Maneuver Detection in Orbital Data: Machine Learning Approach

DIGANTARA

Submitted by: Arya Chakraborty
Date: July 19, 2024

Introduction

Detecting maneuvers in orbital data is critical for maintaining the accuracy and reliability of satellite operations. Maneuvers such as engine burns or orientation adjustments need to be identified accurately to ensure the satellite is functioning as expected. This study aims to develop a method for automatic maneuver detection using a specific machine learning approache.

The data used in this study consist of the semi-major-axis (SMA) variations over time. The goal is to detect potential maneuvers using only these variations, without explicit maneuver data. To assess the accuracy of the detection methods, reference graphs and tables with known maneuvers are provided.

Assumptions

Data Quality and Completeness

- The provided semi-major axis (SMA) data is accurate and free from significant measurement errors or missing values that could affect the anomaly detection results.

Maneuver Characteristics

- Orbital maneuvers result in significant changes in the semi-major axis (SMA), making them detectable through variations in the SMA data.
- Maneuvers occur infrequently and are sparse in the dataset compared to normal orbital variations.

Model Assumptions

- The Local Outlier Factor (LOF) algorithm's assumptions about local density and neighborhood relationships are valid for the dataset and the problem of maneuver detection.
- The choice of parameters (e.g., `n_neighbors`, `contamination`) for the LOF model is appropriate for detecting outliers in this specific dataset.

Reason to Choose Machine Learning Approach

Machine learning models, such as the Local Outlier Factor (LOF), can effectively detect anomalies in complex datasets. These models are capable of identifying non-linear patterns and subtle deviations that heuristic methods might miss. Given the challenge of detecting maneuvers like engine burns or orientation adjustments using only the semi-major axis (SMA) variation, ML models are better suited for this task because of their ability to learn from the data and generalize better.

Used Machine Learning Model

Machine learning methods learn patterns from the data to detect anomalies. The specific machine learning model used in this study is:

- **Local Outlier Factor (LOF):** This algorithm identifies anomalies by assessing the local density deviation of a given data point concerning its neighbors. It computes the local density of a point p using the concept of k-distance and reachability distance. The k-distance of p, denoted as k-distance(p), is the distance to the k-th nearest neighbor. The reachability distance of point p from point o is defined as:

$$\text{reach-dist}_k(p, o) = \max(\text{k-distance}(o), \text{dist}(p, o))$$

The local reachability density (LRD) of p is then calculated as:

$$\mathrm{LRD}_k(p) = \left(\frac{\sum_{o \in N_k(p)} \text{reach-dist}_k(p, o)}{|N_k(p)|} \right)^{-1}$$

where $N_k(p)$ is the set of k-nearest neighbors of p. The LOF score of p is the average ratio of the LRD of p and its neighbors:

$$\mathrm{LOF}_k(p) = \frac{\sum_{o \in N_k(p)} \frac{\mathrm{LRD}_k(o)}{\mathrm{LRD}_k(p)}}{|N_k(p)|}$$

A score significantly greater than 1 indicates an outlier.

The experiment involves preprocessing the data, extracting relevant features, and applying machine learning method to detect anomalies. The anomalies are detected using a threshold value and then evaluated using an accuracy metric.

Feature Extraction and Exploratory Data Analysis

To better understand the data and prepare it for analysis, several features were extracted from the 'Datetime' column, and exploratory data analysis (EDA) was performed.

Feature Extraction

The 'Datetime' column was converted from object type to datetime type, and the following features were extracted:

- Year
- Month
- Day
- Hour
- Day of the Week
- Day of the Year

Exploratory Data Analysis

The following graphs were plotted to understand the distribution and patterns in the data:

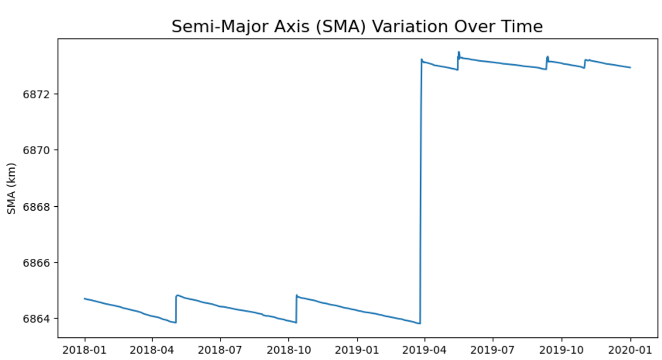

Figure 1: Raw SMA vs. Date Plot

Explanation

The above raw SMA vs. date plot shows variations in the semi-major axis (SMA) over time, highlighting periods of stability and sudden changes, which can indicate maneuvers.

