Anomaly detection of satellite telemetry data based on extended dominant sets clustering

Anomaly detection of satellite telemetry data based on extended dominant sets clustering

Xin Jin, Hui Quan Wang* , Zhong He Jin

Micro-satellite Research Centre, Zhejiang University, Hangzhou 310000, China

*Corresponding author's e-mail: hqwang@zju.edu.cn

Abstract: To mine out anomalies in satellite telemetry data under unsupervised conditions, a cluster-based method is proposed in this paper. Firstly, an extended dominant sets clustering algorithm is proposed to cluster the telemetry data with arbitrary shapes. Secondly, objects that do not belong to any cluster or belong to small clusters are traditionally identified as anomalies. Thirdly, the anomalies in large clusters are detected according to the relative similarity. Finally, the information on anomaly windows in the telemetry sequence is obtained according to the local anomaly rate, which provides more characteristics of the anomalies. Experimental results show that: 1) The proposed extended dominant sets clustering algorithm can deal with the dataset containing multiple and arbitrarily shaped clusters; 2) The introduction of relative similarity increases the AUC values of anomaly detection by 3%~10%; 3) The proposed anomaly detection method can effectively detect the anomalies in magnetometer telemetry data of Tianping-2B satellite. Therefore, the proposed method can achieve anomaly detection of satellite telemetry data under unsupervised conditions, and provide support for improving satellite safety and reliability.

1. Introduction

Once a spacecraft like a satellite is launched into orbit, the telemetry data becomes the only basis for the ground to judge its status. Therefore, the anomaly detection of telemetry data becomes an important way to enhance the reliability and safety of spacecraft in-orbit [1]. Since a spacecraft system is very complex and its telemetry data is highly relevant, professional, multi-modal, and unpredictable, it is extremely difficult and costly to generate an accurate and comprehensive labelled training set for anomaly detection. Therefore, unsupervised anomaly detection algorithms are gradually becoming mainstream in the field [2][3].

Among the many unsupervised anomaly detection algorithms, cluster-based methods are often used for spacecraft telemetry data. For example, NASA has developed a spacecraft system health monitoring software — IMS (Inductive Monitoring System) [4][5] based on cluster-based anomaly detection technology. IMS performs clustering processing on the achieved system data and characterizes nominal interactions between selected parameters [6]. When the system behaviour deviates significantly from the nominal IMS models, an alarm is provided to remind people of potential system failures or precursors of major failures. Inspired by IMS, Gao [7] proposed a normal behaviour clustering-based anomaly detection method and used this method to detect the fault of the satellite power system. Yairi [8] proposed a satellite health monitoring method based on probabilistic clustering and dimensionality reduction and successfully applied it to the small demonstration satellite 4(SDS-4) of the Japan Aerospace Exploration Agency (JAXA). Wang [9] proposed a density peak clustering-based two-stage anomaly detection method and mined out the anomaly of the analogue solar sensor on the ZDPS-1A satellite.

The above cluster-based anomaly detection methods consider objects belonging to large clusters to be normal objects. However, since the goal of clustering is to mine clusters in the dataset, there is no guarantee that all anomalies are excluded from large clusters. In addition, the output of the above method isolated anomalies without further mining the association between them. For these phenomena, we propose a novel cluster-based anomaly detection method, which will be described in detail in Section 2. We will discuss the experimental results and conclude the paper in Sections 3 and 4, respectively.

2. Description of the proposed method

The telemetry data anomaly detection algorithm proposed in this paper mainly includes four parts: sequence segmentation, clustering, anomaly detection, and anomaly window searching:

(1) **Sequence segmentation**: Split telemetry sequence into multiple sub-sequences. For a telemetry sequence with length L, a non-overlapping sliding window with size N is used for segmentation to form $\lceil L/N \rceil$ sub-sequences. For the last sliding window, we fill the first $\lceil L/N \rceil \cdot N - L$ part of it with the last $\lceil L/N \rceil \cdot N - L$ samples of the penultimate subsequence.

