A Report on: Detecting Spacecraft Anomalies Using LSTMs and Nonparametric Dynamic Thresholding

AVISHA BHIRYANI, Technical University Dortmund, Germany

ACM Reference Format:
Avisha Bhiryani. 2022. A Report on: Detecting Spacecraft Anomalies Using LSTMs and Nonparametric Dynamic Thresholding. 1, 1 (March 2022), 11 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Within 84 seconds of the launch of the Columbia STS-107 space shuttle,

> a piece of foam shed from the external fuel tank and struck the edge of the orbiter's left wing which jeopardized the thermal protection system of the orbiter during the re-entry into Earth's atmosphere."[1]

This turned out to be the very reason for the loss of the Orbiter. The Data Processing System part of the Modular Auxiliary Data System(MADS) on the Columbia Orbiter was responsible to detect and bring to attention vehicle system failures and out-of-tolerance system conditions if any. [1]

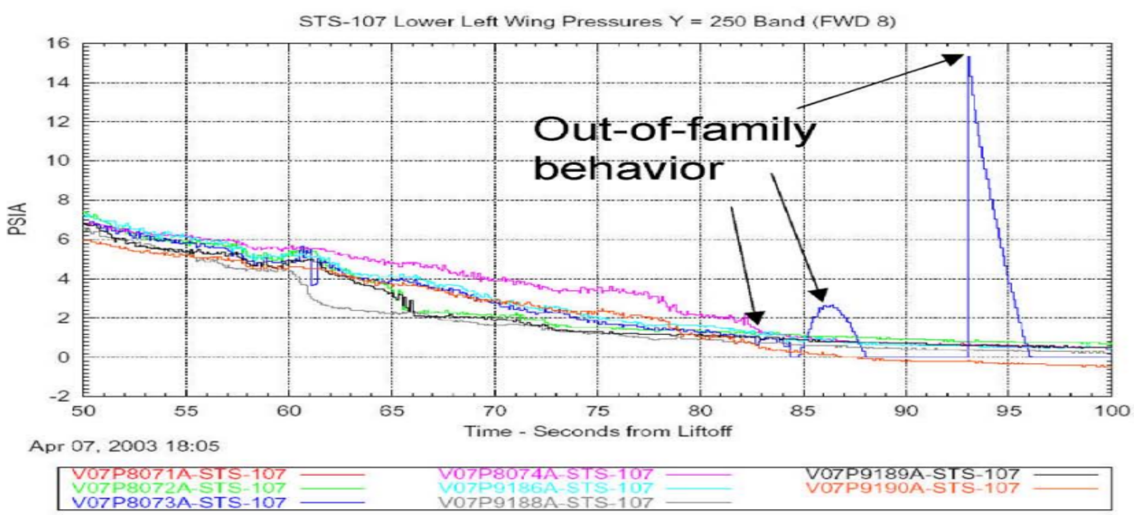

Fig. 1. Anomalous behaviour is clearly noticeable in the pressure sensors of the left wing on being struck but was not communicated by the system. (https://history.nasa.gov/columbia/Troxell/Columbia).

Author's address: Avisha Bhiryani, Technical University Dortmund, Dortmund, Germany, avisha.bhiryani@tu-dortmund.de.

Figure 1 depicts the Pressure taps on the lower left wing of the Orbiter during the launch clearly show that pressure tap V07P8073 among others has been hit at 84 seconds. This data was the first clue of something wrong with the system but was not reported anomalous by the system, so the crew on board as well as the ground crew were unaware about it. Anomaly detection system in spacecrafts are a critical tool to monitor thousands of different telemetry channels measuring temperature, radiation, power, instrumentation, and computational activities and to avoid potential hazards by reducing the amount of manual intervention needed.[1][4]

