Predictive Maintenance for Aircraft Engines

Lihua Pei
Department of Engineering Management and Systems Engineering)
George Washington University
Washington, DC
lihuapei@gwu.edu

Lujin Zhao
Department of Engineering Management and Systems Engineering
George Washington University
Washington, DC
lujin@gwu.edu

Olatunji Akinbule
Department of Engineering Management and Systems Engineering)
George Washington University
Washington, DC
akinbule_ola@gwu.edu

***Abstract*—Prognostics and Health management involves failure prevention and predicting reliability and Remaining Useful Lifetime (RUL) of an aircraft. Using data analytical models to fuel predictive maintenance of aircrafts becomes important not only to foster safety but also to preclude airline downtime. With applied predictive maintenance, an airline can reduce expenses connected with expedited transportation of parts and unplanned maintenance. If a technical problem did occur, maintenance teams could react to it faster with workflow organization software. The solution consists of analyzing data and metadata regarding detected maintenance activity**

Keywords— Principal component Analysis (PCA), Predictive modelling, Logistic Regression

I. Introduction

Coming in after Pilot error, aircraft systems failure is one of the most common reasons for accidents. Catastrophic

equipment failures still account for around 20% of aircraft losses despite the improvements in design and manufacturing quality. It is a fact nowadays that aircraft generate more data than ever. Currently, around 2 million terabytes of data are generated every year by the global aircraft fleet through flight recorders and Aircraft Health Monitoring. By 2026 this may have grown to 98 million terabytes of data. With the cost of sensors, data storage and communication dropped over the years the velocity of incoming data has increased enormously, caused by the advancing information technologies which makes it easier to generate data (van Kempen & van Eijk, 2014; Chen, Mao & Liu, 2014). As a result, Big Data Analytics is even more critical to anticipating aircraft parts and replacements.

In this project, we will implement Logistic regression, Principle Component Analysis (PCA) and K-Means clustering on the dataset that contains engine working data and service statuses. Try to determine whether an engine unit need to go through major service or not. PCA will be used to reduce the dimension for K-Means and use unsupervised learning algorithm to further reduce the dimension into four dimensions. It is critical to perform such dimensionality reduction due to the real world data could be including more data from other parts. Successfully compressed the engine data can help improve run time for other algorithm or help developing an algorithm that can process real-time input.

II. Data Description

The dataset we use contains 25 features such as working cycles, Physical fan speed Engine pressure ratio and Bypass Ratio. There are 45351 observations in total and Each row is a snapshot of data taken during a operational cycle.

engine_id'	cycle'	op_set_1'	op_set_2'	op_set_3'
total_temp_HPC_outliet'	total_temp_LPT_outliet'	fan_inlet_pressure'	total_pressure_bypass_duct'	total_pressure_hpc_outlet'
physical_fan_speed'	physical_core_speed'	engine_pressure_ratio'	static_pressure_HPC_outlet'	fuel_flow_ratio'
corrected_fan_speed'	corrected_core_speed'	bypass_ratio'	burner_fuel_air_ratio'	bleedy_enthalpy'
demanded_fan_speed'	demanded_corrected_fan_speed'	HPT_coolant_bleed'	LPT_coolant_bleed'	fan_inlet_temp'

Table 2-1 Engine Data Features

Summary statistics are provided below, For engine statuses, we define four Time to Failure windows (4 classes):

Class 0: Very urgent maintenance: 0 to 10 cycles remaining before failure.

Class 1: Aircraft maintenance periodic checks need to be deep and more detailed: 11 to 30 cycles remaining before failure.

Class 2: Confident system: We can plan from this period the future maintenance date and provide the needed equipment: 31 to 100 cycles remaining before failure.

Class 3: Very confident system: Only periodic checks are needed: more than 101 cycles remaining before failure.

