

A Clustering Method For Electromagnetic Interference Signals Based On Particle Swarm Optimization

Hongyi Li^{1,2}, Shengyu Chen^{1,2} and Di Zhao^{1,a}

¹*LMIB, School of Mathematics and Systems Science, Beihang University, Beijing 100191, China*

²*School of Software Engineering, Beihang University, Beijing 100191, China*

Abstract. This paper proposes a method which clusters EMI signals based on particle swarm optimization. And the advantages of this method compared to other clustering algorithms are present in this paper. From the clustering experiments on 158 sets of data and comparison with other classical algorithms, the clustering algorithm proposed has certain accuracy and feasibility.

1 Introduction

With the rapidly increase of electronic devices such as mobile phones, the electromagnetic interference (EMI) has become more and more urgent, considering its huge interference to normal electronic communication and signal analysis. It is therefore for of great significance to analyze unknown EMI source signals.

The current state-of-the-art technologies to address this problem include the fast independent component analysis (Fast ICA) [1-2], wavelet transform [2-3], the fast Fourier transform (FFT) [4-5]. An indispensable procedure of these techniques is the clustering step.

Existing clustering method are K-means [12], fuzzy clustering [13], wavelet analysis [3, 14], Fourier analysis [5]. K-means is an intuitive method, and easy to implemented, so that it is one of the most widely used clustering approach. However, it also has obvious defects, that it can only handle numeric attributes and is sensitive to initialization. It also require the number of cluster centroids beforehand, and could only obtain a local optimal clustering results [6]. Similar to K-means, Fuzzy c-means could not overcome the local minimum, and the sensitivity to initialization. Moreover, it is hard to select an appropriate weighting parameter M, which has a significant influence on the clustering performance. Wavelet analysis and Fourier analysis are limited to dimension reduction clustering processing of signal in the time domain.

Due to the high dimension that EMI signals have and its non-linear characteristic [15-17], it's necessary to reduce the dimension of the data in a proper way. In addition, to avoid problems like local minimum discussed above, we decided to use PSO based clustering algorithm which can handle this task well. Therefore, this paper proposes a new algorithm procedure for extracting electromagnetic interference characteristics to reduce the dimension of the data, and uses PSO based clustering algorithm to cluster the low dimensional data. Through

the experiments, this algorithm has a good performance on accuracy and efficiency in signal processing and clustering.

The rest of this paper is organized as follows: in the second section, we a feature extraction method for electromagnetic inference signals. Subsequently, we propose the clustering algorithm based on PSO. Section 4 shows comparison experimental results. Finally, we draw a conclusion in the last section.

2 A low dimensional representation of EMI signals

In this section, we propose a powerful representation of EMI signals with low dimensionality.

EMI signals usually contain thousands of discrete point whether in time domain or frequency domain. Original EMI signal data usually has high dimensions that is from hundreds to thousands. EMI signals we handled in this paper have more than 6000 dimensions. In addition, EMI signal data value changes dramatically. The intensity of EMI signal data is also a significant feature we interested in.

High dimensional data will make cluster algorithm hard to converge and result in deadly problems like dimension crisis. On the other hand, high dimensional data will cost algorithm much more calculation and spend more time.

Because EMI signals have their special characteristics, we need special features to describe them. In this paper, we present a new method to reduce data dimension. EMI signal data are extracted into five aspects of feature: exception, variance, bandwidth, number of thorns and the number of peaks and troughs. Exception value is a macroscopic feature describing the data value on average. And variance describes degree of dispersion on data value. We use bandwidth to show the volatile of the signal. Signal has larger bandwidth may contain more energy. The number of thorns reflects the instability of

^a *Corresponding author(E-mail:zdhy12010@163.com)*

the signal and represent the number of the extreme values. Values in extreme prominent will be recognized as thorns. The number of peaks and troughs shows the overall trend of the signal. This describe the shape of the signal which contains trend information. By doing feature extraction, all information we interested is extracted. And through these features we can discriminate different EMI signals in appropriate time.

3 EMI signal cluster

In this section, we will introduce the PSO clustering algorithm, and describe our proposed method.

