

Analysis of space payload operation modes based on divide-and-conquer clustering

Feng Si^{1,2,a}, Bao Jun Lin¹ and Shan Cong Zhang³

¹Academy of Opto-Electronics, Chinese Academy of Sciences, Beijing, China;

²University of Chinese Academy of Sciences, Beijing, China;

³Technology and Engineering Center for Space Utilization, Chinese Academy of Sciences, Beijing, China

Abstract. With the development of space electronic technology, the space payload operation modes are more and more complex, and manual interpretation is prone to errors for much workload. Generally the space payload's operation modes are reflected by its telemetry data. By analysing the characteristics of the payload telemetry data, it is proposed an automatic analysis method of payload operation modes based on divide-and-conquer clustering. The clustering method combines division and incremental clustering. The principle of the method is introduced and the method is validated using the actual payload telemetry data. Furthermore the improved method is proposed to the problems encountered. Experimental results show that divide-and-conquer clustering method has the feature of calculation simple and classification accurate, when applied to the classification of payload operation modes. Furthermore this method can be applied to the other areas of payload data processing by extending the method.

1 Introduction

The operation modes of space payload are much more complex, as the system size becomes larger [1]. It can be identified by the telemetry data related to the payload in orbit. The mass telemetry data will be produced every day and has large differences between different kinds of payloads in general. Traditional methods either depend on the manual judgment, or the rules according to the prior knowledge of the designer. The rules are more complex to design in the transient between operation modes. Furthermore the rules designed for specific payloads are often difficult to expand to other payloads. In order to reduce the error rate of manual interpretation and complexity of rule-based method, we present a novel method based on data mining technology to find system structure hidden in telemetry data and the operation modes of payload automatically.

There are many kinds of technology in data mining field [2-3]. According to the characteristics of the payload telemetry data, we choose data clustering method to identify the payload operation modes. Data clustering [4-7] belongs to the category of unsupervised learning, which is a discovery process that groups data set such that the intracluster similarity is maximized and the intercluster similarity is minimized. In the absence of any prior knowledge, we don't know the payloads how to work and how many operation modes do they have. In this case, the method of data clustering can help us to find the patterns in the underlying telemetry data.

In this paper, we present a novel clustering algorithm called divide-and-conquer clustering that applies group method and incremental clustering to classify payload telemetry data [8-11]. We demonstrate the effectiveness of divide-and-conquer algorithm in huge actual telemetry data of payloads. Experimental results show that this algorithm can discover the payload operation modes simply and effectively. Finally, we apply this algorithm to other fields of payload data processing.

2 Divide-and-Conquer Algorithm

According to the payload's prior work rule in orbit, much more telemetry data was generated every day (some greater than one hundred thousand records). There will be great amount of calculation if we do data clustering using one day's records directly. In order to solve this problem, we adapt the divide-and-conquer clustering algorithm. Firstly, the telemetry data will be grouped in order and then we calculate the similarity of the records in the group, complete the clustering intra-group. Secondly, within groups after clustering, eigenvector is calculated for each category. Thirdly, the similarity between groups will be calculated using eigenvectors of each category. The process will be iterative until all records are classified.

The process of clustering is as follows:

(1) Data pre-processing

Scan all the parameters in the data set and find out the characteristics of each type of data parameters, such as the parameter's data range, discrete or continuous values,

^a Corresponding author: sifeng04@126.com

or variation of the parameter. For the voltage telemetry, temperature telemetry, or some count values of payload, it can be applied to calculate directly. But for some state variables, the numerical value only represents a different state and these state variables can't be applied for calculation directly. These data should be pre-processed for calculation [12-13].

(2) Normalization

All data will be normalized using the following formula and they will be limited in the range of [0, 1] after normalization.

$$x'_{ik} = \frac{x_{ik} - \min_{1 \leq i \leq n} \{x_{ik}\}}{\max_{1 \leq i \leq n} \{x_{ik}\} - \min_{1 \leq i \leq n} \{x_{ik}\}} \quad (k = 1, 2, \dots, m) \quad (1)$$

(3) Grouping

All the records in the data set will be grouped. In general, the space payload operation modes are continuous and consistent unless in fault condition. The process of grouping can be operated in the sequence of data in the data set.

(4) Clustering intra-group

Calculate the similarity between any two records in the group. We'll cluster the records according to the similarity results. After clustering intra-group, we'll calculate the eigenvector of each category, which represents the category's centroid. The similarity between two records can be analysed in the fractional distance metrics [14]. The formula is as follows:

$$dist_d^f(x, y) = \sum_{i=1}^d [(x_i - y_i)^f]^{1/f}, \quad f \in (0, 1) \quad (2)$$

(5) Clustering inter-group

After clustering in the groups, we'll cluster between groups using the selected eigenvectors of each category. It may be grouped continually if we have many more categories after clustering inter-group. Cluster until the terminal conditions.

(6) Evaluation of clustering result

In order to evaluate the performance of the classification, many other methods may be applied for contrast, such as manual inspection of payload operation modes, rule-based method etc.

