电磁干扰

High Dimensional Electromagnetic Interference Signal Clustering Based On SOM Neural Network

集群,
使聚集

Di Zhao, Shaofeng Xu, Hongyi Li, Pidong Wang and Jiaxin Chen

***Abstract*—In this paper, we study the spectral characteristics and global representations of strongly nonlinear, non-stationary electromagnetic interferences (EMI), which is of great significance in analyzing the mathematical modelling of electromagnetic capability (EMC) for a large scale integrated system. We firstly propose to use Self-Organizing Feature Map Neural Network (SOM) to cluster EMI signals. To tackle with the high dimensionality of EMI signals, we combine the dimension reduction and clustering approaches, and find out the global features of different interference factors, in order to finally provide precise mathematical simulation models for EMC design, analysis, forecasting and evaluation. Experimental results have demonstrated the validity and effectiveness of the proposed method.**

***Index Terms*—EMI, mathematical simulation models, SOM**

Paper Classification
DOI: 10.7251/ELSxxxxxxxx

I. INTRODUCTION

WITH the rapid development of science and technology, the electromagnetic equipment gradually tends to be the large-scale integrated system. The existing measurements are designed for the whole system, and the interference elements are generated by the system components. Thus, the precision and accuracy of the EMI factor analysis will have a certain deviation and uncertainty, and we cannot accurately diagnose various interference factors. Therefore, we need to study mathematical modelling and simulation, namely mathematical representation, of EMI elements from the whole system. Due to the large scale integration of electromagnetic system, EMI signals present high-dimensional, strongly nonlinear and non-stationary characteristics, which existing mathematical models always fail to work. We need to study new mathematical representation methods, that is, to separate and decompose the various interference factors out of the system signal, based on which, we could find out the typical EMI elements such as clustering and so on. Finally, we can provide precise mathematical simulation model for the electromagnetic compatibility design, analysis, prediction and evaluation.

Existing institutions to carry out research in EMC technology can be roughly divided into two categories. One is the national research institution, such as the American NIST, the British NPL and German PTB and so on. They mainly engaged in all kinds of measuring probe antenna, magnetic field, electric field probe calibration and EMC test site research, and provide services. The other is specially engaged in calibration services companies and calibration laboratory of EMC test equipment manufacturers.

On the theoretical research, the IEEE have long noted the problem of radio interference, in October 1957 set up the RF interference professional group, and in May 1959 published the Journal of Radio Frequency Interference (RFI) volumes. In 1964, IEEE proceedings turned the RFI volumes to EMC volumes, the Institute (EMC Society) organizes an important academic conference every year [1].

Several typical existing methods to solve the problem of EMC are the differential equation method based on communication space discretization, the integral equation method based on the scattering body surface or internal discretization and the hybrid method. Sparse matrix can be achieved by using differential equation method, and the field propagation is discrete described in space, which would result in the space dispersion error. Additionally, this kind of methods require initial boundary conditions. Integral equation method analyses the scattering problem by solving the scattering body surface or volume of the induced current, there is no space dispersion error. Differential equation method is often easier to implement than the integral equation method. For uniform background medium open domain problem, we often use integral equation method to solve. For problems that use single integral equation method or differential equation method that are difficult to solve, we often analyse by using hybrid method.

稀疏矩阵

Manuscript received June 23, 2015. Received in revised form April 21, 2015.

The research work was supported by National Natural Science Foundation of China under Grant No. 61379001 and Key Project of Beihang University under Grant No. 201412 ("Exploration and Practice of Innovation Ability Cultivation in Mathematical Education").

Di Zhao is with the LMIB, School of Mathematics and Systems Science, Beihang University.

Shaofeng Xu is with the School of Software, Beihang University.

Hongyi Li is with the LMIB, School of Mathematics and Systems Science, Beihang University (Corresponding author. E-mail: Hongyili_buaa@163.com).

Pidong Wang is with the School of Reliability and Systems Engineering, Beihang University.

Jiaxin Chen is with the School of Computer Science and Engineering, Beihang University.

Common integral equation method includes method of moment (MOM), volume integral equation method (VIEM). Existing methods of differential equations include finite difference time domain method (FDFD), finite element method (FEM), the domain decomposition method (DDM), etc.

This paper focuses on the clustering applications in EMC analysis, on which topic there are few existing works for electromagnetic signals, to the best of our knowledge.

