Anomaly Detection and Identification in Satellite Telemetry Data Based on Pseudo-Period

Haixu Jiang [1,2], **Ke Zhang** [1,2], **Jingyu Wang** [1,2,*], **Xianyu Wang** [1,2] **and Pengfei Huang** [1,2]

[1] School of Astronautics, Northwestern Polytechnical University (NPU), 127 Youyi Xilu, Xi'an 710072, Shaanxi, China; haixuj@163.com (H.J.); zhangke@nwpu.edu.cn (K.Z.); xianyu.wang@mail.nwpu.edu.cn (X.W.); microsofthpf@163.com (P.H.)

[2] National Key Laboratory of Aerospace Flight Dynamics, Xi'an 710072, Shaanxi, China

[*] Correspondence: jywang@nwpu.edu.cn

Featured Application: The method proposed in this paper takes the advantage of pseudo-period to process the massive telemetry data for precise anomaly detection and identification. It can be mainly applied to the fault diagnosis, fault detection and health management in some specific complex control systems such as the industrial application scene.

Abstract: To effectively detect and identify the anomaly data in massive satellite telemetry data sets, the novel detection and identification method based on the pseudo-period was proposed in this paper. First, the raw data were compressed by extracting the shape salient points. Second, the compressed data were symbolized by the tilt angle of the adjacent data points. Based on this symbolization, the pseudo-period of the data was extracted. Third, the phase-plane trajectories corresponding to the pseudo-period data were obtained by using the pseudo-period as the basic analytical unit, and then, the phase-plane was divided into statistical regions. Finally, anomaly detection and identification of the raw data were achieved by analyzing the statistical values of the phase-plane trajectory points in each partition region. This method was verified by a simulation test that used the measured data of the satellite momentum wheel rotation. The simulation results showed that the proposed method could achieve the pseudo-period extraction of the measured data and the detection and identification of the anomalous telemetry data.

Keywords: anomaly detection and identification; satellite telemetry data; data symbolization; pseudo-period; phase-plane trajectory

1. Introduction

The massive amounts of telemetry data transmitted by an in-orbit satellite are the sole observational basis of the satellite's operation. Through the analysis of telemetry data, ground telemetry, track, and command (TT&C) stations can determine the satellite's operational state and detect possible anomalies in a timely fashion, assisting the normal operation of the in-orbit satellite. Currently, the fixed threshold method is the main method for anomaly detection in telemetry data in engineering [1,2]. With the simple mechanism and rapid detection speed, this method can be used to detect anomalies beyond the threshold in a timely manner. However, due to the influence of complex noise in the actual telemetry data, the fixed threshold method is prone to producing false alarms in the detection. In addition, the method cannot detect anomalies within the threshold. The current detection and identification studies are based on the relevant features extraction of telemetry data using data mining methods.

Historical telemetry data have been used for modeling, with the measured data compared with the predicted data of the proposed model in order to achieve the anomaly detection and identification

of the measured data [3–5]. The performance of this method is directly related to the modeling accuracy. When the data type changes, the data model must be updated. Based on the threshold method, Song [6] used the semi-major axis change method (SACM) to extract the mean and standard deviation of telemetry data in different periods as the migration variables for anomaly detection and identification. However, this method showed limited accuracy for anomaly identification. To diagnose the satellite faults, Sherr [7] used empirical mode decomposition to extract the telemetry data features from the time-frequency domain. However, the characteristic frequencies of each component in the data must be determined first, which is difficult in practical application.

Ren [8] used a fuzzy interval-set and tried to detect the potential value of the time series, where the algorithm regarded each subsequence as an interval-set depending on its value space and the subsequence boundary. The experimental results indicated that the algorithm had better discriminative performance than the piecewise aggregate approximation method. However, when the interval approximation was used for anomaly identification, the accuracy of this method was unsatisfactory. Neural networks have been employed to detect and identify anomaly data [9–11]. While these methods are more accurate, their precision mainly depends on the amount of training data. Thus, the performance of these methods becomes significantly worse when insufficient training samples are used. Zhao [12] used the stochastic subspace identification (SSI) method to extract the fault feature vectors of the vibration signal of a wind turbine bearing, and then input the fault feature vectors into the multi-kernel support vector machine (MSVM) for fault pattern identification. The experimental results showed that this algorithm could successfully identify fault types of bearing. However, this method is similar to the neural network method in that a sufficiently large training data set is required to achieve high accuracy, and the number of the parameters that must be set is also large, decreasing the speed of the identification.

With the increasing complexity of satellites, the amount of telemetry data is also expanding. Due to their limited feasibility in use for the diagnosis of anomalous satellite telemetry data, the methods described above cannot meet engineering requirements. A fault diagnosis method based on the phase-plane trajectory chart was proposed in [13–15]. By comparing the difference in the phase-plane trajectory between the measured data and the sample data, rapid detection of anomaly data was achieved, and the related features of the phase-plane trajectory were further extracted to identify the anomaly data. Due to its high calculation speed and high identification rate, this method has been widely used in power system fault detection. However, this method has the same data length selection problem as the existing satellite telemetry data analysis methods. In the analysis of telemetry data, the fixed period obtained either by the expert experience method, or the periodogram method [16] is used as the basic analytical unit. Since the actual telemetry data are mostly non-stationary, truncation error will occur when using the fixed period analysis method. If each dynamic period is taken as the basic analytical unit, the error caused by the fixed period partition can be effectively avoided, thus providing a more effective guarantee for the accuracy of the subsequent analysis. The pseudo-period has been used to analyze the data stream, and the change management of the pseudo-periodic data stream has been achieved by linear regression and error segmentation [17]. In detail, pseudo-period analysis is mainly used in medical signal processing. Voss [18] presented an algorithm to upsample an insufficiently sampled pseudo-periodic signal with the help of a reference signal "hyper sampling". Then, hyper sampling was applied to dynamic magnetic resonance imaging (MRI) data of the human brain; the results showed that the MRI data analysis accuracy was improved [18]. Jin [19] used examples of bilinear and linear time-frequency distribution, and the Cohen class distribution and wavelet transform were discussed and used in the ballistocardiogram pseudo-period segment. The significance and importance of pseudo-periodic analysis were emphasized. Du [20] proposed a method of subsequence analysis of telemetry data based on the pseudo-period. To detect anomalous telemetry data, a detection sequence was dynamically generated by the subsequence segmentation algorithm based on the extreme values in the maximum period window width of the data and mean sequence. However, the pseudo-periodic partition is still

based on the orbit periodic datum. In practice, the classification of different types of telemetry data incurs an extremely large computational cost and shows limited adaptability.

In summary, the pseudo-periodic analysis of satellite telemetry data is still in the preliminary stage. Considering the requirement of fast detection and identification, this paper proposed a novel method of detection and identification of satellite anomalous telemetry data based on the pseudo-period. The similar shape of telemetry data was taken as the basic analytical unit for dividing telemetry data in order to extract the pseudo-period of the data. Then, the phase-plane trajectory characteristics of the pseudo-periodic data were extracted. Through the comparison and analysis with the sample data, the anomaly detection and identification in the pseudo-periodic data were achieved. It is worth noting that when the proposed method was used to process different types of data, the specific parameters should be set separately.

The main contributions of this work are given below:

- A method to extract the salient points of the data was proposed to compress the raw satellite telemetry data.
- The tilt angle of the adjacent data points was symbolized in the compressed data in order to obtain the corresponding string. Then, the pseudo-periods of the measured data were extracted using the standard pseudo-periodic sequence string.
- A fast method based on phase-plane trajectory plots was proposed to identify abnormal satellite telemetry data. In detail, the pseudo-period was taken as the basic analytical unit, and then the phase-plane trajectory plots corresponding to each pseudo-period data were taken as the features for achieving anomaly detection and identification in the satellite telemetry data.

2. Pseudo-Periodic Extraction of Satellite Telemetry Data

2.1. Pseudo-Period Analysis of Satellite Telemetry Data

Satellite telemetry data provide the operational state information for each system on the satellite and are received by the ground TT&C station in the time sequence. The typical telemetry data $x(t_i), i \in [1, N]$ can be described as:

$$x(t_i) = s(t_i) + n(t_i) \tag{1}$$

where t_i is the time point of the telemetry data acquisition, $s(t_i)$ is the ideal telemetry value, and $n(t_i)$ is the noise. The composition of the noise can be expressed as:

$$n(t_i) = \varepsilon(t_i) + d(t_i) + o(t_i) \tag{2}$$

where $\varepsilon(t_i)$ is the random noise widely distributed in the data, $d(t_i)$ is the noise in the numerical trend pattern, $o(t_i)$ is the outlier that only occurs at few time points, and the absolute value of $o(t_i)$ is significantly greater than 0 and significantly deviates from the data.

