



Journal of
**Software
Engineering**

ISSN 1819-4311



Academic
Journals Inc.

www.academicjournals.com



Research Article

A Modified Feature Weighted Clustering Method for Space Payload Operational Mode Analysis

^{1,2}Feng Si, ¹Baojun Lin and ³Shancong Zhang

¹Academy of Opto-Electronics, Chinese Academy of Sciences, 100094 Beijing, China

²University of Chinese Academy of Sciences, 100049 Beijing, China

³Technology and Engineering Center for Space Utilization, Chinese Academy of Sciences, 100094 Beijing, China

Abstract

Background: With the increase in aircraft platform carrying capacity and the development of space information technology, payload operational mode is becoming increasingly complex. In the period of system test and operation control management in orbit, manual interpretation is prone to errors for much of the workload. **Methodology:** By analyzing the characteristics of the payload telemetry data, it's proposed an automatic analysis method for payload operational mode based on improved feature weighted clustering. The method is based on the classical feature weighted algorithm Relief, which is modified according to the characteristics of payload telemetry data. The fractional distance metric is chosen as distance metric instead of the Euclidean distance to calculate the distance between records. In the feature weight value calculation, the initial clustering information is adopted instead of only the record. Furthermore, all the categories are used to calculate the different categorie's weight values instead of relying on sampling in categories. **Results:** The experimental results show that this method has the characteristics of a shorter calculation time, high real-time processing performance, strong autonomy and accurate classification it is when applied to payload operational mode classification. **Conclusion:** Furthermore, the method may be applied to other areas of space payload processing.

Key words: Payload, operational mode, feature weighting, clustering, fractional distance metric, feature selection

Received: April 05, 2016

Accepted: May 10, 2016

Published: December 15, 2016

Citation: Feng Si, Baojun Lin and Shancong Zhang, 2017. A Modified feature weighted clustering method for space payload operational mode analysis. J. Software Eng., 11: 78-85.

Corresponding Author: Feng Si, Academy of Opto-Electronics, Chinese Academy of Sciences, 100094 Beijing, China Tel: +8613811820378

Copyright: © 2017 Feng Si *et al.* This is an open access article distributed under the terms of the creative commons attribution License, which permits unrestricted use, distribution and reproduction in any medium, provided the original author and source are credited.

Competing Interest: The authors have declared that no competing interest exists.

Data Availability: All relevant data are within the paper and its supporting information files.

INTRODUCTION

Space payload has an increasing variety of operational mode as the system size becomes larger and the operational mode become increasingly complex¹. The operational mode of the payload is usually inferred by the relevant telemetry data. During the in-orbit operation of the payload, a large amount of telemetry data is generated every day and the data are significantly different for different types of payload. In previous space engineering projects, the payload operational mode is generally analyzed manually or according to the designer's prior knowledge of the design rules². If the payload has many operational modes, the design of the rules is more complex. In addition, different rules should be designed according to the different characteristics of the payload. The reusability of the decision rules is low. To reduce the error rate of manual interpretation and complexity of the rule-based method, a novel method is presented based on data mining technology to automatically find the system structure hidden in telemetry data and the operational mode of payload.

Data mining^{3,4} is the computational process of discovering patterns in large data sets involving methods at the intersection of artificial intelligence, machine learning, statistics and database systems. The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use. The actual data mining task is the automatic or semi-automatic analysis of large amounts of data to extract previously unknown, interesting patterns, such as groups of data records (cluster analysis), unusual records (anomaly detection) and dependencies (association rule mining).

Data mining involves the use of many technologies. Clustering analysis methods are chosen to perform pattern recognition according to the characteristics of the payload telemetry data. Clustering^{5,6} is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar (in some sense or another) to each other than they are to those in other groups. Clustering is a primary task of exploratory data mining and is a common technique for statistical data analysis. Clustering can be achieved using various algorithms, which differ significantly in their notion of what constitutes a cluster and how to efficiently find them. Popular notions of clusters include groups with small distances among the cluster members, dense areas of the data space, intervals or particular statistical distributions. Clustering can therefore be formulated as a multi-objective optimization problem. The appropriate clustering algorithm and parameter settings (including values such as the distance function to use, a

density threshold or the number of expected clusters) depend on the individual data set and intended use of the results.