3.1. PSO cluster algorithm

In 2002, Omran et al. [10] proposed a swarm optimization based unsupervised image classification algorithm. This is the earliest PSO based clustering algorithm in literature. Most of the PSO clustering algorithms follow the basic framework of this algorithm, where the number of clustering centroids is user defined. Each particle in the swarm contains a vector representing a candidate of cluster centres. The whole particle swarm contains variety of partition of data set. Firstly, a set of random cluster centre is assigned to each particle. For each particle, data are clustered according to the nearest candidate centre. Subsequently, the algorithm utilizes the PSO algorithm to find the optimal particle, of which the position is an approximate optimal partition on a data set.

Based on Omran's work, Merwe et al. proposed basic PSO clustering algorithm for the clustering to general data set [7]. Fitness function of this PSO clustering algorithm is J_e .

$$J_e = \frac{\sum_{j=1}^{N_c} \left[\sum_{z_p \in C_j} d(z_p, m_j) \right] / |C_j|}{N_c} \quad (1)$$

$$m_j = \frac{1}{n_j} \sum_{z_p \in C_j} z_p \quad (2)$$

$$d(z_p, m_j) = \sqrt{\sum_{k=1}^{N_b} (z_{pk} - m_{jk})^2} \quad (3)$$

N_b —dimension of the data; N_c —number of cluster; z_p —data vectors; n_j —number of samples in set C_j ; m_j —mean(centre) of set C_j ;

The outline of the basic PSO clustering algorithm is as follows [7]:

Algorithm 1 PSO Clustering Algorithm

Input: normalized data vectors z_p

Output: the optimal particle z_{pb} that contains best cluster centre

- 1: Randomly chosen cluster centre and assigned to each particle, randomly generated particle velocity.
- 2: Divide each particle on minimum distance principle, and calculate each particle's fitness value by (1) and update individual extreme value.
- 3: According to each particle's extreme value, find the global extreme value and its position.

- 4: Based on the PSO velocity formula, update each particle's velocity, and limit it in VMAX.
- 5: Based on the PSO position formula, update each particle's position.
- 6: Jump to step 2 until satisfy some condition
- 7: Output the position of the optimum particle z_{pb} , which contains the optimal N_c cluster centre.

The basic PSO clustering algorithm provides an effective way to solve the problems of many traditional clustering algorithms. Traditional hierarchical methods need to search and estimate large amount of objects or clusters. This results in bad scalability. In addition, these methods may be hard to deal with overlapping clusters. On the other hand, partition methods like K-means can only divide spherical clusters. And it's easy to get in local optimum. However, PSO clustering algorithm can solve these problems.

3.2 The proposed algorithm

Here, we will present our method. Firstly, we extract a low dimensional feature from source EMI signals, and standardize the extracted features as follows:

$$x_{ij} = \frac{x_{ij} - \min(x_{kj})}{\max(x_{kj}) - \min(x_{kj})} \quad (4)$$

where x_{ij} is the j -th component of the i -th vector, $\max(x_{kj})$, $\min(x_{kj})$ are the maximum and minimum values for the j -th component of all vectors, respectively.

After extraction, all the EMI signal data is presented as vectors as follows:

$$z_{pi} = (x_{i1}, x_{i2}, \dots, x_{in}), \quad (5)$$

Where i is the serial number of the data; n is the dimension of data after feature extraction;

For each particle p_i , it is an $n \times N_c$ matrices, representing the positions of N_c clustering centres, where

$$p_i = (m_{i11}, m_{i12}, \dots, m_{i1n}, \dots, m_{iN_c1}, \dots, m_{iN_cn}) \quad (6)$$

At the beginning, the positions and velocities of particles are randomly initialized. As described in subsection 3.1, we then divide every particle with the minimum distance principle, calculate the fitness and update the individual optimal value as follows

$$J_e = \frac{\sum_{j=1}^{N_c} \left[\sum_{z_p \in C_j} d(z_p, m_j) / n_j \right]}{N_c} \quad (7)$$

$$m_j = \frac{1}{n_j} \sum_{z_p \in C_j} z_p \quad (8)$$

$$d(z_p, m_j) = \sqrt{\sum_{k=1}^{N_b} (z_{pk} - m_{jk})^2} \quad (9)$$

Where N_b is the dimension of data; N_c is the number of clusters; z_p is the data vector; n_j denotes the number of samples in C_j ; m_j stands for centres of samples in C_j .