3 Experimental Results and Discussion

3.1 The original clustering

In this section, we validate our divide-and-conquer algorithm with an actual telemetry data set of one earth observation payload. The data set contains several thousand telemetry records and we choose 300 records for test. Each record includes 206 parameters. Then we'll do the data set clustering in the algorithm above.

We do data pre-processing, normalization and divide the data records into three groups and each group has 100 records. Then we calculate the similarity of records between each other in each group. According to the similarity result, we can find records with the highest similarity to each record. But we can't classify the records only by the similarity result. Figure 1 shows the

first group's 100 records and their partner records with highest similarity.

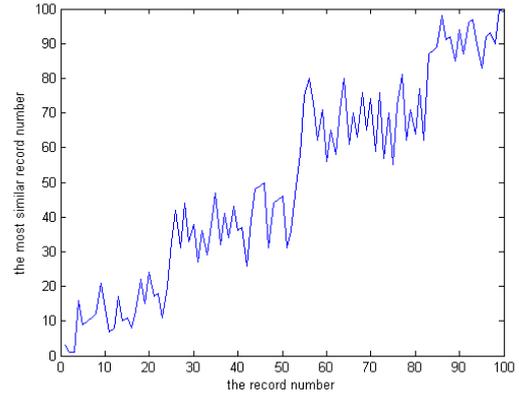


Figure 1. The record and its partner record with highest similarity

3.2 Analysis of clustering failure

By analysis the original records in the first group, we have found two main reasons for clustering failure.

The first reason is disturbance of some irrelevant parameters. Searching all parameters in the record manually, we can find some parameters which have drastic changes between different records. But the drastic changes don't reflect the system operation modes in macroscopic view. For example, the payload heartbeat signal, reversed once between adjacent records, reflects only the normal state not the operation modes change.

The second reason is parameter multiplexing problem. Some parameters are multiplexed in system design, and each cycle represent different physical meanings. Sometimes there is great numerical difference in different cycle, but this difference can't reflect the switch of system mode.

3.3 Modified clustering method

In order to solve the above problems, we'll take the following measures. Firstly, we'll apply weighted parameters instead of original parameters in similarity calculation. According to the characteristics of the parameters distribution, we'll dispatch different weights of parameters. Secondly, we can avoid the parameter multiplexing problem in system planning and design phase. Alternatively we'll use parametric subdivision for the parameters of multiplexing in the original system architecture. The subdivision parameters will reflect the system operation modes.

By taking the above two measures, we re-calculate the similarity of 100 records in the first group. The similarity calculation results show that the similarity of the same mode records is very high and mostly greater than 0.999. By manual analysis of 100 records in the first group, the group can be divided into four clusters, the 1~24 records for the compression-and-storage mode, the 25~50 records for the compression system power-off mode, the 51~82 records for the transmit system power-on mode, the 83~100 records for the system idle mode. The calculation result of modified clustering algorithm is consistent with

manual analysis. Figure 2 shows the first group's 100 records and their partner records with highest similarity.

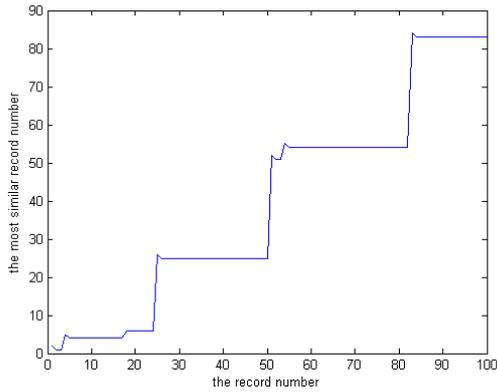


Figure 2. The record and its partner record with highest similarity (using modified clustering algorithm).

Following the same algorithm, we get the other two group records' intra-group similarity. Figure 3 shows the three group records' intra-group similarity. In the figure, we can see that the first group is divided into four clusters, the second group has two clusters and the third group has three clusters.

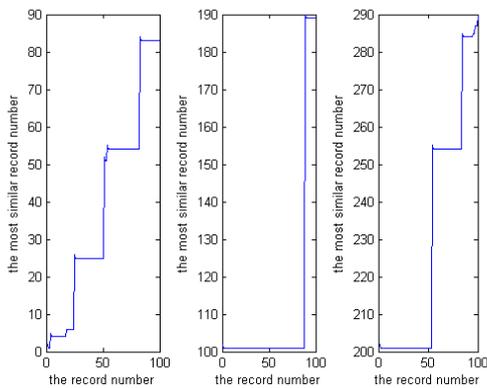


Figure 3. Clustering of every group

After completing the intra-group clustering, we'll calculate the centroid of each category. Using the selected centroid of each category, we'll calculate the similarity between categories. The result shows that the fourth category and fifth category can be merged into one cluster, sixth category and seventh category can be merged into one category. Thus the original nine categories can be combined into seven categories. Figure 4 shows the result of second round calculation.

