Applied Mechanics and Materials Vol 328 (2013) pp 463-467
Online: 2013-06-27

doi:10.4028/www.scientific.net/AMM.328.463

The Improved K-means Cluster Analysis on Diagnosis Data Fusion of The Aero-engine

Xiaobo Liu[1,a], Beibei Deng[2,b], Liangni Shen[3,c]

[1]Key Laboratory of Nondestructive Testing (Nanchang Hangkong University), Ministry of Education, Nanchang 330063, China;

[2]Aeronautical Manufacturing Engineering College, Nanchang Hangkong University, Nanchang 330063, China;

[3]Aircraft Engineering College, Nanchang Hangkong University, Nanchang 330063, China;

[a]liuxiaobo2000@21cn.com, [b]be_ilong@126.com, [c]shenliangni66@163.com

Keywords: Aero-engine, initial clustering center, hierarchical clustering algorithm, improved K-means clustering algorithm.

Abstract. Aiming at the problem about initial clustering center was randomly assigned in K-means clustering algorithm, the improved K-means clustering algorithm based on hierarchical clustering algorithm and K-means clustering algorithm was proposed in this paper. In the improved algorithm, first of all K was calculated by hierarchical clustering. When K was determined, K-means clustering was implemented. The results of the aero-engine vibration data clustering shown that not only the k value was to quickly and accurately determined, but also the number of clusters can be reduced and higher computing efficiency can be attained by the improved K-means clustering algorithm.

Introduction

Aero-engine as the core component of the aircraft, the stability and reliability direct influence on the aircraft's flight safety. A large number of documents and facts show that: the aero-engine failure is the main reason resulting in aircraft accidents and vibration fault occupied a very important position in the aero- engine fault. Aero-engine manufacturing and assembly quality problems can lead to excessive vibration of aero-engine. Excessive vibration of aero-engine may result in the following situations: additional vibration is too large, the rotor and stator rub-impact, bearing load is too large, resulting in the aero-engine condition is not within the normal range, greatly shorten the aero-engine life. What's more, the safety of aircraft and staff on workers will be endangered so that the right to life could not be guaranteed [1, 2, 3]. Therefore, in order to quickly and accurately identify the aero-engine failure, general sensors are installed in several parts of the aero-engine to acquisition of signal. However, when multiple sensor data support with the same function, it is necessary for data fusion which had been measured by multiple sensors[4]. In order to find out the fault feature, improve the certainty and reliability of the information and realize the aero-engine fault diagnosis.

When each sensor can not fully determine their own judgment, using the D-S evidence based on statistical methods for data fusion. However, how to acquire the basic probability assignment is one of the most difficulties in target recognition fusion based on the D-S evidence theory.

K-means clustering algorithm is a relatively simple and fast clustering algorithm. However, it exists problem about the initial clustering center must be set in advance, and the choice of the initial clustering center will seriously affect the clustering results(different k values will get different results). Aiming at the problem about initial clustering center is randomly assigned in K-means clustering algorithm, the improved K-means clustering algorithm based on hierarchical clustering algorithm and K-means clustering algorithm is proposed in this paper. k is calculated by hierarchical clustering. When K is determined, then K-means clustering algorithm will be implemented. The cluster number can be reduced by the improved K-means clustering algorithm.

Basing on the improved K-means clustering analysis, the basic belief function of vibration data will be determined. It can not only effectively solve the problem of the basic probability assignment in D-S evidence fusion, but also improve the computing efficiency.

Improved K-means Clustering Analysis

Clustering analysis can automatic data divided into different categories, it neither need the prior knowledge about the sample nor need samples training. Therefore, it is widely used in target recognition, data fusion,etc[5, 6, 7, 8, 9]. However, how to determine the initial clustering center of K-means is a difficulty. In the past, the initial clustering center is randomly assigned in K-means clustering algorithm, result in different k values will get different clustering. By the hierarchical clustering, the similar data are gathered together to get the best fit category (multicategory). Thus, the K-means clustering's initial center is determined, then K-means clustering could be implemented. Therefore, the improved K-means clustering algorithm is used in this paper. K is calculated by hierarchical clustering. When K is determined, K-means clustering algorithm is implemented. The fault symptom data is determined by the improved K-means clustering analysis. Make for the calculation of the fault symptom data more accurately.

Hierarchical clustering process

Hierarchical clustering process:
1. Calculate the distance between the pairwise independence in n objects.
2. Construct n member cluster, such as $C_n, C_{n-1}, \cdots, C_1$.
3. Find the two nearest cluster and merge the two cluster, such as C_i and C_j. The number of cluster is reduced by one.
4. Calculate the distance of the generated cluster with other clusters. If the obtained result meet the function S which has been obtained the minimum, then end.

$$S(k) = \frac{\sum_{i=1}^{k} \sum_{x \in c_i} \left| x - \bar{x}_i \right|^2}{\min\limits_{i,j<k(i \neq j)} \left| \overline{x_i} - \overline{x_j} \right|^2} \tag{1}$$

Where $\bar{x}_i$ and $\bar{x}_j$ is the center of the i category and j category. Otherwise, from the third step re-started to find the nearest cluster i and j. The purpose of this step, the K-means clustering initial cluster center is determined.

