# A FAULT DIAGNOSIS METHOD BASED ON SEMI-SUPERVISED FUZZY C-MEANS CLUSTER ANALYSIS

Su-Qun Cao[1, 2, 3], Xinggang Ma[2], Youfu Zhang[4], Limin Luo[1], Fupeng Yi[3]

[1]School of Computer Science and Engineering, Southeast University, Nanjing, China
[2]The affiliated Huai'an Hospital of Xuzhou Medical College, Huai'an, China
[3]Faculty of Mechanical Engineering, Huaiyin Institute of Technology, Huai'an, China
[4]School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing, China

### ABSTRACT

*Machine learning approaches are generally adopted in many fields including data mining, image processing, intelligent fault diagnosis etc. As a classic unsupervised learning technology, fuzzy C-means cluster analysis plays a vital role in machine learning based intelligent fault diagnosis. With the rapid development of science and technology, the monitoring signal data is numerous and keeps growing fast. Only typical fault samples can be obtained and labeled. Thus, how to apply semi-supervised learning technology in fault diagnosis is significant for guaranteeing the equipment safety. According to this, a novel fault diagnosis method based on semi-supervised fuzzy C-means(SFCM) cluster analysis is proposed. Experimental results on Iris data set and the steel plates faults data set show that this method is superior to traditional fuzzy C-means clustering analysis.*

### KEYWORDS

*Fault Diagnosis, Unsupervised pattern, Cluster Analysis, Semi-supervised Learning*

## 1. INTRODUCTION

With the widespread use of modern manufacturing systems, machinery is expected to have stable performance for fully meeting customer requirements. Obviously, unexpected downtime due to machine failure leads to be more costly than before. Therefore, fault diagnosis of machinery is of great practical significance in guaranteeing machinery safety and keeping away from the losses. Traditional intelligent fault diagnosis based on expert system is over-reliance on empirical knowledge, and it is hard for getting knowledge automatically [1]. With the development of pattern recognition and data mining, machine learning based methods have been a research focus gradually in the field of intelligent fault diagnosis.

Many machine learning techniques have been proposed in recent years. Most of them use supervised learning methods such as neural network[2], support vector machine[3] et al. to solve fault diagnosis problems. Jaroslaw Kurek et al. [4] presents an automatic computerized system for the diagnosis of the rotor bars of the induction electrical motor by applying the support vector machine. Yi Lu Murphey et al. [5] present a model-based fault diagnostics system developed using a machine learning technology for detecting and locating multiple classes of faults in an electric drive. Huo-Ching Sun et al. [6] proposed an enhanced particle swarm optimization (EPSO)-based support vector classifier (SVC) that extracts the support vector from databases, in

DOI: 10.5121/ijci.2015.4227 

order to diagnose vibration faults in steam turbine-generator sets. Miguel Delgado Prieto et al. [7] presents a novel monitoring scheme applied to diagnose bearing faults. The development of diagnosis methodologies considering both kinds of bearing faults is, nowadays, subject of concern in fault diagnosis of electrical machines. A novel neural network structure is adopted to perform the classification stage. Kaveh Bastani et al. [8] proposed a fault diagnosis methodology by integrating the state space model with the enhanced relevance vector machine (RVM) to identify the process faults through the sparse estimate of the variance change of the process errors. Muhammad Amar et al.[9] presented a novel vibration spectrum imaging (VSI) feature enhancement procedure for low SNR conditions. An artificial neural network (ANN) has been used as a fault classifier using these enhanced features of the faults. It provides enhanced spectral images for ANN training and thus leads to a highly robust fault classifier.