2 PAPER SUMMARY

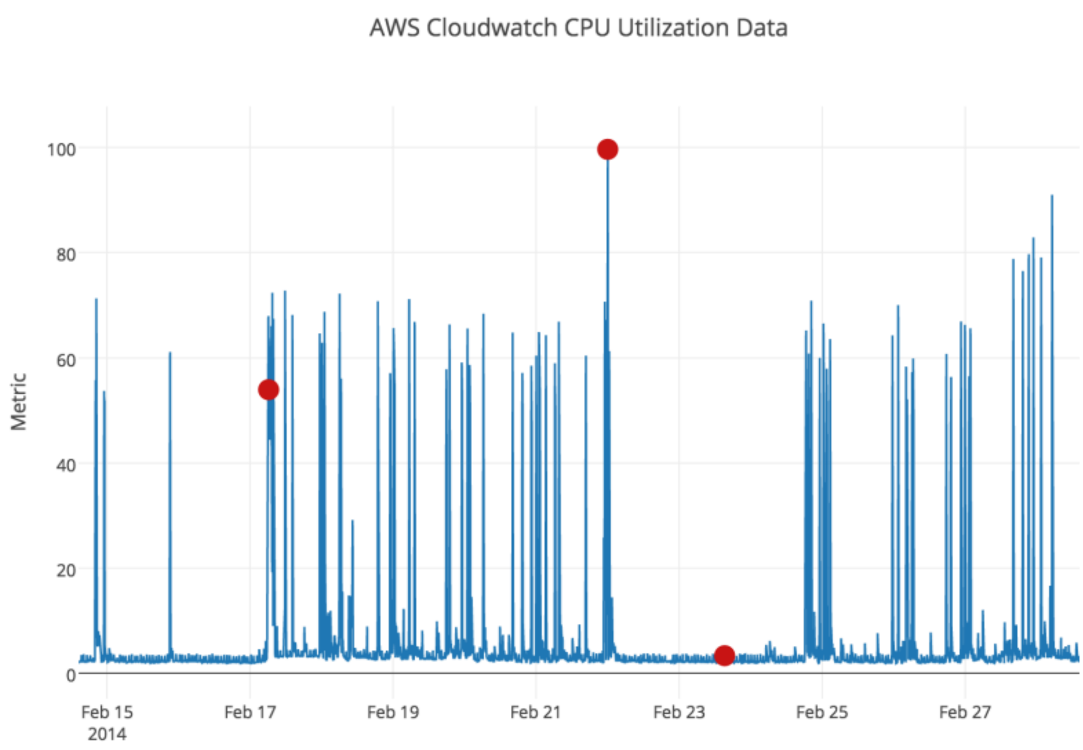

Fig. 2. Red circles represent anomalies. The first anomaly is difficult to detect as the jumps in the data do not return to the baseline as expected, and it emerges as the new normal pattern. The second anomaly is a simple abrupt change in the data after which the system returns to previous patterns. As the jumps normally seen in the data does not return for a long time, the third red circle is considered an anomaly.([5]).

Patterns in data that do not follow the past behavior of the data are called anomalies. As stated by the authors, "As spacecraft telemetry data is time dependent multivariate data, anomaly detection methods for multivariate time series data can also be used for spacecraft telemetry data. Data being monitored are often noisy, high-dimensional, highly non-stationary and dependent on current context." Due to fewer number of anomalies and also a lack of labelled anomalies unsupervised or semi-supervised anomaly detection methods have become necessary to be used. Specific to spacecraft, finding the source channel or system of anomalies and minimal number of false positives is also important.

The authors of the paper have provided a 3-step procedure for anomaly detection in spacecraft telemetry data and they also claim that the same method can be applied to any multivariate time series data. Firstly, LSTMs are used to predict high volume telemetry data per channel using inputs as previous telemetry values and encoded command information for the channel. Secondly, an unsupervised thresholding method is then used to determine whether the resulting prediction errors represent spacecraft anomalies by automatically assessing a high number of telemetry data channels. Lastly, multiple strategies for mitigating false positive anomalies are applied.[4][5]