K-means clustering algorithm

K-means clustering process:
1. Take the K value which has been determined by hierarchical clustering in this previous step as the initial category center.
2. According to the average of objects in the category (i.e., Euclidean distance), if the distance from an object to K category center points is the shortest, the object is given to the nearest category based on the shortest distance principle.
3. Update the average of the category; calculate the average of objects in each category.
4. Calculation principles function E.

$$E = \sum_{i=1}^{K} \sum_{x \in C_i} \left| x - \overline{x_i} \right|^2 \tag{2}$$

Where E is the sum of the squared of all data in the database unit, x is the unit of data in the database, $\bar{x}_i$ is the average of the category C_i.

$$m_i = \sum_{p \in C_i} \frac{p}{|C_i|} \tag{3}$$

Where $|C_i|$ is the number of clustering data units in the category C_i.

5. If E and the result of the last circular compared no longer change significantly, then end. If E and the result of the last circular compared differences significantly, then continue the cycle. (The offset between the cluster center point and the last cluster center point is less the default value, i.e. The default value is 0.02)

The improved K-means clustering implementation steps:

(1)Using hierarchical clustering method to get the initial value.

(2)The initial value of the first step as the initial value of K-means clustering, the K-means clustering is implemented to get the final result.

Example analysis about improved K-means clustering

The training example of aero-engine vibration data are shown in Tab.1. We select two cluster center, using traditional partition clustering algorithm to get the results shown in Tab.2, Tab.3, Tab.4,Tab. 5.

Table 1. Aero-engine vibration data

State pattern	Measure point 1 [mm/s]	Measure point 2 [mm/s]	Measure point 3 [mm/s]	Measure point 4 [mm/s]	Measure point 5 [mm/s]	Measure point 6 [mm/s]
Misalignment	6.11	24.41	6.51	50.32	8.57	23.46
	36.31	37.26	34.96	23.51	21.51	30.21
	24.44	39.45	9.38	37.51	24.76	25.96
	6.89	30.28	11.29	40.01	16.89	33.91
Dynamic unbalance	18.86	24.80	14.57	15.85	20.62	19.98
	38.16	51.93	30.69	16.27	25.15	23.99
	42.91	55.05	30.75	22.52	25.43	24.88
	38.16	51.93	30.69	16.27	25.15	23.99
Rubbing	29.05	37.23	30.59	39.23	24.46	17.84
	7.20	9.70	11.83	12.33	10.98	14.18
	21.93	25.38	24.19	22.29	23.9	18.73
	28.17	22.89	20.25	25.62	24.16	19.28
Failure-free	9.15	9.87	12.63	12.24	14.29	15.69
	23.93	34.80	25.60	30.78	22.37	17.51
	33.40	33.57	30.76	17.17	22.44	29.66
	34.10	34.34	32.34	19.90	22.20	30.81

Table.2 Initial cluster center

	Cluster			
	1	2	3	4
Misalign-ment	39.45	33.91	6.51	50.32
Dynamic unbalance	55.05	23.99	14.57	15.85
Rubbing	25.38	19.28	30.59	39.23
Failure-free	33.57	30.81	12.63	12.24

Table.3 Final cluster center

	Cluster			
	1	2	3	4
Misalign-ment	32.86	30.19	11.26	50.32
Dynamic unbalance	50.46	24.30	25.45	15.85
Rubbing	19.98	16.74	27.26	39.23
Failure-free	34.03	23.26	20.67	12.24

Table. 4 Select the initial cluster center for each category number of samples

Cluster	1	4.000
	2	11.000
	3	8.000
	4	1.000
Valid		24.000
Missing		36.000

Table. 5 Optional initial value to analysis of variance

	Cluster Mean Square	df	Error Mean Square	df	F	Sig.
Misalignment	898.352	3	43.030	20	20.877	.000
Dynamic unbalance	785.146	3	43.569	20	18.021	.000
Rubbing	278.980	3	34.224	20	8.151	.001
Failurefree	211.212	3	53.867	20	3.921	.024

The improved K-means clustering algorithm steps :(1)Select a set of sample data.(2)Calculate K value by hierarchical clustering.(3)Implement K-means clustering based on K value. Using SPSS [10] software can get S_1 dendrogram as shown in Fig.1. The data is divided into two groups from figure 1 and K value is 2. The result as shown in Tab.6, Tab.7 and Tab.8.