Some apply unsupervised learning methods such as cluster analysis [10] to detect and diagnose the failure data. Wentao Sui et al. [11] proposed a new method of bearing fault diagnosis based on feature weighted FCM. G.Yu et al. [12] presented a cluster-based feature extraction from the coefficients of discrete wavelet transform for machine fault diagnosis. Qing Yang et al. [13] proposed an ensemble fault diagnosis algorithm based on fuzzy c-means algorithm (FCM) with the Optimal Number of Clusters (ONC) and probabilistic neural network (PNN), called FCM-ONC-PNN. Abdenour Soualhi et al. [14] presented a new approach for fault detection and diagnosis of IMs using signal-based method based on signal processing and an unsupervised classification technique called the artificial ant clustering. Debin Zhao et al. [15] presented a novel intelligent fault diagnosis method based on ant colony clustering analysis and it is verified by vibration signals acquired from a rotor test bed. Yaguo Lei et al. [16] presented a new hybrid clustering algorithm based on a three-layer feed forward neural network (FFNN), a distribution density function, and a cluster validity index to solve fault diagnosis problem of rotating machinery. Hesam Komari Alaei et al. [17] proposed a new on-line fuzzy clustering-based algorithm which is developed using integration of an adaptive principal component analysis approach with a weighted fuzzy C-means (WFCM) methodology for process fault detection and diagnosis (FDD) applications.

But because the monitoring signal data is numerous, only typical fault samples can be labeled. Thus, how to apply semi-supervised learning technology in fault diagnosis is important for guaranteeing the equipment safety. According to this, a novel fault diagnosis method based on semi-supervised fuzzy C-means(SFCM) cluster analysis is proposed in this paper.

## 2. FUZZY C-MEANS CLUSTERING ALGORITHM

FCM algorithm was introduced by J.C.Bezdek[18]. It is intended to obtain clustering result of a data set by minimizing of the basic c-means objective function:

$$J(Z;U,V)=\sum_{i=1}^{c}\sum_{k=1}^{N}u_{ik}^{m}D_{ik}^{2}$$

Where:

$\mathbf{U}=[u_{ik}]$ is a fuzzy partition matrix of Z;

$\mathbf{V}=[\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_c],\mathbf{v}_i\in\mathbf{R}^n$ is a vector of cluster prototypes;

$D_{ik}=\|\mathbf{z}_k-\mathbf{v}_i\|$ is Euclidean distance between the sample $\mathbf{z}_k$ and the center $\mathbf{v}_i$ of the cluster $i$;

$m\in(1,\infty)$ is a parameter which decides the fuzziness of the resulting clusters.

The minimization of $J(\mathbf{Z};\mathbf{U},\mathbf{V})$, under the constraint $\sum_{i=1}^{c} u_{ik} = 1$, $u_{ik} \in [0,1]$, leads to the iteration of the following steps:

$$\mathbf{v}_i = \frac{\sum_{k=1}^{N} (u_{ik})^m \mathbf{z}_k}{\sum_{k=1}^{N} (u_{ik})^m}, \; 1 \le i \le c$$

and

$$u_{ik} = \frac{1}{\sum_{j=1}^{c} (D_{ik} / D_{jk})^{2/(m-1)}}$$

## 3. Semi-supervised Fuzzy c-means Clustering Algorithm (SFCM)

The semi-supervised fuzzy c-means (SFCM) algorithm was first introduced by Pedrycz [19]. The objective function is shown as follows.

$$J_2^s(U,V;X,F) = \sum_{i=1}^{c}\sum_{k=1}^{n} u_{ik}^2 d_{ik}^2 + \sum_{i=1}^{c}\sum_{k=1}^{n} (u_{ik} - f_{ik})^2 d_{ik}^2$$

Pedrycz and Waletzky introduced a binary vector $b$ ($b_k$ is equal to 1 if the sample $x_k$ has been already labeled and $b_k = 0$ otherwise) to distinguish whether the sample is supervised and proposed an improved semi-supervised FCM algorithm, which has the following objective function:

$$J_{m\cdot a}^s(U,V;X,F) = \sum_{i=1}^{c}\sum_{k=1}^{n} u_{ik}^2 d_{ik}^2 + \alpha \sum_{i=1}^{c}\sum_{k=1}^{n} (u_{ik} - f_{ik} b_k)^m d_{ik}^2$$

Here, $\alpha$ is used to maintain a balance between supervised and unsupervised optimization mechanism and is proportional to the ratio of the number of all samples and the number of labeled samples.