The satellite runs in its own orbital cycle, so the telemetry data transmitted by the satellite mostly follows the cyclic law of approximate periodic change, which is defined as the pseudo-period [16]. Due to the non-stationarity of the pseudo-periodic change, the change law cannot be accurately determined and has no rigorous and unified mathematical definition. To date, there have been few pseudo-period-based studies of telemetry data. According to the characteristics of the telemetry data, the pseudo-period of the satellite telemetry data $T_{fi}, i \in [1, N]$ is defined as:

$$T_{fi} = T_r + \Delta t_i \tag{3}$$

where T_r is the ideal period of the telemetry data, Δt_i is the time deviation of the different pseudo-period whose value fluctuates in a certain range, $\Delta t_i \in [\tau_a, \tau_b]$ and $\Delta t_i \ll T_r$. If $\Delta t_i = 0$, then $T_{fi} = T_r$, namely, the period is an ideal period. For simplicity, a single ideal period is defined as a standard pseudo-periodic sequence $\overline{X}_{fT}$. Figure 1 shows the measured data of the bus current of the solar panels of a satellite.

These data exhibit a typical law of approximate periodic variation. According to the turning point of the data shape, 13 approximate periods can be determined, as shown in Figure 1.
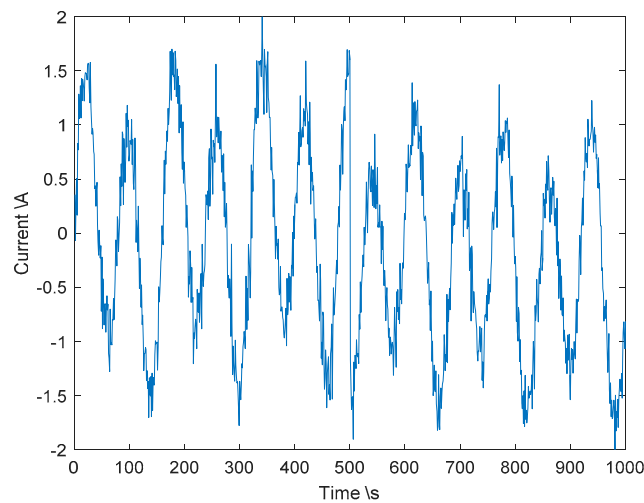
Figure 1. Measured data of the solar panel current of a satellite.

2.2. Symbolization Method of Satellite Telemetry Data

Since satellite telemetry data usually have a high sampling frequency, pseudo-periodic analysis of the raw telemetry data incurs a heavy computational cost. To improve the efficiency of pseudo-periodic extraction, the raw telemetry data should be compressed. Symbolic processing is a compression method based on the data shape features. This method has a high compression rate and fast compression speed and can reduce the noise and outliers' interference in the data.

Symbolic aggregate approximation (SAX) [21] is a data symbolization processing method based on piecewise aggregate approximation (PAA). The principle of SAX is as follows: the data are divided into segments of equal lengths, and then the mean value of each segment data is converted into the corresponding discrete characters. Regardless of the specific characteristics of the data, SAX exhibits strong robustness and a high compression rate. However, SAX uses equal length symbolization partition, and therefore, its shortcomings are similar to those of the fixed segmentation method. Furthermore, the use of mean approximation data can lead to the loss of the local shape information of the data, making SAX vulnerable to the interference of trend noise.

The data in Figure 1 are symbolized by SAX. The symbol set is chosen as **{a, b, c, d}**. By referring to the period of the data, the fixed segment length of data is set to be 50. The number of segments is inversely proportional to the compression; namely, the data compression ratio is 20:1, and the data symbolization sequence is **{dbdbdbcbcbbbcabcabca}**. The data in Figure 1 show a trend change after 500 s. The data symbolization sequence is divided into Sequence 1 **{dbdbdbcbcb}** and Sequence 2 **{bbcabcabca}** at 500 s. A comparison of the two sequences shows that the character sequence of Sequence 1 is more obvious, while there are only local, regular changes in Sequence 2. This result occurs because SAX retains the trend-type noise of the raw data, which interferes with the pseudo-periodic extraction of the data.

In light of the above limitations, this paper proposed a symbolization method based on the tilt angle in terms of the data temporal shape. The method had three main components: data filtering, shape salient points extraction, and data symbolization.

Since the raw telemetry data are quite noisy, which affects the accuracy of data shape extraction, first, the raw data should be filtered. First-order digital attenuated memory filter is a kind of a time-domain filter. Compared with the classical frequency filter, this approach could rapidly filter

noise with different frequencies and has a better robustness with respect to different types of data. The first-order digital attenuated memory filter is described as follows:

$$\widetilde{x}_{i+1} = \widetilde{x}_i + G[x_{i+1} - \widetilde{x}_i] \tag{4}$$

where x_i is the raw data, $\widetilde{x}_i$ is the filtered data, and $\widetilde{x}_1 = x_1$, G is the memory length of the filter, and $G \in [0, 1]$. The value of G is inversely proportional to the memory time. If G is too large, the ideal filtering result cannot be obtained. If the value is too small, a phase delay of the filtering result will occur. The confirmation of G value always depends on the experience. Therefore, G was set to be $G \geq 0.8$ in this paper. A first-order digital attenuated memory filter was used to filter the data in Figure 1, and the filtering results are shown in Figure 2.
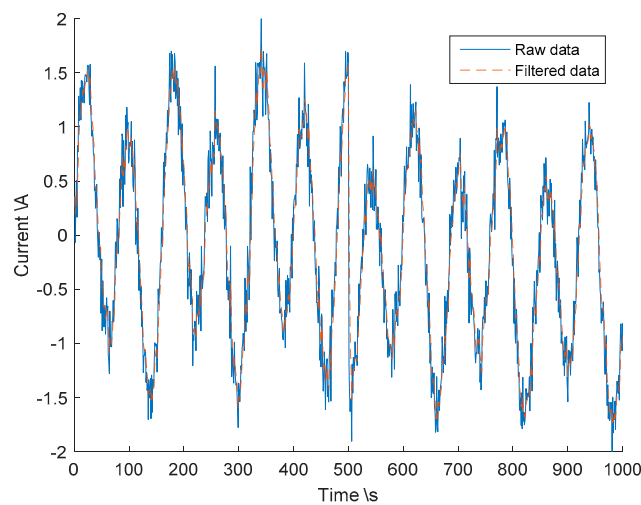

Figure 2. Filtering results of the measured current data of a solar panel.

The salient points of the data include the starting and ending points, turning points, and local extremum points. The essence of salient point extraction is to achieve data compression. Fu [22] proposed the method of extraction of the important points based on the data shape. Yan [23] improved the method of extracting the important points and proposed an extraction method based on the key points (KP) with an improved compression rate. However, the key points with a time interval that is too short may lead to a redundant number of salient points. Based on the KP method, this paper proposed a method to extract salient points of data shape (SP). By judging the interval between adjacent key points, which are extracted by the KP method, salient points can be filtered to obtain the final extracted salient points. The judgment function is given by:

$$\varphi_\lambda\left(\hat{x}_j\right) = \begin{cases} \breve{x}_{i-1} & \tau_i \leq \lambda \\ LE\left(\breve{x}_{i-1} : \breve{x}_i\right) & \tau_i > \lambda \end{cases} \tag{5}$$

where $\hat{x}_j$ is the final extracted salient point, $\breve{x}_i$ is the first extracted salient point, λ is the time-length judgment threshold, and τ_i is the time interval from $\breve{x}_{i-1}$ to $\breve{x}_i$. $LE(\cdot)$ is a function that obtains the local extremum points between $\breve{x}_{i-1}$ and $\breve{x}_i$.

