SciencePG

Science Publishing Group

Conjunction Detection and Series Clustering of Telemetry Data from Spacecraft in Orbit

Hu Shaolin[1, 2], Fu Na[2]

[1]School of Automation, Foshan University, Foshan, China

[2]Key Laboratory of Fault Diagnosis & Maintenance of Spacecraft in Orbit, Xi'an, China

Email address:

hfkth@126.com (Hu Shaolin)

Abstract: To explore the conjunction of abnormal changes among different processes is a key and challenging technique problem in processes monitoring, in faults analysis, and in faults location. In this paper, an indication series is used to symbolize the abnormal change in sampling series, two kinds of conjunction test indices are constructed to measure the conjunction degrees, which rely on the indication series of multidimensional synchronization sampling series and the abnormal change percentage series of multidimensional asynchronous sampling series separately. What is more, these conjunction-test indices are successfully used to set up the clustering algorithms of abnormal changes in multidimensional series. Some Monte Carlo results show that algorithms given in this paper are efficient. The idea and technological methods of this paper are helpful for us to get viable approaches to analyze abnormal changes and to diagnose faults in large-scale dynamic system.

Keywords: Conjunction Analysis, Stationary Processes, Abnormal Changes

1. Introduction

In order to know the in-orbit states of a spacecraft and to make sure its safe spaceflight, there are quite a lot of telemetry data flow must be checked and analyzed, which received from on-orbit spacecraft. For example, there are one hundred thousand various telemetry devices, which are fixed on spacecraft in space station and used to obtain the information of operation states. From spacecraft safety point of view, the telemetry data flow is the most reliable the source of information to directly monitor states and to find out faults of spacecraft which is running in orbit. In a certain extent, telemetry-data flow is the most dependent window that can help us to understand the true states and to manage its health [1] of spacecraft in orbit.

But, it is very sorry that telemetry data flow is not absolutely reliable. In other words, it is possible that there are mishaps taking place at any steps or nodes of the whole process from measuring, coding, transmission, and receiving to decoding of the spacecraft telemetry data flow, such as abnormal changes and faults of spacecraft as well as telemetry device, which may cause abnormalities of multi-dimensional telemetry data flow [2, 3].

Some research progress reveals that it is very important how to scientifically analysis such numerous telemetry data flow so as to discover whether the spacecraft works in normal states or not, which is not only a significant technical problem in engineering fields, but also a scientific problem in academic researches, which need to innovate some new approaches to do further analysis. The first reason is that the complexity of telemetry data flow from spacecraft in orbit: telemetry data flows are high dimensional, time-varying, heterogeneous, nonlinear time series with disturbances and the telemetry data time series is far more complicated than the sampling of industrial and economic fields; the second reason is that the space environment of generating spacecraft telemetry data is complicated and indeterminate, such as impurities and asynchronous sampling and data flow disconnection etc. These reasons stated above directly affect the usability of conventional methods of time series analysis.

In order to analysis relationship of abnormal changes taking place at different data sequences among multidimensional

telemetry data flow, there are some conventional approaches which are based on the correlation analysis and regression analysis of stationary series and mainly focus on exploring the connection among the different data series value [1-3]. Maybe, some quantitative or qualitative analysis methods [4-6], which are based on data mining and knowledge discovery technologies, are usable to explore the inherent regularity or characteristic information, and some fault detection and diagnosis algorithms [7-11], which are based on the threshold monitoring as well as data driven detection, may be directly realized from telemetry data. But, there were few studies discussed the results about inherent relevance between the homologous data abnormal changes in a multidimensional telemetry data flow, which may be give us some valuable inspiration where the faults arise and which kind of faults result in abnormal changes in part of the multidimensional telemetry data flow. Because of the reasons stated above, some practical problems, such as clustering of abnormality changes in telemetry data flow and monitoring of health states and so on, are combined into consideration and new clustering methods oriented at process abnormal changes are proposed in the following sections.

