Anomaly Detection Method for Spacecraft System Based on High Dimensional Space Mapping

Xiangyan Zhang*, Zhiqiang Li 
Beijing Institute of Spacecraft System Engineering, Beijing, 100094, China
* Corresponding author’s e-mail: 2207344@qq.com

Abstract. Spacecraft system, as a typical complex system, is also a high-risk field. Even small faults in local links may bring huge losses or disasters. Analyzing the system level associated anomaly detection method caused by the interaction of multiple telemetry parameters can provide technical means for the rapid and accurate detection of system level in orbit associated faults. In this paper, firstly, a system anomaly detection method based on high-dimensional space mapping is proposed. Secondly, based on the preprocessing of telemetry data, the division of telemetry subsystem is realized. Thirdly, Based on the dimension reduction of subsystem space, the system anomaly detection method is given, and the technical process of the method is given. Finally, through the verification of on orbit actual data, it shows that the system anomaly detection result is accurate.

1. Introduction

As a large and complex system integrating various advanced technologies, spacecraft has been applied to various fields and occupies an irreplaceable position [1]. However, when the spacecraft is on orbit, whether it can successfully complete the related tasks depends on whether the system works normally. Because the spacecraft is exposed to the harsh space environment for a long time on orbit and is interfered by a variety of external factors, the performance of the system will be changed or the spacecraft will be interrupted or even fall, resulting in very serious impact and high loss[2]. It can be seen that on orbit anomaly detection of spacecraft is of great significance. With the complexity of spacecraft functions and the diversification, there are too many cases on orbit that one product has a small problem, which leads to associated failures, leading to the failure of whole spacecraft. Therefore, in order to provide a higher level of a correlation model that can analyse the system level correlation anomaly caused by the interaction of multiple telemetry parameters.

At present, there are four kinds of anomaly detection method for spacecraft: artificial monitoring combined with threshold, expert system, expert experience based model building and data-driven anomaly detection [3]. In China, more attention has been paid to the analysis of spacecraft telemetry data, such as parameter prediction, fault diagnosis, abnormal mutation detection, and health assessment [4-6]. For example, Southeast University proposed to use the autoregressive moving average hybrid model to model the angle burst of geosynchronous satellite control solar panel[7]; Harbin Institute of Technology divided the telemetry data into subsequences by using orbit angle, and detected abnormal segments of temperature, voltage, current and other telemetry data of FY-2 and FY-3 meteorological satellites by using LSSVM[8] and K-nearest neighbour of DTW[9]; Furthermore, it proposed an inductive system anomaly monitoring method based on data correlation analysis for flywheel anomaly detection[10]. However, these studies are lack of detection of system association anomalies.
In this paper, an association anomaly detection method for spacecraft system based on high-dimensional space mapping is proposed, and the technical process of the method is given. Through the experimental verification, it shows that the anomaly detection results are accurate.

2. Anomaly Detection Method Based on High Dimensional Space Mapping

2.1. Spacecraft telemetry data preprocessing

2.1.1. Data completion
For all the telemetry data obtained, the time interval is not fixed. The on orbit telemetry data are supplemented with the minimum time interval. The data analysis needs a fixed sampling frequency, that is, the time interval is the same, so the data should be supplemented with the minimum time interval. The data is resampled by removing the value from the telemetry data file every second. For each telemetry parameter, take the start time point as the starting point, take the value of the corresponding time point in the data file. If there is no sampling value at the current time, take the last sampling value as the current value, and create a new variable to save the current time and value.

2.1.2. Data compression and smoothing
Telemetry data may be damaged or modified during the data transmission. Therefore, it is necessary to eliminate the seriously wrong data. For these data, we have to analyse whether the telemetry data error or the real fault occurs. If it is false data, it is necessary to correct these data. Therefore, a data smoothing algorithm is designed to smooth the data, that is:

\[ x_{new} = \frac{\sum_{i=1}^{k} x_i}{k} \]  

Where: \( x_{new} \) is the telemetry data after smoothing; \( x_i \) is the i-th data to be smoothed; \( k \) is the number of all data to be processed.

