

A Method for Knowledge Mining of Satellite State Association

Jing Zhao*, Tianshe Yang, Ke Zhang, Xingmiao Liu, Na Fu and Xuhua Wang

Xi'an satellite control center, Xi'an, China, 710043

*Corresponding author

Abstract—This paper presents a new scheme for data mining in spacecraft state association and abnormal detection. A method which includes state association knowledge mining (SAKM) and Similar Density Merge Clustering (SDMC) is developed. Data from the satellite are the most critical thing for analyzing satellite state and abnormal detection, therefore correct analyze of the regulation is quintessential for the detection of satellite abnormal detection. The associations in any of the subsystems are known when satellite was designed; however, the associations between any two of the subsystems are to be found. In such cases a SAKM algorithm becomes essential for obtaining the correct regulation. Clustering algorithm such as K-means is important in the SAKM algorithm, although it has its own limitations; however in this paper a novel method has been proposed for conditioning of K-means method. So that it can be effectively used for regulation discovery of data and satellite abnormal detection. The SDMC approach involves determination of classes' centers from the classes established by K-means method. In this way the points in one class are close enough. The high similarity points fall into the same class while the low similarity points fall into the different class based on the SDMC algorithm, the SAKM algorithm use Apriori algorithm mining frequent item sets and association rules of parameter feature characters. The experimental results show that as compared to the K-means algorithm the SDMC method can effectively cluster the data, the SAKM algorithm can correctly mining the satellite association knowledge.

Keywords—spacecraft state association; state association knowledge mining (SAKM); Similar Density Merge Clustering (SDMC); spacecraft abnormal detection

I. INTRODUCTION

Satellite have been used for various purposes, such as remote sensing, television, telephony, data communication, etc. these important services cannot be affected or degraded by faults that may occur in components, actuators, and sensors. Hence, a spacecraft fault detection and identification play an important role in space field.

Spacecraft in orbit return a large number of data to the ground control center every day. Day and months multiplying, data and information are quite large. The analysis and feature extraction of satellite data is also difficult. The abnormal determine and fault diagnose, usually use human experience related to the possible data analysis to determine the change of satellite. In such cases research on the state association knowledge mining method and discovery the regulation among data will play the most significant part in correctly positioning the abnormal in spacecraft.

Current approaches to satellite association knowledge mining are based on the human experiences and require long-time testing with a great amount of data. Fault detection capability of the current strategies is also fairly weak. Motivated by these considerations, a new method based on K-means algorithm is proposed in for clustering the high similarity points to one class. The low similarity points will typically lead too far away.

In this paper an improved K-means method (SDMC) has been proposed the aim of which is to overcome the inherent limitations of K-means and then correctly discover the state change regulations of satellite. For this purpose, a large amount of data has been used for clustering. As in the entire in SDMC algorithm the data points are mapped into high dimension space using a kernel function, where it then adopts the Apriori algorithm method for mining the association knowledge and frequent pattern. the exact class centers are determined using the classes which is established before using history data in satellite.

II. SAKM ALGORITHM

First, extract the window feature of all the parameters, then use SDMC algorithm to cluster each parameter separately, and mark the cluster results with feature characters. The largest cluster, which contains the most of the window, is represented by 'a'. the second largest cluster is represented by 'b', and so on. But if the support of the cluster is less than a given threshold, it represented by '?'.for all the small probability of the cluster all use '?'represented. The null data that is not recorded in the range or interval is used '#'feature to express.

After completion of the clustering process, the sequence data of the parameters are converted to a characteristic character sequence. Apriori algorithm is then used to mine frequent item-sets and association rules in parameter feature strings. There are a lot of parameters in the satellite system, too many association regulations produced, and redundancy, this is not conducive to the judgment of experts. So the association regulations need to be reduced. In order to ensure the consistency of the knowledge, avoid circular reasoning, to detect whether there is a ring in the association rule, and finally eliminate the ring.

