IAF SPACE OPERATIONS SYMPOSIUM (B6)
Ground Operations - Systems and Solutions (1)

Author: Dr. Jakub Nalepa
KP Labs, Poland

Mr. Pawel Benecki
KP Labs, Poland
Mr. Jacek Andrzejewski
KP Labs, Poland
Dr. Daniel Kostrzewa
KP Labs, Poland
Mr. Michal Myller
KP Labs, Poland
Mr. Szymon Piechaczek
KP Labs, Poland

EVALUATING ANOMALY DETECTION IN SATELLITE TELEMETRY DATA

Abstract

Detecting anomalies in telemetry data captured on-board a spacecraft is a critical aspect of its safe operation, and it allows us to respond to various failures and hazards. There exist three main types of anomalies that should be considered for complex missions – in point anomalies, telemetry values fall outside the nominal operational range. The collective anomalies refer to the overall sequences of consecutive telemetry values that are anomalous, whereas in contextual anomalies, the single values are anomalous within their local neighborhood. The out-of-limit detection engines can easily spot point anomalies, whereas expert systems have been proposed to cover other events as well. Since generating new representative ground-truth sets is extremely costly, unsupervised algorithms have become the mainstream in the field. Additionally, as we have been witnessing an unprecedented success of deep learning in virtually all domains of science and industry, such methods are being consistently applied to detecting spacecraft anomalies too. In long short-term memory (LSTM) networks, the expected telemetry values are extracted at first, and this step is followed by an unsupervised thresholding of differences between the expected and actual values which is used to detect anomalous events. As separate LSTMs can process different telemetry channels, such data-driven techniques may easily offer traceability and interpretability which are pivotal in Space applications. In this paper, we show that assessing the quality of anomaly detectors is a critical aspect of their verification and validation, as the metrics that capture temporal aspects of detected anomalies (with respect to the ground truth, if available) convey very important information about such systems. We investigate both classical out-of-limit and deep learning algorithms (also coupled with evolutionary algorithms utilized to optimize their hyperparameters) for anomaly detection from time-series data and apply them to benchmark telemetry sets coupled with the corresponding ground-truth events. Our experiments indicate that the basic performance metrics, directly extracted from the confusion matrices may not be enough to fully understand the abilities of the detection algorithms. We thus propose new metrics that better capture the temporal aspects of the detector's output, as responding to an event (possibly even before it has started happening, hence predicting that they are likely to happen) is critical for ensuring the safety of the spacecraft. Finally, we present our anomaly visualization toolbox (called the Antelope Toolbox) and discuss the importance of the qualitative analysis of the results obtained using automatic anomaly detection from telemetry data.