PRIORITY SCORES BASED ON NOVELTY DETECTION TO IMPROVE THE EFFICIENCY OF GROUND-OPERATIONS

Jacopo Biancat, Chiara Brighenti, Attilio Brighenti
S.A.T.E. Systems and Advanced Technologies Engineering S.r.l., Venezia, Italy

José-Antonio Martínez-Heras, Alessandro Donati, David Evans
ESA/ESOC, HSO-OSA Human Spaceflight and Operations Department

ABSTRACT

It is well known that in space ground-operations the number of spacecraft telemetry parameters and telecommands keeps increasing while the manpower for operating the ground systems keeps being reduced [1]. This means that it is not possible for Flight Control Engineers (FCEs) to analyze manually each parameter or telecommand to verify their consistency with the overall status of the satellite. Therefore there is the need of automatic methods and/or systems that identify in these large datasets, including today over 15.000 parameters and telecommands for one satellite, which parameters should be the focus of FCEs.. Indeed FCEs should analyse first those showing the higher level of inconsistency with respect to the expectations or to the past behavior.

This paper presents some of the methods and software tools developed by the authors in two studies, under contract with the European Space Agency-European Space Operations Centre (Darmstadt-Germany) regarding the Automatic Behaviour Detection and Interpretation from Low Level Data Sets such as Telemetry and the Automatic spacecraft status characterization by data mining mission history[1], that allow the automatic extraction of the nominal behavior and the reliable detection of novel behaviors, by generating *priority scores*, which are associated to the degree of novelty computed as degree of violation of the detected knowledge, which is extracted in several formats, either for *check* definition or *pattern* analysis.

Index Terms — Knowledge extraction, novelty detection, pattern analysis, checkability

1. INTRODUCTION

The detection of a novel or anomalous behaviour and its possible causes by visual analysis of the acquired raw data is an extremely demanding task for FCEs, given the large amount and variety of data and the possibly hidden cause-effect relations among subsystems parameters. Among the several possible approaches, the most promising ones are those based on data analysis, relying mainly on data mining techniques, because they allow the interpretation of large amounts of heterogeneous data using no or very little a priori knowledge. Usually these approaches elaborate acquired data, automatically characterizing the *nominal* behavior (knowledge extraction) against which, then, any possible *new* phenomenon or unknown relationship among subsystems or processes is identified. Indeed this automatic extraction of the *nominal* behavior from data allows avoiding the step, usually required in a diagnostic system, of knowledge extrapolation and representation by field experts, e.g. for the definition and implementation of a model (*model-based* approach) [2]. It is beyond doubt that this step requires resources and time that are often not available or not compatible with the constraints defined for the system to which diagnostics is applied, especially in industrial fields such as the space sector. In alternative, *threshold-based* approaches, which consists in checking measurable variables for upward or downward transgression of fixed limits, may be used. However, the major drawback of this technique is the need to set wide threshold limits to avoid false alarms, with the consequence that only sudden major faults or long-lasting gradually increasing faults can be detected (eventually with relevant time delays) [3].

Therefore the use of historical process data, collected during the system operation, to generate references of normal conditions without the need of a priori knowledge by field experts, appears to be the most promising solution for analyzing spacecraft telemetry parameters and telecommands.

However the generation of this *nominal* behavior either to apply *pattern* analysis or predefined *checks*, is not trivial because this must be *useful*, i.e. significant and reliable. Significant means that it provides valuable information, allowing at least novelty detection, i.e. the determination of the presence of a novel behavior in a system while reliable means that it does not generate too many false positive, represented by the identification of novel behaviors that are not such.

[1] The view expressed herein can in no way be taken to reflect the official opinion of the European Space Agency.

doi: 10.2788/854791

2. PRIORITY SCORES

Priority scores are synthetic measures proposed by the authors to quantify how different are the behaviors of parameters and telecommands in two different time periods. The greater the difference, the higher the priority score, which suggests that the parameter/telecommand should be further analyzed by the FCEs to investigate if failures occurred to a component or part of the satellite or to confirm whether the observed changes are associated to specific commands sent from the ground stations or generated on board.

Therefore these *priority scores* allow the generation of priority lists, i.e. lists of parameters ordered according to the *priority scores*, indicating which parameters should be further analyzed first.

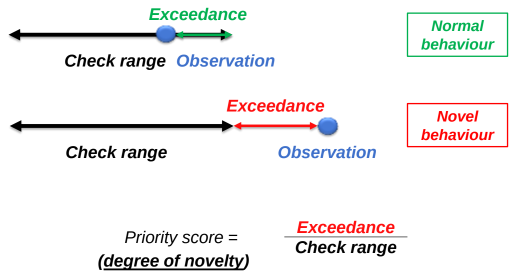

Figure 1 - Example of priority score.