The pseudo code of the salient points extraction algorithm for the data shape is given below:

Input: Data sequence $X = \{x_1, x_2, \cdots, x_n\}$, Threshold of the turning point th_e, time-length judgment threshold λ

Output: Salient point sequence $\hat{X} = \{\hat{x}_1, \hat{x}_2, \cdots, \hat{x}_l\}$

$\widetilde{X} \leftarrow f(X)$;//Data filtering

$\hat{x}_1 \leftarrow x_1$;//Put the starting point of the data sequence into the salient point sequence

for $i \leftarrow 2, n-1$ do

    $if((\widetilde{x}_i > \widetilde{x}_{i-1}) and (\widetilde{x}_i > \widetilde{x}_{i+1})) or ((\widetilde{x}_i < \widetilde{x}_{i-1}) and (\widetilde{x}_i < \widetilde{x}_{i+1})) then \overrightarrow{X} \leftarrow \widetilde{x}_i$;//Extract key points

    end if

    $if \left| \widetilde{x}_i - \left(\frac{\widetilde{x}_{i-1} + \widetilde{x}_{i+1}}{2} \right) \right| > th_e$ then $\dot{X} \leftarrow \widetilde{x}_i$;//Extract turning points

    end if

 end for

$\hat{x}_m \leftarrow x_n$;//Put the endpoint of the data sequence into the salient point sequence

$\breve{X} \leftarrow \dot{X} \cup \overrightarrow{X}$;//The first extracted salient points

for $j \leftarrow 2, m-1$ do

    $timeL_j = \widetilde{x}\_time(\widetilde{x} == \breve{x}_j) - \widetilde{x}\_time(\widetilde{x} == \breve{x}_{j-1})$;//Extract the time intervals of the salient points

    $if \ timeL_j > \lambda$ then $\hat{X} \leftarrow \breve{x}_j$;

    $\Delta\widetilde{x}_j \leftarrow (\widetilde{x}_i : \widetilde{x}_{i+timeL_i})$;//Data extraction in the time intervals

    $if abs\left(\max(\breve{X}) - \max(\Delta\widetilde{x}_j) \right) < abs\left(\min(\Delta\widetilde{x}_j) - \min(\breve{X}) \right) then$

    $\hat{X} \leftarrow find(\max(\Delta\widetilde{x}_j))$;//Retain the local maximum points as the final salient points

    else

    $\hat{X} \leftarrow find(\min(\Delta\widetilde{x}_j))$;//Retain the local minimum points as the final salient points

     end if

    else $\hat{X} \leftarrow \breve{x}_{j-1}$

    end if

 end for

The algorithm of the salient points extraction for the data shape has two parameters: threshold of a turning point and time-length judgment threshold. When processing different numerical characteristics of telemetry data, the SP algorithm parameters need to be adjusted. The threshold of the turning point setting method is described in [23]. The time-length judgment threshold is set empirically, while its value in the real process of using can be determined by means of quantile method [24]. After the time differences of each group of adjacent salient points in the data sequence $\{\tau_j\}$ are calculated, the τ_j corresponding to the first quartile would be the reference value.

Feature points were extracted by SP and KP from the filtered data of Figure 2, and the results are shown in Figure 3.

As shown in Figure 3, both methods could extract the data shape. The KP method could extract 335 data feature points with the compression rate of 2.98, and the SP method could extract 107 data feature points with the compression rate of 9.34. Therefore, the SP method had a higher compression rate. Line segments were used to connect the extracted salient points in time order, and then, the compressed shape of the data could be obtained.
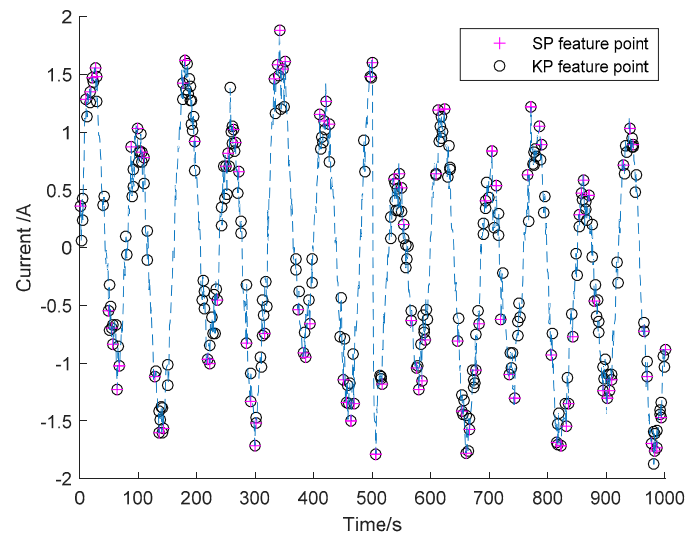
Figure 3. Comparison of the SP (salient points) and KP (key points) extracted data feature points.

The slope of the compressed data was used as the feature of the data shape, and the data were symbolized. For the simplicity of the symbolization, the slope values between the adjacent salient points were converted to the tilt angle values.

$$\theta_j = atan(k_j) * 180/\pi \tag{6}$$

where $k_j = \hat{x}_{j+1} - \hat{x}_j$ is the slope between the adjacent salient points. Based on the tilt angle, the number of intervals was determined, and a query table for the symbol conversion of the tilt angle in the range of $[-90° \sim +90°]$ was established. There was no overlap of the intervals in the table, and the corresponding symbols of each interval were usually expressed by letters **{a, b, . . . , z}**; for example, if there were two intervals, the symbols set was **{a, b}**, and the rest could be expressed in the same manner.

In the symbolization of an angular data sequence, first, the cardinality of the symbol sets must be determined. If the symbol set is too large, the symbol sequence of the data will be too complex, whereas information will be lost if the symbol set is insufficient. The efficiency of data symbolization is directly related to the adoption of an appropriate symbol set in the process of data sequence symbolization. When the size of the symbol set is reasonably set, the ideal data compression effect can be achieved. To obtain a reasonable semiotic set, the sequence of the angle-value data is analyzed by the clustering method, and the number of symbol sets is determined by the number of clusters and each of the clusters corresponding to a symbol.

Currently, there are many different clustering methods, all of which have their suitable application background, and the methods differ from one another in computational efficiency and complexity. Based on the requirements for the rapid telemetry data anomaly detection and identification, clustering by the fast search and finding the density peaks (CFDP) method was adopted in this paper [25]. This method could quickly find the density peak point of any shape data set and take it as the clustering center to achieve an efficient distribution of the logarithm data points and eliminate outliers. Thus, the CFDP method is suitable for the clustering analysis of large-scale data. The method only needs to set the truncation distance, and an empirical method for parameter setting is given in [25].

The data symbolization was performed as follows: (1) the intervals of the tilt angles were searched, and (2) tilt angles into characters corresponding to the intervals were converted. Since the tilt angle could be divided into the basic form of upward, downward, and horizontal, the interval had a basic direction. Upon a directional change in the intervals, the sign should remain the same. Thus, data

symbolization γ could be introduced to correct the interval range. Typical symbols of the tilt angles are listed in Table 1.

Table 1. Query table of the tilt angle partition.

Amount of Partition Angles	Amount of Partition Intervals								
	2	3	4	5	6	7	8	9	10
1	0°	$\gamma - 10°$	−30°	−30°	−30°	−45°	−45°	−60°	−60°
2		$10° - \gamma$	0°	$\gamma - 10°$	−10°	−30°	−30°	−45°	−45°
3			30°	$10° - \gamma$	0°	$\gamma - 10°$	−10°	−30°	−30°
4				30°	10°	$10° - \gamma$	0°	$\gamma - 10°$	−10°
5					30°	30°	10°	$10° - \gamma$	0°
6						45°	30°	30°	10°
7							45°	45°	30°
8								60°	45°
9									60°
symbol set	{a, b}	{a, b, c}	{a, b, c, d}	{a, b, … , e}	{a, b, … , f}	{a, b, … , g}	{a, b, … , h}	{a, b, … , i}	{a, b, … , j}

The data in Figure 1 are symbolized using the SAX method and the method proposed in this paper. Five intervals were selected, and the corresponding character set was **{a, b, c, d, e}**. For comparison, the salient points were extracted by the KP method. The symbolization results obtained by the two methods are shown in Figures 4 and 5.

Comparing the two sets of symbolization results, it was observed that the data symbols obtained using the proposed method had a more significant regularity, and the effect of the trend noise had been effectively eliminated. Therefore, the data symbolization method in this paper was more suitable for extracting the pseudo-periodic law.
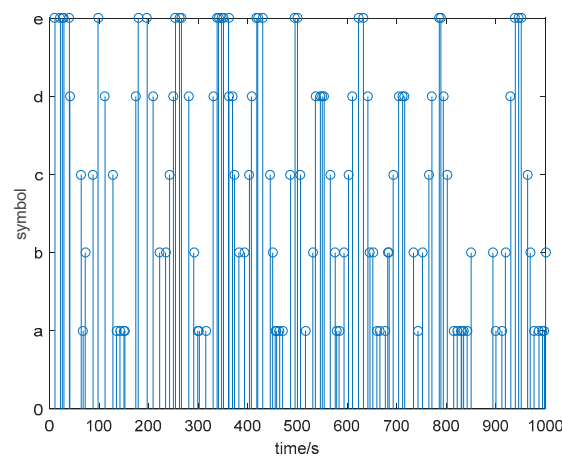
Figure 4. Symbolization results of the SAX (symbolic aggregate approximation) method.
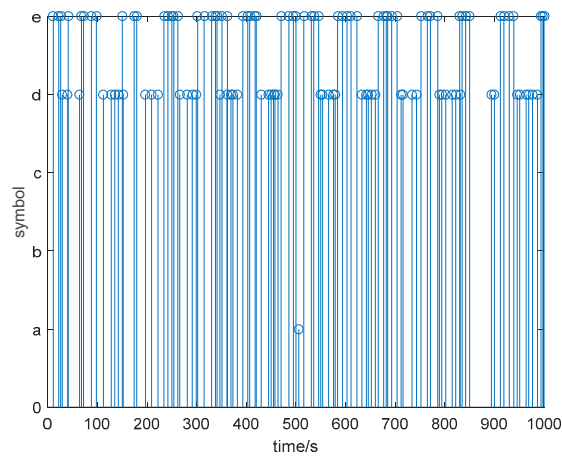
Figure 5. Symbolization results of the proposed method.