2. Conjunction Detection Algorithms

In order to simplify statements, assuming that there are m telemetry devices and m series of telemetry data sequence $S_k = \{f_k(t_i) \mid t_i \in [t_0, t_e], i = 1, \cdots, n\}$ $(k = 1, 2, \cdots, m)$ in a given time segment $[t_0, t_e]$ and that the sampling time of different telemetry devices may be not completely the same or be even hundred-percent different (i.e. $t_{1,i} \neq \cdots \neq t_{n,i}$) and that sampling interval of different telemetry device are not completely equal (i.e. $n_1 \neq \cdots \neq n_m$), the problems are discussed in two different situations respectively.

2.1. Conjunction Detection of Synchronous Telemetry Data

If a spacecraft is provided with m telemeters which synchronously sample telemetry data within a given time interval $[t_0, t_e]$, the m-dimensional measurement data series are represented as

$$S_k = \{f_k(t_i) \mid t_i \in [t_0, t_e], i = 1, \cdots, n\} \quad (k = 1, 2, \cdots, m)$$

In order to analysis the abrupt changes conjunction of the telemetry data series, much attention is paid to the cases whether these series are beyond threshold or not. Assuming that normal range of the k-th dimensional telemetry data of is $[-c_k, c_k]$, the 1/0 type discrete state telemetry representation sequence $I(S_k) = \{T_{k,i} \mid i = 1, 2, \cdots, n\}$ can be generated as follows:

$$T_{k,i}(S_k) = \begin{cases} 1, & |f_k(t_i)| > c_k \\ 0, & |f_k(t_i)| \leq c_k \end{cases} \tag{1}$$

Let's design the Boolean operator Δ

$$1\Delta 1 = 0\Delta 0 = 1, 1\Delta 0 = 0\Delta 1 = 0$$

and use the operator Δ on the set $I(S_k) = \{T_{k,i} \mid i = 1, 2, \cdots, n\}$, a new detection function named as Type-I detection index can be set up as follows

$$\rho_I(u,v) = \frac{1}{n} \sum_{i=1}^{n} \left(T_{u,i} \Delta T_{v,i} \right) \quad (u, v = 1, 2, \cdots, m) \tag{2}$$

Theorem 1: if the m telemetry devices fixed in a spacecraft sample synchronously sampling, then the Type-I detection function $\rho_I(u,v)$ of multi-dimensional time series has following properties:
(a) Symmetrical relation:

$$\rho_I(u,v) = \rho_I(v,u) \quad (u, v = 1, 2, \cdots, m)$$

(b) Normal Bound Inequality

$$0 \leq \rho_I(u,v) \leq 1 \quad (u, v = 1, 2, \cdots, m)$$

(c) Selves Reflexivity:

$$\rho_I(v,v) = 1 \quad (v = 1, 2, \cdots, m)$$

Proof: The correctness of the theorem 1 is obvious. The detailed proof process is omitted.
From formulae (2) you can find out that, if the device u obtaining data sequence S_u is identical or strong conjunct with the device v obtaining data sequence S_v whether these devices are normal or not, then the detection function $\rho_I(u,v) = 1$. is realized; and if the device u are not conjunct with the device v, then the detection function $\rho_I(u,v) = 0$ is realized; and if the device u are partly conjunct with the device v, then the detection function satisfies the following inequality $0 < \rho_I(u,v) < 1$. Statistics $\rho_I(u,v)$ is a good index to reflect the data sequence S_u and data sequence S_v are synchronous conjunction strength whether it is normal or not.

