

Outlier-tolerant K Charts Based Tendency Analysis with Variable Fineness of Big Telemetry Data

Qin Bo, Hu Shaolin, Yang Yumin, Guo Xiaohong

Key Laboratory of Spacecraft Fault Diagnosis and Maintenance, Xi'an, 710043, China

Email: cooler2003@sina.com

Keywords: Big data, Telemetry data, K Charts analysis

Abstract. The spacecraft telemetry data is the only way for operation managers to understand the spacecraft's real working state in space. In face with the increasing amount of telemetry data growing day by day, it is hard to overcome the long data time span, large data volume, dirty data and etc. It is very necessary to explore some effective analytical methods to analysis the big telemetry data. In this paper, the K Charts analysis method is referenced and used to deal with the telemetry big data for safety analysis. Using the transformation and improvement, particle size analysis technique, the fault-tolerant K Charts is proposed and established for spacecraft telemetry data change analysis. K Charts cycle through the transforming of variables, different time granularity, formation, K Charts combination, to realize mass spacecraft telemetry sampling data association analysis of structure the shape and the state. The results of application show that, combinations of different sizes of K Charts can reflect the law of the data intrinsic change more accurately than ordinary timing diagrams.

Introduction*

Telemetry data stream is the most direct means to understand the working state of spacecraft in orbit and its states^[1]. Spacecraft in orbit will constantly transmit telemetry data to the center. Through the analysis and processing of the telemetry data, control personnel can understand each part of the spacecraft's working condition and health, or to find out if the spacecraft running state is abnormal^[2]. It is of utmost importance to ensure safe and stable operation of the spacecraft through fault diagnosis of the transmitted data^[3].

Analysis of telemetry data processing is a complex process. Due to the complexity of the spacecraft structure, its operating environment and the control process of spacecraft, it is common to have unequal frequency intervals, lost frequencies, uneven frequencies or even signal disturbances problems^[4]. All of these problems are making telemetry data analysis and processing more difficult^[5]. The spacecraft telemetry data is obtained from an environment in varying conditions such as high density, long time span, impure data, symbolizing to that of big data feature^[6]. This will indicate that conventional data analysis and treatment methods are no longer relevant or valid, and there is an urgent need to explore and establish a new means of analysis and processing^[7].

In the field of securities investment, there is a kind of widely used sequential data analysis tool -K Charts. The K Charts is also referred as Candle Charts, the Japanese Line^[4], Line of Yin and Yang, Wire Rod, and was first originated from Japan in the eighteenth Century. K Charts theory is first proposed by Japanese as an analysis theory, and discovered the earliest in eighteenth Century for analyzing the Japanese rice market transaction. From then on, K Charts theory has been widely used in domestic and international securities investment analysis and other fields^[4]. The core of K Charts theory is through the form of K - line and a combination of K Charts' shape for prediction of data trend. It has gained good credibility after hundreds of years of market testing and application^[8]. In view of analyzing big data with telemetry characteristics such as long time span, the K charts is a very effective tool. Therefore, this paper is to explore a set of methods for the analysis of telemetry data modeling and variable size that K charts is based on^[9]

* Partly Supported by the National Nature Fund of China (No.61473222)

Telemetry Data K Charts design

The basic structure of K Charts consists of shadow with colored (usually red and blue, this paper use the white and black) wire column, and it reflects changes of the observed object in a period of time. It also has a good ability of information aggregation, especially suitable for big data analysis and mining. Using single K Charts combination column line and color, the time segment measurement of change in data can be generally understood; data sequence with long time span can be divided into a series of segments of data with shorter time span, through a combination of the color with the K, the details of the data can be effectively compressed to grasp the overall data change trend. Data fragments of different sequences can also be distinctly displayed through the K Charts combination.