2.3. Pseudo-Period Extraction of Telemetry Data Based on Symbolization Method

Through the symbolization of the data tilt angle, the satellite telemetry data X_m were converted into a character sequence $\{STR_m\}$. In the extraction of the pseudo-period of the measured data, the standard pseudo-period sequence $\overline{X}_{fT}$ was introduced and transformed into the character sequence STR_{fT}. Then, STR_{fT} was analyzed to determine the key characters. The keywords in the character sequence of the standard pseudo-periodic were defined as follows:

Definition 1. *A character with the least number of occurrences in the standard pseudo-periodic character sequence. If there are minimal characters with equal numbers of occurrences, the earlier characters will be used as the keywords str_{key} in chronological order.*

Taking str_{key} as the index, STR_{fT} was used to divide STR_m into subsequences $\{STR_i | i = 1, 2, \cdots, l\}$, and then the string similarity between STR_i and STR_{fT} of each subsequence was calculated. The rationality of the pseudo-periodic segmentation points was determined by setting a similarity threshold. Thus, the pseudo-period extraction of the measured data was realized; namely, the data pseudo-period extraction problem was converted into the string similarity calculation problem.

Due to the non-stationarity of X_m, it was difficult to align the characters of STR_i and STR_{fT} in each subsequence on the timeline. Therefore, dynamic time warping (DTW) [26] was introduced to measure the similarity. Allowing several-for-one and several-for-several mappings between symbolic sequences, DTW could solve the problem of the time axis distortion in the calculation of similar distances between symbol sequences of different lengths. Therefore, the DTW distance could be used for both raw data and the dimension-reduction data. The principle of the DTW procedure is given as follows:

Assuming that two symbol sequences were $STR_a = \{a_1, a_2, \cdots, a_i, \cdots, a_n\}$ and $STR_b = \{b_1, b_2, \cdots, b_j, \cdots b_m\}$, cumulative distances $D(i, j)$ were constructed as:

$$D(i, j) = d\left(a_i, b_j\right) + \min\left\{D_{(i-1,j)}, D_{(i,j-1)}, D_{(i-1,j-1)}\right\} \tag{7}$$

where $d\left(a_i, b_j\right)$ is the base distance between two points, and the Euclidean distance was used in this paper. When calculating the DTW distance of STR_a and STR_b, the distance matrix **C** between the points of two sequences must be established.

$$C = \begin{bmatrix} d(a_1, b_n) & d(a_2, b_n) & \cdots & d(a_m, b_n) \\ \vdots & \vdots & \cdots & \vdots \\ d(a_1, b_2) & d(a_2, b_2) & \cdots & d(a_m, b_2) \\ d(a_1, b_1) & d(a_2, b_1) & \cdots & d(a_m, b_1) \end{bmatrix} \tag{8}$$

After establishing the matrix, the distance between STR_a and STR_b was transformed into the bending path W from $(1,1)$ to (m,n) in matrix $\mathbf{C}$. If the bending path was $W = (w_1, w_2, \cdots, w_k, \cdots, w_K)$, the element was $w_k = (i, j)$, and the path length satisfied $\max(n, m) \leq k \leq n + m - 1$. The path was constrained as follows:

(1)　W started in matrix $\mathbf{C}$ (1,1), and finally ended in matrix $\mathbf{C}$ (*m,n*);

(2)　The path curve was continuous and monotonic.

The DTW distance calculation between STR_a and STR_b was essentially the task of finding a bending path with the smallest distance in the distance matrix $\mathbf{C}$.

$$DTW(STR_a, STR_b) = \underset{W}{\mathrm{argmin}}\left\{\sum_{(i,j)\in W} w_k\right\} \tag{9}$$

In the extraction of the pseudo-periods of the satellite telemetry data, the DTW distance was used to calculate the distance between each subsequence STR_i and the standard pseudo-period sequence STR_{fT}. By setting the DTW distance detection threshold, the rationality of the preliminary partition point of the pseudo-period was determined. If the distance was less than the set threshold, the right boundary of the data corresponding to the subsequence was used as the pseudo-periodic partition point. If the distance was greater than the set threshold, the correction was performed within k characters of the subsequence, and the DTW distance of STR_{fT} and each modified character sequence were calculated separately. If the distance was less than the correction threshold, the right boundary of the data corresponding to the minimum correction character was taken as the partition point. If the distance was still greater than the correction threshold, the segment was retained as the anomaly partition point, and the right boundary of the data corresponding to the minimum correction character was still taken as the partition point. The locations of the pseudo-periodic partition points were extracted as follows:

$$\begin{cases} pr_i = X_m\{STR_i\}, D\left(STR_i, STR_{fT}\right) \leq th_S \\ pr_i = X_m\{STRC_i\}, D\left(STR_i, STR_{fT}\right) \leq h * th_S \\ pw_i = X_m\{STRC_i\}, D\left(STR_i, STR_{fT}\right) > h * th_S \\ STRC_i = \min\left(D\left(\{STR_i \pm k\}, STR_{fT}\right)\right) \end{cases} \tag{10}$$

where pr_i is the location of the pseudo-period partition point in normal data, pw_i is the location of the pseudo-period partition point in the anomaly data, $D\left(STR_i, STR_{fT}\right)$ is the DTW distance of the subsequence STR_i and STR_{fT}, and k is the character offset correction range. th_S is set at the DTW detection threshold, and h is the amplification coefficient of the DTW detection threshold after the character offset correction; usually $h \in [1, 2]$.

Pseudo-periodic partition was performed for the characterizing result of the data in Figure 1. Given the sequence of the standard pseudo-periodic **{eeeedcdd}**, the pseudo-periodicity was divided into:
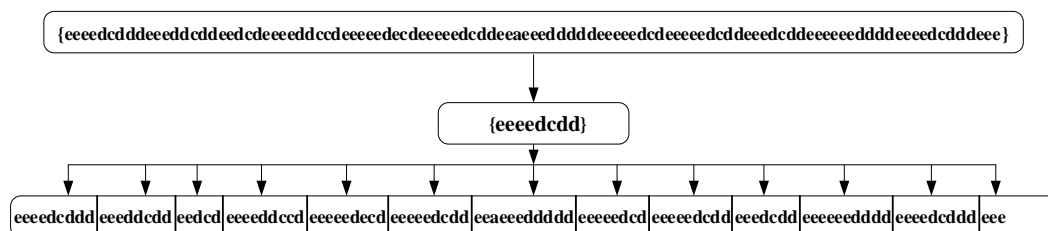
The corresponding pseudo-period extraction points were: [64 135 222 299 385 464 506 579 663 743 821 901 983].

According to the results of the pseudo-period extraction, DTW could effectively extract the pseudo-period of the data. Sequence segmentation using the keywords of the standard pseudo-periodic character sequence could avoid the high complexity of the DTW distance calculation.

3. Anomaly Detection and Identification of Satellite Telemetry Data Based on Phase-Plane Trajectory

Phase-plane analysis is a graphical analysis method for solving low-dimensional dynamic systems, which is usually applied to system analysis, for a second-order system, such as:

$$x'' + f(x', x) = 0 \tag{11}$$

where $f(x', x)$ is a linear function or nonlinear function of x' and x. The phase-plane of data is a two-dimensional plane in which the data value x is taken as the horizontal axis and the derivative value x' as the vertical axis. The phase-plane trajectory shows the motion of the system under different conditions. Suppose the second-order system differential equation is:

$$x'' + \omega_n{}^2 x = 0 \tag{12}$$

The solution to the equation is:

$$\frac{dx'}{dx} = -\frac{\omega_n{}^2 x}{x'} \tag{13}$$

And the phase-plane trajectory equation:

$$x^2 + \left(\frac{x'}{\omega_n}\right)^2 = A^2 \tag{14}$$

where $A = \sqrt{x_0'^2/\omega_n^2 + x_0^2}$. Equation (14) is an elliptic equation. The parameter A depends on the initial value of x' and x. If the value of parameter A takes different values, the phase-plane trajectory of elliptic will be different.