Figure 1 shows an example of priority score computation for a predefined *check*. When the new *observation* falls within the *check range* a negative *exceedance* is computed which is associated to a *normal behavior* of the system (null priority score). When the new *observation* falls out of the *check range* then a positive *exceedance* is computed which is associated to a *novel behaviour* and its *degree of novelty* is quantified as the ratio between this exceedance and the check range.

3. FEATURES

One of the most common approach in data mining is the computation of *features* [4], i.e. characteristics of a set of data or of a signal, which can be measured or calculated over a specific time interval (e.g. the mean, the standard deviation and the minimum values of a signal or parameter in a given time window are features of the signal). Features are often used to reduce the volume of data, while preserving relevant information. In this application context, features can also be used to characterize the *nominal behavior* of a parameter, through the characterization either of the *single features* or of the combination of different features computed over the same time window, representing the so called *feature-based patterns*. Different checks can be defined for both types of approaches, to perform *novelty detection* and *priority scores* computation.

4. SINGLE FEATURE - FETCH

Two alternative approaches were investigated for the definition of *useful* checks (i.e. significant and reliable) of a *single feature*. The first one is based on a density-based measure called *entropy* of a feature. This measures the order of the data in the feature's space. The idea is that a low entropy value corresponds to a set of data with clusters (i.e. well separated data), whereas a high entropy value corresponds to a set of data without clusters [5]. For this reason entropy could be exploited for the selection of checks whose performance may be influenced by the presence of cluster in the data. However the results showed that the entropy does not prove to be a discriminant factor for the reliability of the check.

Indeed the use of *features* for check definition poses the issue of defining a suitable time interval, named Time Window Duration (*TWD*), for their computation. Therefore an innovative method, developed by the authors, called FEaTure CHeckability (***FETCH***), allows identifying the shortest time window beyond which a robust nominal behavior of the feature may be generated so that a *useful* check can be done.

Different kind of checks have been defined to verify that the behaviour of a feature does not change over time from different points of view, including:

1. <u>Domain check</u>: to verify the range or values of the feature.
2. <u>Variability check</u>: to verify the trend of the feature.
3. <u>Distribution check</u>: to verify the distribution of the feature.
4. <u>Frequency check</u>: to verify the frequency content of the feature.
5. <u>Inter period check</u>: to verify the modes of a feature.
6. <u>Few samples check</u>: this check applies to the parameters that do not have enough samples to compute features. It consists in verifying if the parameter values change over time with respect to the reference ones, computed from a nominal dataset.

One of the main outcome of the studies performed by the authors is that features checkability is strongly related to the *TWD* used for the computation of the feature or its characteristics (e.g. distribution, frequency).

Indeed the results showed that the same feature computed for the same parameter may have different behaviors which lead to different checks to be applied.

For example, considering the parameter shown in Figure 2 and the feature *minimum*, it is observed that using a *TWD* equal to 1 hour the feature shows a recognizable periodicity (Figure 3) which allows performing a *frequency check*, to verify that the frequency content remains the same; instead, using a TWD equal to 36 hours (Figure 4), the feature has a constant behavior which allows applying a *domain check* to verify that the values of the feature do not change over time.

doi: 10.2788/854791

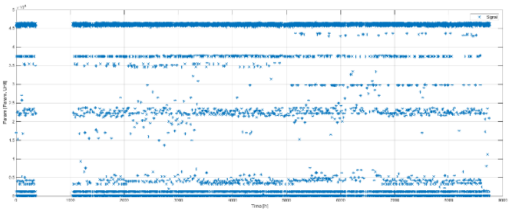

Figure 2 - Example of one satellite telemetry parameter.

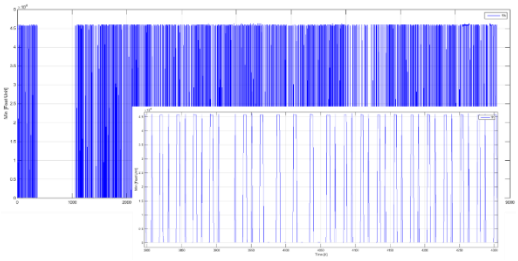

Figure 3 - Feature *minimum* computed with *TWD* = 1h.