2.2. Conjunction Detection of Asynchronous Telemetry Data

Because sampling times and sampling density of different telemetry devices may be different, it is impossible to directly compare with two or more than two series of samples. In order to judge whether the telemetry data obtained from different telemetry device are relevant or not when abnormal changes take place. So, it is necessary to process the telemetry data into several segments in appropriate way. Relying on these segments, it is able to establish the decision algorithm on conjunction detection of telemetry data about abnormal changes.
The basic idea stated above is equally to divide the time segment $[t_0, t_e]$ into w pieces. Counting $n_{k,s}$ which is the number of telemetry data which locate in the piece $Q_s = [t_{s-1}, t_s)$ of $[t_0, t_e] = [t_0, t_1) \cup [t_1, t_2) \cup \cdots [t_{w-1}, t_e]$ and $b_{k,s}$ the number of telemetry data which are beyond the

threshold c_k, we have the following formulae:

$$\begin{cases} n_{k,s} = \sum_{i=1}^{n_k} \left\| \{t_{k,i} \mid i=1,2,\cdots,n_k\} \cap Q_s \right\| \\ b_{k,s} = \sum_{i=1}^{n_k} \left\| \{t_{k,i} \mid f(t_{k,i}) \le c_k, i=1,2,\cdots,n_k\} \cap Q_s \right\| \end{cases} \quad (3)$$

where the threshold c_k is a practical discriminant bound.

Using formulae (1), number of the sampling points in multi-dimensional telemetry data series can be counted in each sampling times and the index (2) can be modified as follows:

$$d_{k,s} = \frac{n_{k,s}}{b_{k,s}} = \frac{\sum_{i=1}^{n_k} \left\| \{t_{k,i} \mid i=1,2,\cdots,n_k\} \cap Q_s \right\|}{\sum_{i=1}^{n_k} \left\| \{t_{k,i} \mid f(t_{k,i}) \le c_k, i=1,2,\cdots,n_k\} \cap Q_s \right\|} \quad (4)$$

Thus, a series of new detection statistics are established for multidimensional telemetry data series and shown in table 1.

Table 1. *Statistics of Multi-dimensional Time Series Beyond Bounds.*

	$[t_0, t_1)$		$[t_1, t_2)$		……	$[t_{k-1}, t_k]$	
S_1	$n_{1,1}$	$d_{1,1}$	$n_{1,2}$	$d_{1,2}$	……	$n_{1,k}$	$d_{1,k}$
…	…						
S_m	$n_{m,1}$	$d_{m,1}$	$n_{m,2}$	$d_{m,2}$	……	$n_{m,k}$	$d_{m,k}$

Using the statistical results shown in table 1 and referencing the concept of correlation coefficient in statistics, a modified formula (5) is constructed to detect sequence correlation of multi-dimensional telemetry data when there are abnormal changes taking place

$$\rho_{II}(u,v) = \frac{1}{w} \sum_{i=1}^{w} \left\{ 1 - \left| d_{u,i} - d_{v,i} \right| \right\} \quad (u,v=1,2,\cdots,m) \quad (5)$$

Formula (5) is called as type II detection function. It is not difficult to verify the correlation detection function $\rho_{II}(u,v)$ is symmetrical.

Theorem 2: For m-dimensional time series in time interval $[t_0, t_e]$ of telemetry data from spacecraft in space, the type-II conjunction detection function determined in formula (5) has the following properties:

(a) Symmetry: $\rho_{II}(u,v) = \rho_{II}(v,u) \quad (u,v=1,2,\cdots,m)$

(b) Polarity: $0 \le \rho_{II}(u,v) \le 1 \quad (u,v=1,2,\cdots,m)$

(c) Reflexivity: $\rho_{II}(v,v) = 1 \quad (v=1,2,\cdots,m)$

Proof: The correctness of the theorem 2 is obvious. The proof process is omitted.

Comparison type I detection function $\rho_I(u,v)$ with type II detection function $\rho_{II}(u,v)$, the following results can be obtained: $\rho_I(u,v) = \rho_{II}(u,v)$ if sampling processes are synchronous. So, the type II detection function can be used in all kinds of sampling series whether they are synchronous or not, but the type I detection function is just only suitable for synchronous sampling series.