Clustering belongs to the category of unsupervised learning, which does not know the search target and only obtains the common features of data from the algorithm. In the absence of any prior knowledge of the payload, it's hard to know how the payload operates or how many operational modes it has. In this case, clustering analysis methods can be used to find the clustering feature of the telemetry data.

By analyzing the characteristics of the payload telemetry data, a method is proposed based on feature weighted clustering for automatic analysis of the payload operational mode. Feature selection is the process of identifying the most effective subset of the original features to use in clustering. The method is based on the original feature weighted algorithm Relief⁷. The algorithm Relief is a very efficient algorithm for estimating features. The key idea of Relief is to estimate features according to how well their values distinguish the records that are near each other. For that purpose, Relief for a given record searches for its two nearest neighbors: One from the same class (called nearest hit) and the other from a different class (called nearest miss). The rationale is that a good feature should differentiate among records from different classes and should have the same value for records from the same class. Original Relief can identify discrete and continuous features and is limited to only two-class problems. In addition, original Relief cannot be extended to process incomplete data or address problems of more than two classes. In this study, the algorithm is modified according to the characteristics of payload telemetry data. In addition, the algorithm is verified by the synthetic data and the actual telemetry data of a payload in orbit. The experimental results show that the algorithm can obviously reduce the computation time compared with other methods without loss of clustering accuracy. Thus, the algorithm can be applied to a real-time application in orbit. Furthermore, the algorithm may be applied to other areas of payload data processing.

MATERIALS AND METHODS

Characteristics of payload telemetry data: According to the operation rules of the payload in orbit, a large amount of telemetry data is produced every day (some payloads have more than one hundred thousand records). Furthermore, for different payloads, many different types of telemetry data exist. In general, the telemetry data can be divided into numerical and categorical data. The numerical data contain the voltage value, current value, temperature value and sequence number. The categorical data include the Boolean value, work state and periodical value.

In addition to the massive quantity and many types of data, the payload telemetry data have the characteristics of a high number of dimensions. A telemetry data record usually has hundreds or even thousands of features and the domain for each feature can be very large. It is not meaningful to seek clusters in such a high dimensional space because the average density of points anywhere in the data space is likely to be quite low. Compounding this problem, many dimensions or combinations of dimensions can have noise or values that are uniformly distributed. Therefore, distance functions that use all of the dimensions of the data may be ineffective. Moreover, several clusters may exist in different subspaces that are composed of different combinations of features.

The problem of high dimensionality is often addressed by requiring the user to specify the subspace (a subset of the dimensions) for cluster analysis. However, the user identification of subspaces is quite error-prone. There are two methods to solve the problem of high-dimensional data clustering^{8,9} feature conversion and feature selection. Feature conversion generally uses principal component analysis or a singular value decomposition algorithm to reduce the number of dimensions and find clusters in a low-dimensional space. The new space has the property that the distances between points remain approximately the same as before. Feature selection generally selects effective features to form a related subspace and cluster in the subspace¹⁰⁻¹². The Relief algorithm is a classical algorithm in the feature selection field, which is a feature subset selection algorithm for weight search. Weight search method gives a weight to each feature and the weighted value reflects the correlation between the feature and the clustering category. In this study, the classical Relief algorithm has been modified according to the characteristics of payload telemetry data.

Background of the relief algorithms: The Relief algorithm is a feature weighted algorithm that was proposed by Kira and Rendell⁷. The algorithm randomly selects a record x_i in the data set and determines the nearest neighbors in the same class and in the different classes, which are represented by Hit and Miss, respectively.

