AIRBUS

Anomaly Detection in Time Series Scientific Challenge

Airbus Defense & Space – Anomaly Detection in Satellites parameters - Specific Rules

LEGAL NOTICE – issued 18 December 2018

Challenge in brief
AIRBUS platforms (aircraft, helicopters, satellites, etc.) generate a huge quantity of sensor data that is collected on big data infrastructures. Time Series Analysis has been identified as a technology that will be used in future years by AIRBUS to extract value from this huge source of knowledge on our operated platforms.

AIRBUS is launching this scientific challenge on anomaly detection in time series data to:

- scout for top players in the field of Time Series Analysis;
- encourage the research community to tackle the specific issues of related to the data generated by the aerospace industry during tests and in operations.

Why AIRBUS context is challenging?
The challenge is to work in an unsupervised context. Almost all data collected from our platforms are considered as normal. The occurrences of faults and failures are very rare and our customers would not appreciate to wait hundreds of faults before being able to anticipate them. We are thus interested in looking for changes in the behavior of the systems we monitor through the sensors installed on the platforms.

What are the tasks?

- Detect abnormal behaviour in sensor data

Who can participate?
The challenge is intended to be of interest to companies, research labs, schools or individuals. There is no cost to participate.

AIRBUS

Content

1. Registration and Terms of Use 3
2. What are the tasks? 3
3. Who can participate? 3
4. Challenge process and schedule 3
5. Invitation extended to top-ranked Participant 4
6. Support during challenge 4
7. Participants privacy and personal data protection 5
8. Detailed datasets description 6
9. Results submission rules and procedure 7
10. Detailed evaluation methodology and Participant ranking procedure 8
11. Challenge- level ranking and winner designation procedure 8
12. Confidentiality – Intellectual Property 8
13. Governing Law and Severability 8

AIRBUS

1. Registration and Terms of Use

When accessing to the AIRBUSAI GYM platform and registering to the challenge you agree with the rules for the ANOMALY DETECTION IN TIME SERIES CHALLENGE as published on the AIRBUSAI GYM Platform and this Legal Notice.

Registration for this challenge (ADS use case) can be done on AIRBUSAI GYM platform: https://aigym.airbus.com/

Registration for the two other Times Series use cases is also available on the platform. You will have to register on each challenge you want to participate in.

2. What are the tasks?

There is one task: *anomaly detection*.

2.1 Anomaly detection

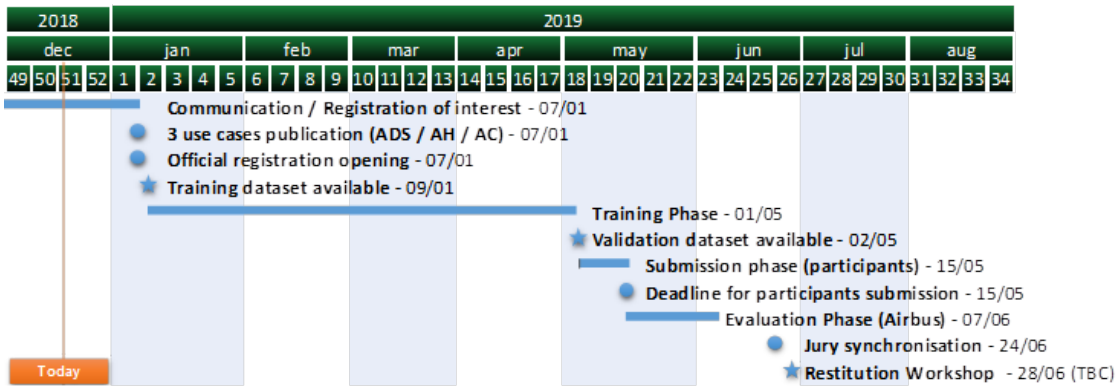
Given an input sequence containing sensor data collected on one sensor during a previously fixed period of time, the Participants' system shall output 1 if this sequence is considered abnormal and 0 if this sequence is considered normal.

3. Who can participate?

A Participant is defined as an individual participating solely, a team of individuals, a small, medium or large company, or a research laboratory, a University, college, or institute (hereinafter the "Participant").
Creating a consortium or sharing human resources is not permitted in order to allow the evaluation of each Participant separately.
Only one participation is allowed per entity.
Participants will provide a focal point for any communication with the organizers.

Teams who have provided at least one submission are entitled to request a 1h company pitch, including a marketing and technical presentation split evenly. Nevertheless, no information about the challenge will be provided during this meeting.

4. Challenge process and schedule



4.1 Registration

Participants can register on the AIRBUSAI GYM platform from 07-JAN-2018 on this Airbus Defence & Space use case.