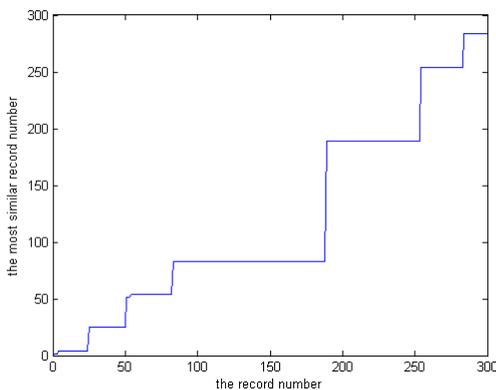


Figure 4. Clustering of three groups

Contrast to the result of the clustering algorithm, the result of manual analysis is basically the same. The following table shows the results obtained by manual analysis.

Table 1. The Result of Manual Analysis

Clustering Number	Record Number	Operation modes
1	1~3	Data Download from Channel 1
2	4~24	Channel 1 Close
3	25~50	Compression Stop
4	51~53	Compression Power Off
5	54~82	Transmission Close
6	83~188	Camera 1 Power Off
7	189~253	Compression and Storage
8	254~283	Idle
9	284~300	Data Download from Channel 2

When the clustering algorithm above is applied to the rest of the data set, we can also obtain a good classification and get the accurate payload operation modes.

3.4 Discussion on how to divide data set

In the above chapter, we have discovered that the payload operation modes can be automatically analysed using divide-and-conquer clustering to the payload telemetry data. But how to choose the best dividing method is still a problem. We'll discuss the best group method below.

We'll take into account the processing time and the clustering accuracy as the evaluation criterion. Using the data set above, we get the result of different dividing method. Figure 5 shows the calculation result. In the figure, the relative processing time is calculated by the ratio of the spending time of all dividing methods and the un-dividing method. The clustering accuracy is got by the ratio of successful clustering records and all the records.

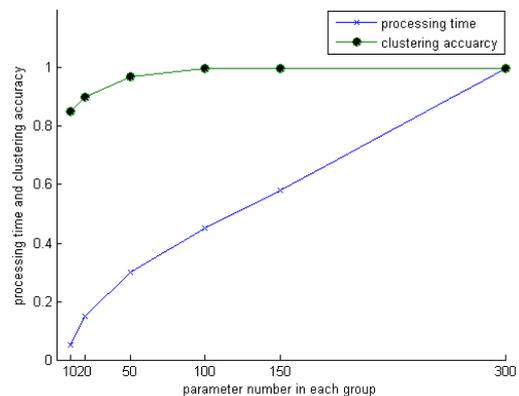


Figure 5. Relative processing time and clustering accuracy

From the figure, we can see that with the increasing records number in each group, the processing time is dramatically increased, but with the more clustering accuracy. Furthermore, the increase of the processing time is synchronous with increase of the records number in each group. But the clustering accuracy is no longer

increased when the clustering accuracy reaches a balance point. So we can get the transition point by the balance of the processing time and clustering accuracy. From figure 5, we can see that when the data set is divided into 3 groups and each group includes 100 records, the clustering accuracy will reach 100% and the processing time is only half of the un-divided method. It's a better choice for the payload operation modes clustering.

4 Conclusions

In this paper, we have presented a divide-and-conquer clustering algorithm for analysing the payload operation modes. The algorithm takes into account the group method, weighted parameter and parametric subdivision, which successfully solves the mass data, disturbance parameter and parameter multiplexing problem. Experimental results on actual telemetry data set show that the modified algorithm can successfully identify the payload's operation modes automatically. The method may be extended to the real-time processing of payload by creating the payload characteristics library. Furthermore when we have acquired all the operation modes and characteristics of payload, we can do some fault detection and fault diagnosis for the payload.

References

1. Y.F. He, G.H. Zhao, C.M. Lv, Journal of University of Chinese Academy of Science, **28**, 4(2011)
2. J. Han, M. Kamber, *Data mining: concepts and techniques*, 187~198(2006)
3. A.K. Jain, M.N. Murty, P.J. Flynn, ACM Computing Surveys, **31**, 3(1999)
4. P. Hansen, B. Jaumard, Math. Program., **79**, 1-3(1997)
5. A.K. Jain, R.C. Dubes, Englewood Cliffs Prentice Hall, **32**, 2(1988)
6. L. Kaufman, P. Rousseeuw, *Finding groups in data: An introduction to cluster analysis*, 1990
7. C.H. Cheng, IEEE Transactions on computers, **24**, 9(1975)
8. M.N. Murty, G. Krishna, Pattern Recognition, **12**(1980)
9. W.F. Eddy, A. Mockus, S. Oue, Computational Statistics & Data Analysis, **23**, 1(1996)
10. G.A. Carpenter, S. Grossberg, Neur Netw, **3**, 2(1990)
11. S. Asharaf, M.N. Murty, Pattern Recognition, **36**, 12(2003)
12. H.L. Chen, K.T. Chuang, M.S. Chen, ICDM, 2005
13. H.L. Chen, M.S. Chen, S.C. Lin, IEEE Transaction Knowledge and Data Engineering, **21**, 5(2009)
14. G. Karypis, E.H. Han, V. Kumar, IEEE Computer, **32**, 8(1999)