Suppose $U = \{x_1, x_2, \dots, x_n\}$ is the data set for clustering, which includes n records and each record includes m feature parameters:

$$x_i = \{x_{i1}, x_{i2}, \dots, x_{im}\} \quad (i = 1, 2, \dots, n) \quad (1)$$

The difference between two records in the feature t (taking the numerical data as an example) is defined as:

$$\text{diff}(t, x_i, x_j) = \left| \frac{x_{it} - x_{jt}}{\max_t - \min_t} \right| \quad (2)$$

where, \max_t and \min_t are the maximum and minimum values, respectively, of the feature t in the data set.

In the process of clustering, the weight value of the feature t is updated using the following formula:

$$w_t = w_t - \text{diff}(t, x_i, \text{Hit}) + \text{diff}(t, x_i, \text{Miss}) \quad (3)$$

The theoretical basis of the algorithm is the information theory to evaluate the feature, in which the mutual information between the correlation feature and the category should be high. In the iterative process, the strongly distinguished features should correspond to larger differences in the different categories and smaller differences in the same categories. After many rounds of iteration, the weight values of the features that are close to the cluster are gradually increased; however, the weight values of features that are irrelevant to the clustering are decreased, even to zero.

The Relief algorithm can only address two-class classification problems. Kononenko¹³ proposed the feature weighted algorithm ReliefF to address the multi-class classification problem on the basis of the work of Kira. The algorithm divided the multi-class classification problem into a number of two-class classification problems. The ReliefF algorithm selects d records of the nearest neighbors of x_i from each category of the data sets. The nearest neighbors in the same class with x_i are marked as $h_j, j = 1, 2, \dots, d$. The difference between the nearest neighbors is defined as follows:

$$\text{diff_hit} = \frac{1}{d} \sum_{j=1}^d \frac{|x_i - h_j|}{\max(X) - \min(X)} \quad (4)$$

The nearest neighbors in the different classes with x_i are marked as $m_j, j = 1, 2, \dots, d$. The difference between nearest neighbors is defined as follows:

$$\text{diff_miss} = \frac{1}{d} \sum_{l \in \text{class}(x_i)} \frac{P(l)}{1 - P(\text{class}(x_i))} \sum_{j=1}^d \frac{|x_i - m_j|}{\max(X) - \min(X)} \quad (5)$$

In the above formula, $P(l)$ is the probability of occurrence of the l class.

The feature weight values are updated using the following formula:

$$w_t = w_t - \text{diff_hit} + \text{diff_miss} \quad (6)$$

Because the ReliefF algorithm requires calculating the difference between each record and any other records in different classes, the amount of calculation required is very large when the data set contains massive numbers of records. Ye *et al.*¹⁴ proposed a modified feature weighted algorithm, called Multi-Relief, which is based on multiple sampling of Relief. The algorithm randomly selects two categories from multiple categories and then selects a fraction of the features from the two sample categories to form the training set that is used to calculate the feature weight. The calculation results of the multiple feature weights are obtained using the iteration process. The weight value of each calculation is fused with the following formula to obtain the final weight value of every feature:

$$w_t = \begin{cases} \frac{1}{N^+} \sum_e \{W_{-e}(t) > 0, \forall e\}, & N^+ > 0; \\ \frac{1}{N^-} \sum_e \{W_{-e}(t) < 0, \forall e\}, & N^+ = 0 \wedge N^- > 0 \\ 0, & N^+ = 0 \wedge N^- = 0; \end{cases} \quad (7)$$

Where:

$$N^+ = \{W_{-e}(t) > 0, \forall e\}, N^- = \{W_{-e}(t) < 0, \forall e\}$$

In the definition of w_t , only those runs for which Relief assigned a positive weight to feature t are considered. In this manner, w_t assigns a high score to position t only if it discriminates at least two classes. In particular, a maximum weight will be assigned if t fully discriminates two specific classes but does not differentiate (i.e., weight less than or equal to zero) any other pair of classes.