Commonly used clustering algorithms are K-means, KNN (K-Nearest Neighbour) and FCM (Fuzzy C-Means), they all can be directly used for EMC analysis. However, they have some disadvantages as follows:

K-means algorithm is not suitable for a large class cluster found in the non- convex shape or a big difference size between the clusters. In signal clustering, the number of different types of signals can vary widely, so the effect of K-means algorithm will be affected.

KNN algorithm has a major drawback, that is, its computation consuming is rather large, since in classification, the distance between any data pair needs to be calculated to obtain its K nearest neighbour points. [2] While, the EMC signals are always with high dimensional, KNN is not suitable for clustering on EMC signals.

FCM algorithm itself has two fatal weaknesses: Firstly, fuzzy clustering objective function is a non-convex function, there are a lot of local extreme points, where improper initialization will lead to convergence to the local extreme value point and get the optimal fuzzy partition of a data set; Secondly, large amount of data takes algorithm seriously, restrict its practical application.

In this paper, for multiple sets of EMI signals, we use a self-organizing feature map (SOM) for clustering. The topology structure of SOM algorithm is simple, and its self-stability and clustering results can be visualized. Meanwhile, the two- dimensional plan after SOM Fabric class also reflect the relevance of different types of signals.

Before using SOM neural network clustering, it is necessary to adopt some dimension reduction method of the signal of the high-dimensional data into low- dimensional that can cluster the data, which is also discussed in this paper.

The remaining part of the organizational structure of this paper is organized as follows: Section 2 describes the basic principles of SOM neural network, Section 3 describes the SOM neural network applications in high-dimensional signal processing, and Section 4 demonstrates the experimental results and analysis to verify the performance of the proposed method.

II. Signal dimension reduction techniques

A. *High-dimensional clustering research background*

Clustering is an important data analysis tool, which in accordance with certain requirements and rules for data set to distinguish and classify data objects, which then make a no category labelled data set divided into several subsets (classes) in accordance with certain criteria. Similar data objects are classified as a class, and dissimilar data objects are classified as different classes. Cluster analysis can effectively find the data distribution which is implicit in the data set, so as to lay a good foundation for further full and effective use of data. Meanwhile, with the rapid development of information technology, clustering is not only facing the problem of increasing amount of data, but also encounter with the high- dimensional problem.

In applications that high-dimensional data related to, the curse of dimensionality is a very common phenomenon. The term was first proposed by Bellman, it refers to a series of problems due to the excessive variables (properties) encountered in the data analysis. Since then, many researchers have done a lot of research dedicated to reduce or even eliminate the impact of the disaster on the dimension of high-dimensional data processing [3, 4].

B. *Two dimension reduction techniques*

When clustering high-dimensional data sets, traditional clustering method mainly encounter two problems. First, there are a lot of unrelated attributes in high-dimensional data set, which makes the possibility of clusters in all the dimensions almost zero. Second, data distribution in high-dimensional data space is sparser than in lower-dimensional space, distance between the data is almost equal is generally phenomenon. Moreover, since traditional clustering methods are based on distance, it is then unable to build clusters based on the distance in high-dimensional space.

There are generally two ways to solve the problem above:

1) Feature Transform [5, 6]: This kind of methods includes many traditional methods, including PCA [7] and SVD [8]. They transform the original data set to a k-dimensional new space by linear combination, making such a class of traditional clustering algorithm can be effective in this new data space, and finally achieve the purpose of reducing the dimension. But the disadvantage of this approach has three points: k is difficult to be determined; there are a lot of unrelated categories in high-dimensional space and it brings difficulties to clustering; it's prone to produce meaningless clusters. [9]
2) Feature Selection: Different from feature transform, feature selection performs mining tasks only on those relevant subspaces. Therefore, it is more efficient than the feature conversion to reduce dimensions. Feature selection generally uses greedy strategies to search different feature subspace, and then use some criteria to evaluate these subspaces, in order to find the desired clusters.

And we are using the second method - feature selection for dimension reduction of high-dimensional signal, enabling clustering in low-dimensional space.

III. The basic principle of SOM neural network

SOM clustering algorithm was first put forward by Finn Kohonen in 1982 [10], it is a kind of unsupervised training of neural network [11], and the self-organizing process in fact is unsupervised, which automatically clusters input patterns without label information.