2.1 LSTM - type of Recurrent Neural Network

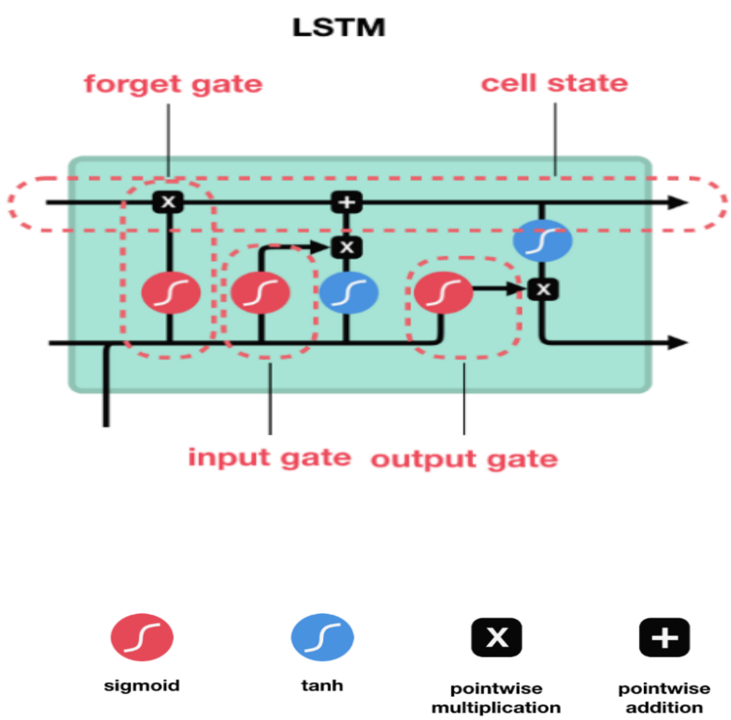

Fig. 3. Long Short Term Memory ([8]).

At the heart of a LSTM is the presence of cell state and the different gates. Information from earlier time steps, if relevant, can be made available throughout the processing of the input data through the cell state. So by using cell state, LSTMs overcome the short term memory issue of RNNs. The Forget gate discards information that is not useful from the prior time steps, the input gate adds the information that it considers to be useful for prediction from the current time step and the output gate determines what the next hidden state should be. The accuracy of an anomaly detection method highly depends on the time window or the number of previous values considered. Because LSTMs can learn long term correlations in the input data and overcome Vanishing Gradient problem, there remains no need to decide a time window for the prior values to be considered.[8] Due to the above mentioned properties of LSTM, a single model is created for each telemetry channel using LSTM which predicts values for the respective channel.

2.2 Telemetry Value Prediction using LSTM

Every channel being monitored is modelled separately because LSTMs struggle to accurately predict high dimensional data. Traceability of an anomaly to its subsystem level of origin is important as it speeds up the process to handle the anomaly. Modelling each channel independently also allows this level of traceability. "If that was not the case, for example, operations engineers would still need to review a number of different channels and alarms across the entire subsystem to find the source of the issue."[4]

$$\hat{y}^{(t)} = \left\{ \begin{bmatrix} x_1^{(t-l_s)} \\ x_2^{(t-l_s)} \\ \vdots \\ x_m^{(t-l_s)} \end{bmatrix}, \ldots, \begin{bmatrix} x_1^{(t-1)} \\ x_2^{(t-1)} \\ \vdots \\ x_m^{(t-1)} \end{bmatrix}, \begin{bmatrix} x_1^{(t)} \\ x_2^{(t)} \\ \vdots \\ x_m^{(t)} \end{bmatrix} \right\} \tag{1}$$

$$y^{(t)} = \begin{bmatrix} x_1^{(t+1)} \\ x_2^{(t+1)} \\ \vdots \\ x_m^{(t+1)} \end{bmatrix} \tag{2}$$

$$e^{(t)} = [\hat{y}^{(t)} - y^{(t)}] \tag{3}$$

$$e = [e^{(t-h)}, \ldots, e^{(t-l_s)}, \ldots, e^{(t)}] \tag{4}$$

The input variables represent a time series such as $X = \{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(n)}\}$ where each step $\mathbf{x}^{(t)} \in R^m$ in the time series is an m-dimensional vector $\{x_1^{(1)}, x_2^{(2)}, \ldots, x_m^{(t)}\}$. Length of the previous data used for prediction is represented by l_s. The number of time steps to be predicted ahead is represented by l_p, which in this case is set to 1, to limit to one time step ahead prediction which decreases the processing time. The number of dimensions used for prediction is represented by d, which is set to 1 here as each telemetry channel is being modelled individually. As a result, $\hat{y}^{(t)}$ represents a single valued scalar prediction at each step t and is used to calculate the prediction error represented by $e^{(t)}$. Each of these prediction errors is combined into a one dimensional vector of errors as in Equation 4, where h is the number of historical error values used to evaluate current errors.[4]