Improved feature weighted clustering algorithm: In this study, an improved feature weighted clustering algorithm has been proposed referring to the ReliefF algorithm and multi-Relief algorithm, according to the characteristics of payload telemetry data. The improved feature weighted clustering algorithm is as follows:

- The Euclidean distance is usually chosen as the distance measure. However, the Euclidean distance is not suitable for the distance between high-dimensional data. The fractional distance measure is much better¹⁵. In this

study, the fractional distance metric has been chosen as the distance metric instead of the Euclidean distance the formula to measure the distance is defined as follows:

$$\text{dist}_d^f(x, y) = \sum_{i=1}^d [(x_i - y_i)^f]^{1/f}, f \in (0, 1) \quad (8)$$

- The ReliefF algorithm calculates the feature weight at the base of each record without considering the initial clustering, which leads to a massive calculation for feature selection when the data set contains many records. In this paper, not all of the records are selected to use in the feature weight calculation. For large categories with many records, the nearest neighbors of the category center are chosen to participate in the calculation. However, for small categories with very few records, all of the records are chosen in the category. The result is to maintain the balance among the different categories in the calculation process. The pseudo-code of the modified ReliefF algorithm is as follows:

Input: The data set U for clustering;
 Output: The weight value w_t of each parameter;
 1. set all weights $w_t = 0$;
 2. for $i = 1$ to iter_count
 3. select a record x_i from the neighbor of cluster center;
 4. find k nearest hits h_j ;
 5. for each class $C \neq \text{class}(x_i)$
 6. from class C find k nearest misses m_{ij} ;
 7. for $t = 1$ to m
 8. $w_t = w_t - \text{diff_hit} + \text{diff_miss}$;
 9. end;

The Relief algorithm, which is not limited to two class problems, is more robust and can process incomplete and noisy data. However, the Relief algorithm has high computational complexity. Different from the ReliefF algorithm, the modified algorithm randomly selects a record x_i from the neighbor of the cluster center and then searches for both k of its nearest neighbors from the same class (called nearest hits h_j) and k of its nearest neighbors from each of the different classes (called nearest misses m_{ij}). The algorithm updates the quality estimation w_t for all features, depending on their values for x_i , hits h_j and misses w_{jit} . The update formula is similar to that of Relief, except that the contributions of all of the hits and misses are averaged. The contributions of hits and misses should be in $[0, 1]$ in each step and should be symmetric and the misses' probability weights should be summed to 1. The process is repeated for iter_count times. Selection of the hits and misses is the basic difference between ReliefF and Relief and ensures greater robustness of the algorithm concerning noise.

- In multi-Relief algorithm, some records are sampled in the calculation of the different categories' weight values. The categories with larger records are sampled at a higher probability and those with fewer records are sampled at a lower probability and may even be abandoned. Without considering the contribution of small categories, the results are not accurate. According to the payload operation in orbit, the operational mode occurs only a few times every day, at most several dozen times. However, there are operational modes with relatively fewer records. All the categories are chosen to calculate the different categories' weight value. Because the number of categories is not so large, the amount of calculation may not be increased significantly

RESULTS

A combination of synthetic and real datasets for one space payload in orbit has been chosen to test the accuracy and efficiency of the modified algorithm discussed in this study. In the synthetic datasets, the performance of the modified algorithm has been evaluated by controlling the number of data points, the dimensionality and the number of clusters. Because synthetic datasets are usually quite different from real datasets, the modified algorithm has also been compared with other clustering algorithms to test its performance.

Test results of the synthetic datasets: The data points of each synthetic dataset follow a series of normal distributions. The size of the datasets ranges from 10-100 K points, the dimension ranges from 10-100 and the number of clusters ranges from 5-20. To reflect the evolution of the data, the mean and variance of the normal distribution have been changed in each 10 K points.