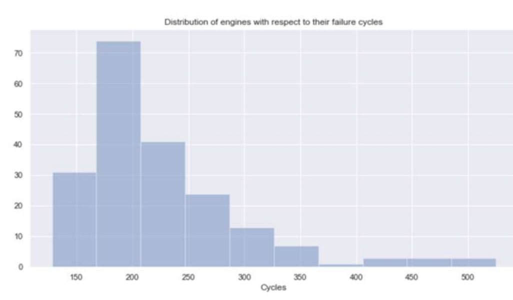


Fig2-1 Distribution of Engine Cycles

Fig 2.1 shows the distribution of engine failure with respect to cycles. Based on our dataset, more than 75% of the engine failed before reaching 256 cycles.

The correlation graph shows little correlation or no correlation between multiple features. The no correlation, according to our dataset is due to missing value. We delete those features that having such problems, making the dataset shrinks from 25 features to 22 features.

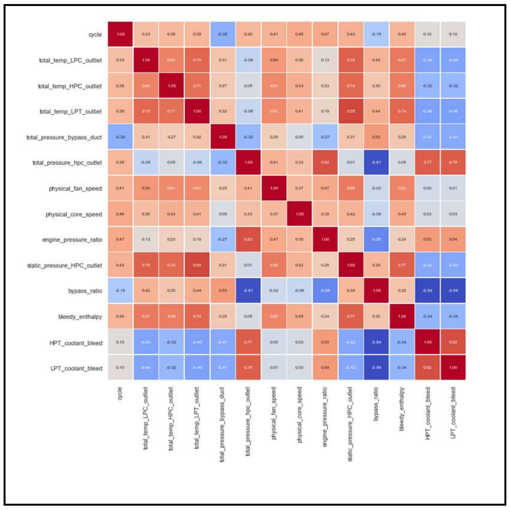


Fig2-2 Correlation Matrix

There are also some features that have a close to 0 correlation which can be also excluded from the analysis since adding or removing it will cause little variation in the dataset. Variables such as fan_intel_temp and burner_fuel_air_ratio have 0 correlation with all other variables but having collinearity issue between them. Removing those two variables will not affect prediction or clustering outcome. After removing such features, the dataset is further reduced to 17 columns.

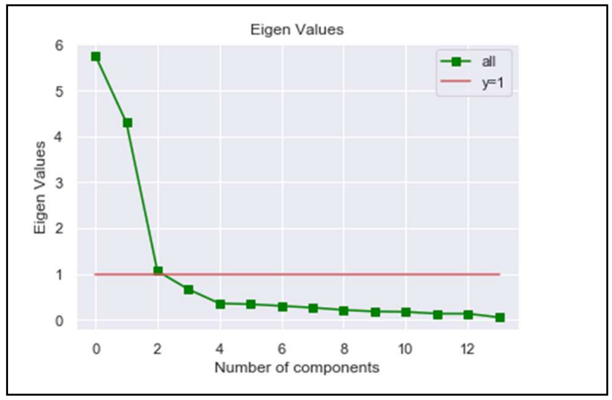
The correlation graph also indicates several highly correlated features. using the same logic, we exclude three features and our final dataset has 14 columns and Figure 2-2 shows the final dataset we will be using for all the analysis.

III. PCA Analysis

As the number of dimensions of any data increases, there is inherent difficulty to visualize, interpreted or perform computations. Hence there is a need to reduce the dimensions of a data by removing redundant dimensions and keeping only the most important ones. To interpret the data in a more meaningful form, it is necessary to reduce the number of variables to a few, interpretable linear combinations of the data with each linear combination corresponding to a principal component. PCA constructs orthogonal – mutually uncorrelated – linear combinations that explains as much common variation as possible in a given dataset. In other words, PCA seeks the direction that maximizes the variance providing independent pieces of the information puzzle represented in the larger dataset.

In our data with initially 26 independent variables we seek to reduce the number of dimensions while retaining as much variation in the data as possible. The scree plot shown below is the metric to choose the number of principal components that we hope to retain. Some features were dropped because they had no correlation with the target variable. The elbow of

Fig3-1 Scree Plot



the Scree Plot helps in determining the number of Principal Components to retain in our analysis. By inspecting the differences between eigenvalues, we see that the first inflection point occurs between the third and fourth eigenvalues. Furthermore, there is an almost linear behavior of the eigenvalues after the third Principal Component highlighted with the red reference line. The Kaiser-Guttman criterion selects eigenvalues greater than the average eigenvalue (i.e, X > 1) because those axes summarize more information than any single ordinary variable (Jackson, 1993).