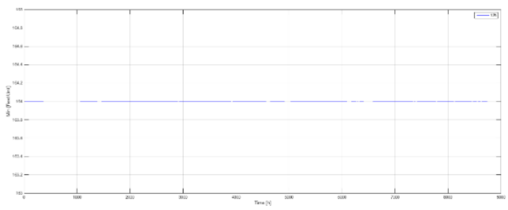

Figure 4 - Feature *minimum* computed with *TWD* = 36h.

5. FEATURE BASED PATTERN - KETTY

Features-based patterns were implemented in a software prototype, named ***KETTY*** (Knowledge ExctracTion from TelemetrY), performing a characterization of the behavior of a parameter synthesizing it into a finite number of states associated to *patterns*.

Several alternative approaches were considered for the analysis of these *patterns* including the detection of the appearance of new *patterns*, the changes in the values of the *patterns*, the analysis of the degree similarity of a new observation with respect to the identified *patterns* and the analysis on the number of occurrences of the patterns. This last approach was selected because it provided the most robust and reliable results, allowing both a static and dynamic characterization of the behaviour of a parameter in a given time period (*reference dataset*). This characterization may be used to detect novelties in the static or dynamic behaviour of the same parameter in another time period (*comparison dataset*), by comparing the occurrences in the *reference dataset* with those in the *comparison dataset*.

In the following some examples are provided using the blue colour to show the behaviour of the parameter in the *reference dataset* and the red colour to show the behaviour of the parameter in the *comparison* dataset which is detected by ***KETTY*** as being characterized by novelties.

The first example is related to a parameter having shown novelties detected by the long-term analysis which aims at the detection of novelties of long duration within a long time interval (e.g. one month). The parameter is categorical, taking the two states ON and OFF. It is clear from Figure 5 the different behaviour of the parameter in the two datasets.

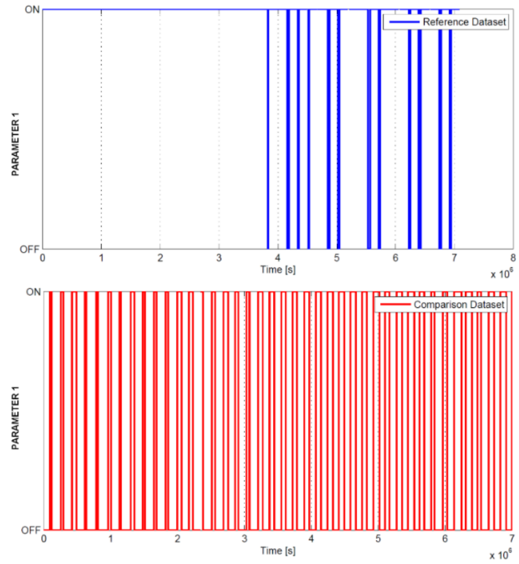

Figure 5 - Long-term novelty analysis - Categorical parameter (top blue plot for the *reference* behaviour; bottom red plot for the behaviour being characterized by novelties).

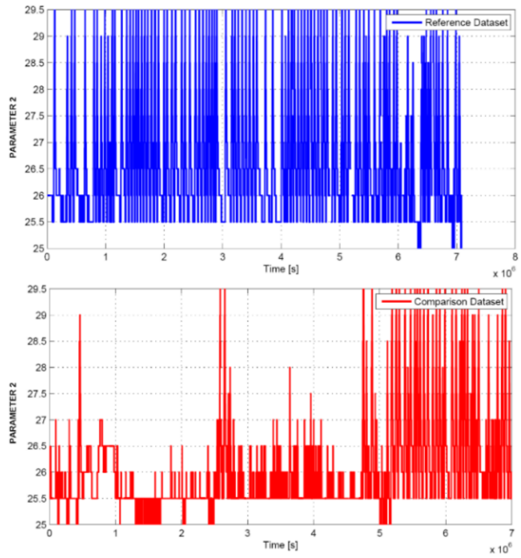

Figure 6 - Short-term novelty analysis - Numerical parameter (top blue plot for the *reference* behaviour; bottom red plot for the behaviour being characterized by novelties in the first part).

doi: 10.2788/854791

The second example is related to a parameter having shown novelties detected by the short-term analysis which aims at the detection of novelties of short duration within a long time interval. The parameter is numerical and also in this case it is clear (Figure 6) that the parameter behaviour is different (with respect to the *reference dataset*) in the first part of the *comparison dataset*, becoming similar in the last part.

For this analysis the method foresees the partition of the time interval of the *comparison dataset* into short time windows, named periods (e.g. 48 hours, see Figure 7 including 40 periods) and the computation, for each window of one priority score associated to the degree of novelty observed in the parameter behaviour in that specific window. Therefore, it is possible to rank the windows and to detect the periods in which the parameter behavior differs most from the reference one.