As the most real satellite telemetry data is nonstationary, the differential equation corresponding to the certain problem will be more complicated under different circumstances. So, there are more significant differences in the different temporal shapes of telemetry data. The phase-plane trajectory could be applied in abnormal data detection and identification for satellite telemetry data. When the phase-plane was used to map the satellite telemetry data, the actual telemetry data were taken as the mapping function. The phase-plane trajectory map of the satellite telemetry data could be obtained by mapping the satellite telemetry data on the phase-plane. Figures 6–8 present the time-domain diagrams of three different shape data and their corresponding phase-plane trajectories, respectively.
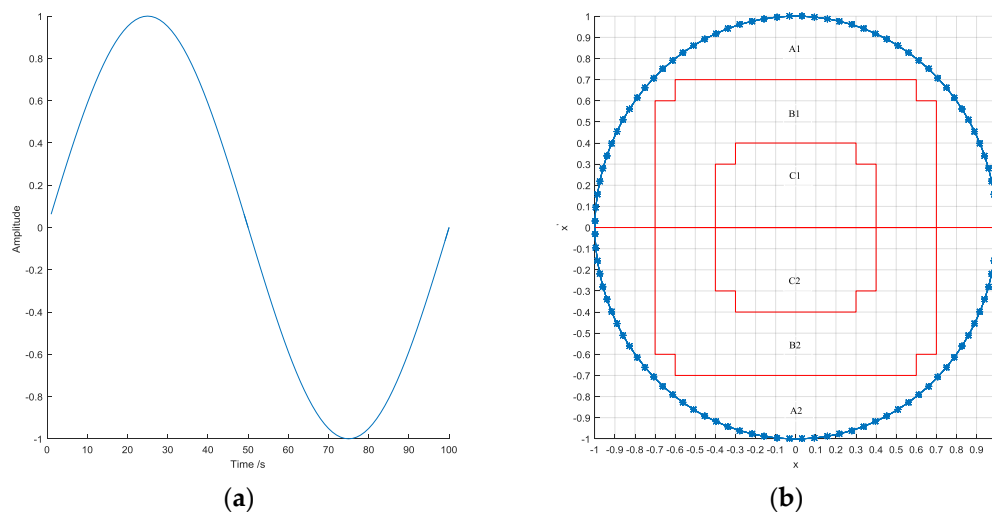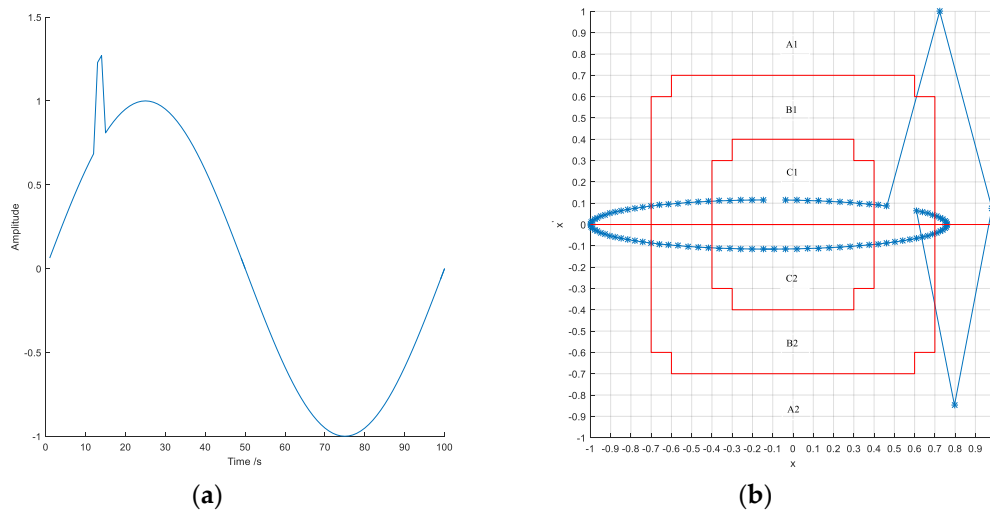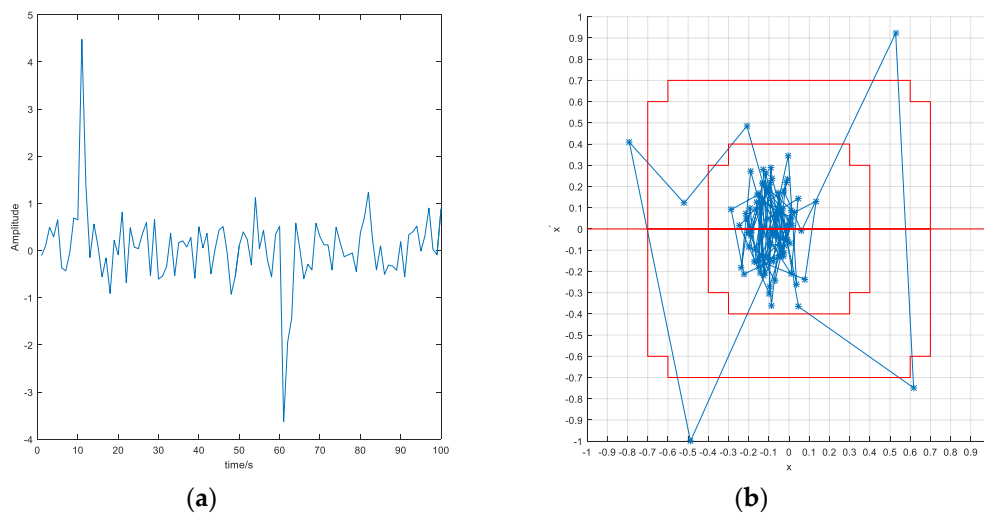

(a)   (b)

Figure 6. (**a**) Standard sinusoidal data; (**b**) Phase-plane trajectory of the standard sinusoidal data.

Figure 7. (**a**) Sinusoidal data with hopping interference; (**b**) Phase-plane trajectory of the sinusoidal data with hopping interference.

Figure 8. (**a**) Pulse data with noise; (**b**) Phase-plane trajectory of the pulse data with noise.

Figure 6b shows the phase-plane trajectory of sinusoidal data for which the trajectory shape was approximately circular. The phase-plane trajectory points were mainly distributed in the boundary area of the phase-plane plots. Figure 7b shows the phase-plane trajectory of the sinusoidal data with hopping noise from time 12 to 16 in Figure 7a, where the trajectory was close to an ellipse. The phase trajectory points were mainly distributed in the horizontal axis area of the phase-plane plots, and the corresponding phase trajectory of the hopping noise points was located in the boundary area of the phase-plane plots. Figure 8b shows the phase-plane trajectory of the impulse data containing the noise, where the trajectory was mainly concentrated in the central area of the phase-plane. The phase trajectory of the pulse points was located in the boundary area of phase-plane plots. It was observed from the distribution characteristics of the above data that the phase-plane trajectories were highly sensitive to the change in the data shape. The data with different shapes had different phase-plane trajectories. A weak change in the data shape was clearly reflected in the phase-plane trajectories. Therefore, by comparing the phase-plane trajectories of the normal satellite telemetry data with the measured telemetry data, the identification of the anomaly data could be achieved.

To detect and identify the anomalous satellite telemetry data based on the phase-plane trajectory, the phase-plane trajectory characteristics of the data should be further extracted in order to quantify the difference in the phase-plane trajectory. Regarding the phase-plane trajectory as the distribution of data

on the phase-plane, the methods based on statistics could be applied to the analysis. The first-order moments between the center point and each phase trajectory point depicted the density of distribution of phase trajectory points; however, it lost the morphological information of data. Considering the accuracy and computational complexity of feature extraction of the phase-plane trajectory, and based on the distribution of the overall phase-plane trajectory of the data, the phase-plane was divided into six statistical regions [A1 A2 B1 B2 C1 C2], as shown in Figures 6b and 8b. The distribution points of the phase-plane trajectories of the different shape data were different in each statistical region. The number of the phase-plane trajectories in each statistical region was taken as the feature to quantitatively describe the phase-plane trajectories of the data, and then the eigenvectors were established as:

$$V_i = \left[N_i^{A1}, N_i^{A2}, N_i^{B1}, N_i^{B2}, N_i^{C1}, N_i^{C2} \right] \tag{15}$$

where N_i^{A1}, N_i^{A2}, N_i^{B1}, N_i^{B2}, N_i^{C1}, N_i^{C2} are the fractions of the number of the phase-plane trajectories in each statistical region out of the total number of points. The eigenvector V_i reserves the morphological information of data; therefore, eigenvectors of data with different morphological features are quite different from each other.