A. Meaning of Telemetry data in K Charts

Similar to the wide application K Charts theory has in the securities investment analysis, for one-dimensional long timing telemetry data $S = \{y(t_n), n=0,1,2,3,\dots\}$ in numerical sequence, $S_i = \{y(t_i), i=i_1, \dots, i_s\} \subset S$ changes in any given time fragment, it can be got from four features: start point b_i , period end point e_i , maximum value u_i and minimum value d_i . Using the above four features with the corresponding color change, one can draw a line under the shadow line / column, as shown in Fig.1. In Fig. 1, the blackbody column line read from top to bottom is called the Yin line; while the white body column line, known as the Yang line is read from bottom to top.

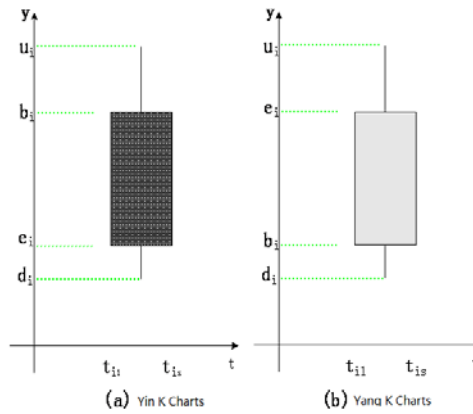


Fig. 1 Schematic diagram of the K Charts structure

B. Realization of the K Charts Algorithm

For big data sequence with long time span, it comprises of a module structure with 2 functions and a main control process to make the drawing of K-Charts possible. Specifically, for the convenience of narration, telemetry data sequence can be derived through

$$S = \{y(t_n), n = 0, 1, 2, 3, \dots\}$$

Beginning from t_0 , increasing along the time axis of length as h for equal interval segmentation, i.e.

$$S = \bigcup_{i=1,2,3,\dots} (\{y(t), t_0 + (i-1)h \leq t < t_0 + ih\} \cap S) \quad (1)$$

It is not difficult to know, the data collection

$$S_i = \{y(t), t_0 + (i-1)h \leq t_j < t_0 + ih\} \cap S$$

is symbolized by, and data points $s = \|S_i\|$ being. For drawing the K charts, it requires two function module structures:

a). Rank and value module

For any real sets $V = \{x_1, \dots, x_s\}$, flow completion of the set V of each element is shown in Fig. 2 in ascending order. Being maximum output as y_{\max} , minimum output as y_{\min} , upper quartile as Q_3 and lower quintile as Q_1 :

b). K Charts drawing module

In a $t-y$ Cartesian coordinate system, at any time $[t_a, t_b]$ the K Charts of the telemetry data is determined by 5 factors: the start point $P_a(t_a, y(t_a))$, end point $P_b(t_b, y(t_b))$, the highest

point $P_{\max}((t_a + t_b)/2, y_{\max})$, the lowest point $P_{\min}((t_a + t_b)/2, y_{\min})$ and column line color. The color factor depends on the relative position between P_a and P_b . As we can see, each K Charts is uniquely determined by the four factors mentioned above. It is not difficult to complete the basic drawing module, and the process as shown in fig 3.