2.3 Dynamic Error Thresholding

With spacecraft anomalies, we are only interested in abnormally high or low anomalies among all the anomalies detected. As an LSTM model is not the best at predicting sudden changes in values, this can result into sharp rise in prediction errors. These sudden changes in errors have to be smoothed because it could be normal system behaviour and not an anomaly.[2][4]

$$e_s = [e_s^{(t-h)}, \ldots, e_s^{(t-l_s)}, \ldots, e_s^{(t-1)}, e_s^{(t)}] \tag{5}$$

To generate these smoothed errors Exponentially Weighted Average method is used as in Equation 5.Exponentially Weighted Average method as a smoothing filter performs a weighted average of the input values with the weights decreasing exponentially.[7]. Once we obtain the smoothed errors, each of these is compared to a threshold value and is classified as an anomaly only if the smoothed error value is greater than the threshold.[4]

$$\epsilon = \mu(e_s) + z\sigma(e_s) \tag{6}$$

$$\epsilon = argmax(\epsilon) = \frac{(\Delta\mu(e_s)/\mu(e_s)) + (\Delta\sigma(e_s)/\sigma(e_s))}{|e_a| + |E_{seq}^2|} \tag{7}$$

$$s^{(i)} = \frac{max(e_{seq}^{(i)}) - argmax(\epsilon)}{\mu(e_s) + \sigma(e_s)} \tag{8}$$

To determine the threshold:

- In Equation 6, in order to consider a range of possible values for the threshold, a set of positive values for standard deviation represented by z are considered above the mean $\mu(e_s)$. According to experimental results, values between 2 and 10 are considered for z.
- From Equation 7, out of all the threshold values a final threshold value is selected that would decrease the mean and standard deviation of the smoothed errors es by the maximum. The denominator penalizes large numbers of anomalous values $|e_a|$ and anomalous sequences $|E_{seq}|$ to mitigate greedy behaviour.
- In Equation 8, each anomaly is assigned a normalized score based on the distance between the anomaly and the chosen threshold which indicates the severity of the anomaly.[4]

2.4 Mitigating False Positives

The amount of previous data used for prediction in turn hugely decides the accuracy of the anomalies detected. Large amount of historical data creates a bottleneck in terms of processing time and power required but can output precise results. Whereas, if insufficient amount of historical data is used it can lead to anomalies being detected that are considered anomalous only because a narrow context is being considered.

$$\begin{gathered} e_{max} = [0.01396, 0.01072, 0.00994] \\ p = 0.1 \\ d^{(1)} = 0.23 > p \\ d^{(2)} = 0.07 < p \end{gathered} \tag{9}$$

In order to achieve a trade-off between false positive and memory usage and computation cost, a pruning method is employed. Firstly, maximum values from all the anomalous error sequences are sorted in descending order and represented in e_{max}. e_{max} also contains a non anomalous smoothed as the last entry because it is below the threshold. Secondly, each value from e_{max} is considered and a percent difference between consecutive values is calculated. If this percentage difference exceeds a minimum percentage decrease p, then the error and its corresponding error sequence stays categorized as anomalous. If it does not exceed p, then the corresponding error sequence is reclassified as non-anomalous. Also, once we obtain a small amount of data classified as anomalous, we can use this information under the assumption that similarly valued anomalies do not occur repeatedly within the same channel. So minimum score s_{min} is used to reclassify anomalies that have a magnitude $s < s_{min}$.[4]

