DrMUST – a Data Mining Approach for Anomaly Investigation

José-Antonio Martínez-Heras[1]
Black Hat S.L., Córdoba, 14002, Spain

Alessandro Donati[2]
European Space Agency, Darmstadt, 64293, Germany

Bruno Sousa[3]
European Space Agency, Darmstadt, 64293, Germany

Jörg Fischer[4]
Vega Space GmbH, Darmstadt, 64293, Germany

DrMUST is a data mining MUST client that can support flight control engineers in their anomaly investigation tasks. It performs pattern matching and correlation analysis. The pattern matching functionality can be used to find occurrences of a certain behavior (e.g. to know when certain anomaly happened in the past and went unnoticed). The correlation analysis can find which parameters are involved in a certain event of interest (e.g. anomaly). The correlation analysis is based on statistical features and is more robust and efficient than the classical mathematical correlation, allowing to perform correlations in linear time. DrMUST can also be used to characterize how a spacecraft is affected by environmental changes (e.g. solar flares). This work describes the technology behind DrMUST and provides real examples from ESA operated missions.

I. Introduction

So far, when an anomaly happens, ESOC engineers mainly rely on their knowledge and experience. They would hypothesize what could be the reason for the anomaly and then verify in telemetry to prove or discard their hypothesis. The limitation of this approach is that the cause of the anomaly can only be found if somebody had the idea of looking in this direction.

In order to help ESOC's Flight Control Teams in anomaly diagnosis, DrMUST has been researched and developed. DrMUST is a data mining application that can systematically scan every telemetry parameter and check if it is involved in the anomaly under study (e.g. cause or effect). DrMUST changes the anomaly investigation paradigm: before engineers had to hypothesize possible causes and then verify them, now DrMUST reports which are the verified possible causes and engineers have to pick the relevant ones.

In addition, DrMUST can perform pattern matching. This feature allows finding similar behaviors to a given telemetry profile in the past. This is useful to know if the same anomaly happened in the past and went unnoticed.

DrMUST's biggest added value is in the area of anomaly investigation. However, it can also be used for other tasks like event characterization and time series searching.

II. Technology

In order to support anomaly investigation, DrMUST makes use of intensive data analysis techniques. The precondition in this case is fast data access to all housekeeping telemetry and ancillary data. DrMUST uses MUST[1,2] (Mission Utilities & Support Tools) as data provider. MUST is a client – server architecture that enables fast access

---

[1] Software Engineer / Researcher, HSO-OSA, Robert Bosch Strasse 5, Darmstadt Germany.
[2] Advanced Mission Concept & Technology Manager, HSO-OSA, Robert Bosch Strasse 5, Darmstadt Germany.
[3] Deputy Spacecraft Operations Manager, HSO-OPV, Robert Bosch Strasse 5, Darmstadt Germany.
[4] Spacecraft Operations Engineer, HSO-OPV, Robert Bosch Strasse 5, Darmstadt Germany.

to engineering telemetry and ancillary data. From the MUST point of view, DrMUST is a MUST client that manipulates mission data made available through the MUST API (Application Programming Interface).

DrMUST can support anomaly investigation by performing pattern matching in a given parameter or by performing correlation analysis across all parameters. These two technologies are discussed in this section.

A. Pattern Matching

It is often the case that the same anomaly happened in the past and went unnoticed because it remained within limits. Finding these occasions can be helpful to better characterize the anomaly, understand why it happened and prevent it from happening again.

So far it was not easy to find similar behaviours. The usual practice was to develop of ad-hoc scripts for recognizing each particular type of behaviour. This required lots of work and lots of visual inspection. In order to improve this process we have deployed DrMUST[3,4] at ESOC. DrMUST contains a pattern matching functionality that allows finding the most similar patterns in a generic and systematic way.

In our experience, similar patterns are almost never exactly the same. Therefore, we need to use a technique that allow us to recognize similar patterns as similar even in the presence of small deviations (either in amplitude or in time). For this reason we choose the Dynamic Time Wrapping (DTW) technique. DTW is often used in the area of speech recognition to recognize words spoken at different speeds by different speakers[5,6]. The problem is that the computational cost of DTW is very high. Fortunately, there are lower bounds techniques[7] that make possible efficient nearest neighbors searches.

