ANNOTATING LARGE SATELLITE TELEMETRY DATASET FOR ESA INTERNATIONAL AI ANOMALY DETECTION BENCHMARK

Krzysztof Kotowski[1], Christoph Haskamp[2], Bogdan Ruszczak[1], Jacek Andrzejewski[1], Jakub Nalepa[1]

Introduction

Early detection of anomalies in satellite telemetry *plays an important role in the daily practice of spacecraft operations engineers (SpaceOps). Many institutions from the space sector (e.g., ESA, NASA, CNES, and Airbus) have been researching anomaly detection algorithms in recent years [1]-[4]. However,* ***there are no publicly available large-scale datasets*** *to test advanced algorithms in real-life scenarios. The* ***ESA Anomalies Dataset for International AI Anomaly Detection Benchmark*** *aims to solve this problem. Here, we present the collection and annotation of this dataset by our interdisciplinary team of* ***machine learning engineers and SpaceOps.***

Dataset release planned in December 2023

Data collection

In cooperation with the ***European Space Operations Centre (ESOC)****, we collected data from 3 ESA missions of different types and orbital periods. Based on the* ***Anomaly Report Tracking System (ARTS)****, we selected the most interesting continuous time periods and subsets of channels for anomaly detection. The collected data contains* ***raw telemetry channels, telecommands, mission plans, and anomaly reports.***

- **9 years of telemetry from 3 different ESA missions**
- **225 channels from 12 different subsystems**
- **~1.5 million occurrences of 700 different telecommands**
- **12 GB of compressed data**

Each mission uses slightly different data formats, subsystems, and anomaly reporting standards, so our ML team established ***a common dataset structure*** *for all missions with data stored in compressed pandas DataFrames.* ***The data will be shared publicly in the anonymized version.***

Data annotation – manual and algorithms-aided

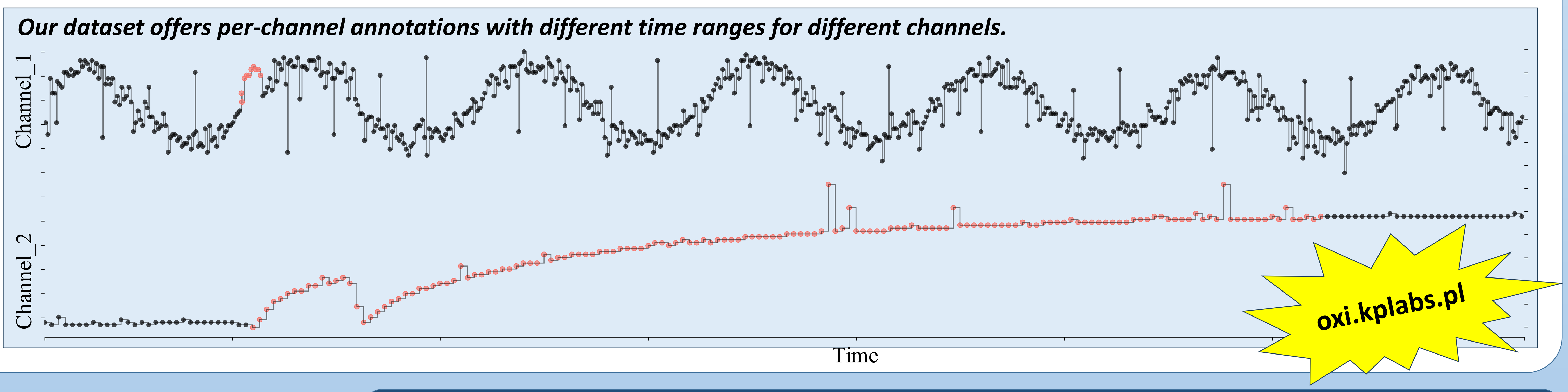
It turned out that anomaly ***reports from ARTS are far from being usable for machine learning purposes****. They are intended for human use and there is no structured way of reporting anomalies. We had to analyse the reports manually to establish an initial version of anomaly annotations using our* ***open-source OXI labelling tool*** *for satellite telemetry (oxi.kplabs.pl) [5]. Based on mission plans, we annotated also* ***rare nominal events****, e.g., commanded manoeuvres, so they are not mistaken for anomalies.*