## 4. Experiments

### 4.1 Experiment on Iris data set

In this experiment, the UCI data set [20], Iris, is used to do clustering by FCM and SFCM. The following Rand index [21] is adopted to evaluate the clustering effectiveness of these two methods.

$$Rand(P_1, P_2) = \frac{a+b}{n \times (n-1)/2}$$

Where $P_1$ means the clustering result of Iris data set, $P_2$ means the standard clustering result obtained by the labels of Iris data set, $a$ denotes the number of any two samples in Iris data set belonging to the same cluster in $P_1, P_2$, $b$ denotes the number of any two samples in Iris data set belonging to two different clusters in $P_1, P_2$, and $n$ is the number of all samples. Obviously,

$Rand(P_1, P_2) \in [0,1]$. And $Rand(P_1, P_2) = 1$ when $P_1$ is the same as $P_2$. The smaller $Rand(P_1, P_2)$ is, the bigger the difference between $P_1$ and $P_2$.

The Iris data set consists of 150 samples which are belong to three classes of Iris such as Iris setosa, Iris virginica and Iris versicolor. Each sample is measured by four features: the length and the width of the sepals and petals. Table 1 illustrates the basic information of the data set.
FCM is applied in Iris data set and the clustering result plotted by first two features is shown in Figure 1. The Rand index of FCM clustering result is 0.9495. We set 30 samples as labeled samples and then apply SFCM in Iris data set. The clustering result plotted by first two features is shown in Figure 2. The Rand index of SFCM clustering result is 0.9575. This means that SFCM has better clustering result than FCM.

Table 1. Class distribution and features of Iris data set

| **Class** | **Number of Samples** | **Features** | |
|---|---|---|---|
| | | **No.** | **Name** |
| Class 1:Iris setosa | 50 | 1 | Sepal length |
| Class 2:Iris virginica | 50 | 2 | Sepal width |
| Class 3:Iris versicolor | 50 | 3 | Petal length |
| | | 4 | Petal width |