First, the modified feature weighted algorithm has been compared with the K-means algorithm based on Euclidean distance metrics. The clustering accuracy has been chosen as the evaluation criteria, which is the ratio of the number of correctly clustered points to the total number of points in the datasets. The parameter f of the fractional distance metrics (formula 8) is chosen as 0.5. The test results are as follows:

- Figure 1 shows that the two algorithms both achieve good clustering accuracy, up to 98%, when the datasets have low dimensions. With an increased number of dimensions, the clustering accuracy of the modified algorithm discussed in this study, which used the fractional distance metrics, decreases only slightly. However, the clustering accuracy of the K-means algorithm, which uses the Euclidean distance metrics,

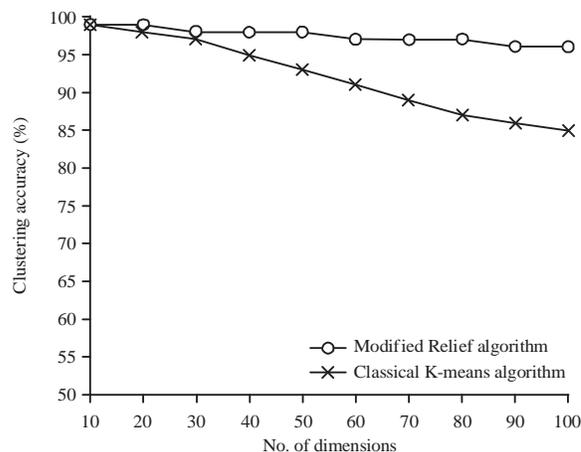


Fig. 1: Comparison of the clustering accuracy

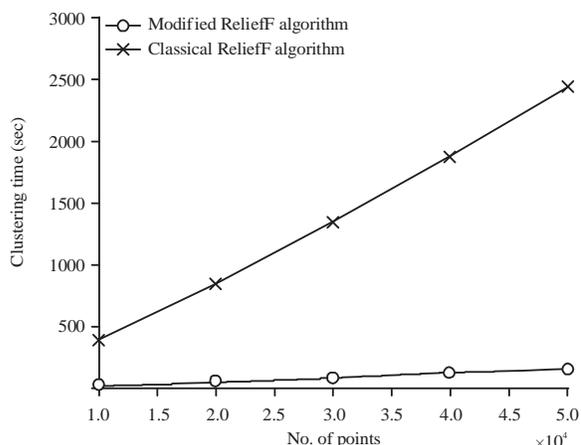


Fig. 2: Comparison of the clustering process time

decreases very quickly from 98-85%. It can be concluded that the fractional distance metrics can achieve better clustering accuracy for high-dimensional data clustering than can the Euclidean distance metrics

- Second, the modified feature weighted algorithm has been compared with classical Relief algorithm in process time. When the initial clustering is created, the modified algorithm uses only the nearest neighbor points of the clustering center for feature weight calculation. The range of the nearest neighbor is chosen as 10% of the number of clustering data points. If there are fewer than 10 chosen points, then all the points will be selected for the calculation. The test results are shown below
- Figure 2 indicates that the process time spent in the modified Relief algorithm is less than that of the classical Relief algorithm

Test results of real datasets: A data set of actual telemetry data of one space payload in orbit has been chosen for the

Table 1: Clustering results of different distance metrics

	Euclidean distance	Fractional distance
Category quantity	2	4
Clustering accuracy (%)	33	67
Processing time (sec)	15	16

Table 2: Feature weight values of the parameters in multiple iterations

No. of parameter	Weight value after the first iteration	Weight value after the second iteration	Weight value after the third iteration
19	0.80	0.90	0.90
21	0.82	0.91	0.91
96	0.15	0.60	0.60
98	0.18	0.66	0.68
174	0.61	0.75	0.76
176	0.60	0.70	0.70

test. The data set includes 45256 records and each record contains 206 parameters. In the dataset 100 records have been chosen for the calculation and obtain the comparison results between the improved algorithm and other algorithms. Some useful conclusions are obtained via clustering analysis.