(2) **Clustering:** Clustering the objects in each subsequence.

(3) **Anomaly detection**: Find out all anomalies in each subsequence.

(4) **Anomaly window searching**: Search the starting positions and lengths of all anomaly windows in the entire telemetry sequence.

The content of clustering, anomaly detection, and anomaly window searching will be described in detail below.

2.1. Extended dominant sets clustering

According to the principle of dominant sets clustering [10], for a dataset $\boldsymbol{D}$ of N size, the similarity matrix on $\boldsymbol{D}$ is defined as $\boldsymbol{A}=(a_{pq})$, $p,q \in \boldsymbol{D}$, where a_{pq} represents the weighted similarity between object p and q, and a_{pq} is given by:

$$a_{pq} = \begin{cases} e^{-\frac{d(p,q)}{\sigma}} & p \neq q \\ 0 & p = q \end{cases} \tag{1}$$

In (1), $d(p,q)$ means the Euclidean distance of object p and q; σ is the attenuation factor, which is taken as σ=0.5median(d). The dominant set of $\boldsymbol{D}$ is represented by $\boldsymbol{S}$, where $\boldsymbol{S}$ is a non-empty data set and $\boldsymbol{S} \subseteq \boldsymbol{D}$. To save space, the solution process of $\boldsymbol{S}$ is not described here, and readers can be referred to in [10]. In this paper, we take $\boldsymbol{S}$ as the cluster centre set, and name p as the centre member if $p \in \boldsymbol{S}$. To complete a cluster, we define the ε-similarity neighbourhood of p as:

$$N_\varepsilon(p) = \{q \mid q \in \boldsymbol{D}, a_{pq} \geq \varepsilon\} \tag{2}$$

Here ε is given by:

$$\varepsilon = \max\left(\text{median}\left(\boldsymbol{A}(p,\boldsymbol{S})\right), p \in \boldsymbol{S}\right) \tag{3}$$

If q is an ε-similarity neighbour of p, then p is an ε-similarity neighbour of q. Furthermore, p and q are considered to have the ε-similarity relationship. Based on the dominant set $\boldsymbol{S}$ and the ε-similarity relationship, a cluster is defined as the largest object set derived from the centre set $\boldsymbol{S}$ through the ε-similarity relationship. Formula (8) gives the recursive definition of the cluster:

$$\boldsymbol{C}^{[i]} = \begin{cases} \boldsymbol{S} & i = 0 \\ \boldsymbol{C}^{[i-1]} \cup N_\varepsilon\left(p \in \boldsymbol{C}^{[i-1]}\right) & otherwise \end{cases} \tag{4}$$

Here i is the growing step of cluster $\boldsymbol{C}$, and $\boldsymbol{C}$ will stop growing once $\boldsymbol{C}^{[i]}=\boldsymbol{C}^{[i-1]}$ occurs.

The pseudocode in Table 1 describes the process of extended dominant set clustering.

Table 1. Extended dominant set clustering

Input	***D***	/* data set*/
Output	***ClusteId***	/* cluster id of each object in ***D*** */
	$\boldsymbol{S}_{size}$	/*size of each dominant set */
(1)	k=1; N=|***D***|; ***Cluster Id*** = -1*ones(N,1)	
(2)	Calculate ***A*** according to formula (1);	
(3)	**While** |***D***|>0.0001*N **do**	
(4)	Get the current dominant set: ***S***;	
(5)	Record dominant set size: $\boldsymbol{S}_{size}(k)$=|***S***|;	
(6)	Get k^{th} cluster according to formula (4): ***C***;	
(7)	Record the cluster id of each cluster member: ***ClusterId***(***C***)=k;	
(8)	k++;	
(9)	**end While**	

2.2. Anomaly detection

The cluster-based anomaly detection algorithm considers objects that do not belong to any cluster or belong to a small cluster as anomaly objects [11][12]. According to Table 1, an object with a cluster id of -1 is an object that does not belong to any cluster, and thus it can be treated as an anomaly.

