Signal Clustering Algorithm and Its Application Progress in Signal Clustering

Deng Chujie, Qi Jing, Li Mei*, Luoxuan Chicheng
School of Information Engineering, University of Geosciences (Beijing), Beijing, China
Email: maggieli@cugb.edu.cn

Abstract. Clustering algorithm, which is a statistical analysis for research of classifications, plays an important role in data mining algorithm. Clustering algorithm based on similarity, and is easy to combine with other methods in optimization. In this review, signal clustering algorithm is introduced by discussing of the clustering parametric in different signal clustering algorithms. In order to develop traditional algorithm, we introduce a series of improvement, development and application of the methods in recent years. Finally, we make an outlook of the future direction and content of the research in this field.

Keywords: Signal; clustering algorithm; K-means; FCM

1 Introduction

Cluster analysis was originated in taxonomy. With the development of science and technology, people have increasingly requirements towards classification. So they refer the mathematical tool in taxonomy, forming numerical taxonomy. Multivariate analysis technology was appointed into numerical taxonomy, forming the clustering algorithm.

From a practical point of view, traditional clustering algorithms are varied, it can be a stand-alone tool to obtain the distribution of data. So that we can observe each cluster data's feature, and make further analysis on a specific set of cluster. And it can also be used as a pretreatment step to other algorithms.

With the development of data, traditional clustering algorithm has been improved, which is simple and intuitive. It is widely used in data mining analysis. Combing and analyzing the literature in this field, we write this review aimed at a comprehensive understanding of the signal clustering algorithm. We have made series of ways to improve, develop and apply signal clustering algorithms. We made an outlook of the future direction and content of research in the field ultimately.

There is a large amount of sources transmit signal everywhere and every moment. In order to improve the efficiency of data applications, and discover valuable signal, we need to study and judge the radiation source in advance. As a result, it is necessary to study valid signal recognition methods, such as signal clustering. To fulfill the urgent needs, many scholars [1~4] dedicated to research on large amounts of data for a fast-identification, high-accuracy, and clear-semantics clustering algorithm.

2 Signal Clustering Algorithm

Clustering, which is based on the similarity of the data object, polymerizes data into different clusters. In the same cluster, data objects have high similarity, but the different clusters of data objects have greater dissimilarity. The difference between clustering and classification are that clustering algorithms can process data in an unsupervised mode, and you will not to input the data label. Clustering algorithms mainly include partition-based, division-based, density-based and grid-based methods.

2.1 Clustering Algorithm Based on Partition

We assign all objects to several signal clusters, making instances of the same cluster gather around a center, and the distance between them is relatively close. At the same time, the distance between the different clusters examples is relatively far. The main representative of the algorithm are K-means、K-medoids and Clarans.

There is a difference between K-means and K-medoids. The central point of K-means is the average of all current data points, but K-medoids select the point with the closest distance to the average value. The following equations describe K-means and K-medoids.

K-means:

Assign each observation to the cluster whose mean yields the least within-cluster sum of squares (WCSS). Since the sum of squares is the squared Euclidean distance, this is intuitively the "nearest" mean.

$$c^{(i)} = arg\min_{j}\left\|x^{(i)} - u_j\right\|^2 \tag{1}$$

K-means algorithm minimize parameter J.

$$\mathrm{J} = \sum_{\mathrm{n}=1}^{N}\sum_{k=1}^{K}\gamma_{nk}\|X_n - \mu_k\|^2 \tag{2}$$

Calculate the new means to be the centroids of the observations in the new clusters.

$$u_j = \frac{\sum_{i=1}^{n} 1\{c^{(j)}=j\}x^i}{\sum_{i=1}^{n} 1\{c^{(j)}=j\}} \tag{3}$$

The algorithm has converged when the assignments no longer change.

K-medoids:

K-medoids changes a random dissimilarity measure Function-*V* to the Euclidean distance in original objective Function-*J*.

$$\tilde{J} = \sum_{n=1}^{N}\sum_{k=1}^{K}\gamma_{nk}v(x_n,\mu_k) \tag{4}$$

2.2 Clustering Algorithm Based on Density

We assign the field radius and the density threshold. We are looking for a target, in which field radius it contains the number of objects greater than or equal to the density threshold, and we see

the target as a center point. Otherwise, we regard it as a boundary point. The main representatives of the algorithm are DBSCAN and DENCLUE.