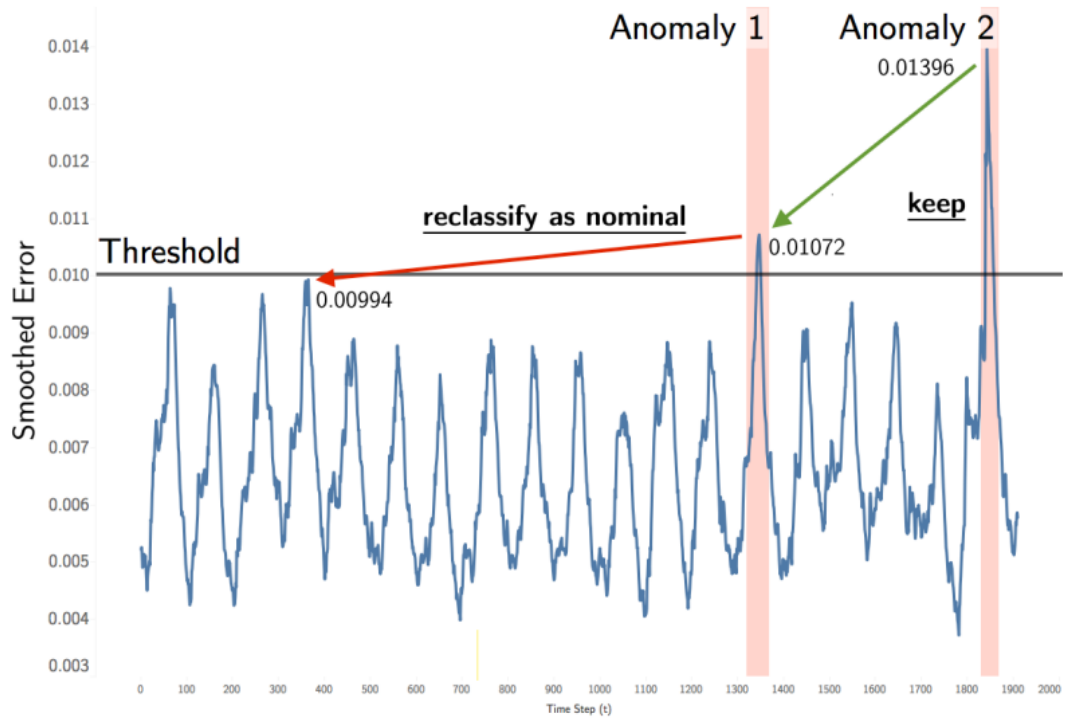

Fig. 4. Pruning Method [4]

2.5 Experiment Results

In order for testing and evaluating the proposed anomaly detection method, data from spacecrafts such as SMAP (Soil Moisture Active Passive) and MSL (Mars Science Laboratory) are used. Data from both these spacecrafts are quite different and have different set of issues when it comes to anomaly detection, and hence perfect candidates to evaluate the proposed method. SMAP has routine and regular behaviour and thus the resulting telemetry can be more easily predicted with less training and less data. Contrary to SMAP, MSL performs a wide variety of behaviours. In order for obtaining some labelled anomalies, manual inspection of Incident Surprise Anomaly Reports (ISA) was performed. It is a possible scenario in which the same anomalous event can reflect in multiple channels as anomalous, so only one was considered for the experiment.[4]. As stated by the authors, A 5 day span is considered around the anomalies, primary anomaly occurring time t_a and the span is $t_s = t_a - 3d$ to $t_f = t_a + 2d$. Final decision for a predicted anomalous event in comparison with the labelled anomalous sequence is made according to the following:

- True Positive: As it might indicate the same anomalous sequence, a single true positive value is recorded even if portions of multiple predicted sequences match a labelled anomalous sequence
- False Positive: All the predicted sequences that do not overlap a labelled anomalous region are marked false positive.

Table 1. LSTM Model Parameters per channel from Experiment Results [4]

Hidden Layers	2
Neurons in each Hidden Layer	80
Sequence Length	250
Training Iterations	35
Drop out	0.3
Batch Size	64
Optimizer	Adam

2.5.1 Model Parameters. For each telemetry channel, values in 1 minute time window are considered and are evaluated together in batches of 70 minutes. A prediction is made using 2100 of prior values and this prediction is then used to evaluate each of the 70 minute batch of values. The p parameter is an important parameter to control precision and recall, and an appropriate value can be inferred when labels are available. According to the experimental results, a p values between (0.05,0.20) was found to be reasonable.[4]

Table 2. Result for each spacecraft [4]