Based on the individual optimal position, we search for a global optimal particle together with its position. According to the velocity formula [11] as below, we update the velocity of particles with a limited maximum value v_{max} .

$$v_{id}(t+1) = \omega v_{id}(t) + c_1 r_{1d}(t)(p_{id}(t) - x_{id}(t)) + c_2 r_{2d}(t)(p_{gd}(t) - x_{id}(t)) \quad (10)$$

$$x_{id}(t+1) = x_{id}(t) + v_{id}(t+1) \quad (11)$$

Where $v_{id}(t)$ is the velocity of the particle i in the t -th iteration; $x_{id}(t)$ is the position of the particle i in the t -th iteration; ω is the inertia weight; and non-negative c_1 and c_2 are learning factors; r_1 and r_2 are independent random functions obeying uniform distribution in the range of $[0, 1]$.

The termination condition can be set as a maximum iteration number, or the residual error.

At last, the optimal position of each particle is output as a cluster centre.

The overall outline of the proposed algorithm is shown in Figure 1.

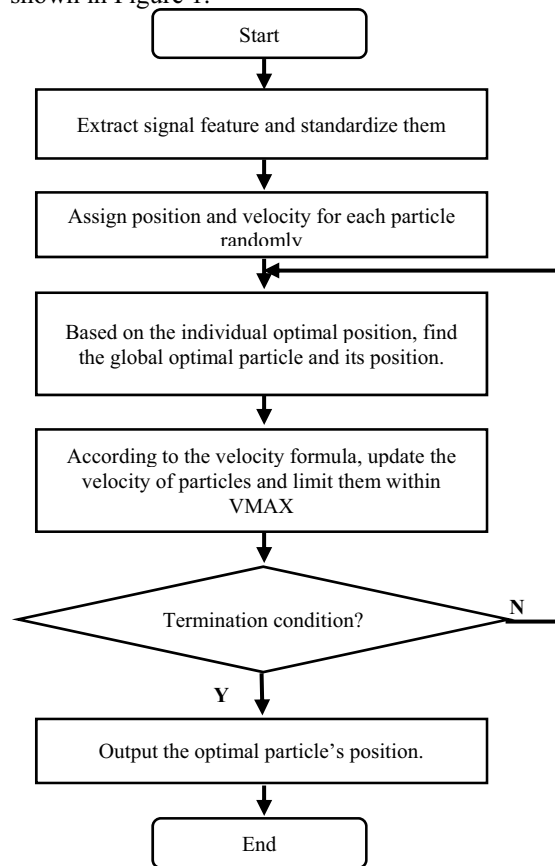


Figure 1. Caption of the Figure 1. Below the figure.

4 Experimental results and analysis

4.1. Data format and other clustering algorithm

The original electromagnetic signal data used in this paper is the frequency domain information in two dimension—frequency (in ascending order) and

voltage. We have totally 158 groups of categorized data and each group contains 2431 to 3069 two-dimensional vectors.

For other different mixed electromagnetic signals, such as time domain signals or mixed signals, they can also be used as the input to our method. When the signals are mixed, this method can cluster closely related signal together, which will be helpful for further separation and analysis.

For some commonly used clustering algorithm, PSO algorithm has obvious advantages. Compare to methods based on neural networks, PSO clustering algorithm has a more simple structure, which means it is easy to converge. In addition, PSO clustering algorithm is not sensitive on initialization, and it has a better scalability.

4.2. Parameter settings

In all our experiments, the number of particles N is set as 85. c_1 and c_2 are set to 1.2, while the inertia weight range W_{MAX} is chosen as 0.9, and W_{MIN} is set to 0.4. the maximum iteration M is 200. Finally, we set the number of clusters K as 7.

4.3. Experimental results

4.3.1 Groups of samples clustering

Here, we first demonstrate the low dimensional feature extracted by using the method described in section 2, as shown in Table 1.

Table 1. Data after feature extraction.