2.3 Clustering Algorithm Based on Grid:

We separate the possible values of each attribute into a number of adjacent spaces to create a collection of the grid cell. And every object falls into a grid cell, the corresponding attribute range of grid cell contains the value of the object.

2.4 Fuzzy C-means Algorithm:

We square weighted the objects and its distance from the cluster center with subjection, to get a general description of the fuzzy clustering objective function [5]. FCM receives each sample point membership of the entire clustering center by optimizing the objective function. So that we can find the type of sample points, in order to achieve automatic sample classification. FCM is a clustering algorithm, which is based on membership, to determine which set the data belongs to.

3 Application Progress in Signal Clustering

3.1 Signal Clustering Algorithm in Military Field, Take Radar Signal as Example

Under normal circumstances, we cluster radar signal pulse to classify the radar signal for the following parameter estimation, emitter recognition, threat discrimination and combat situation analyzing by feature space composed by time of arrival (ToA), pulse width (PW), pulse angle of arrival (AoA), pulse amplitude (PA), radio frequency (RF) [6].

We study the different clustering algorithms for radar signal and make the following review.

3.1.1 K-means Clustering Algorithm Improved by LF Intelligent Optimization Algorithm

K-means clustering algorithm has already applied in radar signal sorting, but the clustering results are sensitive to isolated points and noise points. So the scholars usually improve the K-means algorithm, combining with other algorithm, as the literature [7] algorithm: based on ant (LF) intelligent optimization algorithms to improve the K-means clustering algorithm.

First, we use the ant colony algorithm to place a radar signal on a random plane. Then, it will generate a set of virtual code of conduct randomly. According to this criterion, we could get properties of similar signals together into a category eventually, and obtain the initial centroid and the number of clusters. This method can solve the difficulty of determining the number of clusters and cluster centers.

3.1.2 FCM Clustering Algorithm Based on Partition Matrix Optimization

The FCM algorithm exists problems that the number of clusters is difficult to determine, the initial value is unstable and of the local extreme instability and other issues about local extrema. So, Baraldi proposed fuzzy C-means clustering algorithm based on partition optimized matrix (FCM) [8].

To get the best partition matrix according to the principle of division of matrix and cluster center initialization at first and signal samples can be aggregated into different categories by the principle of the maximum degree of membership and use the signal characteristics of radar radiation generated by simulation as separate arrays. Then a higher similarity signal was classified as a

category according to FCM clustering thought. This allows the merging of data about a large number of miscellaneous radar pulses so as to achieve a better clustering effect.

3.2 Signal Clustering Algorithm in Medical Field, Take ECG as Example

Since people raise higher requirements to identify and analyze abnormal ECG, clustering algorithm begin to be used for the sorting of the abnormal ECG process.

ECG feature extraction is mainly the detection of QRS-wave, P-wave, T-wave and R-wave. P-wave represents atrial depolarization vector. T-wave represents the process of both sides of the ventricular repolarization. Accurate detection of the R-wave represents the ability of the accuracy of QRS-wave detection.

3.2.1 Abnormal ECG Classifying Algorithm Based on FCM

The FCM clustering algorithm is widely used in ECG. As a result of lacking prior knowledge, most of fuzzy clustering algorithm needs to pre-set and depends on the number of prototype cluster-c, the weighting factor-m, and fuzzy cluster center initialization parameters. To improve this technology, Yao Cheng, in 2012, put forward a method [9] directly judged by original ECG signal, to improve the accuracy of abnormal ECG classification. First, this method uses the statistic of ECG to reflect heart disease characteristic data. Secondly, we can put forward a standard of the logical ECG by using the abnormal ECG data. Finally, we could utilize logical judgment and clustering algorithm to raise one kind of algorithm for classifying abnormal ECG, which is called LCFCM [10].