4.2 Release of training dataset and training phase

The training dataset is made available to Participants on the **09-JAN-2018** via the AIRBUSAI GYM platform.

The training dataset is the data that is provided to Participants to build their system (normal sensor data). Participants are free to use their own additional data as long as they specify what data they used when submitting results.

The training phase starts from **09-JAN-2019** and ends on **01-MAY-2019**.
During the training phase, Participants build their system using the training data and any other additional data of their choice.

4.3 Release of evaluation dataset

The evaluation dataset will be made available to Participants on the **01-MAY-2019** via the AIRBUSAI GYM platform.
The evaluation dataset is used to perform the final evaluation (ranking of the Participants and designation of best performing method).

4.4 Submission phase

The submission phase will last from **01-MAY-2019** to **15-MAY-2019**.
During the submission phase, Participants run their system on the evaluation dataset and submit the results of their system on the AIRBUSAI GYM platform. Each Participant can submit up to two different results.
Leader board remains open during the submission phase (up to **15-MAY-2019**).

4.5 Deadline for results submission

Participants can submit their results on the AIRBUSAI GYM platform until **15-MAY-2019**.

4.6 Evaluation phase

The evaluation phase starts from (**15-MAY-2019**) to (**07-JUN-2019**). Evaluation is performed by the challenge organizer (AIRBUS). The submissions of the Participants are compared to the ground-truth data (human anomaly detection) and ranked. A verification procedure will be made to ensure that the submitted results were obtained without human intervention. For example, Participants may be asked to send (or bring to Toulouse) a trial version of their system. Any other way of demonstrating that the submitted results were obtained without human intervention will be discussed with the organizers of the challenge.

4.7 Results announcement and restitution workshop

A one-day restitution workshop will take place on (**28-JUN-2019**) (exact date to be confirmed at a later date) for all Participants to meet and share knowledge / experience at AIRBUS premises in Toulouse. Challenge organizers will announce the results and the top-ranked will give a speech during the workshop. A poster session (marketplace) will be organized during the workshop to allow Participants to present their system. A Round table on Time Series Anomaly Detection will be organized during the workshop as well.

5. Invitation extended to top-ranked Participant

The focal point of the top-ranked Participant will be invited to Toulouse to give a speech at the restitution workshop (round trip transportation to Toulouse, meals and hotel accommodation for 2 nights and 2 days in Toulouse for the focal point of the top-ranked Participant, will be paid by Airbus).

6. Support during challenge

A challenge forum has been created to ensure that all questions and answers are made public and shared among Participants. Access will be granted to focal points of each Participant.

AIRBUS

7. Participants privacy and personal data protection

7.1 Participants privacy

Participation to the challenge is not made public unless the Participant agrees to make it public.
This means that:

- The leader board results (ranking, scores) are anonymized except for Participants who agree to have their name disclosed.
- The challenge results (ranking, scores) are anonymized except for Participants who agree to have their name disclosed.
- Participants are free to publish or report their own score and ranking but they cannot publish other Participants' scores and ranking unless they were authorized to do so by the relevant Participant.

7.2 Personal data protection

Participants' personal data will be collected during challenge registration (name of organization, email and name of focal point, timestamp of your registration and submission). This data will be stored on the AIRBUSAI GYM platform and will be protected in accordance with the Airbus Personal Data Protection Policy.
The collection and processing are handled by Airbus and are in full compliance with the European Directive 95/46/EC (the "Directive 95/46/EC") of the European Parliament and of the Council of 24 October 1995 on the protection of individuals regarding the processing of Personal Data and on the free movement of such data.
You may object such collection of your personal data for legitimate reasons. In such case, Airbus cannot grant access to the AIRBUSAI GYM platform.
You can exercise these rights by contacting the Data Protection Officer at dataprotection@airbus.com specifying which of the aforementioned rights you wish to exercise and attaching a copy of a document justifying your identity.

7.3 Cookies on AIRBUSAI GYM platform

The AIRBUSAI GYM platform may send to your computer small text files called "cookies" which are stored on your computer and can be read by the webserver to ease your internet browsing and to customize your experience. You are free to accept or decline it. As browsers usually tend to automatically accept cookies, if you do not want to accept then you should configure your browser to decline cookies. Please note that if you decline cookies, you may not be able to access all interactive features of the AIRBUSAI GYM, Airbus websites and services.

AIRBUS

8. Detailed datasets description

We are interested in the detection of abnormal sensor behaviour. Sensors are recorded at 1Hz and we provide sequences of variable length. Contestants are free to cut, aggregate the data for their training but the validation will be made on variable length sequences.

8.1 Dataset split

Training data
The training dataset is composed of 4,008 sequences of variable lengths. Each sequence is composed of 35 parameters, some are measures (p1 to p34) and one is computed (p35). All sequences are considered as normal and should be used to learn normal behaviour of antenna data.