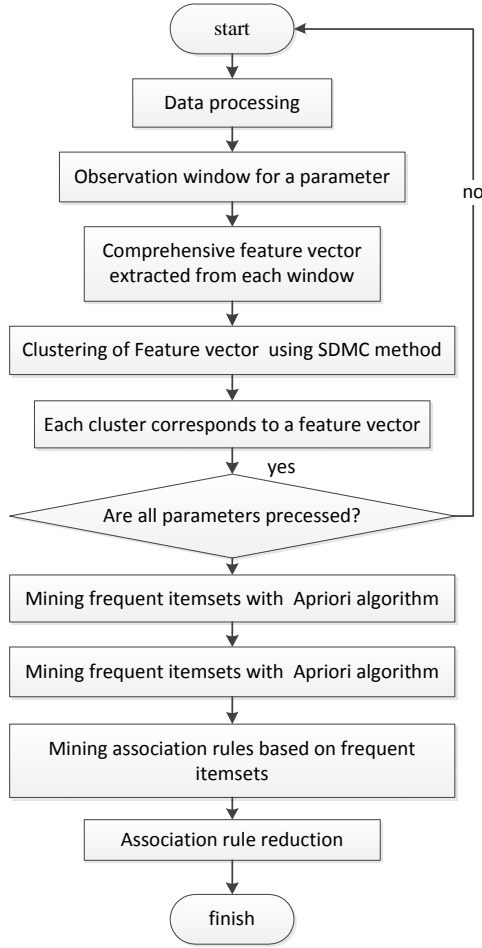


FIGURE I. SAKM ALGORITHM FLOW

III. STATE ASSOCIATION KNOWLEDGE MINING THEORY

The knowledge mining of the state association is actually the correlation between the different parameters of the state. For instance: if $A1=0$, $A2=6$, $A3=a$ have a high frequency, then there is an association among $A1$, $A2$, $A3$, and also an association among state values, that is $A1=0$, $A2=6$, at the same time $A3=a$.

The problem is converted into a frequent pattern of different lengths. There are some classical algorithms for mining frequent patterns and association rules, such as: Apriori and FP-Growth algorithm. Consider the big number of satellite system parameters, constructing FP tree on large data sets may face the problem of insufficient memory, use the Apriori algorithm mining frequent item sets and association regulation. Generally speaking, Apriori processing is transactional data, such as purchase records used in shopping basket analysis. But time series data is high dimensional continuous, need to be converted to a form similar to transactional data. So the continuous parameter should be chartered. Get the transaction data as shown in table I.

TABLE I. TRANSACTION DATA

Time	A1	A2	A3	A4	A5	A6
T1	0	7	a	b	c	a
T2	1	6	a	a	d	a
T3	0	8	b	b	c	b
.....

The problem has to be solved is 1. Character the number continued; 2. Mine frequent pattern and association regulation of different length.

In order to solve the problem 1, character use the cluster method, firstly, the parameter data of continuous value is divided into a fixed length window, comprehensive feature vector is extracted from each window:

[mean value, variance, wavelet coefficient, Fourier coefficient, Fourier frequency]

Feature vector is the basic unit of cluster. The result of cluster will produce several cluster, each cluster has expressed with different character, the points in one cluster have the same character with the cluster they in, then the original data will be chartered.

In order to solve the problem 2, use Apriori algorithm mining frequent item-sets and association regulation. Because of too many parameters of satellite system, redundancy rules will produced, also some similar redundancy, all of the rules should be simplified. In order to ensure the consistency of knowledge, avoid the circular reasoning, detect if there is the ring in the association rules, like $A \rightarrow B$, $B \rightarrow C$, $C \rightarrow A$.

IV. APRIORI ALGORITHMS

Apriori algorithms use the prior knowledge of frequent item-sets, and a method called iterative method of layer by layer search, k item-set used to calculate $k+1$ item-set. Firstly, scan transaction log, find out all the frequent 1 item-set, the set expressed as L_1 , then use L_1 to find the frequent 2 item-set, which is expressed as L_2 , use L_2 to find out L_3 , so until we can't find any more frequent k item-set. Finally, find out the association regulations in all the frequent item-sets which the user interested in.

The non-empty subset of any one of the frequent item-sets must be frequent, otherwise, if a candidate non-empty set is not frequent, then the candidate must not be frequent, so as to remove it from C_k .

The algorithm: Apriori use iterative method according to candidate item-sets to find out frequent item-set.

Input: transaction data base D ; the minimum support count threshold \min_sup .

Output: the frequent item-set L in D .

Technique:

Begin

```

 $L_1$  =find_frequent_1-itemsets(D); // find out frequent 1
item-set

For(k=2;  $L_{k-1}$  !=null;k++)
{
 $C_k$  =apriori_gen(  $L_{k-1}$  ); // produce candidate, reduce the
one
For each
{ // scan D and count the candidate
 $C_t$  =subset(  $C_k$ , t); // get the subset of t

For each candidate c belongs to  $C_t$ 
c.count++;
}

 $L_k$  ={c belongs to  $C_k$  | c.count>=min_sup}
}

Return  $L$ =all the frequent item-sets;

Procedureapriori_gen(  $L_{k-1}$ , min_sup)

For each item-set  $L_1$  belongs to  $L_{k-1}$ 

For each item-set  $L_2$  belongs to  $L_{k-1}$ 

If((  $L_1[1] = L_2[1]$  )&&(  $L_1[2] = L_2[2]$  )

&&(  $L_1[k-2] = L_2[k-2]$  )&&(  $L_1[k-1] = L_2[k-1]$  ))

then{
c=  $L_1$  connect  $L_2$  //produce candidate

if has_infrequent_subset(c,  $L_{k-1}$ ) then

delete c; //delete non frequent item-set candidate

else add c to  $C_k$  ;

}

Return  $C_k$  ;

Procedure has_infrequent_sub(c,  $L_{k-1}$  )

For each(k-1)-subset s of c

If s  $\notin L_{k-1}$  then

Return true;

Return false;

End