Suppose $\boldsymbol{C} = \{\boldsymbol{C}_1, \boldsymbol{C}_2, ..., \boldsymbol{C}_k\}$ is the set of clusters in the sequence that $|\boldsymbol{C}_1| \geq |\boldsymbol{C}_2| \geq |\boldsymbol{C}_k| > 1$, and k is the number of clusters in ***D***. To identify small clusters, we define b as the boundary of a large and small cluster if one of following formulas holds:

$$\begin{cases} \left|\boldsymbol{C}_{b-1}\right| / \left|\boldsymbol{C}_b\right| \geq 5 \\ \left|\boldsymbol{C}_b\right| \leq \boldsymbol{S}_{size}(b-1) \end{cases} \tag{5}$$

Here $\boldsymbol{S}_{size}$ records the size of the dominant set of each cluster. Then the set of a large cluster is defined as: $\boldsymbol{LC} = \{\boldsymbol{C}_i, 1 \leq i < b\}$; and the set of the small cluster is defined as: $\boldsymbol{SC} = \{\boldsymbol{C}_i, b \leq i \leq k\}$. Objects belonging to ***SC*** are treated as anomalies.

To detect the anomalies in ***LC***, we calculate the similarity of p about its m-distance neighbourhood:

$$sim(p) = \frac{1}{m} \sum_{q \in \boldsymbol{M}_m(p)} a_{pq} \tag{6}$$

Here $\boldsymbol{M}_m(p)$ represents the set of m objects with the highest similarity to p in ***LC***, and m is calculated by Formula (7):

$$m = \max\left(\left\lceil \frac{|\boldsymbol{C}_i|}{\boldsymbol{S}_{size}(i)} \right\rceil \right),\ 1 \leq i < b \tag{7}$$

Based on (6), we define the relative similarity of p about its m-distance neighbourhood as:

$$rsim(p) = \frac{\frac{1}{m} \sum_{q \in \boldsymbol{M}_{m(p)}} sim(q)}{sim(p)} \tag{8}$$

The *rsim* value reflects the local anomaly degree of large cluster members: the more p deviates from its neighbours, the larger the $rsim(p)$ value is.

According to the above anomaly detection process, the anomaly label of object p is given by:

$$AnomalyLabel(p) = \begin{cases} 0 & rsim(p) < thresh;\ p \in \boldsymbol{LC} \\ 1 & otherwise \end{cases} \tag{9}$$

Here *thresh* is the anomaly filtering threshold, and is calculated from the mean and standard deviation of *rsim* values:

$$thresh = rsim_{mean} + 5 \cdot rsim_{std} \tag{10}$$

2.3. Anomaly window searching

Suppose the set of anomalies in the telemetry sequence is formed as $\boldsymbol{O} = \{o_1, o_2, \ldots, o_n\}$, where o_i represents the serial number of the i^{th} anomaly in the telemetry sequence, $1 \le i \le n$ and n are the total numbers of anomalies. If i, j holds Formula (11), then $[o_i, o_j]$ is treated as an anomaly window.

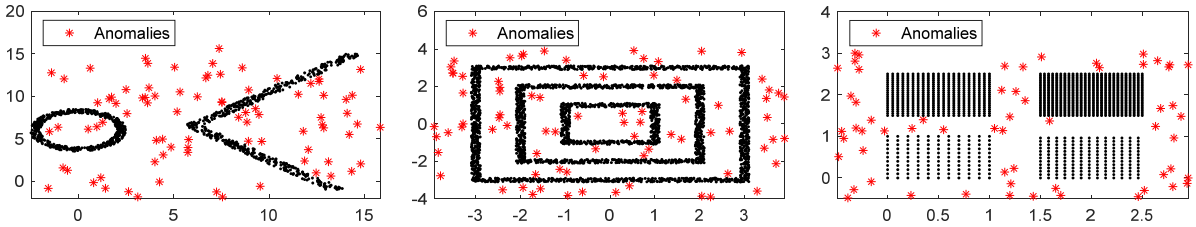
$$\begin{cases} \dfrac{j-i+1}{o_j - o_i + 1} \ge Rate_{th} \\ \dfrac{j+1-i+1}{o_{j+1} - o_i + 1} < Rate_{th} \end{cases} \tag{11}$$