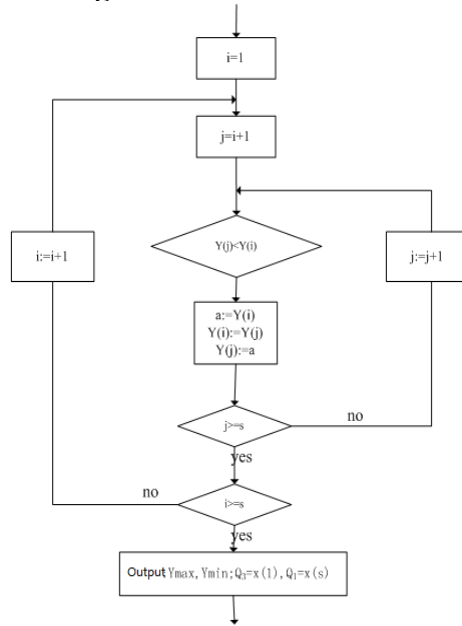


Fig. 2 the values of data sorting module

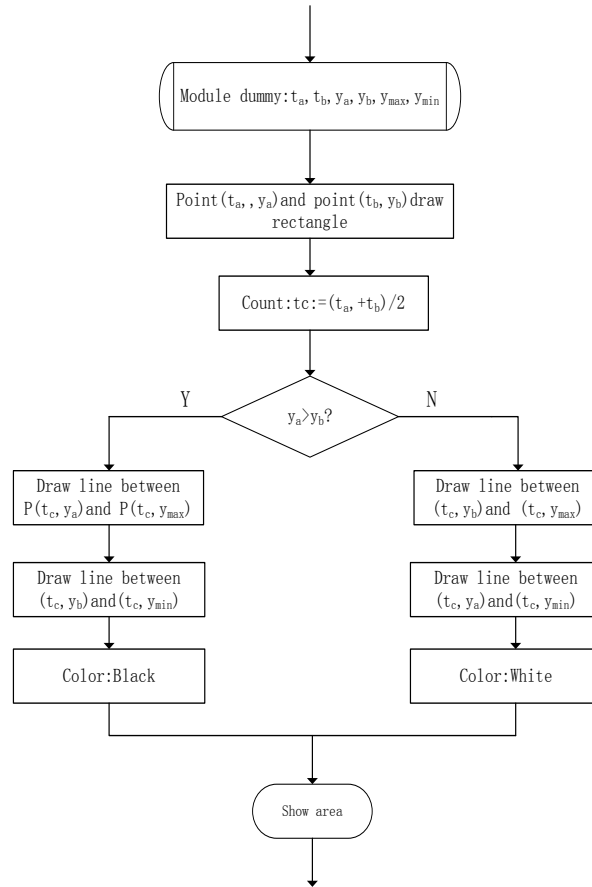


Fig. 3 K Charts drawing module

By using the two modules, it is not difficult to complete the K charts drawing and processing. In fact, for large data sequence with long time span $\{y(t_i), i=1,2,3,\dots\}$, rendering the complete K Charts only requires cutting the long periods of data at a fixed window width into several segments, and then calls the modules by section to finish the whole K Charts drawing.