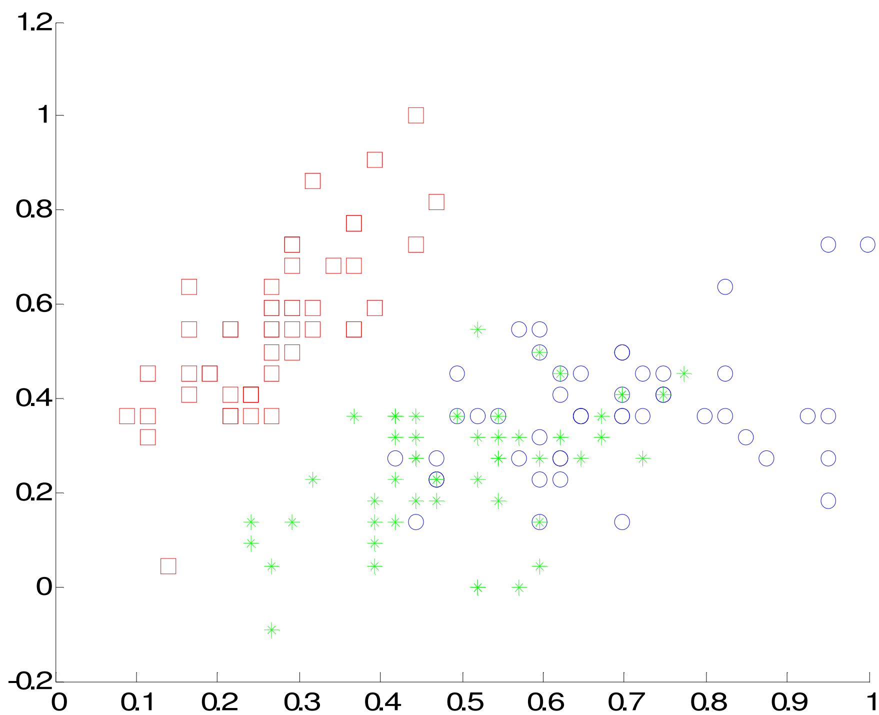


Figure 1. FCM Clustering on Iris data set

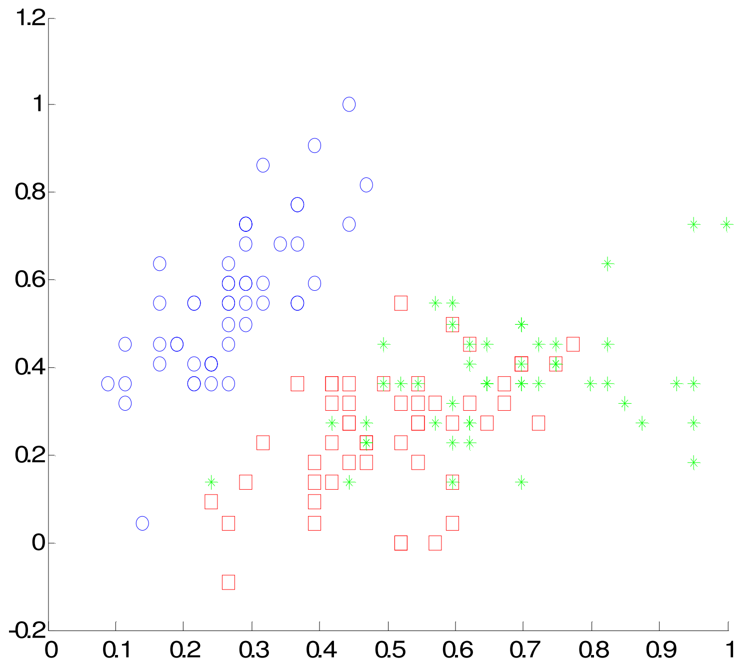


Figure 2. SFCM Clustering on Iris data set

### 4.2. Experiment on the steel plates faults data set

The Steel Plates Faults Data Set was given by Research Center of Sciences of Communication, Rome, Italy [22]. There are 7 different steel plates' faults in this data set: Pastry, Z_Scratch, K_Scatch, Stains, Dirtiness, Bumps and Other_Faults. It includes 1941 samples and each sample is described by 27 independent features.

Table 2 shows the basic information of the data set. 348 samples which belong to Pastry and Z_Scratch faults are chosen as testing data set.

FCM is applied in the steel plates faults data set and the clustering result plotted by first two features is shown in Figure 3. The Rand index of FCM clustering result is 0.5066. We set 80 samples as labeled samples and then apply SFCM in this data set. The clustering result plotted by first two features is shown in Figure 4. The Rand index of SFCM clustering result is 0.5200. This means that SFCM has better clustering result than FCM.

## 5. Conclusions

Because the monitoring signal data is numerous, only typical fault samples can be labeled. Thus, how to apply semi-supervised learning technology in fault diagnosis is important for guaranteeing the equipment safety. This paper presents a novel fault diagnosis method based on Semi-supervised Fuzzy C-Means(SFCM) clustering algorithm. It use labeled samples to guide the clustering process. Experimental results show that it is more efficient than traditional Fuzzy C-Means(FCM) clustering algorithm. In the future, we will make a deep research about how to do fault diagnosis by semi-supervised clustering based on the kernel function[23, 24].