SOM consists of the input layer and competition layer (output layer). The number of input layer neurons is N, the competitive layer is one-dimensional or two-dimensional planar array consisting of m neurons, which can also be seen as a three-dimensional cubic lattice. The network is fully connected.

IV. A HIGH-DIMENSIONAL SIGNALS CLUSTERING METHOD BASED ON SOM

In the previous section, we introduced signal dimension reduction techniques and SOM theory. Now we propose a high dimensional electromagnetic interference signal clustering method based on SOM neural network, and then find out the global features of different interference factors, in order to finally provide precise mathematical simulation models for EMC design, analysis, forecasting and evaluation technology.

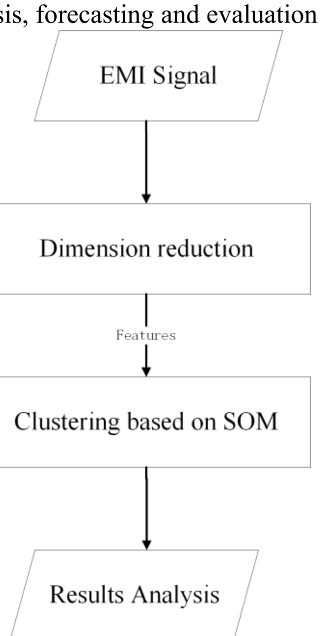

Fig. 1. Flow chart of the proposed method.

As the above chart shows, our method mainly consists of the following three steps.

A. Signal dimension reduction

Based on the spectrum features of the signal, we choose the following five features instead of the original signal high dimensional information: number of crest, number of trough, expectation, variance and bandwidth.

The mathematical definitions of features are defined as follows:

Crest: for each sampling point X, if the function value of this sampling point is greater than the left sample point (if any) and the right sample point (if any), then that point is defined as a crest.

Trough: for each sampling point X, if the function value of this sampling point is smaller than the left sample point (if any) and the right sample point (if any), then that point is defined as a trough.

Expectation: the mean of all sampling points

Variance: the mean of all the square of difference between the actual values and expectations.

Bandwidth: for every 10 sampling points, calculate the mean of crest and trough, meantime subtraction, and then averaged.

B. Clustering based on SOM

Through the choice of above features, we can turn thousands of dimension of signal data of each group into five dimension data of each group, this greatly facilitate the use of SOM neural network clustering. [12]

V. RESULTS AND DISCUSSION

In order to test the performance of the proposed method, we choose 158 groups of frequency domain EMI signal data. EMI signals in this dataset are high-dimensional (with thousands of dimension), and the frequency domain information can help us explore more general features from the signals and get better clustering results.

In this experiment, the signals in the dataset can be reconstructed with the five selected features proposed above and then they are processed as the input of the SOM neural network.

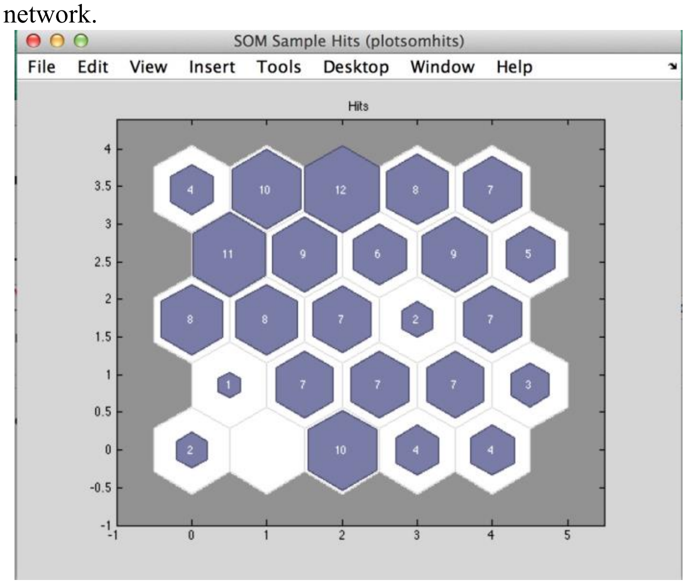

Fig. 2. SOM neural network clustering results.

After a SOM neural network clustering, the 158 groups of signal clustering results are as follows:

The number of the above represent sets of data gathered in this class, the adjacent hexagons shows some features of this two category are similar.

Fig. 3-Fig. 6 show parts of the clustering results.