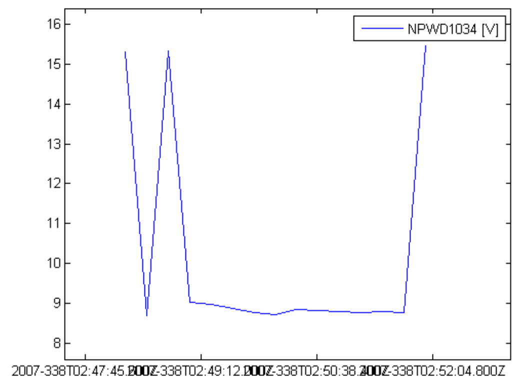

Figure 1. 4-minute pattern to be found in a 3 years time span.

An example of pattern matching in action is given in Figure 1 and Figure 2. In this case we need to find when the Venus Express solar arrays were operating at their maximum power in order to better characterize the battery efficiency. This is a quite rare situation since most of the time Venus Express is fully illuminated or it is in eclipse. The flight control team knew about one occasion at the end of 2007: whenever this happens the control signal provided by the power regulator lies between 7.5 and 10 V. An example time period showing the desired mode transition (to around 9V) over a period of four minutes was used to begin the search (see Figure 1). DrMUST searched for 4-minute patterns in a time span covering 3 years. It successfully found 2 other time periods with similar behavior to the given period (Figure 2). This process took less than 2 minutes on an office PC.

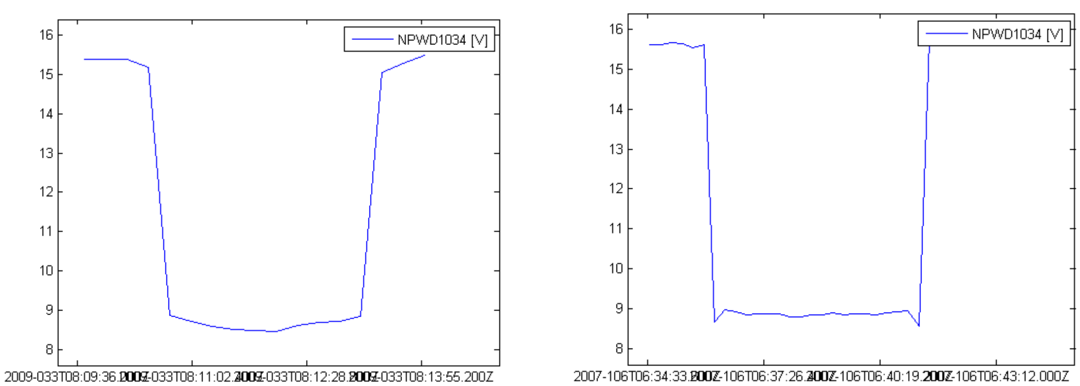

Figure 2. Two matches found by the DrMUST pattern recognition feature. *Note that the patterns are not exactly similar. This kind of flexible pattern matching is possible due to the use of the Dynamic Time Wrapping similarity distance (also used in speech recognition).*

B. Correlation

Whenever an anomaly is detected engineers try to identify its cause so that they can prevent it from happening again in the future. They also need to understand its effects in order to assess its severity. This task is particularly difficult because spacecrafts have in the order of 10,000 – 20,000 housekeeping telemetry parameters. The current approach consists of hypothesizing about possible causes and effects and verifying hypothesis by visual inspection. This approach has the disadvantage that it is not systematic and there are many potential correlations unconsidered.

In order to support engineers in finding correlations we have developed a correlator now built-in DrMUST[3,4]. It makes two basic assumptions:

- If a parameter is connected with the anomaly, it will behave in a similar way during all same-anomaly periods.
- If a parameter is connected with the anomaly, it will behave differently during nominal and anomaly periods.

The user defines several same-anomaly periods and at least one nominal period. In order to find the parameters correlated with the anomaly the correlator checks which parameters fulfill these two assumptions (similar behavior during anomalies and different behavior in nominal and anomaly periods). In order to characterize behavior, the correlator uses several statistical features (e.g. average, range, standard deviation, etc.). This process is depicted in Figure 3.

