BITS PILANI

A Report

On

Hierarchical Fault Diagnosis of Spacecraft Propulsion Systems Using Physics-Informed Machine Learning

BY

Muralidhara Samarth

2023A4PS0418P

Table of Contents

1. Introduction

2. EDA & Feature Engineering

3. Methodology

4. Results and Discussion

1 Introduction

1.1 Problem Context & Motivation

Reliability is of great importance in Spacecraft propulsion systems that operate under extreme conditions for mission success. Common failure modes in these systems include solenoid valve faults which occur as a result of incomplete opening/closing leading to reduced fluid volume and bubble anomalies that occur due to gas contamination, thereby affecting fluid flow. Detecting these faults early is essential to avoid failure during the execution of a mission.

However, a major challenge in deploying Prognosis and Health Management systems arises due to the minute differences between each spacecraft. For instance, a model trained on a 3 spacecraft located in the same place may perform poorly for a completely unseen spacecraft located elsewhere, possibly due to subtle sensor calibration drifts and mechanical variances, a phenomenon in ML formally known as domain shift. Standard data-driven models usually overfit to the training data (3 known spacecraft), and consequently perform poorly in generalizing the test data (1 unseen spacecraft).

1.2 Objective

The objective of this project is to develop a hierarchical and data-driven machine learning framework that is capable of diagnosing faults in a spacecraft's propulsion system. The system therefore aims to perform a 5-level fault diagnosis:

- **Detection:** Identifying if the system is normal or abnormal (Task 1).
- **Classification:** If Abnormal, differentiating between bubble anomalies and valve faults (Task 2).
- **Localization:** Finding the location of the corresponding anomaly/valve fault in the propulsion system (Tasks 3 &4).
- **Quantification:** Estimating a valve's opening ratio, if identified as a solenoid valve fault (Task 5).
- **Anomaly Detection:** Identifying the outliers in the form of Unknown anomalies not present in the training data (part of Task 1).

1.3 Proposed Approach

This project makes use of a Physics-Informed Feature Engineering approach so as to overcome the challenges related to domain shift described previously. While baseline statistical features achieved only 58% training accuracy (which almost certainly guarantees lower generalization accuracy after comparing with ground truth), the proposed approach which makes use of Fast Fourier Transform (FFT) to capture acoustic signatures and Time-Segmentation to capture key information during each valve open-close cycle has achieved a generalization accuracy of 76.09% when compared to the ground truth after testing. The proposed approach further integrated an Isolation Forest Gatekeeper to detect "Unknown" anomalies which is an unsupervised ensemble learning technique to detect outliers in a dataset.

2 EDA & Feature Engineering

2.1 Dataset Description

The given training dataset consists of 177 time-series samples in the form of csv files that represent the pressure fluctuations in 7 pressure sensors for a period of 1.2s, sampled at 1 kHz recorded multiple times for three spacecraft (Spacecraft 1, 2, and 3). The corresponding test dataset consists of 46 csv files belonging to a known Spacecraft 1 and an unseen Spacecraft 4.

2.2 Exploratory Data Analysis

The pressure variations with time for the normal and abnormal (bubbly anomaly and valve fault) cases appeared visually identical in the time domain, dominated by the large pressure spikes, just as the valve begins to close. Since all the plots are visually similar, no valuable insights are seen.

However, by subtracting the mean normal signal from that of an abnormal subtle, high-frequency residuals are obtained (Figure 1). This further indicates that fault information was hidden in the frequency domain (acoustics), rather than in global statistical shifts.

Therefore this rudimentary visualization of pressure variation in EDA tells that fault detection can indeed be done on the dataset by analyzing individual pressure fluctuations with time.