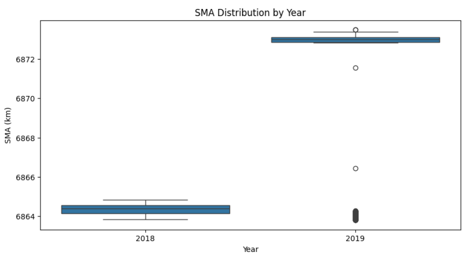

Figure 2: SMA Distribution by Year

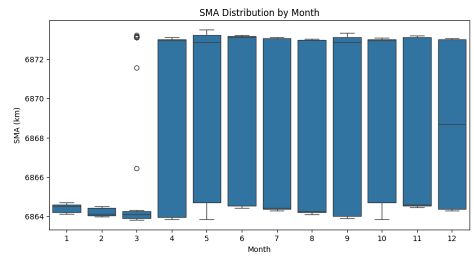

Figure 3: SMA Distribution by Month

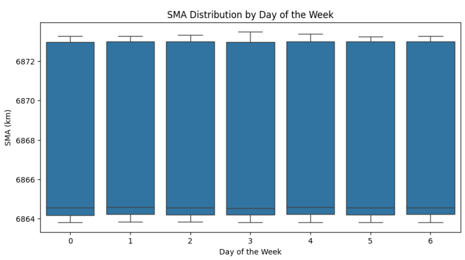

Figure 4: SMA Distribution by Day of the Week

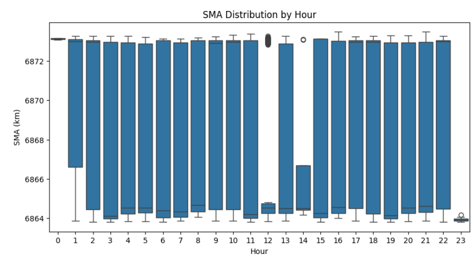

Figure 5: SMA Distribution by Hour

Explanation

SMA Distribution by Year: Shows There are no significant outliers in the 2018 data. In 2019 there are several outliers below the lower quartile and above the upper quartile. These outliers might indicate maneuvers in the orbit of the satellite.
SMA Distribution by Month: Shows Higher variability is observed from April onwards, as indicated by the larger IQR. Several outliers are present, especially in March, suggesting possible orbital maneuvers or anomalies.
SMA Distribution by Day of the Week: Shows that there is not such difference in the SMA distributions by Day of the week.
SMA Distribution by Hour: Indicates that most of the outliers are occurring at a certain time of each day. We can assume that most of the maneuvers occur at that time.

Maneuver Detection using Local Outlier Factor: Result

The final visualization of the performance of the Local Outlier Factor(L.O.F.) model on the orbit data is shown below.

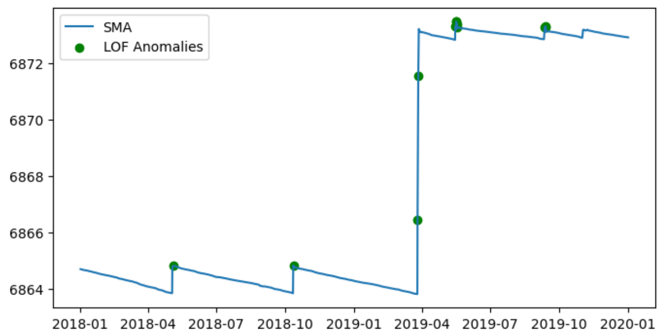

Figure 6: Maneuver Detection using Local Outlier Factor

Now, as we can see, there are multiple **Maneuvers** are detected, we can visualize each anomaly point in the chart below to check the model's performance. Here we have considered five dates before and after the anomaly point and have plotted those in a chart to check the result very closely.

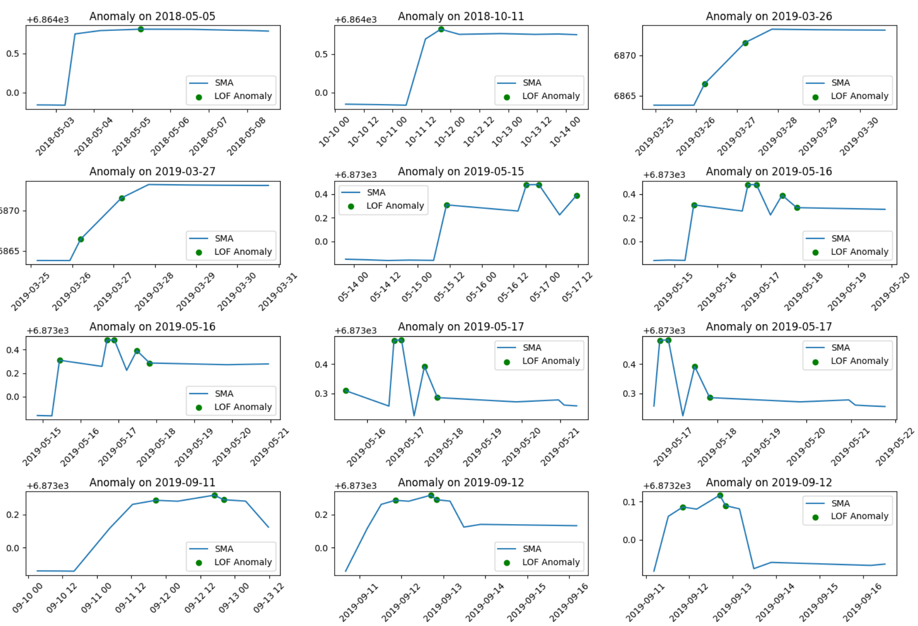

Figure 7: Sub-plots for each Maneuver

Conclusion

The above sub-plots indicate that there are many false positives while detecting maneuver/anomaly. Although the **Local Outlier Factor (LOF)** algorithm detected some maneuvers, its performance was not optimal, possibly due to the limited dataset (2000 points) and the complexity of the data. Future improvements can include exploring techniques like **Autoencoders**, **LSTM**, and **K-Means Clustering** for better anomaly detection. Expanding the dataset to include more years and applying hyperparameter tuning (e.g., adjusting `n_neighbors` and `contamination`) can also enhance the model's accuracy and reliability in identifying maneuvers.