The telemetry parameter R1116 is taken as an example to illustrate this problem: Compressed into a data point in 5 hours. As shown in Figure 1, we can see that a large number of data are in the range of 0.4 ~ 0.6, while only a small number of data are in abnormal fluctuation, and the fluctuation amplitude is obvious. After data smoothing, it is helpful for the subsequent algorithm to detect the anomalies in telemetry data.

2.1.3. Data standardization
In order to mine similar telemetry parameters to form a subsystem, because the value range of each telemetry parameter is not the same, it is necessary to standardize all the data, so that the data between different intervals can be compared. According to the variation characteristics of telemetry parameters on orbit, the first and the second standardization method are selected.

1) The first standardization method
Performing linear transformation on the completed telemetry data, and using the conversion algorithm algorithm to map all data to the [0, 1] interval. The conversion algorithm is:
Figure 1. Telemetry data after compression and smoothing with 5 hours

\[ x_{j,i}^* = \frac{x_{j,i} - \text{Min}}{\text{Max} - \text{Min}}, \quad i = 1,2,\ldots,n \]  

(2)

Where: \( x_{j,i} \) is the complete on orbit telemetry data; \( \text{Min} \) is the expected minimum value of the completed on orbit telemetry data; \( \text{Max} \) is the expected maximum value of the completed on orbit telemetry data; \( x_{j,i}^* \) is the converted on orbit telemetry data.

2) The second standardization method
The conversion algorithm is:

\[ x_{j,i}^* = \frac{x_{j,i} - \mu}{\sigma}, \quad i = 1,2,\ldots,n \]  

(3)

Where: \( \mu \) is the mean value of all data \( x_{j,i} \); \( \sigma \) is the standard deviation of all data \( x_{j,i}^* \).

2.2. Similarity analysis of telemetry parameters
The first similarity method or the second similarity method is used to analyse the similarity between different telemetry parameters. The specific calculation method is as follows.

1) The first similarity method: cosine similarity
Cosine similarity measures the difference between two individuals by cosine value of the angle between two vectors in vector space. Compared with distance measurement, cosine similarity pays more attention to the difference of two vectors in direction than in distance or length. The value range of angle cosine is \([-1, 1]\). The larger the angle cosine is, the smaller the angle between two vectors is. The smaller the angle cosine is, the larger the angle between two vectors is. When the directions of the two vectors coincide, the cosine of the included angle takes the maximum value of 1. The calculation formula of cosine similarity between \( X_j^* = [x_{j,1}^*, x_{j,2}^*, \ldots, x_{j,n}^*] \) and \( X_k^* = [x_{k,1}^*, x_{k,2}^*, \ldots, x_{k,n}^*] \) is as follows:

\[ \cos \theta = \frac{\sum_{i=1}^{n} x_{j,i}^* x_{k,i}^*}{\sqrt{\sum_{i=1}^{n} (x_{j,i}^*)^2} \sqrt{\sum_{i=1}^{n} (x_{k,i}^*)^2}} \]  

(4)

2) The second similarity method: Pearson correlation coefficient
Pearson correlation coefficient is a method to measure the degree of correlation between vectors, and the range of correlation coefficient is \([-1, 1]\). The higher the absolute value of correlation coefficient is, the higher the correlation degree is. When it is linearly correlated, the correlation coefficient is 1 (positive linear correlation) or -1 (negative linear correlation).
\[ \rho_{x_jx_k} = \frac{\text{Cov}(X_j, X_k)}{\sqrt{D(X_j)}\sqrt{D(X_k)}} = \frac{E((X_j - EX_j)(X_k - EX_k))}{\sqrt{D(X_j)}\sqrt{D(X_k)}} \]  

(5)

Where: \( X_j \) and \( X_k \) are the jth and the kth telemetry parameter vector respectively; \( EX_j \), \( EX_k \) and \( E((X_j - EX_j)(X_k - EX_k)) \) are the expectation of \( X_j \), \( X_k \) and \( (X_j - EX_j)(X_k - EX_k) \); \( D(X_j) \) and \( D(X_k) \) are the variance of \( X_j \) and \( X_k \); \( \text{Cov}(X_j, X_k) \) is the covariance of \( X_j \) and \( X_k \).

Using these two different similarity measures, we can calculate the similarity between any two telemetry parameters. With this similarity, we can cluster all telemetry parameters.