Here $Rate_{th}$ represents the threshold value of the local anomaly rate, with the range of (0, 1]. And the larger $Rate_{th}$ is, the narrower the anomaly window generated; especially when the $Rate_{th}$ is set to 1, the objects in the anomaly window generated will be all anomalies.

3. Experimental results and analysis

3.1. Synthetic data sets

To intuitively display the clustering and anomaly detection process, and evaluate the impact of the introduction of relative similarity on the detection results, this experiment uses three synthetic data sets (see Figure 1 and Table 2 for details). And the AUC (Area Under Curve) value under the ROC (Receiver Operating Characteristic) curve is used as the evaluation index.



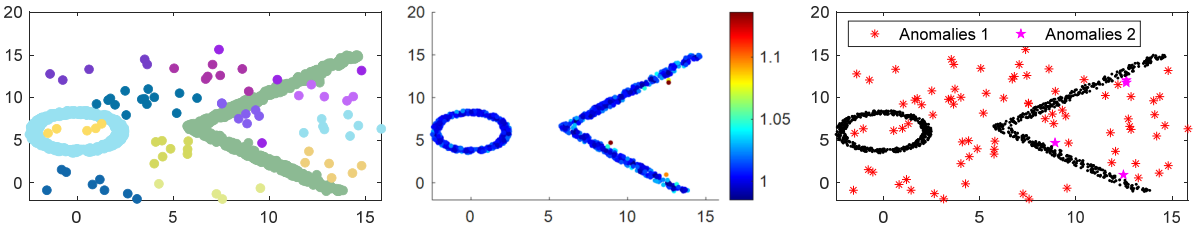
(a) D1 (b) D2 (c) D3

Figure 1. Synthetic data sets

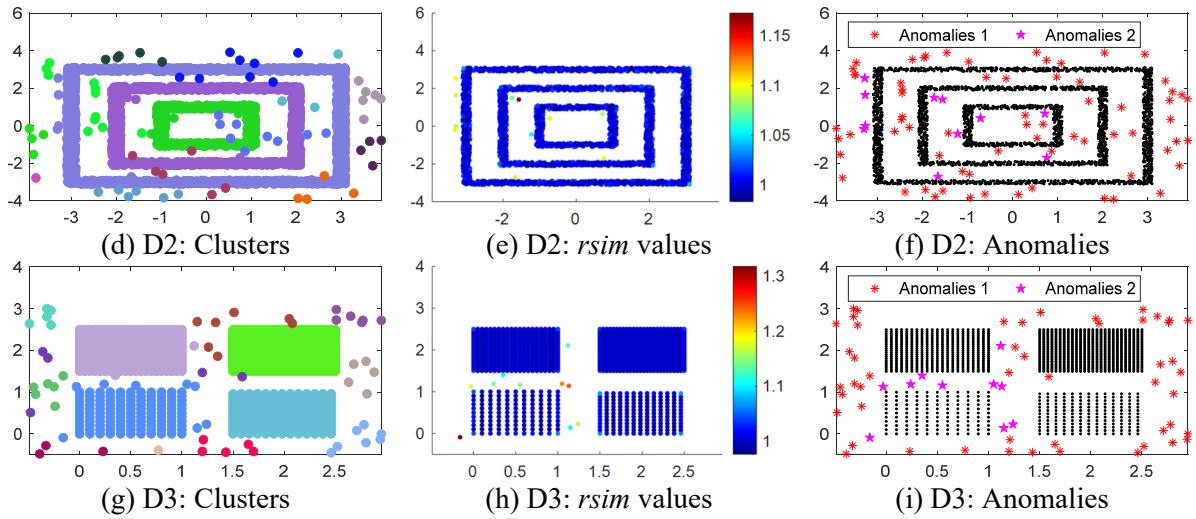
Table 2. Description of synthetic data sets