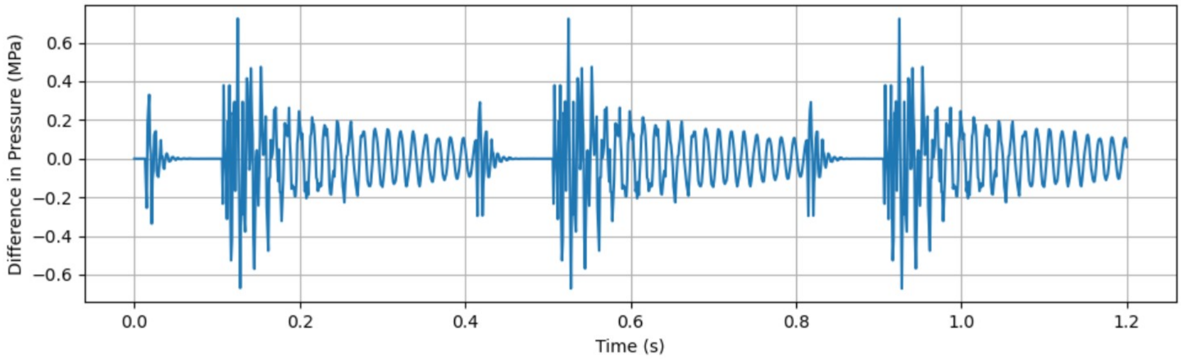


Figure 1: Difference Plot between normal and abnormal pressure data

2.3 Baseline Feature Engineering

The initial baseline feature engineering involved extracting 6 standard time-domain statistical features, namely mean, standard deviation, minimum, maximum, skewness, and kurtosis for each sensor.

This yielded a total of 42 features for all the 7 sensors. However, only a 58.7% training accuracy was obtained in the binary classification problem for normal vs abnormal (both fault types included) cases, which almost guarantees a lower testing accuracy. Moreover, the low training accuracy of 58.7% is even lower than random guessing for the normal case (35/59 = 59.32%)

This signifies that the computed global statistics failed to capture the subtleties of the data in the form of acoustic vibration of bubbles or the transient dynamics of valve cycles.

2.4 Physics-Informed Feature Engineering

To overcome the limitations faced in the baseline feature engineering, a Physics-Informed feature engineering approach was developed which took into account data that was related to the physics of the propulsion system:

2.4.1 Time-Segmentation

Since the solenoid valves actuate three times once each at around 100ms, 500ms and 900ms respectively, a global maximum pressure, which was calculated in the baseline approach for the entire 1.2s period can mask a fault that only occurs in one open-close cycle. To account for this phenomenon, 1.2s signal was segmented into three distinct 400ms windows and calculated the local maximum pressure standard deviation for each segment independently. This was done twice every cycle for each sensor, thereby creating 42 features totaly. This therefore allows the model to detect transient mechanical failures that may arise in any of this individual cycles.

2.4.2 Frequency-Domain Analysis

It is widely known that bubbles can alter the speed of sound in a fluid flow by altering the acoustic modes of the system. In addition, it is also known that such acoustic variations, which are very subtle to identify in time domain data are much easier to identify in the corresponding frequency domain. Therefore, the Fast Fourier Transform (FFT) is applied to convert the data from time-domain to frequency-domain and to obtain:

- **Dominant Frequency Power:** The magnitude of the primary resonance (eg. P1_fft_max).
- **Energy Bins:** Energy was separated into low-frequency (0–50 Hz) which represents the water hammer thud and high-frequency (50 Hz) which represents acoustic ringing (eg. P1_fft_low and (eg. P1_fft_high).

These three numbers was calculated for each sensor, creating another 21 features. We also include the 42 basic statistical features used in the baseline feature engineering approach. Therefore the final training dataset has a total of 105 (42+21+42) features for each of the 177 samples.

Since the physics-informed feature engineered training and test datasets capture more relevant useful information of the system, the hierarchical ML model described below uses the same for training, validating and testing instead of the baseline feature engineered datasets. This is further proven by the validation accuracies of each task described later in the report.