First, in the process of initial clustering, two distance metric methods are applied to the data set. The results of clustering show that the fractional distance metric is better than the traditional Euclidean distance for high-dimensional data. The results are shown in Table 1.

When the 100 records are analyzed manually, the result is that the two categories identified by Euclidean distance method are compression storage mode (Records 1~24) and idle mode (Records 25~100); the four categories identified by the fractional distance method are compression storage mode (Records 1~24), compression stops mode (Records 25~50), transmission mode (Records 51~82) and idle mode (Records 83~100). From the Table 1, it is concluded that the fractional distance method has a higher degree of recognition than does the Euclidean distance method for high-dimensional data without a significant increase in processing time.

Second, using the result of the original clustering, all of the feature weight values can be obtained in multi-Relief method. Then the dataset is clustered again using the weighted parameters. After several iterations, the payload operational mode can be obtained.

The Multi-Relief algorithm selects the categories randomly, selects the records randomly and then calculates the feature weight. The experiment shows that this algorithm may miss the small categories. The clustering result of the Multi-Relief algorithm is the same as that of the fractional distance method, which found four categories. In the improved method, all the categories are adopted in the calculation rather than sampling randomly when the

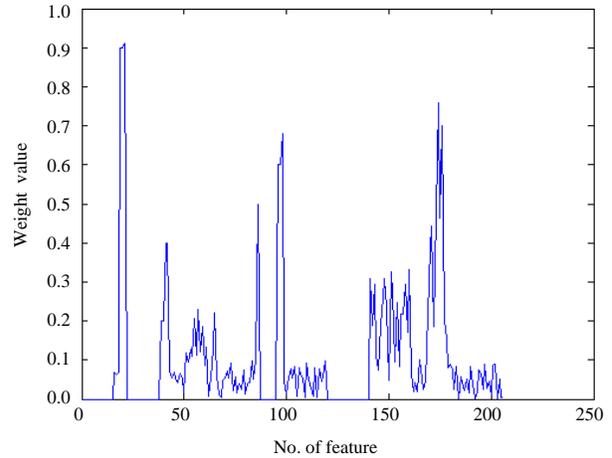


Fig. 3: Feature weight values of the data set

difference calculation is performed. In this way, another two categories are found that the Multi-Relief algorithm cannot identify. The improved algorithm identifies records 1~3 as the idle mode, records 4~24 as the compression storage mode, records 25~50 as the compression power down mode, records 51~53 as the power down mode, records 54~82 as the transmission mode and records 83~100 as the idle mode. Because of the excessive number of parameters, Table 2 only lists the feature weight values that have higher values than the others (the results have been normalized).

The two categories identified by the improved algorithm reflect the payload's transmission state, which is closely related to parameters 96 and 98 in Table 2. In Table 2, parameters 19 and 21 are closely related to the payload's compression state and parameters 174 and 176 are closely related to the payload's storage state.

From Table 2, it's discovered that all of the feature weight values are gradually stable after three iterations. The distribution of the feature weight values is shown in Fig. 3. Figure 3 shows that the feature weight values of several segments are zero. After analyzing the original data, it's found that the parameters are voltage or temperature values and do not exhibit any change in the selected data set. Thus, they do not have any contribution to the clustering and their weight value is zero. If some threshold is established for the feature weight value (e.g., 0.1) and the parameters do not involve a calculation that is less than the threshold, then the processing time can be further reduced.