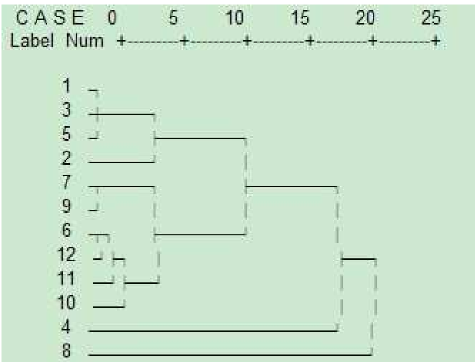

Fig.1 Hierarchical clustering dendrogram

Table.6 Improved the final clustering centers

	1	2	3
Misalignment	17.10	33.80	42.61
Dynamic unbalance	24.24	42.09	18.21
Rubbing	22.58	16.89	29.05
Failure-free	21.70	30.92	16.44

Table.7 Each category number of samples about the improved K-means clustering center

Cluster	1	14.000
	2	7.000
	3	3.000
Valid		24.000
Missing		36.000

Table. 8 Improved K-means clustering algorithm analysis of variance

	Cluster Mean Square	df	Error Mean Square	df	F	Sig.
Misalignment	1172.550	2	57.646	21	20.341	.000
Dynamic unbalance	932.387	2	64.859	21	14.376	.000
Rubbing	167.393	2	56.507	21	2.962	.074
Failure-free	289.664	2	53.888	21	5.375	.013

Conclusion is obtained by analysis of variance about the results of improved K-means clustering and traditional clustering: the results of the improved K-means clustering are more reasonably and accurately. If data unit is too much in database, using hierarchical clustering will result in lacking of scalability. In hierarchical clustering, the last processing step can't be revoked, the unit of data which have been processed can't be exchanged in different category. In K-means clustering, the initial value have a higher dependence. If the selected K value is improper, the final result may be resulted in unsatisfactory. However, the improved K-means clustering only take a part of data as representative from large database, it can solve the lacking of scalability defect when hierarchical clustering have to handle a large number of data units. The K value is determined by hierarchical clustering, so as to solve the initial value problem of K-means clustering algorithm, reducing the probability of the results are unsatisfactory in K-means clustering algorithm.

Conclusions

Aiming at the problem about initial clustering center is randomly assigned in K-means clustering algorithm, the improved K-means clustering algorithm is proposed in this paper. It not only absorbs the advantages hierarchical clustering and K-means clustering algorithm but also avoids there deficiencies, so that it can extremely improve the quality of K-means clustering. Through an example shows that: the improved K-means clustering algorithm can not only quickly and accurately determine the initial clustering center, reducing the number of iterations, but also remove some interference data and fully mining the data of similarity, reducing the number of clusters.

Acknowledgments

This study was principally supported by Aeronautical Science Foundation of China(2009ZD56007), JiangXi Science and Technology Project(GJJ12414) and Key Laboratory of Nondestructive Testing of the Ministry of Education (Nanchang Hongkong University) open fund(ZD201129003).

References

[1] Chen Guang. Aero-engine failure analysis. Beijing: Beijing University of Aeronautics and Astronautics Press, 2001

[2] Randall Bickford, Donald Malloy. Development of a real time turbine engine diagnostic system. AIAA2002-4306.

[3] Fan Zuomin, Sun Chunlin, Bai Jie. Introduction to aero-engine fault diagnosis. Beijing: Science Press,2004.

[4] Yang Wanhai. Multi-sensor Data Fusion and Application. Xi'an: Xi'an Electronic and Science University Press,2004.

[5] Brian S. Everitt. Cluster Anlysis. Halsted Press, Third Edition, 1993.p.767

[6] M Fillippone, F Camastra, F Masulla, etal. A survey of kernel and spectral methods for clustering. Pattern Recognition. 2008.p.176

[7] S Krinidis, V Chatzis. A robust fuzzy local information c-means clustering algorithm. IEEE Trans Image Process. 2010.p.1328

[8] Pan Bin, Shu Ning. Cluster Analysis for Selection of Time Series Interferometric SAR Imagery. Journal of Applied Sciences. 2010.p.501

[9] Liu Bing, Xia Shixiong, Zhou Yong, Han Xudong. A Sample-Weighted Possibilistic Fuzzy Clustering Algorithm. Acta Electronica Sinica. 2012.p.254

[10] SPSS Inc. SPSS Advanced Stasties. USA: Marija J.Norusis. 2000.p.64

Mechanical Science and Engineering III
10.4028/www.scientific.net/AMM.328

The Improved K-Means Cluster Analysis on Diagnosis Data Fusion of the Aero-Engine
10.4028/www.scientific.net/AMM.328.463