As shown in Figures 6b and 8b, the distribution of the phase-plane trajectories in six statistical regions indicated that the main points of the phase-plane trajectories of the standard sinusoidal data were distributed in the A1 and A2 regions, the phase-plane trajectories of the sinusoidal data with the hopping disturbance were mainly distributed in the B1 and B2 regions, and the phase-plane trajectories of the impulse data with noise were mainly distributed in the C1 and C2 regions. Therefore, different types of data could be distinguished through the data distribution of each statistical interval. The similarity between the phase-plane eigenvector of the measured data and that of the sample data was further calculated. The threshold of the similarity judgment was set to identify the anomaly type of the measured data. To highlight the distribution characteristics of the data shape in each statistical interval, cosine similarity [27] was selected to calculate the similarity between the vectors.

If the phase-plane trajectory eigenvectors of each pseudo-periodic data and the sample data were V_i and Vs_i, the chord similarity would be expressed as:

$$sim(V, Vs) = \frac{\sum\limits_{i=1}^{l} V_i Vs_i}{\sqrt{\sum\limits_{i=1}^{l} V_i^2} \times \sqrt{\sum\limits_{i=1}^{l} Vs_i^2}} = \frac{\sum\limits_{i=1}^{l} V_i Vs_i}{\|V\| \cdot \|Vs\|} \tag{16}$$

When the value of $sim(V,Vs)$ approached more to 1, the angle between the two eigenvectors was more close to 0, indicating that the states, respectively, corresponding to V and vs. were more similar to each other. If vs. referred to the *i-th* anomalous condition, the conclusion would be drawn that there existed the *i-th* anomaly on the data within the corresponding pseudo period.

4. Anomaly Detection and Identification of Satellite Telemetry Data Based on Pseudo-Period

Phase-plane analysis can obtain the shape characteristics of the data in the time domain from a global perspective, avoiding the influence of data values. This approach also has a high computational speed and is robust with respect to different data types. Based on the principle of phase-plane analysis, the length of the analyzed data directly affected the phase-plane trajectory analysis results. Therefore, in the anomaly identification of the satellite telemetry data, the pseudo-period was taken as the basic analytical unit, and the anomaly identification was achieved by comparing the differences between the data in each pseudo-period and the sample data, as mentioned in Section 1. In this paper, an identification method for the satellite anomalous telemetry data was proposed based on the pseudo-period. Figure 9 shows the flow chart of the proposed method.
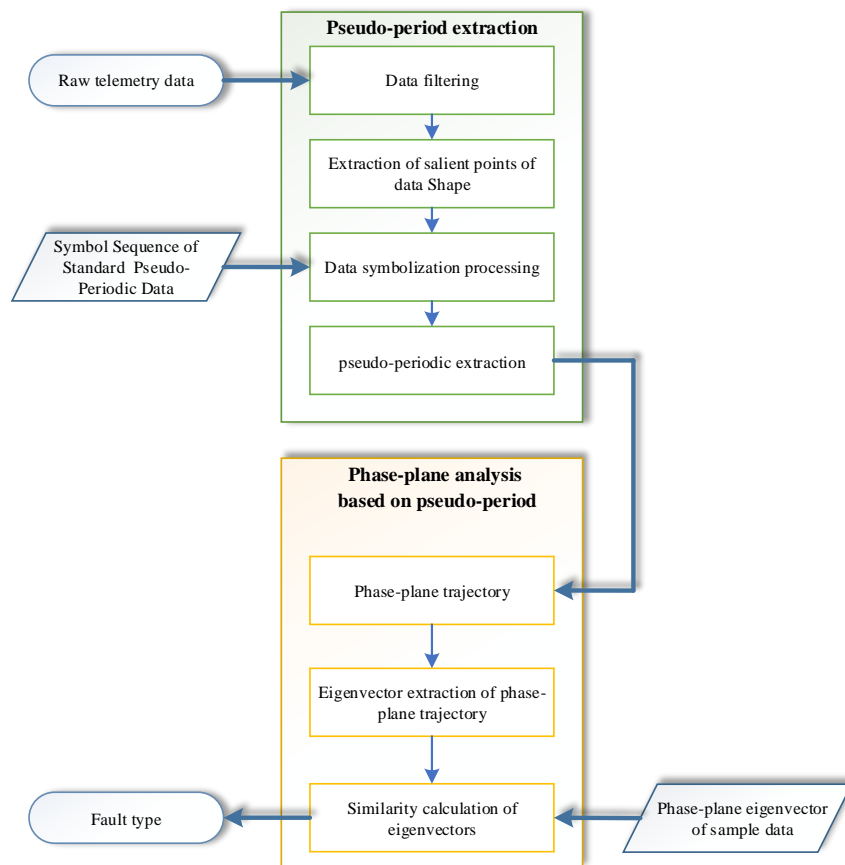
Figure 9. Flow chart of detection and identification of the satellite anomalous telemetry data based on the pseudo-period.

5. Simulation Results

To verify its effectiveness, the proposed method was tested using the measured data of the momentum wheel speed of a satellite for which the data length was equal to 1658. The measured data are shown in Figure 10.
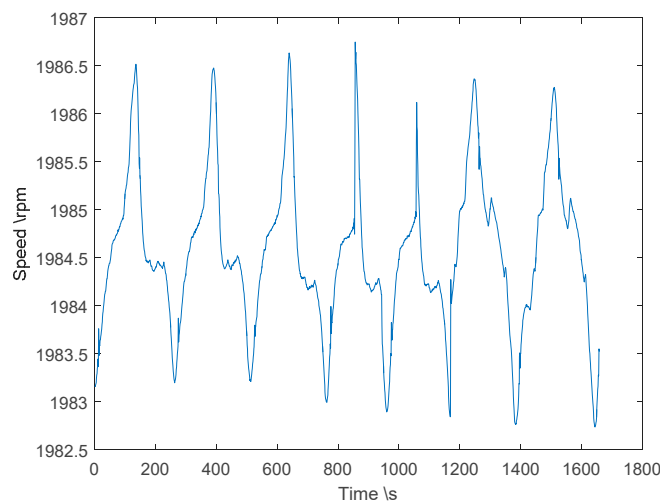

Figure 10. Measured data of the momentum wheel speed of a satellite.

First, the measured data were compressed by the SP method, and the salient points were extracted. The compressed data are shown in Figure 11.
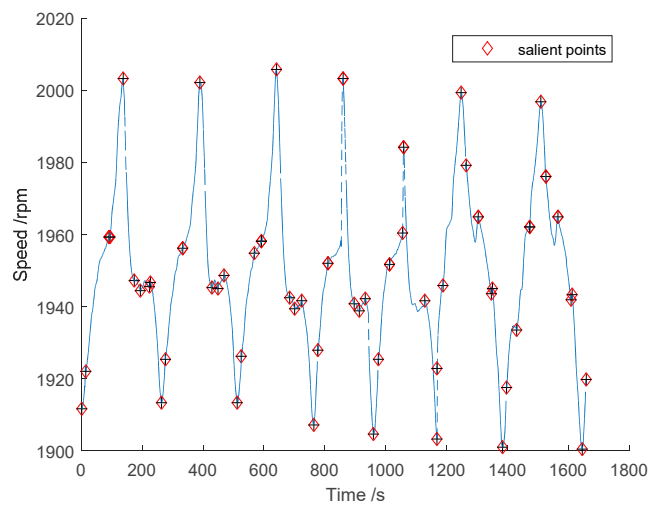

Figure 11. Salient points of the measured data of the rotational speed of a momentum wheel.