Thresholding Approach	Precision	Recall	F(0.5) score
Non parametric with Pruning(p = 0.13)			
MSL	92.6%	69.4%	0.69
SMAP	85.5%	85.5%	0.71
Total	87.5%	80%	0.71
Non parametric without Pruning(p = 0)			
MSL	75.8%	69.4%	0.61
SMAP	43.0%	92.8%	0.44
Total	48.9%	84.8%	0.47
Gaussian tail(ϵ_{norm} = 0.0001)			
MSL	84.2%	44.4%	0.54
SMAP	88.5%	78.3%	0.71
Total	87.5%	66.7%	0.66
Gaussian tail(ϵ_{norm} = 0.01)			
MSL	61.3%	52.8%	0.48
SMAP	82.4%	81.2%	0.68
Total	75.8%	71.4%	0.62
Gaussian tail with Pruning($\epsilon_{norm} = 0.01, p = 0.13$)			
MSL	88.2%	41.7%	0.54
SMAP	92.7%	73.9%	0.71
Total	91.7%	62.9%	0.66

2.5.2 Final Experiment Result. As seen from Table 2, pruning with a non-parametric approach increases precision from 48.9% to 87.5%, that is an overall increase of 38.6%. However, it reduces recall just by 4.8% (from 84.8% to 80.0%). The parametric approach, Gaussian Tail, despite of using different parameter settings gives low precision and recall values. Applying pruning method to the parametric approach, improves precision but with a high recall which results in F score quite lower as compared to the non-parametric approach with pruning.[4]