Table 2. Class distribution and features of Steel Plates data set

| Class | Number of Samples | Features | | | |
|---|---|---|---|---|---|
| | | No. | Name | No. | Name |
| Pastry | 158 | 1 | X_Minimum | 15 | Edges_Index |
| Z_Scratch | 190 | 2 | X_Maximum | 16 | Empty_Index |
| K_Scatch | 391 | 3 | Y_Minimum | 17 | Square_Index |
| Stains | 72 | 4 | Y_Maximum | 18 | Outside_X_Index |
| Dirtiness | 55 | 5 | Pixels_Areas | 19 | Edges_X_Index |
| Bumps | 402 | 6 | X_Perimeter | 20 | Edges_Y_Index |
| Other_Faults | 673 | 7 | Y_Perimeter | 21 | Outside_Global_Index |
| | | 8 | Sum_of_Luminosity | 22 | LogOfAreas |
| | | 9 | Minimum_of_Luminosity | 23 | Log_X_Index |
| | | 10 | Maximum_of_Luminosity | 24 | Log_Y_Index |
| | | 11 | Length_of_Conveyer | 25 | Orientation_Index |
| | | 12 | TypeOfSteel_A300 | 26 | Luminosity_Index |
| | | 13 | TypeOfSteel_A400 | 27 | SigmoidOfAreas |
| | | 14 | Steel_Plate_Thickness | | |