Table 1 shows an extract of the periods ranking for the parameter shown in Figure 7.

Period N°	Period starting date and time	Period ending date and time	Priority score
9	2009-05-01,00:00:00	2009-05-02,23:59:48	0.329
8	2009-04-29,00:00:00	2009-04-30,23:59:48	0.271
10	2009-05-03,00:00:00	2009-05-04,23:59:48	0.255
7	2009-04-27,00:00:00	2009-04-28,23:59:48	0.252
11	2009-05-05,00:00:00	2009-05-06,23:59:48	0.232
12	2009-05-07,00:00:00	2009-05-08,23:59:48	0.232
26	2009-06-04,00:00:00	2009-06-05,23:59:48	0.206
⋮	⋮	⋮	⋮
36	2009-06-24,00:00:00	2009-06-25,23:59:48	0
37	2009-06-26,00:00:00	2009-06-27,23:59:48	0
38	2009-06-28,00:00:00	2009-06-29,23:59:48	0
40	2009-07-02,00:00:00	2009-07-03,23:59:48	0
29	2009-06-10,00:00:00	2009-06-11,23:59:48	0

Table 1 - Ranking of short-time windows, named *periods*, according to the priority score computed for each period.

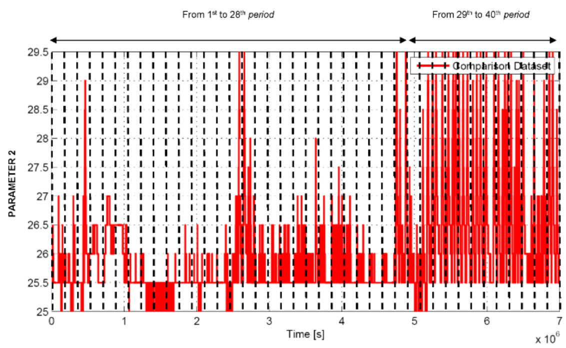

Figure 7 - Short-term novelty analysis - Numerical parameter - Windowing for the *comparison* dataset.

The periods characterized by the highest priority scores are the 8th, 9th 10th and 7th, which are all in the first part of the comparison time interval in which the parameter shows the most different behaviour with respect to the one observed in the *reference dataset*. On the other side the 36th, 37th, 38th, 40th and 29th periods are characterized by a priority score equal to 0, meaning that no novelties are detected in the behaviour of the parameter. Indeed it is clear from Figure 6 that the parameter behaviour in the last part of the *comparison dataset* is similar to the one observed in the *reference dataset*.

6. CONCLUSIONS

The results obtained by the methods included in ***FETCH*** and ***KETTY,*** applied to the telemetry data of two satellites provided by the European Space Agency including over 10.000 parameters each, were validated by visual inspection of the behaviour of the parameters on the top and bottom parts of the priority lists. These were consistent with the expectations proving that the methods are capable of identifying which parameters are characterized by the highest level of novelty and that should be addressed first by FCEs.

It must be highlighted that, although in the examples shown the methods were applied to ex-post analysis, once the diagnostic algorithms are trained they can run continuously, allowing online diagnostics, since their computational constraints are compatible with online operation.

Finally it is worth highlighting that, thanks to the general approach adopted for the definition and implementation of the knowledge extraction techniques, ***FETCH*** and ***KETTY*** are deemed easily applicable to a wide set of different application domains, which have the following characteristics:

1. Low or basic a priori knowledge about the system,
2. Large amount of heterogeneous data.

7. REFERENCES

[1] JC Gil, N. Narula, T. Lopez, "There can be only one: Heterogeneous satellite fleet automated operations with a single tool and language, the MEASAT case", SpaceOps 2014 Conference, 2014.

[2] A. Brighenti "Model-based diagnostics: a real opportunity for efficient vehicles management", SAE On-Board Diagnostics Symposium, Indianapolis Marriott Downtown, September 22-29, 2009.

[3] J.A. Martinez-Heras, A. Donati, M. G. F. Kirsch, F. Schmidt, "New telemetry Monitoring Paradigm with Novelty Detection", SpaceOps 2012 Conference, Stockholm, Sweden, June 11-15, 2012.

[4] P.-N. Tan, M. Steinbach and V. Kumar, *Introduction to Data Mining*, Pearson Education Limited, United Kingdom, 2014.

[5] M. Dash, K. Choi, P. Scheuermann, H. Liu, "Feature selection for clustering - A filter solution", 2nd IEEE Int. Conf. on Data Mining ICDM 2002, Maebashi Japan, pp. 115-122, 2002