3. Changes Conjunction Based Cluster

Spacecraft telemetry data have the characteristics of variety, complex shape, asynchronous sampling, unequal interval as well as impurities, losing samples, variable dimensions, different physical meaning, and variant amplitude etc. These intricate situations are very difficult to deal with the conjunction detection. In this section, much attention will be paid to set up a new kind of equivalence relationship among different telemetry data series in the case that there are abnormal changes take place in the processes.

Assuming that, in the period of time $[t_0, t_e]$, a spacecraft have m telemetry devices that obtain spacecraft status from different sensors and form m telemetry data segment

$$S_k = \{f_k(t_{k,i}) \mid t_{k,i} \in [t_0, t_e], i=1,\cdots,n_k\} \quad (k=1,2,\cdots,m)$$

The type II detection function $\rho_{II}(u,v)$ can be used to cluster different series as follows:

Step 1: To divide equally the time segment $[t_0, t_e]$ into w pieces, namely

$$[t_0, t_e] = [t_0, t_1) \cup [t_1, t_2) \cup \cdots [t_{w-1}, t_e]$$

Step 2: Using formula (3) to count the numbers of telemetry data points $n_{k,s}$ and $b_{k,s}$;

Step 3: Calculating the maximum and minimum abnormal change percentage:

$$\begin{cases} Max_k = \max \left\{ \frac{b_{k,s}}{n_{k,s}} \mid s=1,2,\cdots,w \right\} \\ Min_k = \min \left\{ \frac{b_{k,s}}{n_{k,s}} \mid s=1,2,\cdots,w \right\} \end{cases} \quad (6)$$

Step 4: To divide the m-dimensional telemetry data sequence $S = \{S_k \mid k=1,2,\cdots,m\}$ into three parts: almost no abnormal change subset S^A, almost entirely abnormal change subset S^B and partly abnormal change subset S^C:

$$\begin{cases} S^A = \{S_k \mid Max_k \le 1\%, k=1,2,\cdots,m\} \\ S^B = \{S_k \mid Min_k \ge 99\%, k=1,2,\cdots,m\} \\ S^C = S - S^A - S^B \end{cases} \quad (7)$$

Step 5: In the partly abnormal change subset S^C, the telemetry data sequence

$$S_k^1 = \{f_k(t_{k,i}) \mid t_{k,i} \in [t_0, t_e], i=1,\cdots,n_k\}$$

is randomly selected and the type II detection function $\rho_{II}(k_1, v)$ are calculated between subset S_k^1 and subset

$S^C - \{S_k^1\}$ among all sequence S_v. Using detection function $\rho_{II}(k_1, v)$ to split the subset of abnormal change relationship

$$\hat{S}_k^1 = \{S_k^1\} \cup \{S_v \mid \rho_{II}(k_1, v) \geq 90\%, S_v \in S^C\} \qquad (8)$$

Step 6: To set $S - S^A - S^B - \hat{S}_k^1 \Rightarrow S^C$;

Step 7: To repeat step 5 and step 6 until $S^C = \Phi$.

For any m-dimensional telemetry data series, using step 1 – step 7 stated above, the telemetry series can be divided into several sub-sequences that have inherent abnormal change correlation as follows:

$$S = \hat{S}_k^1 + \cdots + \hat{S}_k^p + S^A + S^B \qquad (9)$$

The number of abnormal change association type is

$$p \leq m - (\|S^A\| + \|S^B\|)$$

4. Simulation Results Analysis

In order to verify the validity of the algorithm stated above, the Monte Carlo method and two independent standard normal random process models are used with contamination.

$$\begin{cases} S_1(t) = s_1(t) + \delta_1(t) \\ S_2(t) = s_2(t) + \delta_2(t) \end{cases} \quad (s_1(t) \sim N(0,1),\ s_2(t) \sim N(0,1)) \quad (10)$$