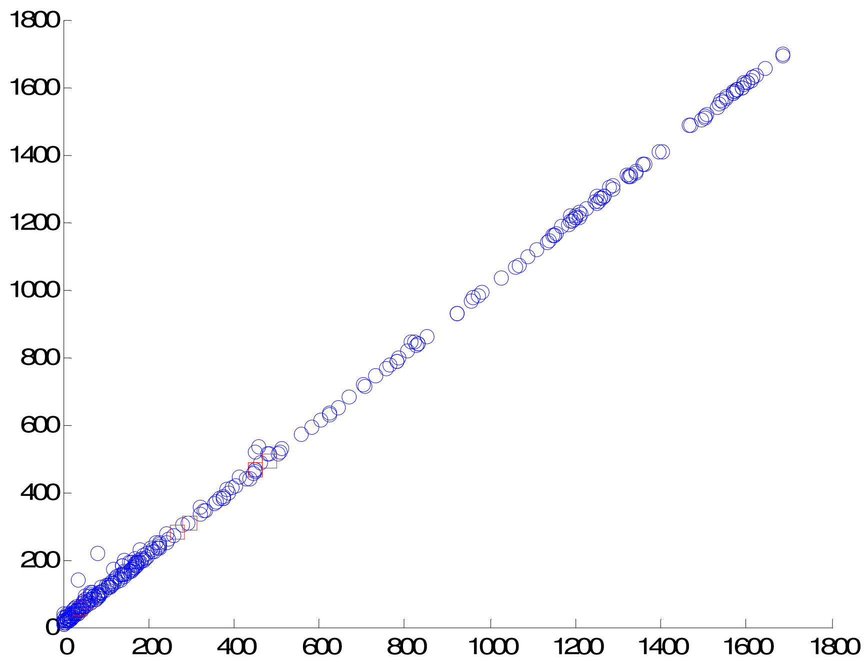


Figure 3. FCM Clustering on the steel plates faults data set

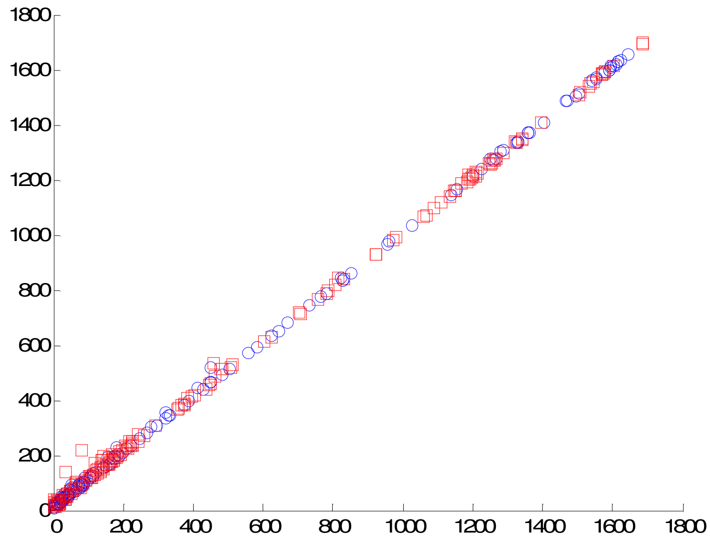


Figure 4. SFCM Clustering on the steel plates faults data set

## ACKNOWLEDGEMENTS

This work is supported by National Natural Science Foundation of China (61402192), the National Sparking Plan Project, China (2013GA690404), the Major Program of the Natural Science Foundation of the Jiangsu Higher Education Institutions of China (11KJA460001, 13KJA460001), the Open Project from the Key Laboratory of Digital Manufacture Technology at Jiangsu Province in China (HGDML-1005), Technology Innovation Project of Science and Technology Enterprises at Jiangsu Province in China (BC2012429), Huaian 533 Talents Project, Huaian International Science and Technology Cooperation Project (HG201308), and Jiangsu Overseas Research & Training Program for University Prominent Young & Middle-Aged Teachers and Presidents, China. Xinggang Ma is the corresponding author(e-mail: hamxg@sina.com).

## REFERENCES

[1] Jingyi Du & Yuanlin Hou, (2010) Fault Diagnosis Theory based on Kernel Method and its Method Study, Network Books, Beijing University Press.

[2] Jianjun Zhang, Li Zhang, Haifang Mu & Zhengyu Liu , Juan Xu (2011) "Improved BP-neural network of the particle swarm optimization in the research on engine fault diagnosis", Journal of agricultural machinery, Vol. 42, pp198-203.

[3] Xuewen He & Haiming Zhao, (2005) "Support vector machine and its application in the mechanical fault diagnosis", Journal of Zhongnan University, Vol. 36, pp97-101.

[4] Jaroslaw Kurek & Stanislaw Osowski, (2010) "Support vector machine for fault diagnosis of the broken rotor bars of squirrel-cage induction motor", Neural Comput & Applic Vol. 19, pp557–564.

[5] Yi Lu Murphey, M. Abul Masrur & ZhiHang Chen, Baifang Zhang, (2006) "Model-based fault diagnosis in electric drives using machine learning", IEEE/ASME Transactions on MECHATRONICS, Vol. 11, No. 3, pp290-303.

[6] Huo-Ching Sun, Chao-Ming Huang & Yann-Chang Huang, (2013) "Fault diagnosis of steam turbine-generator sets using an EPSO-based support vector classifier", IEEE Transactions on ENERGY CONVERSION, VOL. 28, NO. 1, pp164-171.

[7] Miguel Delgado Prieto, Giansalvo Cirrincione, Antonio Garcia Espinosa & Juan Antonio Ortega, Humberto Henao, (2013) "Bearing Fault Detection by a Novel Condition-Monitoring Scheme Based

on Statistical-Time Features and Neural Networks", IEEE Transactions on INDUSTRIAL ELECTRONICS, VOL. 30, NO. 8, pp3398-3407.

[8] Kaveh Bastani, Zhenyu Kong, Wenzhen Huang & Xiaoming Huo, Yingqing Zhou, (2013) "Fault Diagnosis Using an Enhanced Relevance Vector Machine (RVM) for Partially Diagnosable Multistation Assembly Processes", IEEE Transactions on AUTOMATION SCIENCE AND ENGINEERING, VOL. 10, NO. 1, pp124-136.

[9] Muhammad Amar, Iqbal Gondal & Campbell Wilson, (2015) "Vibration Spectrum Imaging: A Novel Bearing Fault Classification Approach", IEEE Transactions on INDUSTRIAL ELECTRONICS, VOL. 62, NO. 1, pp494-502.

[10] Gelan Yang, Xue Xu & Gang Yang, Jianming Zhang, (2010) "Semi-supervised Classification by Local Coordination", Lecture Notes in Computer Science, Vol. 6444, pp517-524.

[11] Wentao Sui, Changhou Lu & Dan Zhang, (2008) "Bearing Fault Diagnosis Based on Feature Weighted FCM Cluster Analysis", International Conference on Computer Science and Software Engineering, pp518-521.