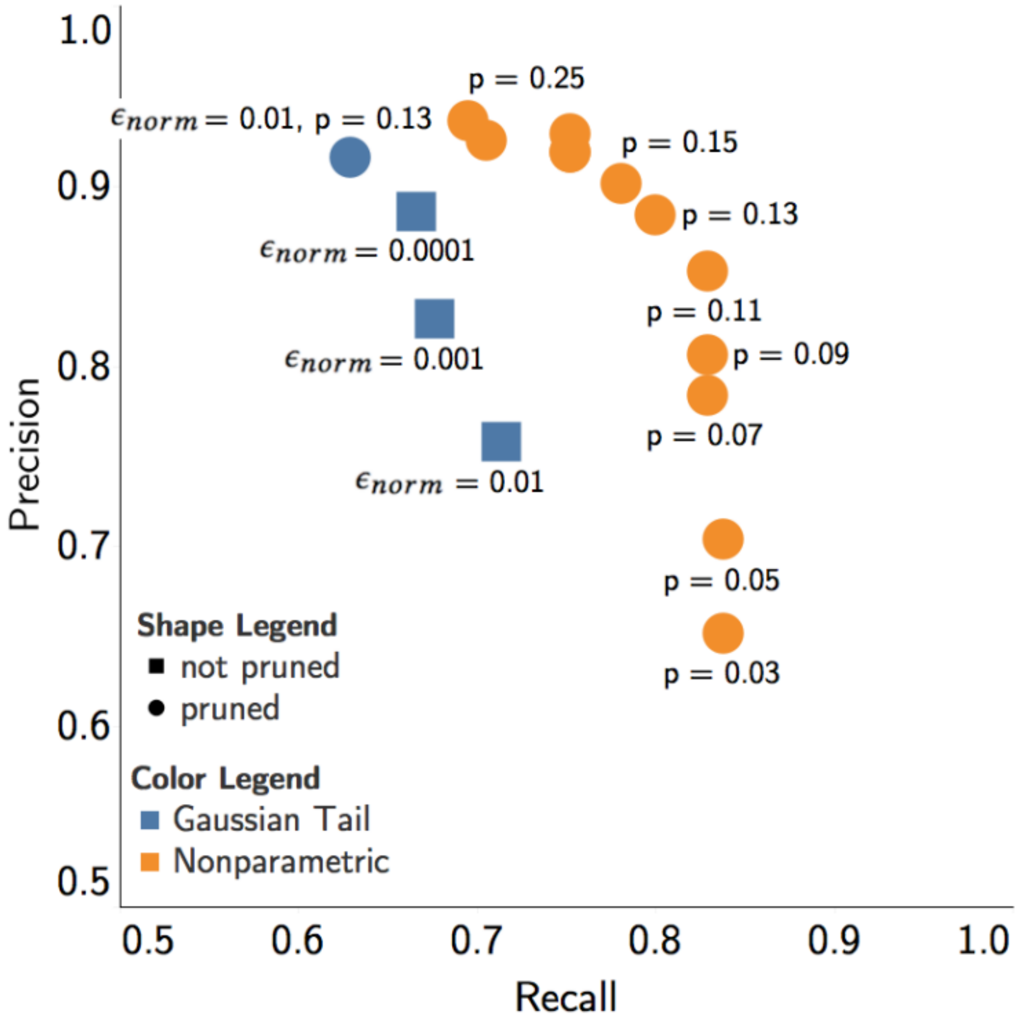

Fig. 5. Overall precision and recall for parametric and non-parameteric approach [4]

3 OTHER ANOMALY DETECTION METHODS FOR SPACECRAFT DATA

3.1 Out of Limits(OOL)

Out of Limits approach is till date a very popular approach because of its simplicity leading to low computational expense. This approach defines a upper and lower threshold for selected parameters. If the measurement goes beyond the threshold an alarm is triggered. Then the engineers would be required to do a manual inspection of the cause behind the triggering of the alarm for the parameter and decide if it was an anomaly or not and what sorts of action to take next.

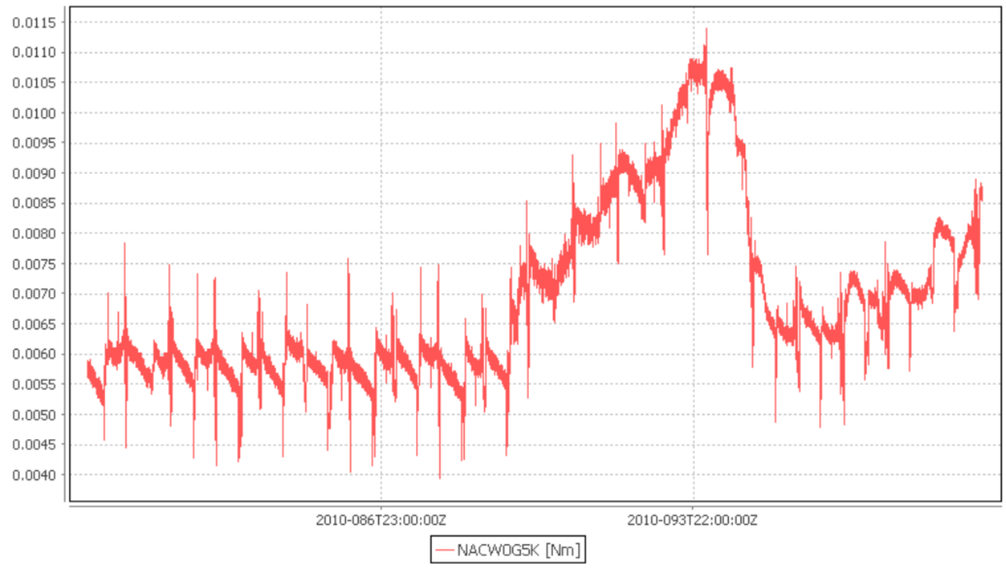

Fig. 6. Anomalous behaviour despite values lying within the limits [6]

Issues with this approach are that as seen in Figure 6,some behaviors are anomalous even if they are within the defined limits. A lot of manual intervention is still needed with this approach.[6]

3.2 Nearest Neighbor Based Approach

Nearest neighbour approach, currently being used to detect anomalies in the Orca system on the International Space Station, calculates the neighbourhood for each point and identifies anomalous points accordingly. Different distance measures are used such as Euclidean distance for continuous valued parameters and Hamming distance for discrete valued parameters.The program outputs a score for each point representing the average distance to the nearest k neighbors in the data set. The value of k is specified by the user. Points that have a larger average distance to their nearest neighbors than most other points in the data set are considered anomalies.[3] Issues with this approach are:

- The number of past values to be considered needs to be decided.
- Distance calculation between each of the points leads to High computational complexity and memory consumption.
- Can only find global anomalies, as it does not consider local neighborhood.