We would like to highlight the fact that a parameter is never compared to other parameters. Only parameter features are compared with same parameter features in different time periods. This is very computationally effective because it allows finding which parameters are involved in an anomaly in linear time. This is remarkable since most correlation approaches grow quadratically in complexity with the number of parameters as they compare the behavior of every parameter with the behavior of every other parameter.

An example of the results found by the correlator is shown in Figure 4. The problem was that from time to time the attitude errors in the Venus Express' Y axis were higher than expected. These time periods were input to the correlator as target periods and periods where the attitude error levels were expected were input as nominal periods. The correlator found, to the surprise of the flight control engineers, that the activation of a payload – the ASPERA scanner – was introducing a torque that was not considered. This proofs the power of the DrMUST correlation in finding correlations systematically; even those that engineers will think a priori that have no connection.

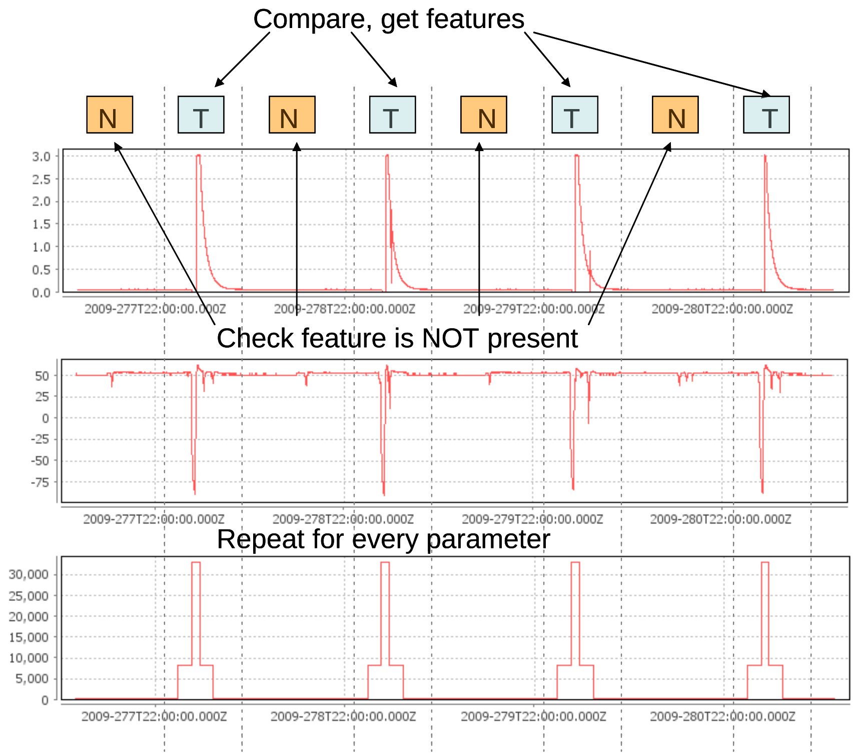

Figure 3. Correlation solving approach. *Users specify target time periods(T) where the event of interest they are investigating (e.g. anomaly) happened and nominal time periods (N). DrMUST correlator then checks which parameters have common features in target periods which are not present in nominal periods.*