```

A. Generation of Frequent Item-sets

It should be point out that frequent item sets are a collection of all the items that support the degree of support is greater than the minimum support threshold. The support degree of one item-set is the ratio of times that this item-set appeared in all the record and the number of records.

$$\text{sup}(A_1, A_2, \dots, A_n) = \frac{\text{count}(A_1, A_2, \dots, A_n)}{\text{count}(\text{sum})}, \text{sup}(A_1, A_2, \dots,$$

$A_n)$ express the item-set support degree of N item-sets(A_1, A_2, \dots, A_n), $\text{count}(A_1, A_2, \dots, A_n)$ express the times candidate item-set(A_1, A_2, \dots, A_n) appeared in all the records, $\text{count}(\text{sum})$ express the number of all records.

Because the sampling interval of the parameters is the same, the time span of each observation window is also equal. In the time period of data monitoring, each window corresponds to a period of time, in the period of time, the transaction set can be obtained by all parameters, as shown in the Table I. Figure II take the above data as an example, explain the process of mining frequent item-sets using Apriori algorithm, which assumes the minimum support degree is 3/count (sum).

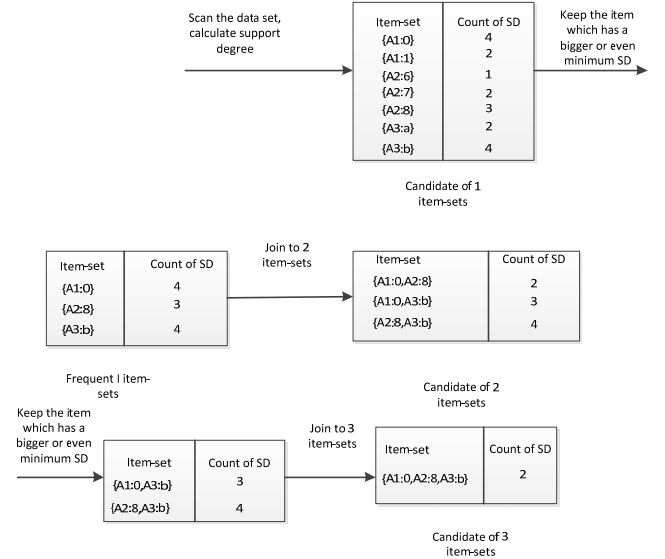


FIGURE II. MINING FREQUENT ITEM-SETS USING APRIORI ALGORITHM

B. Establish of Association Regulation

The frequent item-set is $L = R \cup S$. S is non-empty and proper subset of frequent item L , $S \subset L$, R is the different set which is the frequent item that remove the S , $R = L - S$. The association regulation is: $S \rightarrow R$, confidence of association

$$\text{rules is } \text{conf}(S \rightarrow R) = \frac{\text{sup}(L)}{\text{sup}(S)} = \frac{\text{sup}(R \cup S)}{\text{sup}(S)}.$$

Obviously, a frequent item can generate multiple association rules. Such as frequent item-sets (A, B, C) can produce rules like $A \rightarrow B, C$, $B \rightarrow A, C$, $A, B \rightarrow C$, the

number of association rules about A_3^2 . So we need to reduce association rules, eliminate redundancy.

C. Reduce Association Rules

1) *Reduction principle*: Principle 1: For every frequent K item, just keep the association rules which have the biggest confidence. Principle 2: if two association rules have the same first part, then compare the following part. If there is some inclusion relation in the following part, in the condition that difference between confidences is very small; delete the rules which the following part is small, keep the following parts which have a quantity rules. Principle 3: if association rules have the same following part, then compare the first part. If the first part have the inclusion relation, and the confidence differences between associations is a little, delete the rules which have a big first part, keep the rules which have a small first part. Principle 4: in order to ensure the consistency of knowledge, avoid circular reasoning, we need to detect whether there is a ring in association rules or not.

In order to detect inference ring, the directed acyclic graph should be used. We use directed acyclic graph to express rules set, one node express first part and one node express the following part, and the two nodes are connected by a directed edge. Detect the association rules according to confidence that is in descending order one by one.