Third, if the weight values of the parameters are calculated according to the ReliefF algorithm strictly, accurate results can be obtained. However, the amount of calculation is very large when the record size is massive, thus making it

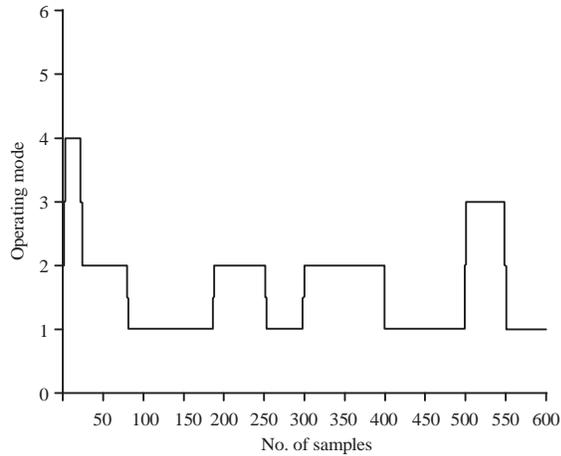


Fig. 4: Variation curve of the payload operational mode

Table 3: Comparison of two algorithms in processing time and clustering accuracy

	Processing time (sec)	Clustering accuracy (%)
Classical ReliefF algorithm	420	100
Modified ReliefF algorithm	32	99

impossible to obtain final results. In this study, the nearest neighbors of the clustering center are selected instead of all of the records to achieve a calculation that clearly reduces the processing time. In the data set 10,000 records are randomly selected for clustering analysis and obtain the comparison results of the processing time and the clustering accuracy for the two algorithms. The comparison results are shown in Table 3. The table indicates that the modified ReliefF algorithm can greatly reduce the computation time with a lower loss of clustering accuracy.

By analyzing the payload operational mode, the change trend of the payload can be described. Taking the example of 600 records in the data set, the operational mode curve of the payload is shown in Fig. 4, where the value of 1 represents the idle mode, 2 represents the transmission mode, 3 represents the compression storage mode and 4 represents the maximum mode, which includes the compression storage and transmission mode.

Figure 4 shows that after the payload powers on, it enters the image compression storage and transmission mode. After the payload powers off, there are three data transmission modes. Finally, the payload powers on and enters the image compression storage and transmission mode again.

DISCUSSION

Compared with Nearest Neighbor, K-means, Relief, ReliefF and Multi-Relief algorithm, the improved feature weighted

clustering algorithm is more efficient in payload telemetry data analysis. It has the characteristics of a shorter calculation time, high real-time processing performance and accurate classification.

The Nearest Neighbor clustering algorithm as a hierarchical method, with large computation complexity, is suited for the small dataset. Compared with the amount of payload telemetry data, the Nearest Neighbor clustering algorithm is not efficient. The improved algorithm mentioned in this study as a partition algorithm, can gradually improve the clustering quality through iterative optimization process.

The K-means algorithm adopts the Euclidean distance metrics in data difference calculation. It's not efficient when the data dimension is very large. By verification in synthetic datasets and real datasets, it can be concluded that the fractional distance metrics can achieve better clustering accuracy for high-dimensional data clustering than can the Euclidean distance metrics.

The ReliefF algorithm calculates the feature weight at the base of each record without considering the initial clustering, which leads to a massive calculation for feature selection when the data set contains many records. The ReliefF algorithm may be impossible for real-time application. The initial clustering information is adopted and the balance among the different categories is maintained in the calculation process according to the cluster's scale. The experimental results show that the modified algorithm is efficient in processing the payload telemetry data.

Multi-Relief algorithm may miss small cluster for its random sample strategy. In the improved method, all the categories are adopted in the calculation rather than sampling randomly when the difference calculation is performed. For the number of payload telemetry data category is not so large, the amount of calculation is not increased significantly.