[12] G.Yu, C.N.Li & Sagar.Kamarthi, (2008) "A Cluster-based Wavelet Feature Extraction Method for Machine Fault Diagnosis", Applied Mechanics and Materials, Vol. 10-12, pp548-552.

[13] Qing Yang, Jingran Guo & Dongxu Zhang, Chang Liu, (2011) "Fault Diagnosis Based on Fuzzy C-means Algorithm of the Optimal Number of Clusters and Probabilistic Neural Network", International Journal of Intelligent Engineering & Systems, Vol. 4, pp51-59.

[14] Abdenour Soualhi, Guy Clerc & Hubert Razik, (2013) "Detection and Diagnosis of Faults in Induction Motor Using an Improved Artificial Ant Clustering Technique", IEEE Transactions on INDUSTRIAL ELECTRONICS, VOL. 60, NO. 9, pp4053-4062.

[15] Debin Zhao & Jihong Yan, (2013) "Ant colony clustering analysis based intelligent fault diagnosis method and its application to rotating machinery", Pattern Anal Applic, Vol. 16, pp19-29.

[16] Yaguo Lei, Zhengjia He & Yanyang Zi, Qiao Hu, (2008) "Fault diagnosis of rotating machinery based on a new hybrid clustering algorithm", Int J Adv Manuf Technol, Vol. 35, pp968–977.

[17] Hesam Komari Alaei, Karim Salahshoor & Hamed Komari Alaei, (2013) "A new integrated on-line fuzzy clustering and segmentation methodology with adaptive PCA approach for process monitoring and fault detection and diagnosis", Soft Comput, Vol. 17, pp345-362.

[18] J.C. Bezdek, (1981) Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum Press.

[19] Pedrycz W, (1985) "Algorithms of fuzzy clustering with partial supervision", Pattern Recognit Lett, Vol. 3, No. 1, pp13-20

[20] C.L. Blake & C.J. Merz, (1998) "UCI repository of machine learning databases", [Online], Available: http://www.ics.uci.edu/~mlearn/MLRepository.html.

[21] Rand W., (1971) "Objective Criteria for the Evaluation of Clustering Methods," Journal of the American Statistical Association, Vol. 66, No.336, pp846-850.

[22] (2011) "Center for Machine Learning and Intelligent Systems", [Online], Available: http://archive.ics.uci.edu/ml/datasets/Steel+Plates+Faults.

[23] Dao-Qiang Zhang & Song-Can Chen, (2004) "A Novel Kernelized Fuzzy C-means Algorithm with Application in Medical Image Segmentation", Artificial Intelligence in Medicine, Vol. 32, pp37-50.

[24] Guoxin Zhang & Zuoshi Liu, (2006) "The method of fuzzy clustering for bearing fault diagnosis", Journal of Jiangxi University of Science and Technology, Vol. 27, pp15-17.

## AUTHORS

Su-Qun Cao was born in Huai'an City, Jiangsu Prov., China in 1976. He received the B.E. in Mechanical Engineering from Jilin University of Technology, China, in 1997, the second B.E. in Software Engineering from Tsinghua Uni versity, China, in 2002, and the Ph.D. degree in Computer Engineering from Jiangnan University, China, in 2009. Since 2010 he has been an associate professor at the Faculty of Mechanical Engineering, Huaiyin Institute of Technology, China. From 2012 to 2013, he was a visiting scholar at the University of Melbourne, Australia. Now he is a post-doctor at Southeast University, C hina. His current r esearch interests include machine learning, pattern recognition and fault diagnosis.

Xinggang Ma is a professor and supervisor of postgraduate at Xuzhou Medical College.He is also a chief physician at the affiliated Huaian Hospital of Xuzhou Medical College. His current research interests include gastroenterology and medical informatization.

Youfu Zhang was born in Huai'an City, Jiangsu Province, China in 1994. From September 2012, He began studying for an undergraduate degree in Information Engineering at Beijing University of Posts and Telecommunications in Beijing. His current research interests include information system and software development.