3.3 Clustering Based Approach

The IMS(Inductive Monitoring System)(currently being used on ISS) tool uses clustering approach to analyze historical spacecraft data and identifies normal interactions between selected parameters. This characterization, or model, is compared with real time or archived system data to detect abnormal or anomalous behavior. [3]

Table 3. LSTM-NDT(Independent Channel) and MTAD-TF(Dependent Channel) [9]

Dataset	Precision	Recall	F1 score
LSTM-NDT			
MSL	0.5944	0.5374	0.5640
SMAP	0.8965	0.8846	0.8905
SMD	0.5684	0.6438	0.6037
MTAD-TF			
MSL	0.9043	0.8988	0.9015
SMAP	0.9779	0.8192	0.8916
SMD	0.9045	0.9048	0.8940

4 CRITICISM

Limitations with the approach:

- MTAD-TF method employs the use of dependencies between the different features to predict anomalies in multi variate time series data. From Table 3, when comparing to the method MTAD-TF, LSTM-NDT has a high score on SMAP, but it performs poorly on MSL and SMD, reflecting that the model is very sensitive to different scenarios[9]. As one of the goals of the authors of the paper was to build an anomaly detection system that can be very well applied to not only different spacecrafts data but also for any multivariate time series data. This goal is not being met.
- Along with the telemetry data, channel specific command information is also an input to the model that the model is very much dependent on. Refining it would lead to more accurate anomaly detection.
- The dependencies between different channels is not taken into account.

5 SUMMARY

An anomaly detection system that caters to the particular need of a spacecraft system considering the time and precision critical environment of a spacecraft has been proposed in the paper I have reported on. Using state of the art deep learning algorithm Long Short Term Memory, gives this approach an advantage of not being heavily dependent on the amount of historical data that can be used to predict telemetry data. A dynamic thresholding approach has been proposed that does not make any assumptions about the prediction error distribution. This makes this approach superior to others that assume data to be normally distributed. Considering the key areas of improvement as those mentioned in the Criticism section, will be included in the future variations of the approach which can lead to this approach being applicable for a large number of varied anomaly detection tasks in multivariate time series data.

REFERENCES

[1] 2003. *Columbia Accident Investigation Board Report, Volume 1.* Technical Report. NASA.

[2] Paolo M. Piselli Dominique T. Shipmon, Jason M. Gurevitch and Stephen T. Edwards. 2017. Time Series Anomaly Detection; Detection of anomalous drops with limited features and sparse examples in noisy highly periodic data. (2017). https://doi.org/10.48550/arXiv.1708.03665

[3] David Iverson. 2008. Data Mining Applications for Space Mission Operations System Health Monitoring. *SpaceOps 2008 Conference* (2008). https://doi.org/10.2514/6.2008-3212

[4] Christopher Laporte Ian Colwell Tom Soderstrom Kyle Hundman, Valentino Constantinou. 2018. Detecting Spacecraft Anomalies Using LSTMs and Nonparametric Dynamic Thresholding. *KDD '18: The 24th ACM SIGKDD International Conference on Knowledge Discovery Data Mining* (2018), 9. https://doi.org/10.1145/3219819.3219845

[5] A. Lavin and S. Ahmad. 2015. Evaluating Real-time Anomaly Detection Algorithms – the Numenta Anomaly Benchmark. *14th International Conference on Machine Learning and Applications (IEEE ICMLA'15)* (2015). https://doi.org/10.1016/j.neucom.2017.04.070

[6] Jose MartÃŋnez-Heras and Alessandro Donati. 2014. Enhanced Telemetry Monitoring with Novelty Detection. 35 (2014).

[7] Richard A. Davis Peter J. Brockwell. 2016. *Introduction to Time Series and Forecasting.* Springer.

[8] Michael Phi. 2018. *Illustrated Guide to LSTM's and GRU's: A step by step explanation.* https://towardsdatascience.com/illustrated-guide-to-lstms-and-gru-s-a-step-by-step-explanation-44e9eb85bf21

[9] C.L. Zhang and H. Y. Wang Q. He, Y. J. Zheng. 2020. MTAD-TF: Multivariate Time Series Anomaly Detection Using the Combination of Temporal Pattern and Feature Pattern. *Hindawi Complexity Volume 2020* 2020, Article 8846608 (2020), 9 pages. https://doi.org/10.1155/2020/8846608