Exception	Variance	Bandwidth	Thorns	Peaks & trough
38.1077	119.2337	6.672511	12	13
37.3756	94.9141	6.594406	11	11
46.6688	19.2154	3.672786	3	1
65.3996	101.838	3.144071	17	2
37.8156	131.037	9.222	0	2
36.5505	129.9987	9.270943	0	0
-1.8437	78.2903	6.438411	13	7
-4.6797	46.608	3.317348	13	10
23.5268	102.1685	3.465579	0	1
39.9496	176.6618	4.889012	0	6
38.4838	139.4067	4.538699	2	9
37.9304	145.5721	4.744465	1	5
15.2725	180.7833	5.200211	0	2
31.7017	396.7352	6.818715	12	8
40.0647	157.7335	6.711977	1	6
40.8305	117.545	6.203311	0	5
9.2589	232.1797	9.173041	0	2

We then normalize the feature, and conduct clustering by using PSO clustering algorithm. The result is shown in Table 2.

Table 2. Cluster result by using the low dimensional feature and PSO clustering algorithm.

Class	Data serial number
1	17
2	5,6
3	1,2
4	9,10,11,12,13,15,16
5	14
6	7,8
7	3,4

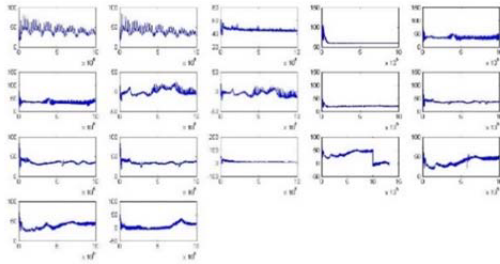


Figure 2. Curves of the original EMI signal data.

The final output clustering centres can be seen in Table 3.

Table 3. Output centres using the proposed method.

Class	Clustering centre
1	(0.1989, 0.5641, 0.9840, 0, 0.1538)
2	(0.5974, 0.2948, 0.9960, 0, 0.0769)
3	(0.6053, 0.2327, 0.5695, 0.6765, 0.9231)
4	(0.5480, 0.3350, 0.3205, 0.0336, 0.3736)
5	(0.5191, 1.0000, 0.5998, 0.7059, 0.6154)
6	(0.0202, 0.1145, 0.2830, 0.7647, 0.6538)
7	(0.8664, 0.1094, 0.0431, 0.5882, 0.1154)

According to the comparison with original data images (see Figure 2), we can draw the conclusion that the algorithm has good effect on clustering.

Acquiring the clustering centre, we can compare each feature's difference between two classifications based on their clustering centre. For example, the first class and the second class are similar in bandwidth and the number of thorns, but are different in exception and variance. Furthermore, we can also calculate the difference accurately.

For the same data, this paper also carried on comparison experimental by using the SOM. The results are as follows:

The final clustering result is as shown in Table 4:

Table 4. Clustering result by using SOM.

Class	Data serial number
1	11,13,15,16
2	17
3	5,6
4	1,7

5	2
6	14
7	8
8	3,4,9,10,12

As can be seen from the result, by using SOM, data 1 and data 7 are divided into the same class. Data 2 and 8 are divided into different classes. According to the classification label, PSO clustering algorithm has a better solution that data 1 and 2 are divided together and data 7 and 8 are in the same class (see Figure 2).

Besides, the result of SOM clustering is influenced by subjective factors. On the contrary, PSO clustering can draw a clear clustering results and output each classes' clustering center so as to quantitatively reflect the difference between each category.

4.3.2 Comparison of 158 groups of data clustering algorithms

Here we use 158 groups of categorized EMI signal data as test data. To present the advantages of our method, we use 4 other common clustering algorithms as comparison. To show the significance of dimension reduction, we also use original data as input vector as a comparison. During experiment, it spends hours of time to calculate. After doing the same number of recursion, results are disappointing. Clusters are hard to converge, which means almost all the signal are divided into a same class.

The comparison clustering results by using different approaches are shown in Table 5.

Table 5. Comparison results by using different approaches.