CONCLUSION

According to the analysis of the telemetry data, a new method has been proposed for analyzing the payload operational mode. The method is based on the feature weighted clustering algorithm and increases the clustering accuracy by using the fractional distance measure. The modified algorithm has a reduced amount of calculations, high real-time processing performance and accurate classification. It performs an automatic analysis of the payload operational mode. The method can be applied to the following areas:

- **Real-time processing:** Through the training of a large number of records, a library of characteristics can be

established for the payload operational mode. This library of the operational mode includes all of the feature vectors of this payload. When a new record is obtained, it's only needed to perform similarity matching between the record and the feature vectors in the feature library to identify the operational mode of the coming record in real time

- **Fault diagnosis:** After analyzing the payload operational mode, fault diagnosis may be performed using the analysis results. Generally, the payload operates in idle mode in the beginning and the operational mode is switched in the external command. For example, the ground observation payload switches from idle mode to camera-on mode and the scientific payload switches from idle mode to experimental process. When the operational mode changes, the telemetry data generally exhibit an obvious change; this change is reflected by one or more telemetry data points. By monitoring the status of these telemetry data, many methods can be performed including fault detection, fault diagnosis and fault recovery

REFERENCES

1. He, Y.F., G.H. Zhao, C.M. Lv and L.L. Guo, 2011. Improved OOPN-based approach for analyzing system operating modes. *J. Astronaut.*, 32: 1163-1170.
2. Wang, F., 2007. Research on space payload fault diagnosis system based on fault tree. Master Thesis, University of Chinese Academy of Sciences, China.
3. Han, J. and M. Kamber, 2006. *Data Mining: Concepts and Techniques*. 2nd Edn., Morgan Kaufmann Publisher, San Francisco, USA., ISBN-13: 978-1558609013, Pages: 800.
4. Jain, A.K., M.N. Murty and P.J. Flynn, 1999. Data clustering: A review. *ACM Comput. Surv.*, 31: 264-323.
5. Hansen, P. and B. Jaumard, 1997. Cluster analysis and mathematical programming. *Math. Program.*, 79: 191-215.
6. Kaufman, L. and P.J. Rousseeuw, 1990. *Finding Groups in Data: An Introduction to Cluster Analysis*. John Wiley and Sons, New York, ISBN: 0471878766, Pages: 342.
7. Kira, K. and L.A. Rendell, 1992. A practical approach to feature selection. *Proceedings of the 9th International Conference Workshop on Machine Learning*, July 12-16, 1992, Aberdeen, Scotland, pp: 249-256.
8. Parsons, L., E. Haque and H. Liu, 2004. Subspace clustering for high dimensional data: A review. *ACM SIGKDD Explorations Newslett.*, 6: 90-105.
9. Karypis, G., E.H. Han and V. Kumar, 1999. Chameleon: Hierarchical clustering using dynamic modeling. *Computer*, 32: 68-75.
10. Agrawal, R., J. Gehrke, D. Gunopulos and P. Raghavan, 1998. Automatic subspace clustering of high dimensional data for data mining applications. *ACM SIGMOD Rec.*, 27: 94-105.
11. Cordeiro, R.L.F., A.J. Traina, C. Faloutsos and C. Traina, 2010. Finding clusters in subspaces of very large, multi-dimensional datasets. *Proceedings of the IEEE 26th International Conference on Data Engineering*, March 1-6, 2010, Long Beach, CA., USA., pp: 625-636.
12. Singh, B., N. Kushwaha, O.P. Vyas, 2014. A feature subset selection technique for high dimensional data using symmetric uncertainty. *J. Data Anal. Inform. Process.*, 2: 95-105.
13. Kononenko, I., 1994. Estimating attributes: Analysis and extensions of relief. *Proceedings of the 7th European Conference on Machine Learning*, April 6-8, 1994 Catania, Italy, pp: 171-182.
14. Ye, K., K. Feenstra, J.A. Heringa, A.P. Ijzerman and E. Marchiori, 2008. Multi-RELIEF: A method to recognize specificity determining residues from multiple sequence alignments using a machine-learning approach for feature weighting. *Bioinformatics*, 24: 18-25.
15. Aggarwal, C.C., A. Hinneburg and D.A. Keim, 2001. On the surprising behavior of distance metrics in high dimensional space. *Proceedings of the 8th International Conference on Database Theory*, January 4-6, 2001, London, UK, pp: 420-434.