Validation Data
The validation dataset is composed of 1,336 sequences of variable lengths. Each sequence is composed of 35 parameters, some are measures (p1 to p34) and one is computed (p35). They are to be tested with the normal behaviour learnt from the training data to detect abnormal behaviour. The amount of abnormal sequences in the validation dataset is a priori unknown.

8.2 Dataset usage

The dataset is made available to the Participants of the challenge from January 2019 to June 2019.

8.3 Dataset format description

Datasets are provided in a HDF5 format that can be decoded by many standard machine learning modules (like pandas for instance):

- In the training dataset, the dataframe is called “dftrain”.
- In the validation dataset, the dataframe is called “dfvalid”.

Each dataframe has 40 columns corresponding to the sequence id, the satellite number, its mission number, the orbit, the 35 parameters and the time. The number of rows is naturally depending on sequences lengths inside each dataframe. The training dataset has 2,135,131 rows and the validation dataset 699,595.

9. Results submission rules and procedure

9.1 Results submission rules

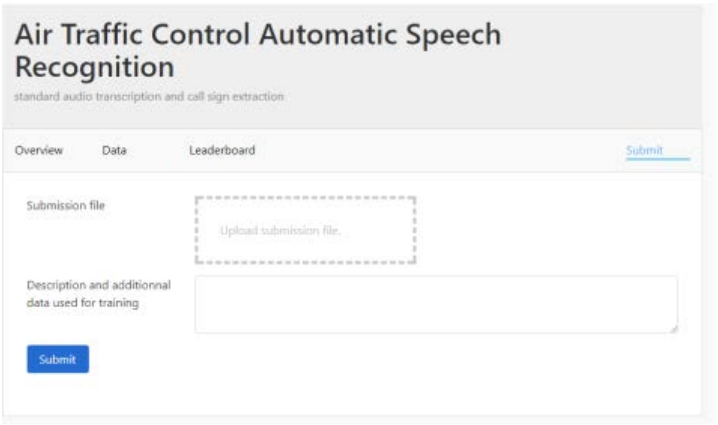
Participants run their system on the evaluation dataset and submit the results of their system to the AIRBUSAI GYM platform for evaluation.

The Participants agree to process the data in accordance with the following rules:

- The Participants agree to abide by the terms guiding the training conditions (any additional data used for training the system shall be mentioned);
- The Participants agree not to have the evaluation data processed or improved via manual/human means;
- The Participants are allowed to use any automatically derived information for training;
- The Participants are allowed to submit up to two results for the final evaluation;
- Only one participation is allowed per entity/organization (University, School, Company, one or a team of individuals);
- Participants agree to have their system screened if required by challenge organizers to ensure that the evaluation data has not been data processed or improved via manual/human means.

9.2 Results submission procedure

The system’s results should be uploaded on the using a web form similar to:



Submission form

The system’s results are uploaded in the form of a tab-delimited .csv file.

9.3 Results submission format

Participants are expected to deliver their results in the form of an UTF-8 encoded .CSV (comma-separated values) file.
This .CSV file has two columns, each separated by a comma:

1. id: the number of the sequence
2. anomaly: 1 if the sequence is abnormal, 0 if not.

10. Detailed evaluation methodology and Participant ranking procedure

3 indicators will be built from the submission results:

- Precision
- Recall
- $F_\beta score = \frac{(1+\beta^2)*precision*recall}{\beta^2 precision+recall}$, $\beta = 0.1$

11. Challenge- level ranking and winner designation procedure

11.1 Participants' ranking

Submissions will be ranked using only $F_\beta score$. This indicator favors a lot precision over recall as we want to have as few false detections as possible.

11.2 Participants' system screening for winner designation

A verification procedure will be made to ensure that the submitted results were obtained without human intervention. For example, Participants may be asked to send (or bring to Toulouse) a trial version of their system. Any other way of demonstrating that the submitted results were obtained without human intervention will be discussed with the organizers of the challenge.

11.3 Designation of top ranked Participant

The winner of the challenge is the Participant whose system has been screened by the challenge organizers and whose submission is ranked 1.

12. Confidentiality – Intellectual Property

Results submitted by Participants will remain the property of each Participant. However, each Participant authorizes challenge organizers to use the results for evaluation purposes.

13. Governing Law and Severability

This Legal Notice shall be governed by and construed in accordance with the laws of France, subject to any local applicable non derivable public policy rule.
Should one of the provisions of this Legal Notice or the TIME SERIES ANOMALY DETECTION CHALLENGE become partly or entirely invalid, the validity of the remaining provisions shall not be affected thereby. The invalid provision shall be replaced by a valid one best achieving the economically desired result. The same applies in the event of gaps or omissions.