Name	Size	Number of large clusters	Number of anomalies	Anomaly rate
D1	1137	2	85	7.48%
D2	3031	3	79	2.61%
D3	1467	4	60	4.09%

Figure 2 shows the clustering, *rsim* calculation, and anomaly detection results of the proposed method on synthetic data sets. In Figure 2, 'Anomalies 1' indicates the anomalies detected traditionally, which are objects that do not belong to any cluster or belong to small clusters, while 'Anomalies 2' indicates the anomalies detected according to the relative similarity index.



(a) D1: Clusters (b) D1: *rsim* values (c) D1: Anomalies



(d) D2: Clusters (e) D2: *rsim* values (f) D2: Anomalies

(g) D3: Clusters (h) D3: *rsim* values (i) D3: Anomalies

Figure 2. Results on synthetic data sets

Experimental results show that: 1) The extended dominant sets clustering method proposed in this paper can handle data sets with different cluster numbers and cluster shapes; 2) The AUC values obtained by traditional means are 0.9706, 0.9241 and 0.9083; 3) The introduction of relative similarity increase AUC values by 3.03%~10.09%.

3.2. Magnetometer telemetry data

This experiment uses the magnetometer telemetry data of the real in-orbit satellite named Tianping-2B, as shown in Figure 3. The data is generated from 16:09:00~22:00:00 on July 5, 2022, with a total of 21,010 samples. Tianping-2B satellite has a 600km sun-synchronous orbit, and the orbital period is about 96 minutes. Besides, the magnetometer telemetry data has 3 dimensions, namely X-axis, Y-axis and Z-axis; and the update frequency is 1Hz. According to the above physical characteristics, the sliding window size is set as N=5760, that is, the divided subsequence contains the data amount of one orbital period.

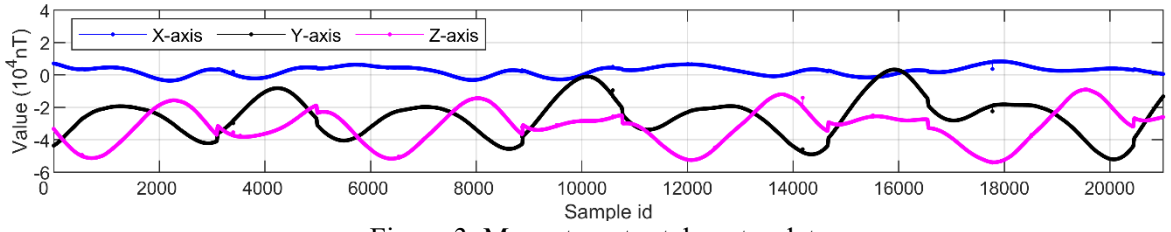


Figure 3. Magnetometer telemetry data

Figure 4 shows the result of the anomaly detection method proposed in this paper on magnetometer telemetry data. According to the result, this magnetometer data contains 1.07% of anomalies, which form 19 anomaly windows. Information on these anomaly windows is shown in Table 3.