Case I

Setting the pollution components $\delta_1(t_k) = (-1)^k 6.1$ and $\delta_2(t_k) = (-1)^{k-1} 5.9$ ($k = 15i$ *where* $i \in I$ *and* $k \leq 100$) separately. Using the Monte Carlo method, two groups of abnormal data series with 100 random samples are generated, named as S_1 and S_2, and plotted in Fig 1. Obviously, the pollution components are different in directions, the magnitudes as well as times of the pollution components are similar.
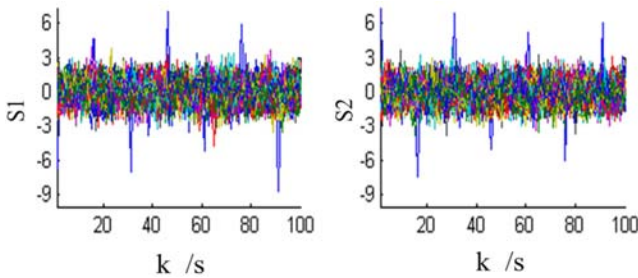
Figure 1. *Two Series of Conjunctive Data Contaminated by Outliers at the Same Times and Opposite Directions.*

Using formula (1), the characterization of sequence T_1 and T_2 are calculated and plotted in Fig 2, which are almost the same.
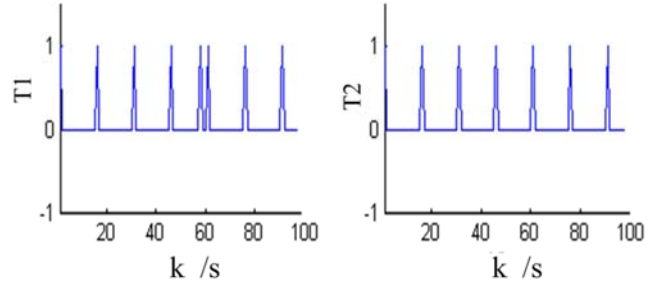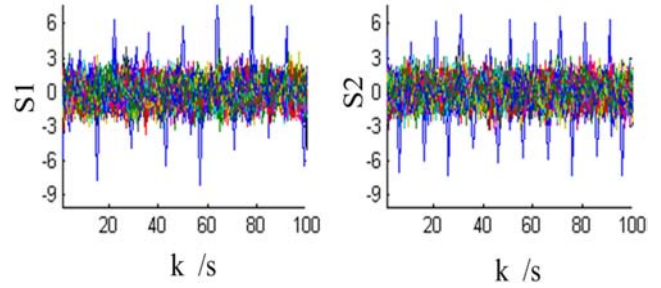

Figure 2. *Abnormal Index Series of The Two Series in Fig. 1.*

Using formula (2) to calculate the type I detection function, correlation index is $\rho = 0.99$. This result illustrates that the abnormal changes in these two sequences are intrinsic relevant.

Case II

Setting the pollution components $\delta_1(t_k) = (-1)^k 6$ ($k = 7i, i \in I; k \leq 100$) and $\delta_1(t_k) = (-1)^k 5.9$ ($k = 5i, i \in I$ *and* $k \leq 100$), the pollution amplitude is almost the same but the pollution times are different, which are plotted in Fig. 3.

Figure 3. *Two Series of Non-conjunctive Data Contaminated by Outliers at Different Times and the Same Directions.*

Using formula (1) to calculate characterization sequence T_1 and T_2, which are plotted in Fig. 4.
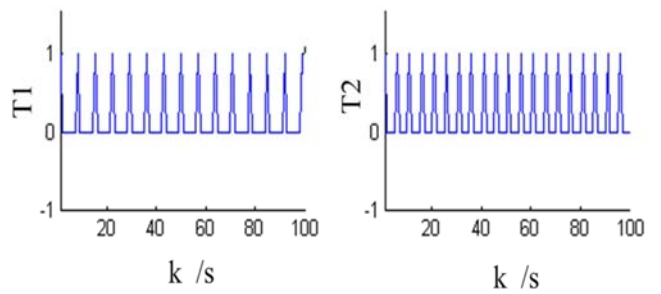
Figure 4. *Abnormal Index Series of The Two Series in Fig. 3.*

Using formula (2) to calculate the type I detection function, the correlation index is $\rho = 0.71$. This result illustrates that abnormal change in these two sequences have no relevance.