2) *Step of reduction*: First, reduction according to principle 1, and sort the association rules by confidence degree in descending order; second, reduction according to principle 2; third, reduction according to principle 3; fourth, reduction according to principle 4.

A method of satellite state association knowledge mining is presented, practice has proved that it is a very effective method, according to the author's research and application of this method to a satellite battery capacity alarm data, through the original data of 13 days and 12 parameters, as shown in Figure III, mined 64 frequent 2 item-sets, 96 frequent 3 item-sets, 64 frequent 4 item-sets, 25 frequent 5 item-sets, produced 34 association rules by reduction. The Figure IV represent the result of the third association rule.

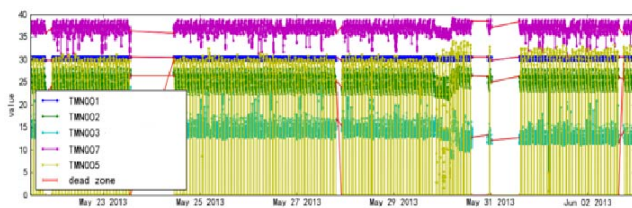


FIGURE III. FIVE PARAMETRIC CURVES OF SATELLITE BATTERY CAPACITY ALARM DATA

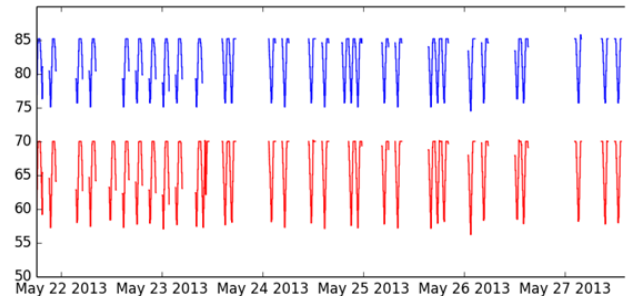


FIGURE IV. REPRESENTATION ACCORDING TO THE THIRD ASSOCIATION RULES

V. CONCLUSION

Data of satellite is large and complex, it is difficult to find the potential abnormal information contained in the human experience, this paper present a satellite state association knowledge mining method, research on data feature of normal state based on the discovery of frequent item-sets, the main content include window feature extraction of all parameters, SDMC clustering algorithm, production of frequent item-sets, association rules and reduction of association rules. Satellite is a complex system, running in the space environment there is a lot of uncertainty, it will break down at times, the method of satellite state association knowledge mining make the discovery of characteristics of data in normal state possible. The practice proves that this method can reveal the normal data more effectively, is conducive to early detection of abnormal changes in satellite, and support to the measurement and control of satellite.

REFERENCES

- [1] F Guil,R.Marin.Extracting Uncertain Temporal Relations from Mined Frequent Sequences. Proceedings of the Thirteenth International Symposium on Temporal Representation and Reasoning, pp. 152-159, 2006.
- [2] Frans Coenen,Paul Leng,Shakil Ahmed. Data Structure for Association Rule Mining:T-Trees and P-Trees. IEEE transactions on knowledge and data endingeeing . vol. 16, No.6, JUNE ,2004.
- [3] Yuqiang Shi, Qibin Liao, Hongxu Wang. Again discussion for definition of similarity measures between Vaguesets. Computerscience. Vol.39,No.12, PP.255-259, 2012.
- [4] Lai Ping Ho,Sung Nok Chiu. Using Weight Functions in Spatial Point Pattern Analysis with Application to Plant Ecology Data[J]. Communications in Statistics - Simulation and Computation . Vol.38,No.2, PP.269-287, 2009 .
- [5] Huisheng Zhu, Changxia Ma. Two improved algorithm for mining association rule. Computer Applicationsand software. Vol.27, No.8, PP.7-9, 2006.
- [6] Xingjie Feng, Zhun Zhou.Improvement for the Apriori algorithm. Computer Engineering. Vol.S1, No.31, PP.172-173, 2005.
- [7] Pei J, Han J,Mao R. Closet: An efficient algorithm for mining frequent closed itemsets. SIGMOD Int'l Workshop on Data Mining and Knoeledge Discovery(DMKD'00), Dallas, TX, PP.11-20, May,2000
- [8] G Grahne, L Lakshmanan, X. Wang. Efficient mining of constrained correlated sets. Proc. 2000 Int. Conf. Data Engineering (ICDE'00) . San Diego, CA, PP.512-521, Feb.2000.
- [9] Lal K ,Mahanti N C. A novel data mining algorithm for semantic Web based data cloud. International Journal of Computer Science and Security(IJCSS) . Vol.4, No.2, PP.160-175,2010.