As shown in Figure 11, the number of salient points extracted from data was 67, and the corresponding data compression rate was 24.74. The data were further symbolized. The number of the symbols was determined as 5 by the CFDP method. Therefore, the corresponding characters **{a,b,c,d,e}** were selected. Then, the strings to be measured were obtained as:

{dddcdaccdaddddacdadddcdabcaedcdeacdaedcdbeaaeeddadadaeddcdadadadae}

The standard pseudo-periodic sequence string was given as **{ddcdacda}**. The standard pseudo-periodic sequence string was used to segment the pseudo-periodic. The pseudo-periodic was divided as follows:

{dddcdaccdaddddacdadddcdabcaedcdeacdaedcdbeaaeeddadadaeddcdadadadae}

{ddcdacda}

| **{dddcdaccda** | ddddacda | dddcdabcae | dcdeacdae | dcdbeaaee | ddadadae | ddcdadada | dae}** |
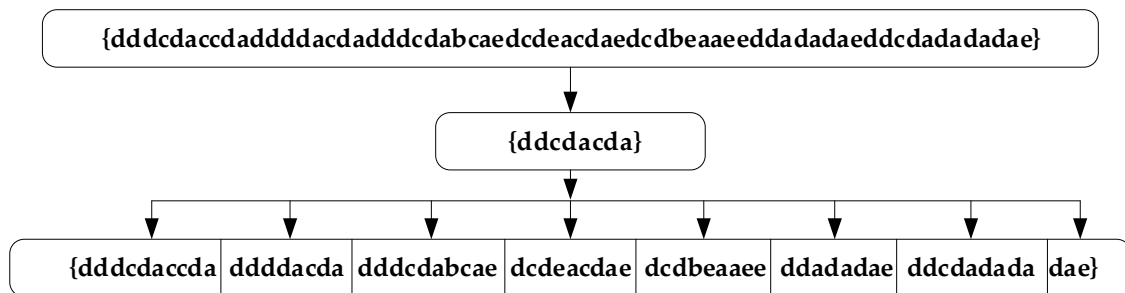
To verify the accuracy of the pseudo-periodic partition, the standard pseudo-periodic sequence string was used as the reference, and the exact pseudo-periodic in the data was manually determined. The results were compared with the results obtained by the pseudo-periodic extraction algorithm proposed in this paper. Table 2 shows the analysis results.

Table 2. Analysis of the pseudo-period extraction for the data.

Partition Points Extraction Method	1	2	3	4	5	6	7
Proposed method	263	512	763	960	1169	1383	1645
Manual method	264	513	763	961	1169	1385	1644
Error analysis	-	−1	0	−1	0	−2	+1

In Table 2, "−" indicates the advanced error, and "+" indicates the delayed error. As shown in the error analysis, the proposed pseudo-period extraction method based on the data shape showed high accuracy, making this method suitable for application to the actual telemetry data pseudo-period extraction.

The anomaly data in each pseudo-period were further identified, and the phase-plane trajectory of the normal standard pseudo-period data of the measured data was given, as shown in Figure 12.
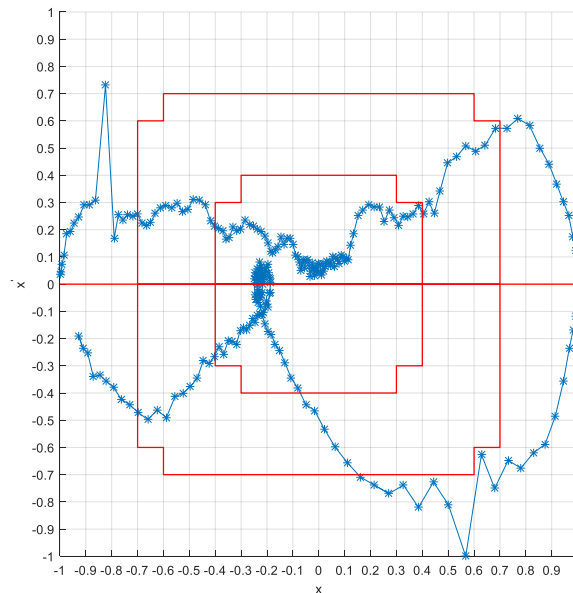
Figure 12. Phase-plane trajectory of the standard pseudo-periodic reference data for the rotational speed of a momentum wheel.

The phase-plane was divided into six statistical regions. The data points of each region were counted as the normal data of the standard pseudo-period. The eigenvectors of the phase-plane trajectory were obtained, as shown in Table 3.

Table 3. Eigenvectors of the phase-plane trajectory for the normal data at the standard pseudo-periodic.

Statistical Interval	A1	A2	B1	B2	C1	C2
Amount of quantitative points	0.114	0.087	0.091	0.061	0.389	0.255

The pseudo-periodic phase-plane trajectories of the measured data were obtained, and the phase-plane trajectory eigenvectors were obtained by statistical analysis in the same statistical region, as shown in Table 4.

Table 4. Eigenvectors of the plane trajectories of the pseudo-periodic phases for the measured data.

Statistical Interval / Pseudo-Period	A1	A2	B1	B2	C1	C2
1	0.116	0.116	0.100	0.064	0.428	0.176
2	0.096	0.088	0.096	0.068	0.371	0.278
3	0.068	0.096	0.088	0.104	0.400	0.244
4	0.056	0.102	0.051	0.061	0.403	0.326
5	0.058	0.058	0.033	0.082	0.391	0.377
6	0.070	0.107	0.079	0.107	0.243	0.392
7	0.100	0.103	0.100	0.096	0.257	0.342

The cosine similarity distances between the pseudo-periodic vectors presented in Table 4 and the standard pseudo-periodic vectors presented in Table 3 were calculated. The following results were obtained [0.971 0.949 0.986 0.837 0.789 0.779 0.774], and the threshold value of the similarity judgment was set to 0.940. It was found that pseudo-periods 1, 2, and 3 had high similarity with normal data, and the eigenvector distribution in each pseudo-periodic data interval was close to that of the normal data. Therefore, the data of pseudo-periods 1, 2, and 3 could be regarded as normal data. On the other hand, low similarities were obtained between pseudo-periods, 4, 5, 6, and 7, and normal data. Compared with the normal number and the eigenvector distribution in each pseudo-periodic data interval, there were differences in the A1, A2, B1, and B2 regions. Therefore, pseudo-periods 4, 5, 6, 7 were regarded as anomaly data.

To further determine the possible fault types, pseudo-periods 4, 5, 6, and 7 were analyzed. In comparison with the three typical anomaly data templates of the measured data, the same statistical interval was used for statistical analysis, as shown in Table 5.

Table 5. Eigenvector of the trajectory plane for typical anomaly data.

Statistical Interval Abnormal Data Templates	A1	A2	B1	B2	C1	C2
1	0.097	0.051	0.162	0.151	0.348	0.288
2	0.062	0.074	0.070	0.078	0.364	0.349
3	0.232	0.150	0.188	0.098	0.153	0.176

The similarity values between the 4, 5, 6, and 7 data and the typical anomaly data are shown in Table 6. The obtained results showed that the anomaly data evolved from Type 2 to Type 1.

Table 6. The similarity between the 4, 5, 6, and 7 datasets and typical anomaly data.

Cosine Similarity Pseudo-Period	Abnormal Data Templates			Identification Result
	1	2	3	
4	0.940	0.994	0.724	Type 2 fault
5	0.960	0.996	0.698	Type 2 fault
6	0.984	0.965	0.787	Type 1 fault
7	0.986	0.975	0.841	Type 1 fault

To verify the applicability of the proposed method, five groups of data with 1000 different period lengths were used to test the accuracy of the pseudo-period partition and anomaly identification, respectively. These data were generated in the same system of the measured data, as described above. Table 7 shows the results obtained by the anomaly identification using the first pseudo-period data of each group.

As shown in Table 7, the symbolization method based on the data shape could accurately extract the pseudo-period of the satellite telemetry data. The statistical results of each pseudo-period obtained by the phase-plane trajectory statistics method could identify different abnormal states of the telemetry data. Therefore, the proposed method could achieve the anomaly detection and identification of satellite telemetry data.

Table 7. Test data pseudo-period partition and anomaly identification results.

Data Number	Verification Method Pseudo-Period Extraction Pseudo-Period Extraction Results	Partition Error of Pseudo-Periodic Point	Abnormal Identification Statistical Value of Each Interval of Phase Trajectory	Identification Result
1	4	[+1 +2 +3]	[0.134 0.152 0.139 0.094 0.296 0.185]	Normal
2	5	[+2 0 +2 +1]	[0.117 0.120 0.109 0.158 0.266 0.231]	Normal
3	5	[+1 +1 –2 +2]	[0.105 0.068 0.102 0.183 0.330 0.211]	Normal
4	5	[0 +2 +3 –1]	[0.149 0.096 0.192 0.181 0.215 0.161]	Type 3 fault
5	5	[+1 +3 0 +2]	[0.036 0.101 0.087 0.099 0.325 0.352]	Type 2 fault

To compare the accuracy and speed of the anomaly identification between the method in this paper and the MSVM method, 600 groups of the telemetry data of the momentum wheel speed of the same satellite were used for testing. Each group had a length of 1000. The dataset contained four types of anomalies. For the MSVM method, 200 and 500 groups of data were used as the training samples, and 100 groups of data were used as the test samples. The comparison results are shown in Tables 8 and 9.

Table 8. Test results obtained with 200 groups used as the training sample.