2.3. Division of telemetry parameter subsystem

A similarity matrix based cluster algorithm (SMC) is designed to cluster all telemetry parameters. All telemetry parameters can be divided into different subsets to form a series of subsystems. All telemetry parameters in the subsystem have similar behavior characteristics. The division process of telemetry parameter subsystem based on SMC clustering is as follows:

1) According to the specified correlation index, the similarity between all telemetry parameters is calculated to form the similarity matrix.
2) An empty dictionary clustereddict is established to store all the telemetry parameters that have entered the subsystem.
3) Select a telemetry parameter sensorcandidate randomly from all telemetry parameter list sensorlist to judge whether the telemetry parameter has entered the subsystem. If not, jump to 4); If yes, go to 8).
4) Establish a subsystem for sensorcandidate.
5) The similarity between sensorcandidate and other sensors in the sensorlist is found from the similarity matrix.
6) If the similarity exceeds the specified threshold, the sensor is added to the current subsystem, the sensor is deleted from the sensor list, and the sensor is added to the clusteredDict.
7) Determine whether the number of telemetry parameters in the subsystem of sensorcandidate exceeds the specified threshold. If so, output the subsystem; If not, all telemetry parameters in the subsystem of sensorcandidate will be added to the sensorlist again.
8) Judge whether there are any unprocessed telemetry parameters in the sensorlist, if so, jump to 3); Otherwise, jump to 9).
9) The end of the algorithm.

Using SMC clustering algorithm, all telemetry parameters can be divided into different subsystems. The anomaly detection of these subsystems is different from single telemetry parameter. When each telemetry parameter is normal, the subsystem may also be abnormal.

2.4. Dimension reduction of high dimensional space based on contribution degree

Subsystem is generally a high-dimensional space. To analyze the data in the high-dimensional space as a whole, it must be mapped to a new space, and the dimension of the new space must be lower than that of the original high-dimensional space. In order to reduce the dimension of the original high-dimensional spatial data, the principal components (eigenvectors) and their weights (eigenvalues) of the data are obtained by eigen decomposition of the covariance matrix. The result can be understood as explaining the variance in the original data: which direction of the data value has the greatest impact on the variance. In other words, if the component corresponding to the smallest eigenvalue is removed from the original data, the resulting low dimensional data must be optimized (that is, if the dimension is reduced in this way, the information loss must be the least). It can be summarized as follows:

1) The covariance between any two dimensions is calculated, and the covariance between all dimensions constitutes the covariance matrix of the whole data.
2) All eigenvalues and eigenvectors are solved for the covariance matrix.
3) The eigenvalues and their corresponding eigenvectors are sorted according to the size of the eigenvalues.
4) The feature vectors corresponding to the first k eigenvalues are selected as the mapping of the original data in the new space.

Since the covariance matrix is a real symmetric matrix, any two eigenvectors obtained must be orthogonal. Therefore, the selected k-dimensional vectors are orthogonal and form a new dimension space. When choosing the contribution degree K, that is:

\[
\text{contribution} = \frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{n} \lambda_i}
\]  

(6)

Where: \( \lambda_i \) is the ith eigenvalue, k is the number of selected eigenvalues, and n is the number of all eigenvalues.

The method of selecting the first k eigenvectors is as follows: the first k eigenvectors with more than 80% contribution are selected, and the selection of K values adopts the principle of minimization. The feature vectors with more than 80% contribution are selected to form a new space, and the data in this space can represent the original data to a great extent.

2.5. System anomaly detection method

For the reduced telemetry data, the Gauss model is established by using the difference between the fitting polynomial and the real data. The model is used to detect the anomaly of each dimension in the new k-dimensional data, and the set of suspected abnormal data points in each feature dimension is obtained. Based on the algorithm of polynomial fitting, the smooth data after dimension reduced is fitted with a specified number of polynomials. A Gauss model is established by using the difference between the fitting polynomial and the real data, and the model is used to detect anomalies. By integrating all the suspected outliers in the k-dimensional data, we can get the pattern anomaly information of the whole subsystem. When integrating the abnormal data of different dimensions, two different ideas are designed:

1) When the spacecraft system anomalies affect the spacecraft payload mission execution, attitude control, energy supply and TT & C uplink and downlink, the intersection of abnormal data of different dimensions is obtained. The physical meaning of doing so is that the subsystem is considered to be abnormal only when there are exceptions in different dimensions at the same time.

