td>
<td>htBleed (Bleed Enthalpy)</td>
</tr>
</tbody>
</table>

- The correlation with cycle number was calculated for each sensor and for each unit.

- The average correlation value for each sensor for the training units was obtained, and we ranked each sensor.

- Sensors with a correlation value above 0.6 were selected and this resulted in the top 7 sensors.
The last 5 flights from each engine were given a label of unhealthy, and the first 5 flights were given a label of healthy.

A linear regression model was trained with the 7 features as input and the health label as an output.

The health index results for a few sample engines are shown on the left, and one can observe a clear health trend.
Degradation Pattern Library

- An exponential fit was performed on the health index values for the training data set.

- This would provide a library of degradation curves which will be used by the similarity prediction method.
Intuition of Similarity Based Prediction Method

Move the block of data along time; find the most probable position with regard to the curve of degradation pattern.

History of a test unit

Remaining life

Degradation pattern extracted from a training unit with run-to-failure data

End life of the training unit

Health Indicator

Time/Cycle

Unit 10: Age=177; RUL=105; Adj RUL=105
Similarity Calculation (Finding Most Similar Engine Degradation Curves)

\[ d(\tau, M_i) = \frac{1}{2N\sigma_i^2} \sum_{j=1}^{N} (y_j - f_i(j + \tau))^2 \]  

Euclidean distance calculation considers the time-lag in the formulation.

\[ RUL_i = T_{Ai} - T_E - \arg\min_{\tau} d_i(\tau, M_i) \]  

RUL is based on the time lag and how many cycles the matched engine unit was used.

\[ W_i = \exp(-d_i^2) \]  
An exponential function is performed on the distance values, to bound the values between 0 and 1.

\[ w_i = \frac{W_i}{\sum_{i=1}^{L} W_i} \]  
The weights for the RUL fusion estimation are based on the distance values, and normalized so they add up to 1.

\[ RUL = \sum_{i=1}^{L} w_i RUL_i \]  
The remaining useful life (RUL) estimation is a weighted sum based on the life estimate from each matched unit.
1. If health index and cycle correlation > 0.75  
   i. Use the 7 most similar degradation patterns

2. If health index and cycle correlation > 0.1 but < 0.75  
   i. Use the 32 most similar degradation patterns

3. If health index and cycle correlation > 0.05 but < 0.1  
   i. Use the 43 most similar degradation patterns

4. If health index and cycle correlation < 0.05  
   i. Use an estimate of the average life (64th percentile) of the engine from the training data

5. RUL estimates from the library that are outside 1.98 STDEV from the mean are removed (outlier removal of unreasonable life estimates).
Based on the current health trend, a similarity score is calculated for that trend in comparison to the library of trend patterns.

In this case, maybe 5-10 units are very similar, and there are some moderately similar units.

How to do the weighting and similarity calculation is one aspect of this method that could be fine-tuned and improved for each application.
Example RUL Prediction Result – Test Unit 32

- In this example, the health trend and the most similar degradation patterns are plotted.

- The estimated end of life is 64 cycles, which is very close to the actual remaining cycles of 59.

- In practice, one might need to provide a lower and upper bound estimate, as opposed to just a point estimate.
My Results

• The actual vs. predicted results were promising using this approach, although there were a few engines in which the error was quite large.

• My mean absolute error score was 14.65, in comparison to the top score of 14.36 and the second best score of 14.55 (this would put me in third place).

• If you remove the bottom 5 test unit errors, my score improves to 11.9, and if you remove the bottom 10, my score improves to 10.3.
Review of Top Winner Methods
(Ensemble with Two Models)
1st Model Descriptions

- In the 1st stage of the 1st model, RUL prediction is performed for each one line. For each flight regime, separate NN models are used.

- In the 2nd stage of the 1st model, each prediction is calculated for multiple lines and the final RUL is calculated by LightGBM. (By using multiple lines, the accuracy is improved and the variations of the engine units are minimized.)

- In the 1st stage of the 1st model, very few pre-processing and feature extractions are performed. The purpose is just to eliminate parameters which have low correlations with the results.

- In the 2nd stage of the 1st model, the parameters from Linear Regression such as gradient and intercept etc. are created as the features.
1st Model Descriptions

RUL Prediction by NN (1st Stage)
Separate Models for Each Flight Regime

Original CSV Data

Create parameters (features) by Linear Regression etc.

Light GBM (2nd Stage)

Final RUL

Estimated RULs from 1st Stage

Estimated RUL from 2nd Stage

The values calculated in the 1st stage are actually treated as parameters that are related to life rather than remaining life itself. (The variations of the engine units are large.)
2nd Model Descriptions

- NN is used as a prediction model. For each flight regime, NNs are used, and then the results from each regime are combined.

- 6 Inputs -> 2-3 Hidden Layers -> Merge -> 2-3 Hidden Layers -> Final Output

- Ensemble with the 1st Model and the 2nd Model

- In the 1st model, there are data leak and information loss from the 1st stage to the 2nd stage. Therefore, the 2nd model is developed to create one consistent model. Eventually, when the data length is not long enough, the accuracy of the 2nd model is not improved. In the end, this results in the ensemble with the 1st model and the 2nd model.
2\textsuperscript{nd} Model Descriptions

This model is created to eliminate data leak and information loss in the 1\textsuperscript{st} model. Actually, when the data length is short, the results are not so good, and in the end, this results in the ensemble of the 1\textsuperscript{st} model and the 2\textsuperscript{nd} model.

Original CSV Data

Separate NN Models for Each Flight Regime

Correspond to the 1\textsuperscript{st} stage in the 1\textsuperscript{st} model

Created as one model.
Top Scoring Method - Comments

• The method has some good characteristics:
  ◦ Using separate machine learning models such as neural networks for each operating regime.
  ◦ The model has two stages and each method uses an ensemble approach.

• We are wondering if some initial feature selection should have been considered to remove variables that had no real correlation with the engine degradation.

• It would also be interesting to see if other machine learning models in place of neural networks would have performed with a similar level of accuracy.
Second Highest Scoring Method - Comments

• The method used the following:
  i. Data normalization to have a single trend (deal with operating regime)
  ii. Window time slicing (to have more data for model training)
  iii. Convolutional neural network and LTSM neural network method

• Adversarial Validation was used to avoid overfitting, and multiple random seeds were used to have multiple models (ensemble).

• I liked how this method used data normalization to remove the operating regime effect.

• Also the window-time slicing is a good approach for time series data.

• We are wondering if using multiple neural network architectures would have performed better.
Perspectives on the Data Challenge
Uniqueness / Creativity of Approach / Method

• We feel that sometimes selecting the top scores might leave out unique and innovative approaches.

• One approach could be the following:

1. Select the top 10 scoring methods (we still want to consider methods that were competitive from an accuracy perspective).

2. Judge the method on the following criteria:
   i. Novelty of algorithm / approach
   ii. Whether the approach is logical
   iii. Could the method be applied to other applications (general purpose)
   iv. The number of parameters to configure
Practical Considerations

It is worth mentioning some other practical aspects for the RUL prediction method to be implemented:

1. Time to develop method / fine-tuning effort
2. RUL Distribution and Lower/Upper Bounds
3. Actionable information / Suggested Decision
Scoring Metric

Lastly, we are wondering if we should revise the scoring method or consider multiple metrics:

1. Mean absolute error after removing bottom 5 or 10.
2. Number of predictions within a certain RUL error (such as 5 or 10).
3. Different weights for an early prediction vs. a late prediction.
4. Prediction accuracy for engines that had a noticeable trend / health value.