Due to the size of the dataset, ***many anomalies were overlooked or only partially annotated in the initial version****. We utilized a set of different* ***machine learning algorithms*** *to identify missing annotations. Due to the weak labelling, we had to start with* ***unsupervised anomaly detection algorithms*** *in the first refinement phase which resulted with tens of additional annotations. In the second refinement phase, we used more advanced* ***semi-supervised algorithms, i.e., Telemanom algorithm by NASA*** *[2], to identify more subtle issues and analyse each channel separately. We performed several iterations of this process including* ***hours of analysis with SpaceOps*** *from each mission.*

We annotated ***more than 1000 anomalies and rare events*** *of several different sources and types (global\local, point\subsequence, and uni-\multi-variate). Some of them are very challenging to detect which should encourage the development of novel algorithms that can handle real-life problems of satellite telemetry.*

>1000 annotated anomalies and rare nominal events curated for machine learning purposes

Our dataset offers per-channel annotations with different time ranges for different channels.



Anomaly detection algorithms

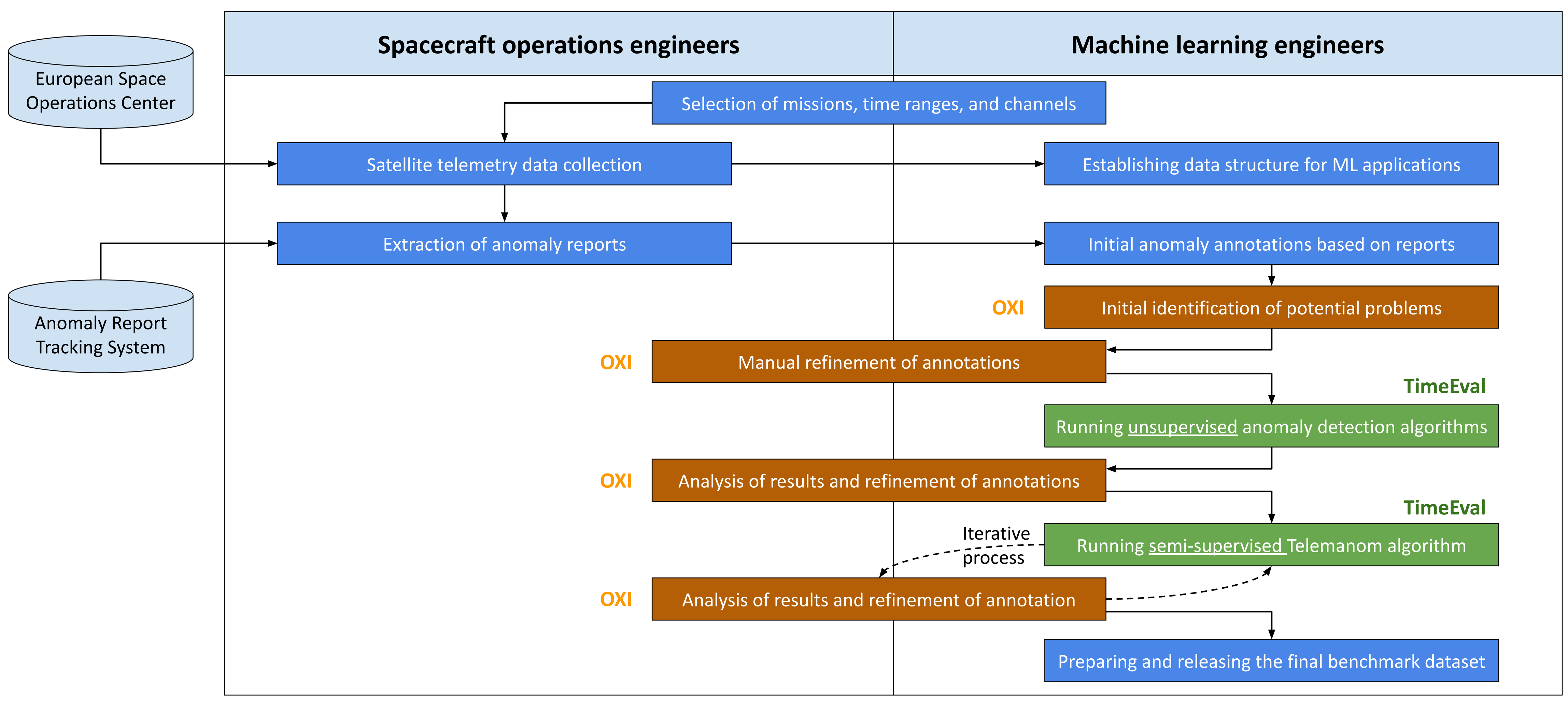
The annotation of such large dataset would be infeasible without the support from anomaly detection algorithms. *We utilized the* ***TimeEval framework*** *[6] to run and select algorithms for our dataset.*