3.2.2 K-means Clustering Method Based on Simulated Annealing Algorithm

K-means clustering algorithm for ECG, the result will be easily influenced by centroid, appearing locally optimal solution and isolated point. A literature [11] raised K-means clustering method based on simulated annealing algorithm to solve above problem.

According to an initial solution, we set an initial target function. After calculating the probability, with continuous iteration, we calculate the difference of function. By progressively determining and discarding, we will receive the optimal solution with traversing a large space.

The algorithm is of strong local search capabilities, being able to reach the global optimal solution.

3.3 Signal Clustering Algorithm in Engineering Field, Take the Rotor Rubbing, Tank Bottom Corrosion Acoustic Emission Signal as Example

There are many characteristic parameters with acoustic emission signals, for example, Waveform Amplitude (A), Rise Time (R), Energy (E), Duration (D), Pulse Index (X), Margin Index (W), Steep Wave Clicks (Z), Event Times (Y), Number of Rings (C) etc. They can be described as characteristic parameters of waveform, which can be used for clustering.

At the same time, the scholar introduces three derived AE parameters [12]: RA (Rise ti me/Amplitude); AF (Number of rings/duration); RD (Rise time/duration).

3.3.1 Transmission Signal of Rotor Rubbing Sound Based on K-means Clustering Algorithm

Jin Zhihao [13] proposed an approach to cluster rotor rub-impact acoustic emission signals by k-means. He selected the waveform amplitude, pulse index, margin indicators and kurtosis indicators as characteristic parameters to cluster and analysis acoustic emission signal. Firstly, using k-means clustering to analysis acoustic emission detection signal under the situation whether happened

collision and friction and obtained the cluster center. With further analysis in conjunction with other algorithms, we finally get acoustic emission characteristics and lay the foundation for Rubbing Fault Diagnosis.

By selecting characteristic parameters, the effect of the threshold value is avoided and acoustic emission signals can be identified efficiently with high accuracy.

3.3.2 Pitting Acoustic Emission Signal Based on K-Means Clustering Algorithm

Bi Haisheng apply acoustic emission technology for monitoring Tank Bottom online and fault identification [14]. For detection of Tank Bottom Corrosion acoustic emission signal, the results analyzed by a single parameter are often quite different from the activity. Therefore, title used seven characteristic parameters of A, E, D, C, RA, AF and RD to cluster. Application of K-means clustering algorithm to identify the acoustic emission signal used distance as an index from the signal target cluster and identifies three categories of typical acoustic emission signals. By clustering and identifying the fault signal, the test results will help improve the reliability and reduce the risk of running the tank, ensuring its safe operation.

This algorithm is based on the need to determine the number k of category in advance. It will make fault detection easier if this algorithm is combined with other algorithms which can change the k value adaptively.

3.4 Clustering Algorithm in FM Radio and Case in the Field of Computer Science.

K-means algorithm is adopted to identify broadcasting and aviation voice signals from interfering signals in space communications through traditional methods. However, the algorithm cannot completely classify signals accurately and automatically. For this situation, it was proposed voice signals (f-kmd) algorithm based on K-medoids and FCM, which focuses on FM broadcast and aviation signal classification [15].

Characteristic parameters of the speech signal include the mean, variance, short-time average energy, average zero-crossing rate, normalized amplitude. Short voice data was clustered by K-means and K-medoids and FCM respectively at first [16]. And then we used those algorithms on entire speech data, which means three kinds of combinations of two clustering algorithms, finding the optimal clustering mode: the first FCM clustering, then K-medoids clustering.

The effect of the conjunction between FCM and K-medoids is better than individual algorithm. F-kmd clustering algorithm has good stability and the correct rate was significantly higher.

4 Development and Prospects

Clustering algorithm of signal varies, which has shown its character and charm in many fields. However, the traditional algorithm and improved algorithm theory is far from mature and practical application is far from showing its true potential. Many challenging issues waiting to be solved are listed below:

1) Strengthening theoretical Study for clustering algorithm [17]. Developing new mathematical analysis and modeling tools further is very important for all kinds of basic theory of algorithms in particular. We have not yet found the mathematical discourse about reasonable choice on the number of cluster centers and clusters in the absence of prior knowledge .Therefore, the mathematical theory based on a clustering algorithm applied to the signal aspect will become an important topic for future research.