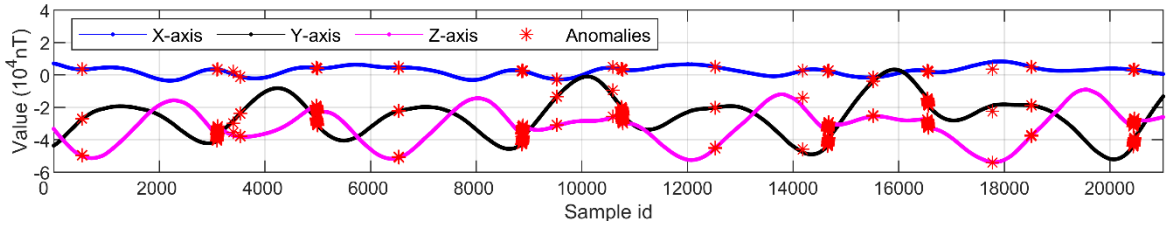


Figure 4. Anomaly detection result on Magnetometer telemetry data

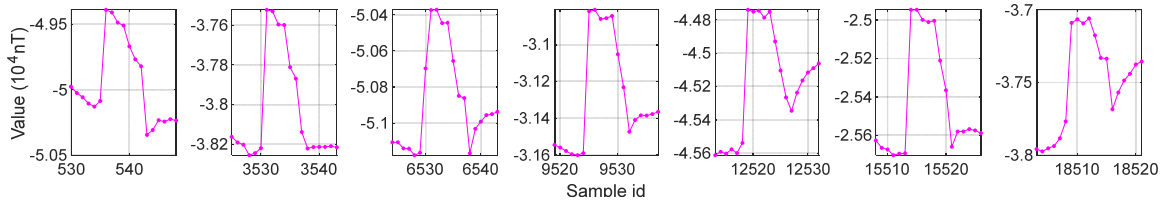
Table 3. Information on each anomaly window

Id	Window type I		Window type II		Window type III	
	Start position	Length	Start position	Length	Start position	Length
1	3398	1	536	7	3089	18
2	8861	1	3531	6	4969	27
3	10587	1	6530	9	8869	21
4	14182	1	9525	6	10749	32
5	17776	1	12524	5	14652	24
6	-	-	15514	8	16542	25
7	-	-	18508	10	20439	26

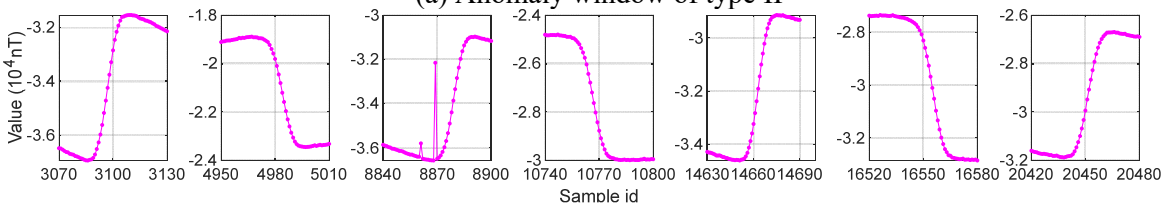
(1) 5 anomaly windows of Type I: This type of window has a length of 1, and anomalies appear randomly. These anomalies are considered to be wild values.

(2) 7 anomaly windows of Type II: The length of such windows is between 5 and 10, and the average difference between the starting positions of adjacent windows is about 2995 sample points. Figure 5(a) shows the Z-axis data of the magnetometer around these windows. It is speculated that the magnetometer will be subject to fixed interference every 49 minutes and 55 seconds on average. Furthermore, according to the expert knowledge of the system, it is determined that the interference comes from the reconfiguration operation of the TT&C transponder A. To resist the space single-particle effect, the TT&C transponder of the Tianping-2B satellite is designed to automatically reconfigure the main FPGA every 50 minutes. During the reconfiguration process, the transponder device has a small current, and the magnetic field generated by this current is weak, which leads to a decrease in the amplitude of magnetometer data.