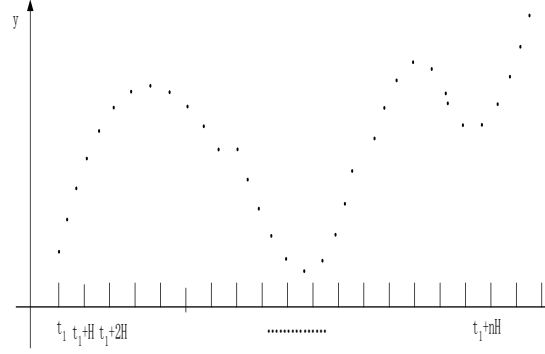


Fig. 4 Schematic diagram of sequential data partition

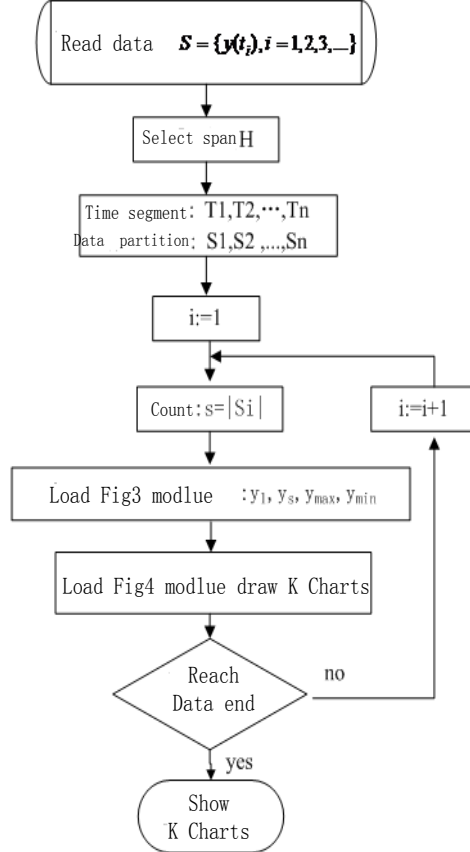


Fig. 5 K Charts drawing process

C. Fault-tolerant K Charts algorithm

The practical significance of each data set in different transaction from the stock market is different. The process of spacecraft telemetry in data acquisition, encoding, transmission, to receive and decode may result in error, with "a few abnormal data that seriously deviated from most of the data presented in the change trend." [5] Such occurrence is called outliers. Due to the existence of outliers, it will result in K Charts with long shadow on or under the shadow line, making the K Charts difficult to visualize and reflect the overall change of the telemetry data and trend. Therefore, this paper will attempt to make adjustment to the fault tolerance of the classical K Charts.

Method 1: Make direct adjustment in each section in $T_i = [t_i, t_i + H)$ of the telemetry data formula

$$S_i = \{y(t_j) \mid y(t_j) \in S, t_j \in [t_i, t_i + H)\} \quad (2)$$

and then draw the K Charts diagram.

$$\tilde{y}(t_j) = m_i + \begin{cases} 4.5\Delta_{j,i}, & y(t_j) - m_i > 4.5\Delta_{j,i} \\ y(t_j) - m_i, & |y(t_j) - m_i| \leq 4.5\Delta_{j,i} \\ -4.5\Delta_{j,i}, & y(t_j) - m_i < -4.5\Delta_{j,i} \end{cases} \quad (3)$$

where $med\{\dots\}_i$ in the formula is a median operator

$$m_i = med\{y(t_j)\}_{t_j \in T_i} \cdot \Delta_{j,i} = med\{|y(t_j) - m_i|\}_{t_j \in T_i}$$

Method 2: For each section in $T_i = [t_i, t_i + H)$ of the telemetry data formula

$$S_i = \{y(t_j) \mid y(t_j) \in S, t_j \in [t_i, t_i + H)\}$$

sort values in ascending order. Record the lower quartile Q_1 and the upper quartile Q_3 respectively. Make rational adjustment to all S_i variables in formula (3), and then draw the K Charts diagram.

$$\tilde{y}(t_j) = m_i + \begin{cases} 2(Q_3 - m_i), & y(t_j) > 2Q_3 - m_i \\ y(t_j) - m_i, & 2Q_1 \leq y(t_j) + m_i \leq 2Q_3 \\ 2(Q_1 - m_i), & y(t_j) < 2Q_1 - m_i \end{cases} \quad (4)$$

The above two methods have good fault- tolerant ability, and can ensure that the rendering of the K Charts diagram will not be distorted due to overly long upper or lower shadow line.

K Charts based Fault-Tolerance Analysis

Each of the K Charts diagram data can have highly concentrated changes over a period of time (start, end, maximum and minimum, and the rising and falling), changes of a combination of several K Charts may highlight a long period of time and changes in the intrinsic characteristics. Securities investment analysis combined morphological and color used multiple K changes, analysis of market price trend and forecast the change tendency, the core of the thought and technology of spacecraft telemetry data, analysis and prediction of evolution process is undoubtedly a reference value. It can get good reflection from variable granularity analysis of multi K Charts combination.

A. Analysis based on combination of variable size K charts

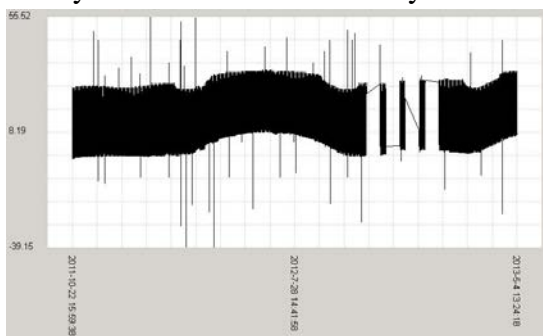
In the field of spacecraft management, because the data coverage of time is over several years or up to 10 years, and unequally spaced sampling contains a large number of impurities or outliers, which makes it difficult for conventional analysis method. Fault tolerant K Charts method can effectively compress the data to effectively avoid the adverse effect of the impurity of data, large span and long period changes of visual angles to show data changes.

