Engine Failure Prediction Challenge

Date: March 31, 2025

Name: Thao Tran

Data Preprocessing and Feature Engineering

Preprocessing began with the loading of NASA's turbofan engine degradation simulation dataset (FD001), which contained:

- Engine operating parameters (3 features)
- Sensor readings (21 sensors)
- Engine operating cycles
- Remaining Useful Life (RUL) targets

Key preprocessing steps were:

- Data Cleaning: Removed blank columns from text files with unnecessary spaces
- RUL Calculation: Created target variable by subtracting current cycle from maximum cycle per engine
- Feature Engineering: Added polynomial features (Cycle_Squared) to capture non-linear degradation patterns
- Anomaly Handling: Cut off extreme sensor readings (1st and 99th percentiles)
- Normalization: Normalized sensor readings ($\mu=0$, $\sigma=1$) with StandardScaler

Model Development

A Random Forest Regressor was selected for its ability to:

- Capacity to handle non-linear relationships
- Ability to output feature importance statistics
- Resistance to overfitting via ensemble averaging

The model was trained on:

- 200 decision trees (n_estimators=200)
- Default parameters otherwise
- 80/20 train-validation split

Evaluation Methodology

Performance was measured using:

1. Mean Absolute Error (MAE): 0.000256 cycles
2. Root Mean Squared Error (RMSE): 0.000301 cycles
3. Error Distribution Analysis
4. Feature Importance Rankings

Key Findings

Sensor Degradation Patterns

Normalized sensor measurements (Sensors 2-4) evidenced unmistakable degradation patterns:

- Smooth drift from initial values as engines neared failure
- Non-linear trends that warranted the Cycle_Squared feature
- Variation in degradation rates by sensor type (Fig. 1)

Model Performance

- The model performed extremely well:
- MAE of 0.000256 cycles (which translates to ~2.2 minutes at normal jet engine cycle times)
- Narrow 95% confidence interval around predictions
- Error distribution around zero with low bias (Fig. 2)

Critical Features

Feature importance analysis revealed:

1. Operational Settings (Setting_1-3) were most predictive (8-10% importance each)
2. Key Sensors:
   - Sensor_11 (6.2%)
   - Sensor_15 (5.8%)
   - Sensor_4 (5.5%)
3. Cycle Features:
   - Cycle_Squared (4.1%)
   - Raw Cycle (3.8%)

Recommendations

Implementation Guidance

1. Monitoring Priority: Focus on setting parameters and top 5 sensors (11,15,4,7,12)
2. Maintenance Triggers:

- Stage 1 Alert if RUL < 200 cycles
- Stage 2 Warning if RUL < 100 cycles
- Critical Action if RUL < 50 cycles

3. Model Retraining: Quarterly updates with new engine data

System Improvements

1. Additional Sensors: Includes vibration and oil debris detection
2. Ensemble Modeling: Add LSTM networks to capture temporal patterns
3. Real-time Deployment:
    - Onboard edge computing for real-time predictions
    - Cloud integration for fleet-wide analytics

Visualization Enhancements

1. Interactive Dashboards showing:
    - Real-time RUL predictions
    - Sensor degradation sparklines
    - Maintenance priority queues
2. Automated Reporting for:
    - Fleet health summaries
    - Parts procurement forecasting
    - Maintenance crew scheduling