(3) 7 anomaly windows of Type III: The length of such windows exceeds 15, and the starting positions of adjacent windows are separated by about 1883 and 3900 sample points, corresponding to the interval of 31 minutes 23 seconds and 65 minutes 0 seconds, respectively. Figure 5(b) shows the Z-axis data around these windows. Since the orbital period is about 94 minutes, it is speculated that this phenomenon may occur during the movement of the satellite in and out of shadow areas. According to the analysis of the satellite mechanical structure and the output of the solar sensors, it is determined that the interference comes from the input current of the expanded solar panel, which is manifested: 1) When the solar panel switches from the illuminated to the unilluminated state, the input current decreases rapidly, and the magnetic field strength generated by the current weakens rapidly, which will cause the amplitude of the X-axis output of the magnetometer to increase rapidly, while the amplitude of the Y-axis and Z-axis output decreases rapidly; 2) When the solar panel switches from the unilluminated to the illuminated state, the strength of magnetic field changes in the opposite way.



(a) Anomaly window of type II



(b) Anomaly window of type III

Figure 5. Details of Z-axis data of magnetometer in anomaly windows

The above results show that the anomaly detection method based on extended dominant sets clustering proposed in this paper can effectively mine the anomalies in the magnetometer telemetry data of Tianping-2B, and can provide data support for the ground to grasp the health status of the magnetometer device, and the changes of the magnetic field inside the satellite.

4. Conclusions

This paper proposes a clustering-based anomaly detection method for telemetry data, which mainly includes extended dominant sets clustering, anomaly detection and anomaly window searching. Experimental results show that this method can effectively mine anomalies in satellite telemetry data under unsupervised conditions, and provide data support for improving the safety and reliability of satellites in orbit.

Acknowledgments

This research was supported by the Fundamental Research Funds for the Central Universities (2021XZZX038).

References

[1] Peng X Y, Pang J Y, Peng Y and Liu D T. (2016) Review on anomaly detection of spacecraft telemetry data. *Chinese Journal of Scientific Instruments*, 9: p1929-1945.
[2] Hundman, K, Constantinou V, Laporte C, Colwell I and Soderstrom T. (2018) Detecting Spacecraft Anomalies Using LSTMs and Nonparametric Dynamic Thresholding. *In Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*, London, p387-395.
[3] Nalepa, J, Myller M, Andrzejewski J, Benecki P and Piechaczek S. (2022) Evaluating algorithms for anomaly detection in satellite telemetry data. *Acta Astronautical*, 2022, 198: p689-701.
[4] Iverson D L. (2004) Inductive system health monitoring. *International Conference on Artificial Intelligence*. p1-7.
[5] Iverson D L. (2008) System Health Monitoring for Space Mission Operations. *IEEE Aerospace Conference*, Montana, p1-8.
[6] Iverson D, Martin R, Schwabacher M, Spirkovska L and Taylor W. (2009) General Purpose Data-Driven System Monitoring for Space Operations. *AIAA Infotech @ aerospace conference*, Seattle, p1-11.
[7] Gao Y, Yang T S, Xu M Q, Xing N. (2012) An Unsupervised Anomaly Detection Approach for Spacecraft. *Fifth International Conference on Intelligent Computation Technology and Automation,* Zhangjiajie, p478-481.
[8] Yairi T, Takeishi N, Oda T, Nakajima Y, Nishimura N and Takata N. (2017) A Data-Driven Health Monitoring Method for Satellite Housekeeping Data Based on Probabilistic Clustering and Dimensionality Reduction. *IEEE Transactions on Aerospace & Electronic Systems*, 53: p1384-1401.
[9] Wang C, Wang H Q, Jin Z H. (2018) Pico-satellite telemetry anomaly detection through clustering. *Journal of Harbin Institute of Technology*, 50: p110-116.
[10] Pavan M, Pelillo M. (2007) Dominant Sets and Pairwise Clustering[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29: p167-172.
[11] He Z, Xu X, Deng S. (2003) Discovering cluster-based local outliers. *Pattern Recognition Letter*, 24: p1641-1650.
[12] Huang J L, Zhu Q S, Yang L J, Cheng D D, W Q W. (2017) A novel outlier cluster detection algorithm without top-n parameter. *Knowledge-Based Systems*, 121: p32-40.