	Accuracy	Computation Time	Whether need clustering number
PSO+ feature extraction	93.04%	257ms	Yes
PSO+ original signals	10.17%	>1h	Yes
K-means	70.89%	985ms	Yes
SOM	91.77%	102ms	No
FCM	86.08%	364ms	Yes

Through repeated experiments for clustering, as can be seen, the algorithm proposed in this paper has high accuracy and efficiency in electromagnetic signal clustering. However, compared to SOM network, this algorithm requires the number of clustering, which is one disadvantage of the algorithm.

5 Conclusion

In order to separate interference factor from the system signal and use clustering to find out typical electromagnetic interference factor, this paper presents a method for feature extraction and clustering of electromagnetic signal data by PSO clustering algorithm,

and describe the advantages of this method compared to traditional clustering algorithms such as K-means, SOM, FCM, which show that our method outperforms the comparison approaches, and require less running time.

Acknowledgement

This work was supported by the National Natural Science Foundation of China (No.61379001) and the Key Project of Beihang University (No. 201412) (“Exploration and Practice of Innovation Ability Cultivation in Mathematical Education”).

References

1. Y. Ye, Z. Zhang, J. Zeng, L. Peng. A fast and adaptive ICA algorithm with its application to fetal electrocardiogram extraction. *Appl. Maths. Comput.*, **205(2)**, 799-806 (2008)
2. M.T. Akhtar, W. Mitsuhashi, C.J. James. Employing spatially constrained ICA and wavelet denoising, for automatic removal of artifacts from multichannel EEG data. *Signal Processing*, **92(2)**, 401-416 (2012)
3. D. Zhao, S. Wu. Method of electromagnetic signal leakage detecting based on wavelet transform. *Elec. test*, **8(1)**, 1-3 (2008)
4. C. Zhang, S. Liu, Q. Yang, L. Jin, L. Feng. Research on signal processing of electromagnetically induced acoustic emission based on the FFT and wavelet packet transform. *CTATEE'09*, 283-287 (2009)
5. S. Yang, Q. Zhong. Low-tension arc fault detection based on FFT and electromagnetic radiation. *Modern elec. Tech*, **35(18)**, 86-88 (2012).
6. Y. Jiang, Z. Zhang, P. Qiu, D. Zhou. Clustering algorithm used in data mining. *Journal of elec. & info. Tech.*, **27(4)**, 655-662 (2005)
7. D.W. Merwe, A.P. Engelbrecht. Data clustering using particle swarm optimization. *Proc of IEEE CEC*, 215-220, (2003)
8. L. Yang, H. Su, Y. Zhang, J. Chu. A method of data mining based on SOM clustering and its application. *Computer eng. & sci.*, **29(8)**, 133-136, (2007)
9. T. Kohonen. Self-Organized Formation of Topologically Correct Feature Maps. *Bio. Cybernetics*, **43(1)**, 59-69, (1982)
10. M. Omran, A. Salman, A.P. Engelbrecht. Image classification using particle swarm optimization. *Proc of the 4th APCSEL*, 370-374 (2002)
11. N. Sun, Q. Zhang. Research on PSO algorithm. *Journal of Hengshui Univ.*, **12(1)**, 25-28 (2012)
12. L. Chen, E. Zhang, R. Shen. The sorting of frequency hopping signals based on K-means algorithm with optimal initial clustering centers. *Journal of national Univ. of defense tech.*, **2**, 70-75 (2009)
13. Z. Ding, Z. Hu, J. Yang, Z. Tang, Y. Wu. A fuzzy-clustering-based approach to image segmentation. *Computer research & devel.*, **7**, 536-541 (1997)
14. Z. Chen, J. Deng, H. Chen, P. Ni. De-noising research of magneto telluric signal based on wavelet transform. *Chinese journal of eng. Geophys.*, **6**, 732-737 (2012)
15. H. Li, C. Wang, D. Zhao. An Improved EMD and Its Applications to Find the Basis Functions of EMI Signals. *Mathematical Problems in Engineering*, (2015)
16. H. Li, Z. Song, D. Zhao, P. Wang, J. Chen. A Single Channel EMI Signal Separation Method Based on Directly-mean Empirical Mode Decomposition. *Journal of Information & Computation Science*, **12(17)**, 6333-6340 (2015)
17. D. Zhao, H. Li, On the computation of inverses and determinants of a kind of special matrices. *Applied Mathematics and Computation*, **250**, 721-726 (2015)