For the long timing data sets, through a series K Charts drawing with different length, it can form the K Charts combination of different particles, it can understand and separates series data in short period changes, medium length and long periods time changes. Then by using the K Charts combination, it can form different size under- standing of long period data sequence variation trend, short-term, medium-term and long- term trends hold the data changes.

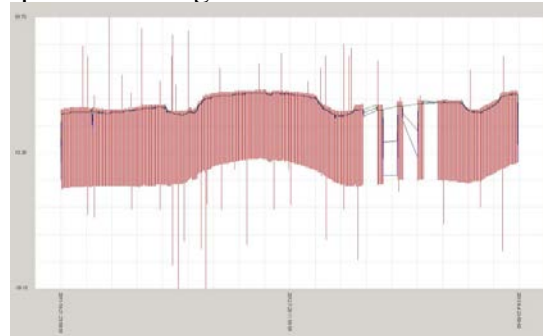
From the angle of quantitative analysis, it can also take the K Charts in shadow on vertex of each K Charts and under the shadow of the end point, the connection or curve fitting through point by point, forming a curve on the two. The two curves are essentially changes of the timing data on the [6] envelope and the lower envelope, pipeline two envelope formations is the telemetry data changes in pipe. Thus one can predict the variation tendency of telemetry data.

B. Fault tolerant K Charts and variable granularity

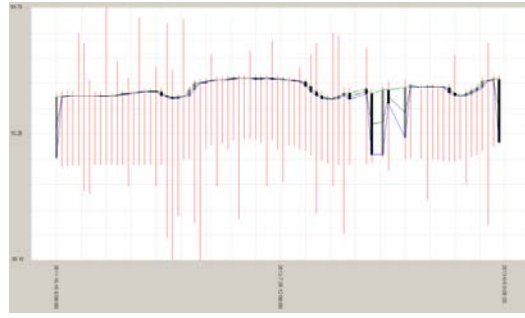
One spacecraft in-orbit for years, Fig 6 (a) is the actual acquisition of telemetry data temperature change situation earth sensor shell of spacecraft on the scatter diagram. Fig 6 (b) – (d) is drawn by using the method of drawing the K Charts by day, week and month. The overall situation from Fig 6 can clearly understand the sensitivity of shell temperature change.