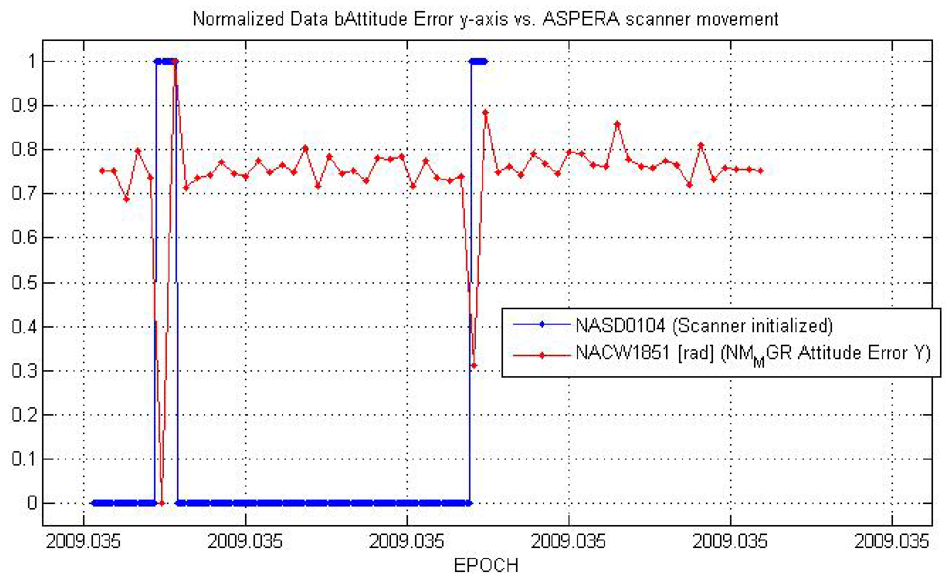

Figure 4. DrMUST Correlator finds the activation of the ASPERA scanner as the cause of the Attitude Errors in the Y axis on Venus Express.

III. Characterization

So far, the main focus of DrMUST has been in supporting anomaly investigation. In this section we highlight an additional application of DrMUST: characterization. By characterization we mean the process of understanding what is typical in the satellite about a particular operational mode or how the spacecraft is affected by an external environmental change. In this section we show how the correlator functionality of DrMUST has been applied to characterize Venus Express in two circumstances.

A. Venus Express under the effect of Solar Flares

We wanted to understand how Venus Express was affected by Solar Flares, so we configure DrMUST correlator to use as target period two times when solar flares affected Venus Express (5th June 2011, 7th March 2012). As nominal periods we selected the previous day for both cases in the same time interval. DrMUST reported correlations in parameters related to the fact that star trackers are blind during solar flares. In addition, a substantial increase of the number of corrupted bits in memory that were repaired both in the attitude and orbit control management system and in the data management system (see Figure 5) were also recognized. The Venus Express flight control team confirmed these findings as typical effects of solar flares in Venus Express.

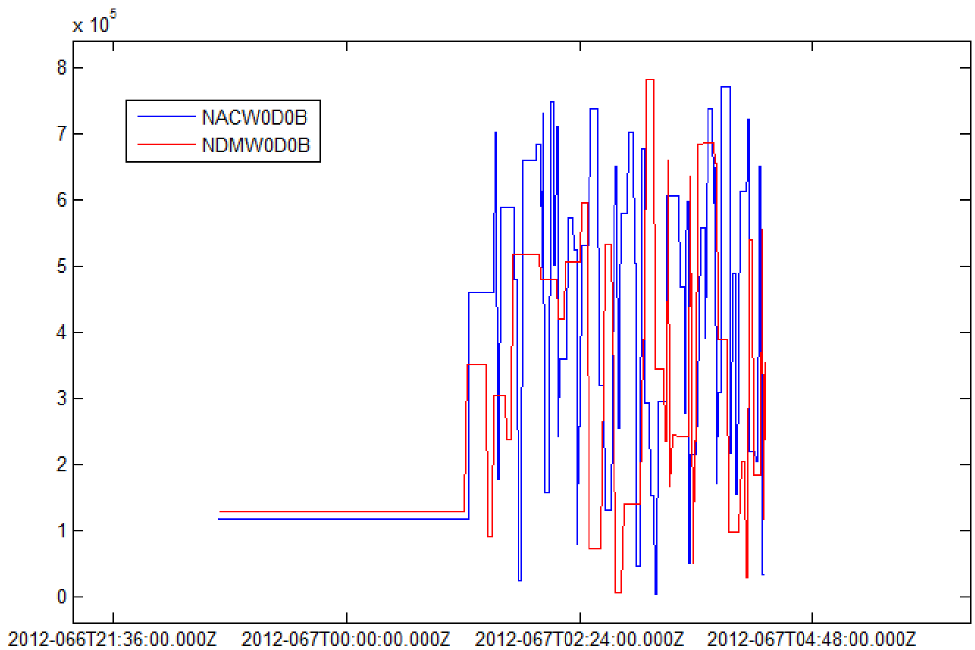

Figure 5. Number of bit errors detected and corrected. *The parameter NACW0D0B measures the number of bit errors detected and corrected. It is related to the Attitude and Orbit Control Management System processor module. The meaning of the parameter NDMW0D0B is the same except that it is related to the Data Management System processor module.*

B. Venus Express High Gain Antenna Switch

Venus Express has two high gain antennas (HGA). In this exercise, we characterize what changes when the HGA2 is used in contrast with the use of HGA1. For this purpose, the DrMUST correlator was configured using as target period several days from several HGA2 periods and as nominal periods we used several days from several HGA1 periods. The flight control team analyzed the correlation results and reported that DrMUST found the obvious characteristics of this phase, namely that the telemetry and telecommanding is configured to route the signal to HGA2, that the attitude and orbit control system is configured to point HGA2 to Earth and the system configuration knows it is using HGA2. It is remarkable the absence of thermal differences in spite of the attitude flip; however it considered to be normal as the Sun remains where it normally is.