2) In other cases, the union of abnormal data of different dimensions is obtained. The physical meaning of doing so is that only when the data in any dimension is abnormal can the subsystem be considered abnormal.

3. Flow chart of spacecraft system anomaly detection

An anomaly detection flow chart based on high-dimensional space mapping is as follows:
Raw telemetry data

Data completion

Data compression and smoothing

Standardize the data according to the specified standardization method

SMC clustering algorithm is used to divide all sensors into subsystems

Select a subsystem

All telemetry data in the subsystem are regarded as a high dimensional spatial data

The contribution of PCA results is calculated

A new k-dimensional space is constructed by selecting eigenvectors that exceed the specified value

Dimension reduction of high dimensional spatial data

Are there any unprocessed dimensions

One dimension of k-dimensional space is selected and the compressed polynomial fitting algorithm is used for pattern anomaly detection

Is there any subsystem that has not been processed

Output the mode exception of the system

Is there a continuous abnormal that exceeds the specified value

Yes

No

Yes

No

end of algorithm

No

Yes

Are there any unprocessed dimensions

No

Yes

All possible outliers of dimension are integrated according to the specified integration method

No

Yes

All outliers of the subsystem are obtained

Is there any subsystem that has not been processed

Yes

No

end of algorithm

4. Experimental verification

The above system anomaly detection method is verified by the actual data on orbit. Firstly, all the telemetry parameters of the whole system are divided into subsystems, and the mode anomaly detection is carried out for each subsystem. Because each subsystem includes a large number of telemetry parameters, only the first and last two telemetry parameters of each subsystem are selected as examples in the visualization process, as shown in Figure 3.

5. Conclusions

This paper implements a clustering algorithm based on similarity, through which all telemetry data in the system can be divided into different subsystems according to their characteristics. Then, the data in the subsystem is mapped to a high-dimensional space, and the mode anomaly detection algorithm of single remote measurement parameter is used to realize the mode anomaly detection of the subsystem. The experimental results show that the algorithm performs well in subsystem partition and pattern anomaly detection.
Figure 3. System mode anomaly detection results

References

[1] XU Y. Research on prediction algorithm of spacecraft on orbit state change trend based on machine learning method[D]. University of Electronic Science and Technology, 2017.

[2] ZHU H. Spacecraft fault diagnosis, prediction and health management [J]. Quality and Reliability, 2013(2):15-19.

[3] PENG X Y, PANG J Y, PENG Y, et al. Review on anomaly detection of spacecraft telemetry data [J]. Chinese Journal of Scientific Instrument, 2016, 37(9): 1929-1945.

[4] FANG H, MA H, HAN L, et al. The study of the spacecraft telemetry data credibility calculation method based on the probability and fuzzy algorithm[C]. National Conference on Information Technology and Computer Science, 2012:628-631.

[5] WU T, SUN D, YU T. Knowledge representation method for spacecraft health status telemetry monitoring [C]. International Conference on Quality, Reliability, Risk, Maintenance, and Safety Engineering, 2013:1771-1773.

[6] ZHANG D, HU SH L, YU N L, et al. Anomaly data detection algorithm for telemetry of spacecraft[J]. Journal of Spacecraft TT&C Technology, 2013, 32(2): 123 - 126.

[7] LI W, MENG Q. Fault Detection for in orbit Satellites Using an Adaptive Prediction Model [J]. Chinese Journal of Space Science, 2014, 34(2):201-207.

[8] SONG AI G, LIANG J, LIU D T, et al. Anomaly detection of condition monitoring with predicted uncertainty for aerospace applications[C]. IEEE International Conference on Electronic Measurement & Instruments, 2015: 248-253.

[9] CHEN J. Similarity measure of time series for satellite telemetry data [D]. Harbin: Harbin Institute of Technology, 2015:39-43.

[10] GONG X B, WANG R X, XU M Q. Abnormality detection for fly wheels based on data association analysis [J]. Acta Aeronautica ET Astronautica Sinica, 2015, 36(3):898-906.