5. Conclusions

To explore the conjunction of multidimensional abnormal series from telemetry components (or abnormal series contaminated by outliers) is a key and challenging

technique problem in process monitoring, in faults analysis, and in faults location. In this paper, it is ingeniously solved by appropriating two kinds of correlation detection function. Based on the correlation detection functions, a large number of different processes can be clustered. After clustering these series, it is easy to distinguish different anomalies in a large-scale complex system, which relies on whether the changes cause the same cause or not. It provides a feasible technical approach for us to discover and to judge causes of faults in the system operation processes. Specifically, in section 2, a set of relevance detection index are established, which are applicable to clear impurity band and to mend the missing telemetry data sequence. The multi-dimensional clustering method of telemetry data are used and several series of are existed. Analysis of the simulation result can be used to verify the availability of this method and algorithm.

Acknowledgment

References

[1] B L Bowerman, R T O'Connell, A B Koehler. Forecasting, Time Series, and Regression: an Applied Approach, 2004, 1-62.

[2] S Menard. Applied logistic regression analysis. Shanghai People's Press, 2002: 2002, 1-139.

[3] Jianqing Fan, Qiwei Yao. Nonlinear Time Series—Non-parametric and Parametric Methods. Beijing: Science Press, 2006, 29-274.

[4] Hu Shaolin, Li Ye, Zhang Dong. Fault-tolerant mining algorithm of sampling data from dynamic system. 25th Chinese control and decision conference. 2013, pp. 4927-4931.

[5] Hu Shaolin, Li Ye, Zhang Wei. Algorithms of Data Mining and Knowledge Discovery of Correlativity in two-dimensional Time Series. Applied Mechanics and Materials, 2013, Vols. 263-266 pp 1844-1848.

[6] S. P. Bingulac. "On the compatibility of adaptive controllers (Published Conference Proceedings style)" in Proc. 4th Annu. Allerton Conf. Circuits and Systems Theory, New York, 1994, pp. 8–16.

[7] J Vega, S Dormido-Canto, T Cruz, et al. Real-time change detection in data streams with FPGAs. Fusion Engineering & Design, 2014, 89 (5): 644-648.

[8] C Svärd, M Nyberg, E Frisk, et al. Data-driven and adaptive statistical residual evaluation for fault detection with an automotive application. Mechanical Systems & Signal Processing, 2014, 45(1): 170-192.

[9] SX Ding, P Zhang, T Jeinsch, et al. A survey of the application of basic data-driven and model-based methods in process monitoring and fault diagnosis. IFAC Proceedings Volumes, 2011, 44(1): 12380-12388.

[10] S Cho, SB Kim. One-Class Classification Methods for Process Monitoring and Diagnosis. Intelligent Systems IEEE, 2015, 30(6): 16-18.

[11] Chai Min, Yng Yue, Xu Xiaohui, et al. The Dimension Reduction Analysis of Spacecraft's Telemetry Data for Fault Diagnosis. Journal of Projectiles, Rockets, Missiles and Guidance, 2014, 34(1): 150-153.

Biography

HU Shaolin is currently working as a Professor in the School of Automation at Foshan University and Xi'an University of Technology, China. He received the B. Sc. Engg degree in Mathematics from Anhui Normal University in 1986. He received Master degree in Mathematics and PhD degree in Control Science and Engineering from Xi'an Jiaotong University in 1989 and 2000 respectively. His research interests include data-driven faults diagnosis for complex system, process monitoring, fault-tolerant control, outlier-tolerant computing and applications.