Through the comparison, we found that the effect of whole SOM neural network for the 5-d eigenvalue is more satisfactory. From Fig. 3 to Fig. 6, it can be seen that signal spectrum are similar in the same class (as shown in Fig. 3 and Fig. 4), the spectrum between two classes have some similar features. Fig. 5 shows that different signal frequency spectrum features are big different in this kind of class. It indicates two aspects, one is the selected five eigenvalues for dimension reduction are not accurate enough, they cannot on behalf of the entire signal features; the other is SOM neural network still has optimize space, by adjusting the node number of competitive layer, changing the final clustering number, through the experiment to find the most suitable clustering number, can also help accurate clustering.

To illustrate the effect of SOM neural network clustering, we use the K-means clustering method on source data again, the results are partly shown as Fig. 7-Fig. 8.

Through the above part of the clustering results we can see that, K-means clustering method although can group similar signal together, but compared with the SOM clustering, two small classes distinct signal are easy to cluster into a broad class, for example, in Fig. 7, the middle line of three signal and other signal features are big different in the graph; In Fig. 8, the features of the last two signals and other signals are big different also. It shows that performance of K-means clustering method is poor compared to SOM neural network.

Generally speaking, since EMI signals are high-dimensional and non-linear, it is thus computationally cost and difficult to directly conduct clustering. However, by using the proposed low dimensional features representation of EMI signals, SOM can treat the five features as normal input to cluster and yield significantly better performance.

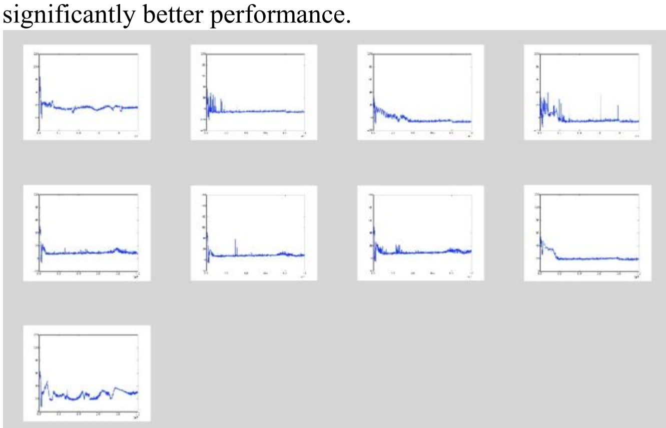

Fig. 3. Class 8.

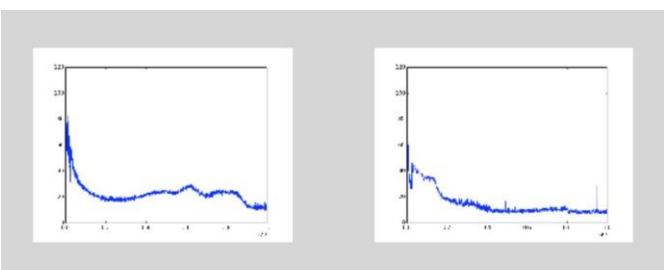

Fig. 4. Class 9.

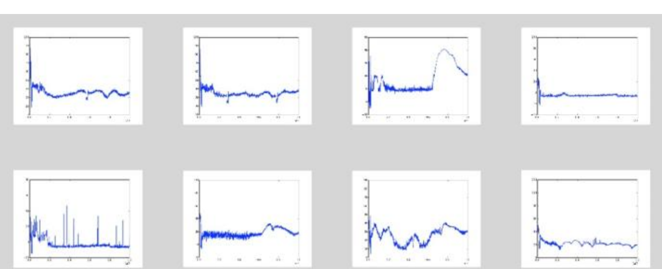

Fig. 5. Class 12.

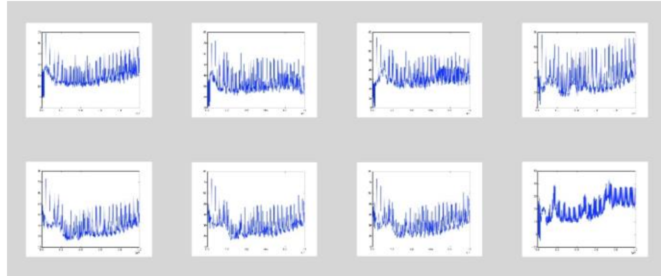

Fig. 6. Class 25.