The cumulative variance explained by the first 3 principal components is about 80% of the variation in the data as shown by the cumulative variance curve in Figure 3-2 below.

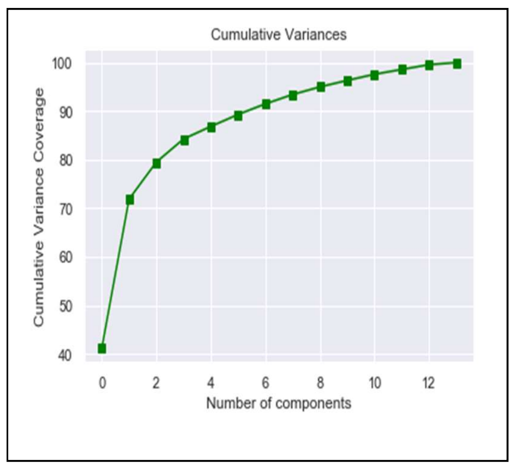


Fig3-2 Cumulative Variance Curve

__Interpreting the Principal Components__

Figure C shows distribution of the 14 possible principal components and the various weightings across all 14 features in consideration.

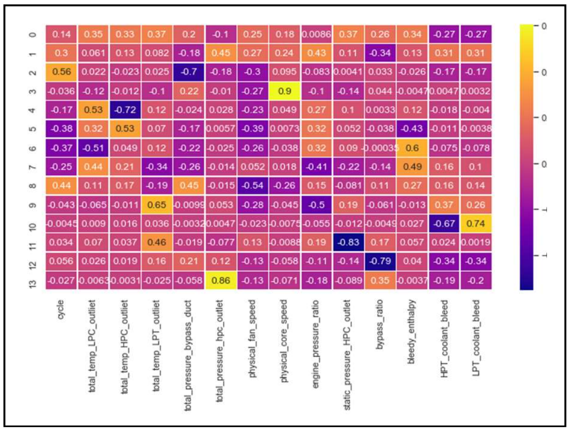


Fig3-3 Principle Components

__1st Principal Component__ **:**

This component is highly positively weighted on the following features 'total_temp_LPC_outlet', 'total_temp_HPC_outlet', 'total_temp_LPC_outlet', 'static_pressure_HPC_outlet', 'bleedy_enthalpy'. And significantly negatively weighted on 'HPT_coolant_bleed' and 'LPT_coolant_bleed'.

This Component can be surmised to be the *__Heat Factor Component__*

__2nd Principal Component__ **:**

This component is highly positively weighted on the following features 'total_pressure_hpc_outlet', 'engine_pressure_ratio' and significantly negatively weighted on the 'bypass_ratio'. This component can be surmised to be the *__Pressure Component__*

__3rd Principal Component__ **:**

This component is highly positively weighted on the 'cycle' feature and negatively on the 'total_pressure_bypass_duct' which primarily indicates the efficiency of the propulsion engine. Our inference on this component is interpreted as the **Aging Component** of the Engine.

To further understand the components, we have identified the top performing engines (Engine 110, 118, 124, 134 and 155*), that is the engines that have maintained Class 3 status (very confident) even after over 350 cycles and plotted all engines' cycle in the 3D principal component space. These engines are indicated in Magenta in Figure 3-4 below. We can infer that these high performing engines are generally closer to PC3 (the Aging Component) and are very likely newer engines.