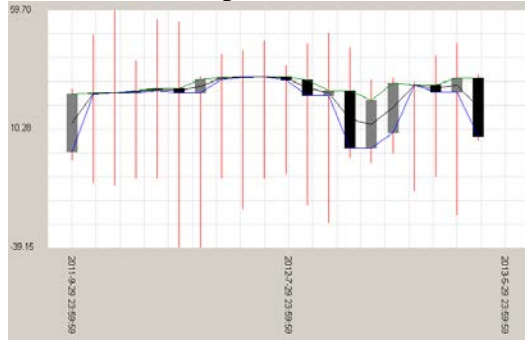
(a) Shell temperature Sandian diagram



(b) Shell temperature k-day charts



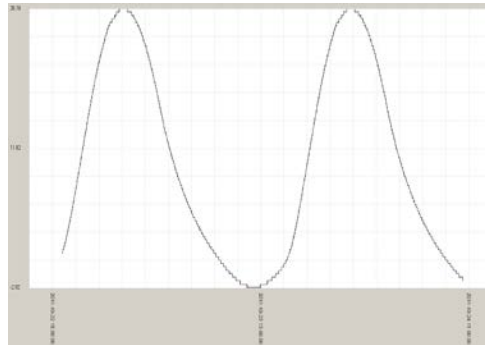
(c) Shell temperature k-week charts



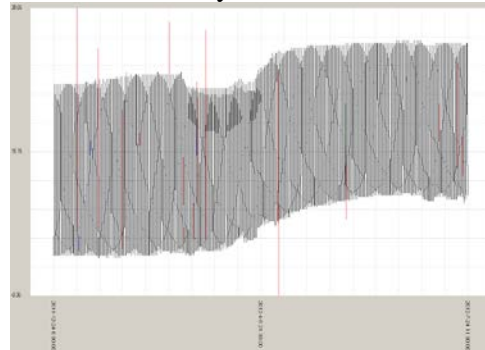
(d) Shell temperature on K Charts

Fig. 6. K Charts combination of of Temperature

Further transform size, are plotted respectively in 6 months, 2 days when the K charts, as shown in fig 7. From Fig 7 we can see clearly: temperature in 28.5 as the central axis changes, changes in amplitude of about 28.5, about 6 month period. A detailed analysis of Fig 7 can also be found that the temperature is in 21 hours or so cycle, to approximate sinusoidal curve, phase deviation is about 2 months.



(a) The K Charts 2 days when the shell temperature



(b) The K Charts of 6 months temperature

Fig. 7 K charts of temperature

Method to analyze the K charts data can help to improve efficiency significantly. For example, the above sampling density of 1 point / sec telemetry data series, annual data points is about $365 \times 24 \times 3600 = 31536000$, when adopting the K Charts day K Charts and compressed data quantity are about 4.3×10^4 and 1.8×10^3 . K Charts graph method can effectively compress the data

capacity, particularly suitable for the analysis and prediction of impurity of telemetry data over a long time.

Conclusion

The process of spacecraft telemetry data acquisition in the long run help us to understand the state of system changes process and trend of space- craft. In order to overcome issues affecting the telemetry data analysis such as long time span, large data volumes, impurities, outliers. Based on the classical K Charts analysis technology, we have established a set of analysis method for impurities, outliers and missing data with fault tolerance. In addition, we have made improvements to the classic K Charts analysis method by improving the fault tolerance level. Through analyzing and processing of the sensitive measured data of a spacecraft, we have verified the validity and practicability of the new method.

Acknowledgment

The paper is partly supported by National Nature Fund of China (No.61473222).

References

- [1] Big data. Baidu encyclopedia..http://wapbai.ki.baidu.com/subview/6954399/13647476.html?fr=Aladdin&ref=wise&tj=Xv_1_0_10_title
- [2] Victor Maier - Schoenberg, Kenneth. Kukeye. The age of big data[M]. Zhejiang people's Publishing House, 2013:1
- [3] Morgan. The box of K Charts theory [M]. Taipei: after the history subset press.2014, 6.
- [4] [4]Xu Qigang, Zhong Lu. Fast package algorithm designed and implemented. Microcomputer development 2002, 04, pp01-04.
- [5] Feng Xiaoqiong, The practice of Spacecraft products identification of state management ,quality and reliability 2013, 04, pp01-02.
- [6] Li Xiaomin,Guo Jing,Liu Junze.Research on the realization method and system for monitoringthe spacecraft state information management technology based on. China astronautics Society Committee of deep space exploration technologyacademic annual meeting. 2010
- [7] The measuring method of relative state of Spacecraft Based on IEKF image sequence analysis. Journal of the Academy of Equipment Command & Technology. 2004, 04, pp. 125-126
- [8] Cao Guorong , Research and application of self health management technology for spacecraft. National Defense Science and Technology University.2008
- [9] Qin Wei, Guo Yongfu,A on orbit satellite fault early warning system based on history of the telemetry data , Spacecraft engineering 2010, 06, pp.03.
- [10] Long Bing, Song Lihui, Jing Wuxing, Jiang Xingwei, Review and Prospect of fault diagnosis technology of spacecraft, Missiles and space vehicles2003, 03, pp. 05.