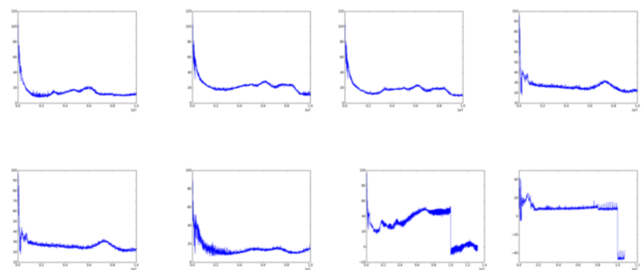

Fig. 7. Class 1.

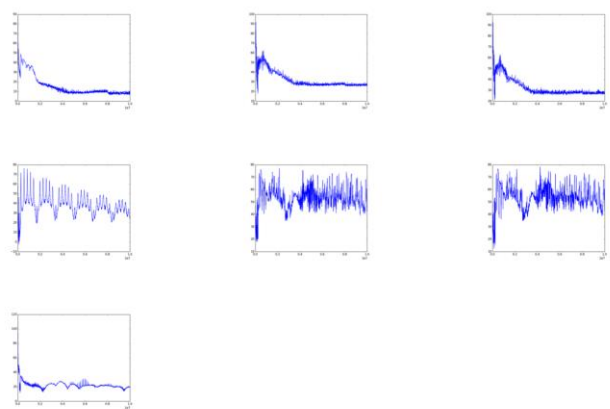

Fig. 8. Class 4.

VI. CONCLUSIONS

In this paper, through the clustering of EMI signal, we have studied the global representation of EMI, and provided research methods for EMC mathematical model building. Meanwhile, to deal with the high dimensionality of EMI signals, we have introduced the dimension reduction approaches, which could further improve the clustering effect of EMI signals.

ACKNOWLEDGMENT

The authors would thank the editor and anonymous reviewers. They have provided valuable suggestions, which significantly help improve the quality of our paper.

REFERENCES

[1] R. Kohavi, G. John, Wrappers for feature subset selection. Artificial Intelligence, (1–2), pp. 273–324, 1997.

[2] D. Zhao, Y. Fu, C. Wang and H. Li, A novel algorithm for classifying measuring curves based on interval estimation. Advanced Materials Research, 734–737, pp. 3061–3064, 2013.

[3] T. Kohonen, The self-organizing map. Proceedings of the IEEE, 78(9), pp. 1464–1480, 1990.
[4] L. Ertöz, M. Steinbach, V. Kumar, Finding clusters of different sizes, shapes, and densities in noisy, high dimensional data. SDM, pp. 47–58, 2003.
[5] J. H. Ham, D. D. Lee, L. K. Saul, Learning high dimensional correspondences from low dimensional manifolds. Proc. of ICML Workshop on the Continuum from Labeled to Unlabeled Data in Machine Learning and Data Mining, pp. 34–41, 2003.
[6] C. Ding , et al, Adaptive dimension reduction for clustering high dimensional data. Proc. of the Second IEEE International Conference on Data Mining, pp. 147–154, 2002.
[7] C. Farhat, J. Mandel, F. X. Roux, Optimal convergence properties of the FETI domain decomposition method. Computer methods in applied mechanics and engineering, 115(3), pp. 365–385, 1994.
[8] S. Wold, K. Esbensen, P. Geladi, Principal component analysis. Chemometrics and intelligent laboratory systems, 2(1), pp. 37–52, 1987.
[9] H. Li, M. Ye and D. Zhao, An improved ICA algorithm based on the negative entropy and simulated annealing algorithm. Applied Mechanics and Materials, 411–414, pp. 1125–1128, 2013.
[10] J. C. Bezdek, R. Ehrlich, W. Full, FCM: The fuzzy c-means clustering algorithm. Computers & Geosciences, 10(2), pp. 191–203, 1984.
[11] H. Li, X. Wu, T. Wang and D. Zhao, A simulation and prediction method of the height of wheat based on BP neural networks and the ant colony algorithm. Advanced Materials Research, 659, pp. 113–117, 2013.
[12] D. Zhao, H.F. Liu and H.Y. Li, A Method Based on Grey Correlation Clustering for Improving the Hierarchy of Analytic Hierarchy Process. Proc. IEEE/WIC/ACM International Conference on WI-IAT, 3, pp. 156-159, 2010.