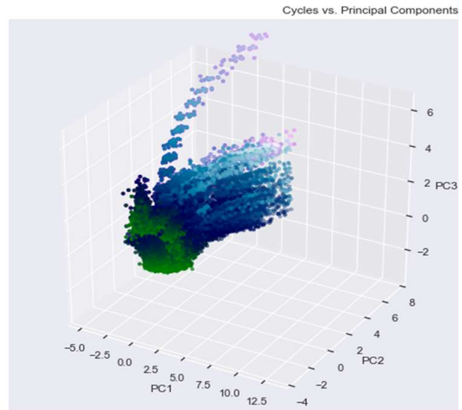


Fig3-4 Plot Engine Cycle in Eigenvector Space

In addition, the PCA loadings show cluster patterns of features (temperature and pressure) as noticed in the right side of the cartesian coordinate in Figure 3-5

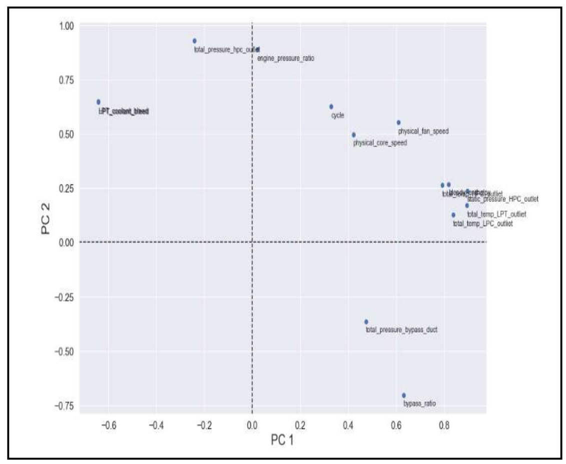


Fig3-5 Engine Features Cluster Patterns on PC1 and PC2

IV. K-MEANS USING PCA

PCA results provides three components for dimension reduction. The first component which is accountable for 41.17% of the variance can be interpreted as Heat factor. The second one and the third components can be interpreted as Pressure factor and Aging factors. All of those three principle components combined can explain 79.51% of the total variances. Projecting the original dataset to the new three-dimensional space created by those three components, we are able to run K-Means clustering algorithm time efficiently on the dataset which previously having over 20 dimensions.

In this dataset, we want to categorize all engines into four groups. Ideally, we can separate nearly broken, brand new, slightly used and heavily used engines into those four groups. We will use transformed data to run K-Means clustering and obtain centroids in form of principle components. Then transfer PCA back to the actual data points and evaluate the result on that.

	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	-0.241697	-0.834953	-0.815022	-0.895947	-0.611483	0.107090	-0.730542	-0.439700	-0.124824	-0.908186	-0.549480	-0.842458	0.531683	0.530711
1	0.694596	1.379188	1.246144	1.466886	0.509197	-0.731131	0.717960	0.629851	-0.232551	1.439344	1.246888	1.287085	-1.327606	-1.325681
2	2.006661	0.437410	0.895191	0.586520	-1.097394	3.104782	1.894435	1.652961	2.962244	0.813541	-2.325019	0.913875	2.154185	2.166216
3	-0.404491	0.092885	0.056521	0.086062	0.510690	-0.293767	0.040861	-0.142498	-0.273627	0.071582	0.344288	0.060717	-0.266448	-0.268355

Table 4-1. Initial K-Means Results.

The initial K-Means results are shown in table 4-1. Those 4 centroids are hard to compare with each other giving there are too many columns. To deal with such situation, we decide to use centroid in PCA form to evaluate the results.

	PCA1	PCA2	PCA3
0	0.404144	-0.730731	-0.433767
1	0.059135	6.893183	-0.140161
2	3.995850	-0.521936	0.531136
3	-2.385915	-0.191155	0.261354

Table 4-2. K-Means Results in PCA

Using PCA results, we have a straightforward output that indicates engine statuses in three major factors. The result is provided in table 2.2.

It is clear that cluster 0's centroid has low Heat Pressure and Aging factors comparing to other engines. Centroids in cluster 0 and 1 are also have negative Aging factor than cluster 2 and 3. The Pressure factor for cluster 1 is significantly higher than cluster 0's but 1's Aging factor is slightly higher than cluster 0's. A reasonable explanation is that new engine needs to go through rounds of testing, some parts of may use full power but some parts do not. As a result, compare to a slightly used engine, a brand-new engine may not be fully used entirely. Thus, the Aging factor is low and working factors such as Heat and Pressure factors will be similar or lower.