3 Methodology

3.1 Architecture Overview

Due to the nature of the given tasks, a divide and conquer hierarchical pipeline is developed instead of a single, universal model that handles all the data at one. This pipeline materializes the natural diagnostic decision process wherein the existence of a fault is first identified before attempting to classify fault types (valve fault or bubble anomaly). That is

Level 1 : Abnormal datapoint detection : Determines if the system is normal or not normal (includes both types of faults along with the unknown case which acts as the outlier).

Level 2 : Anomaly Detection : If a datapoint is classified as not normal from Level 1, does the datapoint belong to the unknown class or does it have some defined fault (valve fault/bubble anomaly).

Level 3 : Fault type, location and information detection : If datapoint is classified to have some defined fault from Level 2, this level answers what type of fault does it have, where it is located and the valve's opening ratio if detected as a solenoid valve fault.

3.2 Abnormal Datapoint Detection

For classifying points into normal and abnormal (includes faults and unknown datapoints), which is part of Task 1, a K-Nearest Neighbors (KNN) classifier with k=7 is used. This choice of k was obtained after hyperparameter tuning by plotting accuracy vs increasing odd k.

The first choice of doing this task was by using the baseline feature engineered features to train a generic Decision Tree (DT). When the training data was split using the hold out principle and made to run on the DT, a training accuracy of 100% was obtained. While this is a welcome result, it can be highly misleading as it may mean extreme overfitting of the model, which may lead to extremely poor testing/generalization performance. In addition, it also does not account for the unseen spacecraft data problem that is mitigated using GroupKFold.

While an ensemble technique like Random Forest (RF) is generally robust, it struggled with the domain shift during training in Task 1. This domain shift arises as GroupKFold (Section 3.5) has been used to decide which models to use for later testing by assessing the validation accuracies. A possible reason for this could be the fact that KNN, unlike RFs relies on geometric distance in the feature space rather than rigid decision boundaries. It thus proved to be more resilient when training data was validated using GroupKFold. Further, a validation accuracy of 80.79% is obtained for the physics-informed feature engineered training data fed to KNN, compared to Random Forest's 46.89%.

However, an interesting fact is that the RF's validation accuracy is higher when it is fed with the baseline feature engineering training dataset than when it is fed with the physics-informed feature engineered training dataset.

3.3 Anomaly Detection

To handle Unknown anomalies, whose instances are not given in the training dataset, an Isolation Forest model has been trained on the non normal dataset (obtained from abnormal datapoint detection) which only includes points that either have some defined fault or have some unknown fault.

Isolation Forest is a model that is an unsupervised learning technique commonly used for anomaly detection wherein the model learns the boundary of Normalcy (behaviour of a non-outlier, not the normal class). Any datapoint falling significantly outside this boundary (Anomaly Score = -1) is classified as Unknown.

3.4 Fault type, location and information detection

For the datapoints not classified in the Unknown class, Random Forest Classifiers (RFCs) have been used to classify the fault type and the corresponding location. To find the opening ratio of a valve affected by a solenoid fault, we employ Random Forest Regressors (RFRs).

Random Forests are known to handle high-dimensional data effectively and are robust to noise. In addition, they have a reduced tendency to overfit as it is a bagging ensemble technique. The RFC used in Task 2 obtained a high validation accuracy of 97.22%, proving to be a good choice for fault type classification. The respective validation accuracies for Tasks 3 and 4 which involved finding the location of the fault were 75% and 83.34% which are decent marks of performance. Since Task 5 is a regression problem, the RFR obtained a validation Mean Absolute Error (MAE) as 11.43% opening ratio which is once more a decent measure of model performance. Therefore, the ensemble nature of RFCs and RFRs allows them to capture complex interactions between sensors without explicit rule programming.

3.5 Validation Strategy

A crucial pivot in this project was to validate the predictions using GroupKFold Cross Validation, grouped by the Spacecraft# column instead of computing regular training accuracy using the holdout principle.

Standard splitting according to the holdout principle would allow the model to memorize specific facts of each spacecraft, thereby performing well during training, but perform poorly during testing when it encounters unseen data of Spacecraft 4. GroupKFold forces the model to train on 2 spacecraft and test on the other, and this process runs thrice, once for each spacecraft, simulating the challenge during testing of the unseen Spacecraft 4. The validation accuracy is then the average of the training accuracies obtained in each training iteration (fold). This gives a realistic estimate of generalization performance.