Among unsupervised approaches, the most valuable results were achieved with ***Isolation Forest*** *and* ***Copula-based Outlier Detector.*** *Other algorithms had problems with memory optimization (i.e., kMeans and kNN) or poor performance (i.e., LOF, PCC, HBOS).*

Among semi-supervised approaches, we focused on the forecasting-based ***Telemanom algorithm*** *[2] which learns nominal characteristics of each telemetry channel using recurrent neural network with LSTM units.*

Details of algorithms, problems we had to overcome, and benchmarking results for the final version of the dataset will be published after the dataset release.

Overview of the workflow



References

[1] B. Ruszczak et al., "Machine Learning Detects Anomalies in OPS-SAT Telemetry," in ICCS 2023, in Lecture Notes in Computer Science. Springer, 2023, pp. 295–306. doi: 10.1007/978-3-031-35995-8_21.
[2] K. Hundman, V. Constantinou, C. Laporte, I. Colwell, and T. Soderstrom, "Detecting Spacecraft Anomalies Using LSTMs and Nonparametric Dynamic Thresholding," in Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Jul. 2018, pp. 387–395. doi: 10.1145/3219819.3219845.
[3] S. Fuertes, B. Pilastre, and S. D'Escrivan, "Performance assessment of NOSTRADAMUS & other machine learning-based telemetry monitoring systems on a spacecraft anomalies database," in 2018 SpaceOps Conference, doi: 10.2514/6.2018-2559.
[4] B. Pilastre, L. Boussouf, S. D'Escrivan, and J.-Y. Tourneret, "Anomaly detection in mixed telemetry data using a sparse representation and dictionary learning," Signal Processing, vol. 168, p. 107320, Mar. 2020, doi: 10.1016/j.sigpro.2019.107320.
[5] B. Ruszczak, K. Kotowski, J. Andrzejewski, C. Haskamp, and J. Nalepa, "OXI: An online tool for visualization and annotation of satellite time series data," SoftwareX, vol. 23, Jul. 2023, doi: 10.1016/j.softx.2023.101476.
[6] P. Wenig, S. Schmidl, and T. Papenbrock, "TimeEval: a benchmarking toolkit for time series anomaly detection algorithms," Proc. VLDB Endow., vol. 15, no. 12, pp. 3678–3681, Sep. 2022, doi: 10.14778/3554821.3554873.

Author information and acknowledgements

[1]KP Labs, Bojkowska 37J, 44-100 Gliwice, Poland. {kkotowski, bruszczak, jandrzejewski, jnalepa}@kplabs.pl
[2]Airbus Defence and Space GmbH, Airbus-Allee 1, 28199 Bremen, Germany. christoph.haskamp@airbus.com

The work was financially supported by ESA as a part of the "ESA Anomalies Dataset for International AI Anomaly Detection Benchmark" project (contract number 4000137682/22/D/SR). We would like to thank all SpaceOps from European Space Operations Centre involved in the project.