IV. Conclusion

DrMUST supports flight control engineers in the process of anomaly investigation with a data mining approach. It major characteristics are:

- It requires very little engineering knowledge (input consists of a set of time periods)
- It makes no assumptions on how data should look like (e.g. other techniques require that data follow a Gaussian distribution)
- It is systematic: it checks every parameter, even the ones that engineers would think that a priori are not at all connected. This allows DrMUST to find surprising correlations.
- Performance: it is much faster than humans in either finding patterns or finding correlated parameters. In addition, once configured it runs unattended.

The Pattern Matching technique enables engineers to find similar behaviors through telemetry history. It can find a needle in a haystack (4-minutes pattern in 3 years of history as seen in section II). This is useful in the context of space operations for performing characterizations and in support to anomaly investigation. The implemented pattern matching techniques uses speech recognition technology in order to allow for some flexibility in time and amplitude.

The Correlator can systematically find which parameters are involved in a certain anomaly or period of interest. This is very useful in supporting anomaly investigation and characterizations since the number of housekeeping parameters is very large, usually in the order of 10,000 – 20,000. Our approach is robust due to the fact that it is based on statistical features rather than using time series values directly. An additional advantage is that the complexity grows linearly with the number of parameters; this is a big difference with classical correlation techniques since their complexity grows quadratically with the number of parameters, making it almost impossible to use it in problems of this magnitude. A further advantage as compared to classical correlation is that our approach can detect non-linear dependencies where the classical correlation only considers linear ones.

We believe DrMUST will play an important role in the way space operations diagnostics are performed. Before DrMUST, engineers had to hypothesize which parameters may have a credible correlation with the anomaly they face and then verify which one holds. With DrMUST, correlations are automatically and systematically found and engineers have to evaluate them. The DrMUST approach can also be extended to other non-space domains as it works with any collection of time series.

We are currently working in integrating DrMUST in a graphical user interface (GUI) via web[8] that will allow users to easily provide their inputs and evaluate DrMUST output visually.

References

[1]Martínez-Heras, J.; Baumgartner, A.; Donati, A.; “MUST: Mission Utility and Support Tools”, Proceedings DASIA 2005 conference, Edinburgh, UK.

[2]Baumgartner, A.; Martínez-Heras, J.; Donati, A.; Quintana, M.; “MUST – A Platform for Introducing Innovative Technologies in Operations”, Proceedings of the ISAIRAS 2005 conference, in Munich, Germany.

[3]Martinez-Heras, J.A., Yeung, K., Donati, A., Sousa, B., Keil, N. (2009); DrMUST: Automating the Anomaly Investigation First-Cut; IJCAI-09 Workshop on Artificial Intelligence in Space. Pasadena, California, USA, July 17 - 18, 2009.

[4]Martinez-Heras, J.A., Donati, A. (2011); Making Sense of Housekeeping Telemetry Time Series; IJCAI-2011 Workshop on Artificial Intelligence in Space. Barcelona, Spain, July 2011.

[5]Itakura F. (1975) Minimum prediction residual principle applied to speech recognition. IEEE Trans Acoustics Speech Signal Process ASSP 23:52–72

[6]Sakoe H. & Chiba S. (1978) Dynamic programming algorithm optimization for spoken word recognition. IEEE Trans Acoustics Speech Signal Process ASSP 26:43–49

[7]Keogh, E. (2002). Exact indexing of dynamic time warping. In 28th International Conference on Very Large Data Bases. Hong Kong. pp 406-417

[8]Oliveira, H.; Lais, A.; Francisco, T.; Donati, A.; "Enabling visualization of large telemetry datasets"; Proceedings SpaceOps 2012 Conference, Stockholm, Sweden, June 11-15 (to be published).