4 Results and Discussion

4.1 Task Validation Accuracies: Baseline vs Physics-informed

The below table summarizes the validation accuracies for Tasks 1-5 between baseline and physics-informed feature engineering techniques for the entire training datasets.

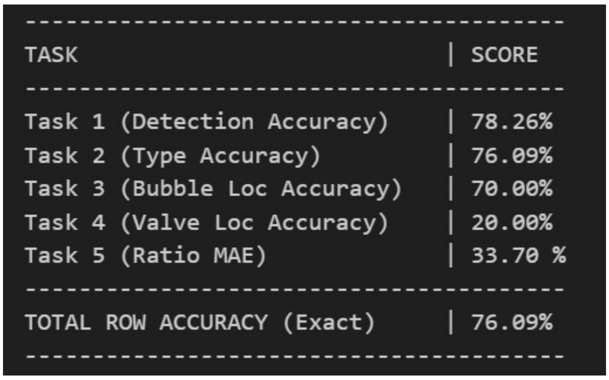
Task	Baseline	Physics-informed
Task 1	58.75%	80.79%
Task 2	98.61%	97.22%
Task 3	91.67%	75.00%
Task 4	93.75%	83.33%
Task 5	ll.84% MAE	ll.44% MAE

Individual tasks are performed by individual models trained on the respective entire feature-engineered training dataset. We observe that although the validation accuracies for tasks 2-5 are greater in the baseline dataset, a similar performance will not be seen in test because in reality, the tasks are done in hierarchy. Since there is a low abnormal datapoint detection accuracy, for the baseline dataset, there is a high chance that the wrong data will get passed to the further tasks in the hierarchy, thereby performing poorly in testing.

However, a higher abnormal datapoint accuracy is seen for the model trained on physics informed dataset with decent validation accuracies for the remaining tasks, hence indicating a better performance during testing.

4.2 Actual Performance during Testing

The following results were obtained after comparing the predicted values given by the hierarchical ML model that uses KNN, RFC and RFR, trained on the physics-informed feature engineered training dataset with the provided ground truth values.



```
----------------------------------------
TASK                            | SCORE
----------------------------------------
Task 1 (Detection Accuracy)     | 78.26%
Task 2 (Type Accuracy)          | 76.09%
Task 3 (Bubble Loc Accuracy)    | 70.00%
Task 4 (Valve Loc Accuracy)     | 20.00%
Task 5 (Ratio MAE)              | 33.70 %
----------------------------------------
TOTAL ROW ACCURACY (Exact)      | 76.09%
----------------------------------------
```

We have obtained a decent testing accuracy of 76.09%. However, we observe a very low accuracy for Task 4 which may be because of a cascading effect wherein the wrongly filtered data reach the model which further wrongly identifies the location of the solenoid valve fault.

5 Conclusion

In this project, a hierarchical, physics-informed diagnostic machine learning framework for spacecraft propulsion systems was developed. By upgrading from simple statistical features to time-segmented and frequency-domain features, abnormal datapoint detection training accuracy rose from a baseline of 58% to a robust 80%. Overall, the hierarchical ML framework produced a good accuracy of 76.02% on completely unseen data.

This project further showcased that how Isolation Forests can identify unknown system behaviors. The performance drop on the unseen test set (compared to training) highlights the critical challenge of domain shift often seen in the aerospace industry.

In the future, the Task 1 validation accuracy for the baseline case should be improved and worked upon as the validation accuracies for all the subsequent tasks are significantly higher than that of the physics-informed case, which may indicate a better performance in testing, although chance of overfitting very much persists. A viable explanation as to why the validation accuracies for the tasks 2-4 are lower for the physics-informed case could be the Curse of Dimensionality as this case contains 63 more features than the baseline case which only contains 42 features.