2) The algorithm can be improved with other types of methods to develop the integrated use of hybrid-optimization method. Many authors in this respect have made a good attempt. It will become the focus of research on signal clustering to combine traditional or improved clustering algorithms with neural networks, fuzzy control, genetic algorithms, and simulated annealing algorithm and so on.
3) The new theory must be tested by practice in which we can find new problems, so as to promote the theory forward. Although clustering algorithms have been promoted to the application in many areas in recent years, most of them just a simple simulation algorithm in the field of application. Therefore, we should fully dig out potential of signal clustering in the practical application. Furthermore, the hardware implementation of the signal clustering algorithm will also become one of the hot research directions.

5 Acknowledgements

This work is financially supported by the National Natural Science Foundation of China (Grant No. 41572347).

References

1. Wang, J., Zhang, B.: A radar signal sorting algorithm based on dynamic grid density clustering. J. Modern electronics technique. 36, 1--4 (2013)
2. Li, X.Y., Yang, C.Z., Qu, W.T.: A radar signal sorting algorithm based on adaptive grid density clustering. J. Aerospave Electronic Warfare. 29, 51--53 (2013)
3. He, X.W., Yang, C.Z., Zhang, R.: A radar signal sorting algorithm based on improved grid clustering. J. Radar & ECM. 31, 43--49 (2011)
4. Zhang, C.C.: Radar emitter signal deinterleaving based on support vector clustering. D. Xian: Xidian University, 2012
5. Xie, T.J.: Clustering Algorithm Summary (in Chinese). D. Beijing: Beijing University of post and telecommunications. 2014.
6. Xiang, X.: Research of unknown radar signal sorting algorithm. D. Xian: Xidian University. 2011.
7. Zhao, G.X., Luo, L.Q., Chen, B.: Improved artificial fish school algorithm applied in radar signal sorting. J. Electronic Information Warfare Technology. 7, 142--146 (2009)
8. Baraldi, A., Blonda, P.: A Survey of Fuzzy Clustering Algorithms for Pattern Precognition-Part1 and Part2. J. IEEE Trans. Systems, Man and Cybernetics, Part B. 29, 778--801 (1999)
9. Yao, C.: The study on the key technology of ECG signal intelligent analysis. D. Jilin: Jilin University, 2012.
10. Lin, Z.T., Ge, Y.Z.: A study on Clustering Analysis of Arrhythmias. J. Journal of Biomedical Engineering. 23, 999--1002 (2006)
11. Zhang, X.R.: Study on the improved methodology of ECG clustering strategy. D. Harbin: University of Science and Technology, 2015
12. Giuseppe, C., Luigi, C., Edoardo, P.: Evaluation of increasing damage severity in concrete structures by cluster analysis of acoustic emission signals. C. European Conference on Acoustic Emission Testing. 29, 8—10 (2010)

13. Jin, Z.H., Tang, F.L., Zhao, C.M.: Extraction based on clustering analyses on rotor rubbing sound emission characteristics. J. Journal of Shenyang University of Chemical Technology. 29, 342--346 (2015)
14. Bi, H.S., Li, Z.L., Hu, D.D.: Cluster analysis of acoustic emission signals during tank bottomsteel pitting corrosion process. J. Journal of China University of Petroleum. 39, 145--152 (2015)
15. Zhang, Z., Ma, F.L., Pei, Z.: Recognition of Aviation Interference Signal Based on K-means Clustering Algorithm. Publishing House of Electronics Industry, Beijing (2013)
16. Hu, A., Pei, Z.: K-medoids and FCM fusion clustering application research on broadcast and aviation speech signal classification. J. Journal of University of Jinan. 30, 1671--3559 (2016)
17. Duan, H.B., Wang, D.B., Huang, X.H.: Development on ant colony algorithm theory and its application. J. Control and Decision. 19, 1322--1326,1340 (2004)