Abnormal Type	Method MSVM Accuracy	Time	The Proposed Method Accuracy	Time
1	0.78	0.0025 s	0.88	0.0020 s
2	0.76	0.0022 s	0.81	0.0032 s
3	0.73	0.0037 s	0.78	0.0021 s
4	0.91	0.0019 s	0.93	0.0016 s

Table 9. Test results obtained with 500 groups used as the training sample.

Abnormal Type	Method MSVM Accuracy	Time	The Proposed Method Accuracy	Time
1	0.9033	0.1700 s	0.8900	0.0022 s
2	0.8525	0.1660 s	0.8267	0.0046 s
3	0.8250	0.1790 s	0.8045	0.0025 s
4	0.9564	0.1560 s	0.93	0.0023 s

According to Tables 8 and 9, when a smaller number of groups was used as the training sample, the proposed method had higher accuracy and approximate speed than the MSVM. When the number of the training sample was increased, while the accuracy of MSVM was improved and was close to that of the proposed method, the identification time of MSVM was also increased and was clearly greater than that of the proposed method. The proposed method did not require a large amount of data for training, and also there was no need to reset the parameters of the method if there existed new kinds of an anomaly. At the same time, what needed to be calculated was just the cosine similarity when dealing with the task of anomaly identification. Besides, the time complexity of the proposed method was lower than the MSVM method; therefore, our method could satisfy the requirement of rapid satellite anomaly detection and identification, and the method presented in this paper had clear advantages in practical application.

6. Conclusions

To effectively detect and identify the anomaly data in satellite telemetry data, the novel detection and identification method based on the pseudo-period was proposed in this paper. Through the telemetry data temporal shape features, the salient points of the data were extracted to compress the raw data. As the characteristic quantity, the tilt angle of the adjacent data points was symbolized in the compressed data in order to obtain the corresponding string. Then, the pseudo-periods of the measured data were extracted using the standard pseudo-periodic sequence string. Taking the pseudo-period as the basic analytical unit, the phase-plane trajectory plots corresponding to each pseudo-period data were obtained. The phase-planes of each pseudo-period data were divided into certain regions, and the statistical values of the trajectory points in each region were taken as the features for achieving anomaly detection and identification in the measured data.

References

1. Chandola, V.; Banerjee, A.; Kumar, V. Anomaly Detection: A Survey. *ACM Comput. Surv.* **2009**, *41*, 15. [CrossRef]
2. Li, Q.; Zhou, X.; Lin, P.; Li, S. Anomaly detection and fault Diagnosis technology of spacecraft based on telemetry-mining. In Proceedings of the 3rd International Symposium on System & Control in Aeronautics & Astronautics, Harbin, China, 8–10 June 2010.
3. Omran, E.A.; Murtada, W.A. Fault Detection and Identification of spacecraft reaction wheels using Autoregressive Moving Average model and neural networks. In Proceedings of the 12th International Computer Engineering Conference, Cairo, Egypt, 28–29 December 2016.
4. Venkatasubramanian, V.; Rengaswamy, R.; Yin, K.; Kavuri, S.N. A review of process fault detection and diagnosis Part I: Quantitative model-based methods. *Comput. Chem. Eng.* **2003**, *27*, 293–311. [CrossRef]
5. Venkatasubramanian, V.; Rengaswamy, R.; Kavuri, S.N. A review of process fault detection and diagnosis Part II: Qualitative models and search strategies. *Comput. Chem. Eng.* **2003**, *27*, 313–326. [CrossRef]
6. Song, W.; Wang, R.; Wang, J. A simple and valid analysis method for orbit anomaly detection. *Adv. Space Res.* **2012**, *49*, 386–391. [CrossRef]
7. Sherr, Y.; Zhang, Y.; Wang, Z. Satellite fault diagnosis method based on predictive filter and empirical mode decomposition. *J. Syst. Eng. Electron.* **2011**, *22*, 83–87. [CrossRef]
8. Ren, H.; Li, X.; Li, Z.; Pedrycz, W. Data Representation Based on Interval–Sets for Anomaly Detection in Time Series. *IEEE Access.* **2018**, *6*, 27473–27479. [CrossRef]
9. Al-Zyoud, A.D.; Khorasani, K. Neural Network-based Actuator Fault Diagnosis for Attitude Control Subsystem of a Satellite. In Proceedings of the 2006 World Automation Congress, Budapest, Hungary, 24–26 July 2006.
10. Li, Z.; Ma, L.; Khorasani, K. Fault Diagnosis of an Actuator in the Attitude Control Subsystem of a Satellite using Neural Networks. In Proceedings of the International Joint Conference on Neural Networks, Orlando, FL, USA, 12–17 August 2007.
11. Wang, Y.; Zhang, F.; Zhang, S. A New Methodology for Identifying Arc Fault by Sparse Representation and Neural Network. *IEEE Trans. Instrum. Meas.* **2018**, *67*, 2526–2537. [CrossRef]
12. Zhao, H.; Gao, Y.; Liu, H. Fault diagnosis of wind turbine bearing based on stochastic subspace identification and multi-kernel support vector machine. *J. Mod. Power Syst. Clean Energy* **2018**, *7*, 350–356. [CrossRef]

13. Yang, S.; Hojo, M.; Zhang, B. A Phase-Plane Trajectory Vector-Based Method for Real-Time Identification of Critical Machines. *IEEJ Trans. Electr. Electron. Eng.* **2018**, *13*, 1578–1585. [CrossRef]

14. Li, Z.; Wu, W. Classification of power quality combined disturbances based on phase space reconstruction and support vector machines. *J. Zhejiang Univ. Sci. A* **2008**, *9*, 173–181. [CrossRef]

15. Park, N.O.; Shin, Y.C.; Ahn, S.P. A Study on the Classification of Arcing Faults in Power Systems using Phase Plane Trajectory Method. *Trans. Korean Inst. Electr. Eng. A* **2002**, *51*, 209–216.

16. Liu, D.; Pang, J.; Song, G.; Xie, W.; Peng, Y.; Peng, X. Fragment Anomaly Detection with Prediction and Statistical Analysis for Satellite Telemetry. *IEEE Access.* **2017**, *5*, 19269–19281. [CrossRef]

17. Tang, L.A.; Cui, B.; Li, H.; Miao, G.; Yang, D.; Zhou, X. Effective variation management for pseudo periodical streams. In Proceedings of the ACM SIGMOD International Conference on Management of Data, Beijing, China, 11–14 June 2007.

18. Voss, H.U. Hypersampling of pseudo-periodic signals by analytic phase projection. *Comput. Biol. Med.* **2018**, *98*, 159–167. [CrossRef] [PubMed]

19. Jingjing, J.; Wang, X.; Wu, Y.; Yanbo, Y. Pseudo-Period Segment of Ballistocardiogram Based on Joint Time-Frequency Analysis. In Proceedings of the IEEE International Conference on Intelligent Computing and Intelligent Systems, Shanghai, China, 20–22 November 2009.

20. Du, Y.; Li, S.; Hu, S.; Zhang, L.; Wang, F. Detection algorithm of abnormal sub-sequence in pseudo periodic time series of telemetry. *Chin. Space Sci. Technol.* **2016**, *36*, 18–24.

21. Lin, J.; Keogh, E.J.; Lonardi, S.; Chiu, B. A Symbolic Representation of Time Series, with Implications for Streaming Algorithms. In Proceedings of the 8th ACM Sigmod Workshop on Research Issues in Data Mining and Knowledge Discovery (DMKD'03), San Diego, CA, USA, 13 June 2003.

22. Fu, T.; Chung, F.; Luk, R.; Ng, C. Representing financial time series based on data point importance. *Eng. Appl. Artif. Intell.* **2008**, *21*, 277–300. [CrossRef]

23. Qiuyan, Y. A Novel SAX Based Time Streams Similarity Approach. In Proceedings of the International Conference on Future Biomedical Information Engineering, Sanya, China, 13–14 December 2009.

24. Chiu-Ching, T.; Yi-Chao, W.; Wei-Shiou, C.; Wen-Tzeng, H. Fault tolerance by quartile method in wireless sensor and actor networks. In Proceedings of the International Conference on Complex, Intelligent and Software Intensive Systems, Krakow, Poland, 15–18 February 2010.

25. Rodriguez, A.; Laio, A. Clustering by fast search and find of density peaks. *Science* **2014**, *344*, 1492–1496. [CrossRef] [PubMed]

26. Berndt, D.J.; Clifford, J. Using Dynamic Time Warping to Find Patterns in Time Series. In Proceedings of the KDD Workshop, Seattle, WA, USA, 31 July–1 August 1994.

27. Dong, Y.; Sun, Z.; Jia, H. A cosine similarity-based negative selection algorithm for time series novelty detection. *Mech. Syst. Signal Process.* **2006**, *20*, 1461–1472. [CrossRef]