Judging from the precious findings, we can infer that cluster 0 and 1 are brand new and slightly used engines, which makes cluster 2 and 3 are for heavily used and nearly broken engines but to decide which cluster is for nearly broken engines, we still need more evidences.

Looking at the centroid for cluster 2 and 3 closely, we can find that although cluster 3 has a lower Aging factor, it has worse Heat factors. This factor is closely related to the combustion chamber and turbine fan's workload. This shows that the engines in cluster 2 are older but more functional than engines in cluster 3. According to this, I would like to label cluster 2 as the heavily used and cluster 3 as nearly broken or in another word, require immediate service.

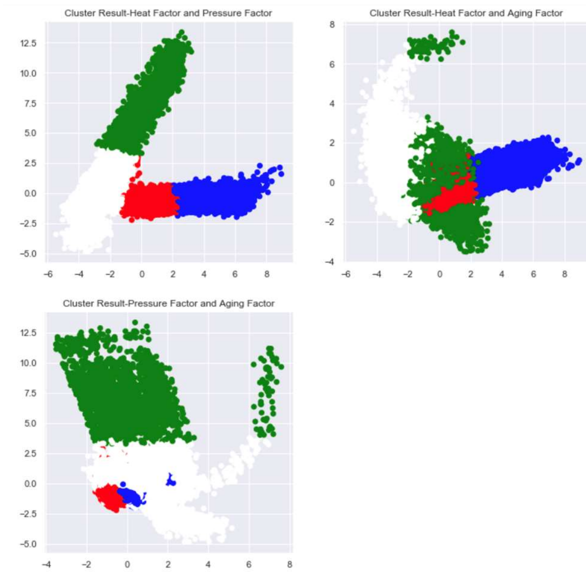


Fig 4-1. Clusters in PCA Space

Fig 4-1 shows the colored clustering result in three-dimensional space. We can find that cluster 2, which is in white color is centered at lower left corner in top-right graph, indicating that it has worst engine performances. Cluster 1 which is in green color has best performance judging by top-right graph and it also has low Aging factor compared to cluster 3.

With prior knowledge of engine statuses, we are able to test the cluster accuracy rate. After transforming all the clusters to corresponding engine statuses, we got an accuracy score of 34%, which is worse than base case. Altering the parameters to let both cluster 0 and 3 identified as statuses 3, it shows a significant improvement to 65%. The test result shows that the clustering model can identify statuses 1 and statuses 2 engines. Those are heavily or slightly used engine. Due to some reason, we cannot distinguish brand new and nearly broken engine very well. One possible explanation for this is that the engine can have failure in any stage, and it is really a matter of probability. New engine and heavily used engine all may have problem so the features in stage 0 engine may include all the features from other groups, which make this harder to be identified among others.

Overall, K-Means combined with PCA help reduce the dimension of the data from over 20 to four, PCA makes the result from K-Mean more interpretable and let K-Means runs faster. In real world scenario, with a more accurate algorithm, using K-Mean can help reduce the dimension of the whole dataset, making real-time analysis become possible.

V. SURPERVISED LEARNING ANALYSIS

As previously mentioned, reliably estimating remaining life holds the promise for considerable cost savings. In fact, that allows us to avoid unscheduled maintenance and to increase equipment usage and ensures the operational safety improvements.

Class	Description	Frequency	Percentage
0	Very urgent Maintenance	2200	4.85%
1	11 - 30 Cycles before failure	4000	8.8%
2	31 - 100 Cycles before failure	1400	31%
3	more than 101 cycles before Failure	25151	55.5%

Table 5-1 Percentage distribution of the target classes

Our initial objective is to classify our models into their various classes. The obvious class imbalance in the percentages of the minority classes (0 and 1) in the training set will pose a challenge for developing a robust model. We will need to recalibrate for a higher recall. Our ultimate objective is to identify the minority class of engines requiring very urgent maintenance and those with a few cycles before failure.
Generally, this problem deals with the trade-off between **recall** (percent of truly positive instances that were classified as such) and **precision** (percent of positive classifications that are truly positive). In this situation we are concerned more so with recall than precision, as it is usually costlier to miss detecting a failing engine than it is to falsely label a viable engine.

Two ways of fixing our imbalanced dataset involve either oversampling instances of our minority ailing engine classes that will enable us to create a balanced dataset that should not theoretically lead to classifiers biased towards one class or the other. A downside to this is that it can lead to model overfitting. On the other hand, we could under sample the majority class leaving out important instances that provide important differences between the multiple classes.

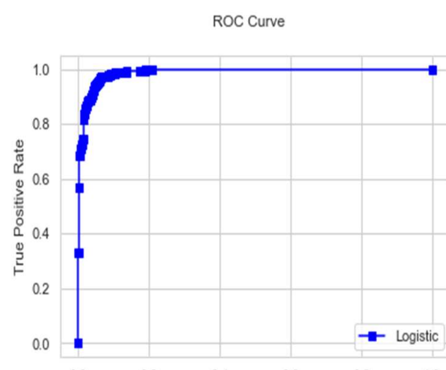
For the purposes of this project, for our first model we will jointly classify both Classes 0 and 1 (requiring maintenance) and Classes 2 and 3 (viable, requiring minimal checks). Subsequently we will proceed to classify our requiring maintenance class to see if we can identify engines that require immediate maintenance or have a few cycles before failure (creating distinction between Classes 0 and 1).

The logistic regression model was used in classifying both classes in both models using Grid Search to obtain the best parameters including the radial basis function kernels to obtain the highest recall score. The results of both models are shown below:

Model 1 Results:

	Good Condition	Maintenance Required
Good Condition	613	10
Maintenance Required	103	520

Metric	**Value**
Recall	0.98
Precision	0.83
F- Score	0.90



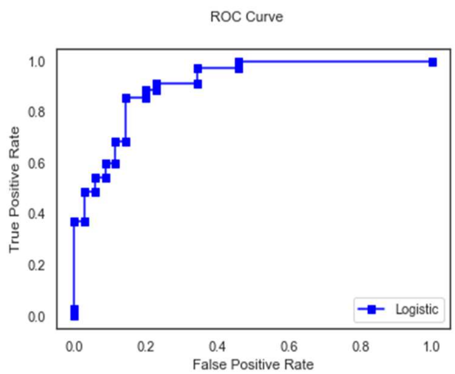
The receiver operating characteristic (ROC) curve is an evaluation method we can use to assess the efficacy of a binary classification algorithm as well as choose the optimal threshold based on our tolerance for false negatives and desire for true positives. In other words, it is a plot of the true positive rate against the false positive rate for the different possible cut points of a diagnostic test.

The ROC curve shows the tradeoff between sensitivity and specificity. The slope of the tangent line at a cutpoint gives the likelihood ratio (LR) for that value of the test and the area under the curve is a measure of text accuracy

Model 2 Results:

	Ready for Maintenance	**Urgent Maintenance**
Ready for Maintenance	31	4
Urgent Maintenance	13	22

Metric	**Value**
Recall	0.63
Precision	0.846
F- Score	0.72



Comparing the ROC curves for model 1 and model 2, model 1 has a higher accuracy with more area under the curve than the area under model 1.

CONCLUSION

Principal Component Analysis is invaluable to making sense of highly dimensional data as a dimension reduction technique to interpret complex data in a more meaningful form. Also, reliably estimating remaining life holds the promise for considerable cost savings. In fact, that allows us to avoid unscheduled maintenance and to increase equipment usage and ensures the operational safety improvements.

REFERENCES

[1] Chen, M., Mao, S. and Liu, Y. (2014) Big Data: A Survey. Mobile Networks and Applications, 19, 171-209.

[2] Salman, J., Aymen, D., Aymen, M and Ahmed A: RAMP on predictive maintenance for aircraft engines